Frameworks for Planning Collaborative Supply Chain Programs

by

Suryanarayanan Gurumurthi

Department of Business Administration
Duke University

Date: ____________________

Approved:

Paul Zipkin, Advisor

Li Chen

Pranab Majumder

Reha Uzsoy

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2011
Abstract

(Business Administration)

Frameworks for Planning Collaborative Supply Chain Programs

by

Suryanarayanan Gurumurthi

Department of Business Administration
Duke University

Date: __________________________

Approved:

______________________________
Paul Zipkin, Advisor

______________________________
Li Chen

______________________________
Pranab Majumder

______________________________
Reha Uzsoy

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2011
Abstract

This dissertation is written in three progressively restrictive parts. Part I is a set of two expansive essays on collaborative supply chain management that proposes several new perspectives and interconnections between current day global business and economic issues, and the evolving supply chain structures and decision-making paradigms that depend on extensive inter-firm collaboration. Part I also develops new guidelines for both practitioners as well as academic researchers in their quest to incorporate collaborative requirements as an explicit component of existing planning frameworks and modeling approaches. Part I further comments on how the technological evolution of manufacturing, service, and general business processes have led to decentralized structures that require a fundamentally collaborative approach to the planning of such processes. We also argue that existing supply chain decision-making and planning approaches are modeled in the fashion of corporate and enterprise resource planning systems, which given their scope, limit the extent of collaboration in both planning and in execution. The arguments and discussion in this part are not specific to any particular supply chain function and is without technological bias. The frameworks presented in Part I are also unified in their approach to managing supply chains of service providers, manufacturing partners, or some combination of both types of activities. This unified presentation is also a fundamental contribution of this first part of the dissertation.
Part II of the dissertation, while still expansive in scope of application and the range of industry sectors and supply chain environments discussed, develops the ideas presented in Part I for more specific (or functional) categories of business processes. A commonly accepted categorization of operational processes, at least in manufacturing settings, is into (i) product design and development or related projects, which are akin to services in the nature and interaction between implied tasks, (ii) procurement, production, and customer service processes, and (iii) logistics and distribution networks. Projects are typically represented as a network of inter-related activities bound by a common purpose, and by a time-line dictated by a finite product or project life cycle; activities are also sometimes defined and created in response to project environments. Processes in the procurement, production, customer service, or logistics domains, on the other hand, are typically modeled as a set of inter-related but more loosely coupled activities that are repeated indefinitely across multiple product or project life cycles. Our primary concern in Part II is to understand environments where projects and processes span multiple firms, and therefore require a collaborative effort, not only for executing the activities entailed, but also in the planning of the tasks and projects.

Modeling of supply chain management problems (such as those discussed subsequently in Part III) assume that the fundamental structure of tasks and processes are at least well-defined for analysis and subsequent design of parameters for optimal performance. Often, however, the inclusion and structuring of these tasks is also a collaborative exercise that requires negotiation and careful consideration of the costs and advantages presented by alternative sets of tasks. The scope of tasks is also frequently determined by their assignment to one or more firms with differing capabilities. For example, the range of logistics activities and services provided by a specialized firm would be greater than a manufacturer assuming additional responsibility.
sibility for the distribution or procurement logistics. Similarly, the capabilities of a supplier would either expand or restrict the range of tasks that would be included in the design and development of a product or a service. Therefore, Part II of the dissertation, consisting of Chapters 4 and 5, develops strategic frameworks that can allow the definition and structuring of tasks and processes in a collaborative setting. These chapters present frameworks for strategy and for defining project or process objectives which are commonly the guideposts for task definition and structuring.

These frameworks presented in Part II can also help determine the degree of collaboration either warranted or indeed suitable for different project and logistics environments. Thus, we propose that some business and technology environments call for more cohesive or coupled structuring of tasks that in turn require collaborative frameworks for planning and execution. Some other environments, either as a result of market forces or technological constraints, are a bad fit for collaborative efforts unless they are seamless and frictionless. Identifying such environments through a small set of market and technological factors is a fundamental contribution of Part II of the dissertation. Similar to our efforts in Part I, we also chart the evolution of collaborative planning and execution environments; here we adopt a more direct case based approach to illustrating issues, and related concepts. Another significant contribution of this second part is to outline how various facets of the operating environment shape the parameters of the collaborative arrangements between partner firms. In particular, we address the environmental and strategic forces that motivate a model of work sharing in environments where collaboration is not a technological requirement. Thus, we address the fundamental value proposition in collaborative logistics management for the outsourcing provider and the contracting firm, and discuss how product or process technology and structure influences such choices by firms.
Part III *which is more restrictive in its statements and conclusions*, is devoted to models of collaborative supply chain management that are motivated by the imperatives outlined in Part I, but whose elements are defined by the strategic frameworks and structuring guidelines of Part II. While Part III derives guidance from Part II in the formulation of its models, it can also be viewed and read independently for its contributions to the (related) academic literature. Part III consists again of two independent modeling exercises. Through either of these exercises, we address two of the most important problems in collaborative supply chain planning: partner selection, or alternatively task and project assignment, and decentralized capacity management in a supply chain or logistics environment. These models describe two different environments where collaborative planning is vital to the success of firms: (i) decentralized and collaborative projects or programs that involve the financial, technological, and human resources of several firms towards a common revenue or savings objective, and (ii) collaborative but decentralized logistics and transportation systems where several firms in a supply chain must invest in common fulfillment or processing infrastructure, and further determine the material flows within that logistics network.

Through both models, we show how decentralized capacity investment can be inefficient relative to centralized planning. We also characterize the decentralized equilibrium behavior of firms under the proportional risk-sharing regimes. We then provide mechanisms that can coordinate the decentralized systems, where coordination is defined here in a more limited sense as the development of decentralized equilibria that match the central planner’s capacity requirement. These co-ordination mechanisms require bilateral incentives offered by the central planner to each participating firm, and therefore require more rigid decision-hierarchies. For example, while the cost structure of each firm is assumed transparent to the entire supply
chain, we assume that the firms are amenable to incentives offered by a central planner. Hence, these coordination mechanisms are not always realistic, and are applicable to only a subset of environments. However, such coordination mechanisms provide some intuition on how to coordinate capacity investments under more general decision-hierarchies, and moreover can be plausible in situations where there is a strong and dominant supply chain agent who can play the role of the central planner.

In the next introductory chapter, we provide a more detailed synopsis of Parts I-III with the objective of identifying the considerable interconnections between the various chapters within the three parts. We also aim to highlight the contributions of the work to various streams of academic literature. Throughout this dissertation, we strive to maintain a dual tone of discussion: One for practitioners and researchers in the field of operations strategy that synthesizes insights on supply chain structure and the crucial elements of collaborative supply chain planning for the sake of managers, and the second theme focusing on more fundamental operations research problems underlying the collaborative planning environment.
To my father, R. Gurumurthi (1941-2009), who was so anxious to see this document, and to my mother, Jayalakshmi, who will no doubt remain anxious even after.
Contents

Abstract iv
List of Tables xxiii
List of Figures xxiv
List of Abbreviations and Symbols xxxi
Acknowledgements xxxvii

1 Introduction: Perspectives on Collaboration at Three Levels 1

1.1 Synopsis of Chapter 2: Supply Chain Collaboration And Our Economic Lives. .................................................. 1

1.2 Synopsis of Chapter 3: Collaborative Planning As A Fundamental Capability Of The Supply Chain. ......................... 9

1.3 Synopsis for Part II: Supply Chain Programs as Vehicle and Platform for Collaboration ............................................. 13

1.3.1 Chapter 4: Functional perspectives of collaborative planning: Innovation and the Project Environment. ................. 13

1.3.2 Chapter 5: Functional perspectives of collaborative planning: Logistics processes. ............................................ 15

1.4 Synopsis for Part III ................................................................. 17

1.4.1 Chapters 6 - 8: Task Assignment and Capacity Decisions in Collaborative Programs for Innovation. ..................... 17

1.4.2 Chapters 9-10: Capacity Planning in Collaborative Logistics Programs. ....................................................... 22
2 Supply Chain Collaboration and Our Economic Lives

2.1 The Global Economic Crisis of 2007-2010 and Supply Chain Strategies
2.1.1 Introduction
2.1.2 Macro-economic outlook and the current economic situation
2.1.3 How have supply chains responded operationally to the crisis
2.1.4 Has an outsourced & vertically disintegrated model of operations staved off a second Great Depression

2.2 Unique Markers of 20th Century Supply Chains
2.2.1 The major technology breakthroughs
2.2.2 What technology has not solved
2.2.3 Supply chain strategies to harness these breakthroughs

2.3 Structural Evolution Of Supply Chains And Collaborative Capabilities In The 20th Century
2.3.1 Supply chain structures for mass manufacturing
2.3.2 The growing importance of the individual consumer and product variety
2.3.3 The multi-national corporation
2.3.4 Air travel as a feasible mode for transport and freight
2.3.5 Telecom deregulation and the Information Age
2.3.6 From multi-national presence to off-shoring and outsourcing
2.3.7 Cross-cultural fusion and convergence of consumers
2.3.8 The process efficiency perspectives of a supply chain
2.3.9 The services revolution
2.3.10 Structural evolution and the missing dimension of 21st century supply chain strategy and collaboration

3 Designing 21st Century Collaborative Planning Frameworks
3.1 Overview and Organization
3.2 What Is Planning? ......................................................... 73
3.3 Planning Approaches For The 20th Century Corporation And Supply Chain. ........................................ 76
  3.3.1 The measurement and division of labor; and management of sequences of tasks. ................................. 76
  3.3.2 Computer aided approaches and the advent of operations research. ..................................................... 77
  3.3.3 Scaling the computer based planning approaches. .................. 79
  3.3.4 Incorporating safety and environment concerns in planning. 80
  3.3.5 Incorporating service tasks and planning frameworks. ..... 82
  3.3.6 Perspectives on the planning function in the Information Age. 84
3.4 Pitfalls And Challenges Arising From 20th Century Approaches. ........................................ 85
  3.4.1 Broad limitations of existing planning approaches. ............... 85
  3.4.2 Potential for ill-defined objectives, responsibilities, and roles. 88
      Hierarchical and cross-functional stake-holder conflict. ....... 88
      Uncertain scope of the planning exercise. .................... 89
      Reinforced bias. ................................................. 90
      Responsibility for outcomes (or lack thereof). ............... 91
      Blurring of boundaries between planning and execution. .... 92
      Procedural versus a creative outlook. ........................ 93
      Planning versus execution authority. ........................ 94
  3.4.3 Lack of standards and uniformity in design, development, and usage. ................................................. 95
      Organic growth of planning systems. .......................... 95
      Disparities in standards within a supply chain. ............... 96
      Rate of innovation in computing and database platforms. .. 97
      Reputation effects on adoption rates. ........................ 97
3.4.4 Corporate versus Supply Chain Focus

Scope limited within firm boundaries.

Lack of visibility into supply chain level information, material, and financial flows.

Lack of flexibility to react to supply chain events.

3.4.5 Competing alternatives.

Automation (procedural) versus manual (creative or responsive) approaches.

Low-technology versus unproven high-technology alternatives.

Specialized legacy applications versus integrative frameworks.

3.4.6 Separate emphasis on manufacturing, services, and on information technologies.

Floundering in the Information Age.

Dealing with human behavior and psychology.

A creative service perspective of the supply chain management function.

Modeling and planning service transactions.

Modeling and planning creative and innovative operations with complex information flows.

3.4.7 Limited collaborative capabilities.

Centralized versus Decentralized Supply Chain Approaches.

Controlled access to firm level data.

Confusing information exchange with collaborative decision-making.

Clout and financial strength determine planning hierarchies.

3.4.8 Limited risk management capabilities.

Plans offer an ex-ante view.

Performance versus flexibility.
Risk is harder to manage across firms. .................... 110
Perverse incentives for bias and distortion. ............. 110

3.5 Expanding The Planning Concept For 21st Century Supply Chains. ... 111

3.5.1 Distributed and decentralized agents and objectives. ....... 112
3.5.2 Dynamic redefinition of activities, resources, agents, and re-

3.5.3 Task specific information sets and agent access. .......... 115
3.5.4 Hierarchies of decision-making authority ............... 117
3.5.5 An incentive framework for agents. .................. 119

3.6 Addressing 20th Century Pitfalls Through The Expanded Planning

3.6.1 Clarifying the objectives, roles, and scope of planning. .... 121
3.6.2 Bridging the manufacturing vs. service divide for the Informa-

A common definition for tasks. ............................ 126
A service oriented and networked organization. .......... 130
Managing knowledge and information processing tasks. .... 132

3.6.3 Building collaborative planning capabilities for the service ori-

Decision-making hierarchies that enable collaboration and co-


4 Collaborative Supply Chain Programs For Innovation ........ 143

4.1 Introduction To Supply Chain Programs For Innovation .......... 143
4.1.1 What is collaborative program management? ............... 143
4.1.2 Collaborative programs as the foundation for supply chains. . 150

4.2 Organizational Structures, Decision Processes, And Their Interactions. 155
4.2.1 Firm roles in program management. ...................... 155
5 Collaborative Supply Chain Programs For Fulfillment

5.1 Overview and Organization

5.2 The Antecedents Of Outsourcing And Inter-Firm Collaboration In Logistics.

5.2.1 Deregulation.

5.2.2 Containerization.

5.2.3 Information and communication technologies.

5.2.4 Realignment of inventory, manufacturing, and logistics management principles.

5.2.5 Specialization and commoditization of logistics services.

5.2.6 Growth in sea and air freight; demand for expedited services.

5.3 A Service (Provider) Based Segmentation Of Logistics Processes.

5.3.1 A capacitated view.

5.3.2 Lead logistics providers.

5.3.3 Systems integrators.

5.3.4 Tiers and pricing of logistics services.
5.4 Value Proposition And Pitfalls In Outsourcing Logistics. 208

5.4.1 The advantages presented by LSPs to lead firms. 209

5.4.2 The risks posed by the outsourcing of logistics. 210

5.5 The Linkage Between Product, Process, And Supply Chain Architectures. 213

5.5.1 Advanced technologies and turnkey programs. 214

5.5.2 Integral products or processes with modular supply chains. 215

5.5.3 Modular products with integral supply chains. 215

5.5.4 Stable technologies and supply chain processes. 217

5.5.5 Logistics in turnkey projects. 217

5.5.6 Modular logistics with integral supply chains. 218

5.5.7 Tightly coupled logistics with collaborative partnerships. 219

5.5.8 Managing modular logistics with collaborative arrangements. 220

5.6 Differentiating Supply Chain Fulfillment Strategies. 221

5.6.1 Inbound versus outbound logistics. 225

5.6.2 Make-to-stock systems. 227

5.6.3 JIT→Push system. 229

5.6.4 Push→Pull system. 231

5.6.5 Build-to-Order system. 233

5.7 Planning Logistics To Support The Supply Chain Fulfillment Strategy. 234

5.7.1 Logistics capacity for make-to-stock systems. 234

5.7.2 Build-to-order systems. 236

5.7.3 JIT→Push strategies. 238

5.7.4 Push→Pull strategies. 240

5.8 Capacity Measurement, Investment, And The Outsourcing Option. 242

5.8.1 Towards a consistent measure of freight capacity. 242
5.8.2  Lead time and schedule based measures of capacity. . . . . . . 245
5.8.3  Combining order size and lead times into a throughput measure. 248
5.8.4  Resource level measures of capacity. . . . . . . . . . . . . . . 249
5.8.5  Inventory storage capacity measures in push systems. . . . . 250
5.8.6  The fundamental trade-offs in logistics capacity planning. . . . 252
      An individual process. . . . . . . . . . . . . . . . . . . . . . . . . . 252
      Serial and parallel processes. . . . . . . . . . . . . . . . . . . . . 254
      Short and long term stability. . . . . . . . . . . . . . . . . . . . . 255
5.8.7  The broader network design problem. . . . . . . . . . . . . . . 255
      Capacity investment constraints and service level trade-offs. . . . 256
      Network operating policies and impact on capacity. . . . . . . . 256
5.8.8  Comparing different capacity cost or pricing structures. . . . 257
      Fixed price and options contracts. . . . . . . . . . . . . . . . . . . 257
      Aggregate capacity contracts. . . . . . . . . . . . . . . . . . . . . 258
5.8.9  Fit between contract type and the fulfillment strategy. . . . . 260
      Push systems. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 263
      Build-to-Order systems. . . . . . . . . . . . . . . . . . . . . . . . . . 263
      Hybrid systems. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 264
5.9   Summary And Further Work. . . . . . . . . . . . . . . . . . . . . . 265

6   Models for Collaborative Programs For Innovation 268
6.1   Modeling literature review. . . . . . . . . . . . . . . . . . . . . . 268
6.2   The Fundamental Elements of Supply Chain Programs . . . . . . 275
       6.2.1  Program Management Models: Tasks. . . . . . . . . . . . . . 275
       6.2.2  Program Management Models: Resources. . . . . . . . . . . 278
       6.2.3  Program Management Models: Firms. . . . . . . . . . . . . . 282
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.4</td>
<td>Program Management Models: Task-Resource Assignment.</td>
<td>283</td>
</tr>
<tr>
<td>6.3</td>
<td>Measures for Task Requirements and Performance</td>
<td>284</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Measuring collaborative requirements across tasks.</td>
<td>284</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Effective processing times for tasks.</td>
<td>288</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Time and delay measures for tasks.</td>
<td>289</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Time and delay measures for programs.</td>
<td>290</td>
</tr>
<tr>
<td>6.3.5</td>
<td>Cost measures.</td>
<td>291</td>
</tr>
<tr>
<td>6.4</td>
<td>The Value of Innovation and Resource Investments</td>
<td>293</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Discounting of program value or revenue functions.</td>
<td>293</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Proportional gain and cost sharing mechanisms for firms and resources.</td>
<td>294</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Impact of risk in the demand and supply variables.</td>
<td>296</td>
</tr>
<tr>
<td>6.5</td>
<td>The Fundamental Trade-offs in Programs for Innovation</td>
<td>299</td>
</tr>
<tr>
<td>6.5.1</td>
<td>Deterministic program environment with processing time-sensitive value.</td>
<td>300</td>
</tr>
<tr>
<td>6.5.2</td>
<td>Deterministic program environment with delay sensitive value.</td>
<td>305</td>
</tr>
<tr>
<td>6.5.3</td>
<td>Task assignment in program environments subject to uncertainty.</td>
<td>312</td>
</tr>
<tr>
<td>6.5.4</td>
<td>Some remarks on globally optimal resource selection strategies.</td>
<td>315</td>
</tr>
<tr>
<td>7</td>
<td>Centralized Capacity Planning For Collaborative Innovation Programs</td>
<td>317</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview and Organization</td>
<td>317</td>
</tr>
<tr>
<td>7.2</td>
<td>Some Convexity And Monotonicity Properties Of Program Performance Measures.</td>
<td>318</td>
</tr>
<tr>
<td>7.3</td>
<td>An Integrated Formulation of the Centralized Assignment and Capacity Problem.</td>
<td>322</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Some remarks on solution algorithms for the central planner's integrated problem.</td>
<td>325</td>
</tr>
</tbody>
</table>
7.4 Centralized, Program Optimal Capacity Investment Given A Fixed Assignment Strategy. .................................................. 329

7.5 Remarks on centralized program optimal assignment in general networks. ................................................................. 332

8 Decentralized Planning & Coordination Mechanisms For Innovation Programs 336

8.1 Overview and Organization .................................................. 336

8.2 Decentralized Resource Level Capacity Decisions For a Fixed Assignment. ................................................................. 338

8.3 The Firm Level Capacity Coordination Problem. ................. 348

8.4 Equilibrium Behavior of Investors in the Decentralized Program Environment. .......................................................... 350

8.4.1 Tasks in series. ............................................................... 351

8.4.2 Tasks in parallel. ............................................................. 354

8.4.3 Illustration of best response functions and decentralized equilibria. ................................................................. 361

8.4.4 Efficiency of equilibrium investments, and equilibria in general program networks. ................................................. 363

8.5 Resource Level Capacity Investment Coordination Mechanisms. .... 372

8.5.1 Why does proportional gain share fail to coordinate? ....... 372

8.5.2 Restructuring the resource level gain share mechanism. ...... 375

8.5.3 Guaranteed revenue or capacity insurance model at the resource level. .............................................................. 376

8.5.4 Non-compliance penalty model at the resource level. ....... 381

8.6 Firm Level Capacity investment coordination mechanisms. ....... 383

8.6.1 Restructuring the firm level gain share mechanism. .......... 383

8.6.2 Guaranteed revenue or capacity insurance model for firms. ... 387

8.6.3 Non-compliance penalty model for firms. ......................... 390
8.7 Conclusions and Summary for Capacity Planning Mechanisms for Collaborative Programs ........................................ 393

9 Planning Capacity within Collaborative Programs for Logistics and Fulfillment .......................................................... 396

9.1 Overview and Organization .................................................. 396

9.2 Model ............................................................................ 397

9.2.1 A Dynamic intermodal logistics or processing network. ..... 397

9.2.2 Orders and their fulfillment strategies. ......................... 400

9.2.3 Illustration of fulfillment strategies. .......................... 402

9.2.4 Revenues and costs in the network. ............................. 404

9.2.5 Shared capacity and collaborative planning framework. .... 407

9.3 Model Literature Review .................................................. 412

9.4 Centralized Planning of a Shared Multi-Commodity Network .......................................................... 417

9.5 Decentralized Capacity Planning ........................................ 421

9.6 Capacity Investment Coordination Mechanisms For A Logistics Network .................................................. 427

9.6.1 Why does proportional revenue sharing fail to coordinate the network? .......................................................... 427

9.6.2 Restructuring the revenue sharing mechanisms. ............ 432

9.6.3 Guaranteed revenue model for firms as a coordination mechanism .......................................................... 433

9.6.4 The risk averse firm and the guaranteed profits model .... 437

9.6.5 Non-compliance penalty model for firms. .................... 437

9.7 Chapter Summary ............................................................ 442

10 Equilibrium Capacity Investment Behavior in Collaborative Logistics Environments ........................................ 447

10.1 Specific Network Structures: Arcs in series .................... 448

10.1.1 $K_{i-} < d$ ................................................................. 449
List of Tables

10.1 Centralized Optimal Capacity Investment in a Make-to-Stock (MTS) System; $r = 15$. Low demand variance: $Pr(d_i = 100) = 0.2; Pr(d_i = 200) = 0.6; Pr(d_i = 300) = 0.2; i = \{1, 2\}$; High demand variance: $Pr(d_i = 100) = \frac{1}{3}; Pr(d_i = 200) = \frac{1}{3}; Pr(d_i = 300) = \frac{1}{3}; i = \{1, 2\}$. 491

10.2 Centralized Optimal Capacity Investment in an MTS System with Higher Revenues; $r = 20$. Low demand variance: $Pr(d_i = 100) = 0.2; Pr(d_i = 200) = 0.6; Pr(d_i = 300) = 0.2; i = \{1, 2\}$; High demand variance: $Pr(d_i = 100) = \frac{1}{3}; Pr(d_i = 200) = \frac{1}{3}; Pr(d_i = 300) = \frac{1}{3}; i = \{1, 2\}$. 492
List of Figures

2.1 Risk factors faced by a stand-alone firm during an economic crisis, with a vertically disintegrated model. 48
2.2 Risk factors faced by a supply chain during an economic crisis, with a vertically disintegrated model. 49
2.3 The foundations of the new strategic imperatives for global firms and supply chains. 59
3.1 Evolution of general planning approaches in the 20th century. 82
3.2 Evolution of select stochastic planning approaches in the 20th century. 83
3.3 Pitfalls of general corporate planning approaches. 86
3.4 A magnified view of the pitfalls of current day planning approaches. 87
3.5 The planning concept for 21st century supply chain environments. 112
3.6 Expanding Planning Concepts to Address 20th Century Pitfalls. 122
3.7 An Illustration of Non-nested Decision-Hierarchies from Distributed Agents. 124
3.8 A service oriented architecture of the organization and supply network. 132
3.9 The central enabling role of information processing tasks in a service oriented architecture. 133
4.1 Highlights of Boeing Program Management Best Practices. 144
4.2 A Process View of a Program Management Framework at IBM Government Services. 145
4.3 Key elements of supply chain programs. 150
4.4 Typical supply chain program structures. 151
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 New product development programs.</td>
<td>152</td>
</tr>
<tr>
<td>4.6 Infrastructure and change programs.</td>
<td>153</td>
</tr>
<tr>
<td>4.7 Global programs.</td>
<td>154</td>
</tr>
<tr>
<td>4.8 Categorizing research issues in supply chain programs.</td>
<td>155</td>
</tr>
<tr>
<td>5.1 Major Developments in the Logistics and Transportation Arena.</td>
<td>183</td>
</tr>
<tr>
<td>5.2 A Classification of Logistics Services.</td>
<td>203</td>
</tr>
<tr>
<td>5.3 Differentiating the Service Providers.</td>
<td>206</td>
</tr>
<tr>
<td>5.4 Value Proposition in Engaging Logistics Service Providers.</td>
<td>211</td>
</tr>
<tr>
<td>5.5 Pitfalls in Outsourcing Logistics.</td>
<td>212</td>
</tr>
<tr>
<td>5.6 Understanding Linkages between Product and Supply Chain Architectures</td>
<td>216</td>
</tr>
<tr>
<td>5.7 Logistics Capacity and Modularity of Processes.</td>
<td>219</td>
</tr>
<tr>
<td>5.8 Order Management to Support Fulfillment Strategies</td>
<td>223</td>
</tr>
<tr>
<td>5.9 A Classification of Fulfillment Strategies.</td>
<td>229</td>
</tr>
<tr>
<td>5.10 Attributes of logistics capacity for various fulfillment strategies</td>
<td>237</td>
</tr>
<tr>
<td>5.11 An Illustrative Approach to Measuring Logistics Process Capacity: A Order Throughput Measure for a Push System</td>
<td>245</td>
</tr>
<tr>
<td>5.12 An Illustrative Approach: A Lead Time Based Capacity Measure for a Pull System</td>
<td>247</td>
</tr>
<tr>
<td>5.13 Fixed versus Options Contracts, Benefits of Aggregation.</td>
<td>262</td>
</tr>
<tr>
<td>6.1 Illustrative tree structure for a program</td>
<td>277</td>
</tr>
<tr>
<td>6.2 Illustrative subsection of a program tree</td>
<td>284</td>
</tr>
<tr>
<td>6.3 Implementing a Decentralized Decision Hierarchy within a Program Environment</td>
<td>299</td>
</tr>
<tr>
<td>6.4 The optimal $K^*_{m}$ with varying $\theta_m$ and $\frac{\lambda_m}{\mu_m}$ in Proposition 6.5.1.</td>
<td>306</td>
</tr>
<tr>
<td>6.5 The optimal processing time $\tau^*_{n}$ with varying $\theta_m$ and $\frac{\lambda_m}{\mu_m}$ in Proposition 6.5.1.</td>
<td>307</td>
</tr>
</tbody>
</table>
6.6 The ratio of the optimal value $V_n^*$ to the optimal value with $\theta_m = 1$; for varying $\theta_m$ and $\lambda_n$ in Proposition 6.5.1. 
6.7 The optimal capacity $\hat{K}_n^*$ for varying $\hat{\tau}_n$ and $\frac{\lambda_n}{\mu_m}$ in Corollary 6.5.6.
6.8 The optimal value $\hat{V}_n^*$ for varying $\hat{\tau}_n$ and $\lambda_n$ in Corollary 6.5.6.

8.1 Example program with two tasks run by two firms in series: $c_1^K = 20, c_2^K = 10; \Pi_0 = 1000; \frac{\beta\lambda_i}{\mu_i} = 100; i = \{1, 2\}$. The firms overinvest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
8.2 Example program with two tasks run by two firms in series: $\Pi_0 = 200; Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25$. The firms under-invest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
8.3 Example program with two tasks run by two firms in parallel: $c_1^K = 20, c_2^K = 10; \Pi_0 = 1000; \frac{\beta\lambda_i}{\mu_i} = 100; i = \{1, 2\}$. The firms overinvest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
8.4 Example program with two tasks run by two firms in parallel: $\Pi_0 = 200; Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25$. The firms under-invest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
8.5 The base series case: $c_1^K = c_2^K = 10; \Pi_0 = 1000; \frac{\beta\lambda_i}{\mu_i} = 100; i = \{1, 2\}$.
8.6 Increased load in series for resource 1: $\frac{\beta\lambda_1}{\mu_1} = 1000$.
8.7 Increased load in series for resource 2: $\frac{\beta\lambda_2}{\mu_2} = 1000$.
8.8 Increased cost in series of resource 1: $c_1^K = 20$.
8.9 Increased cost in series of resource 2: $c_2^K = 20$.
8.10 Impact of lower program gain in series network: $\Pi_0 = 500$.
8.11 Impact of uncertain loads in series network: $Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25$.
8.12 Impact of uncertain loads and lower revenues in series network: $\Pi_0 = 200; Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25$.
8.13 The base parallel case: \(c_1^K = c_2^K = 10; \Pi_0 = 1000; \frac{\lambda_i}{\mu_i} = 100; i = \{1,2\}; \beta = 1\). 367

8.14 Increased penalty for delays in a parallel network: \(\beta = 5\). 368

8.15 Increased load for resource 1 in parallel: \(\frac{\lambda_1}{\mu_1} = 1000\). 368

8.16 Increased load for resource 2 in parallel: \(\frac{\lambda_2}{\mu_2} = 1000\). 369

8.17 Increased cost of resource 1 in parallel: \(c_1^K = 20\). 369

8.18 Increased cost of resource 2 in parallel: \(c_2^K = 20\). 369

8.19 Impact of lower program gain in a parallel network: \(\Pi_0 = 500\). 370

8.20 Impact of uncertain loads in a parallel network: \(Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25\). 370

8.21 Impact of uncertain loads and lower revenues in a parallel network: \(\Pi_0 = 200\). 371

8.22 Example program with two tasks run by two over-investing firms in series: \(c_1^K = 20, c_2^K = 10; \Pi_0 = 1000; \frac{\lambda_1}{\mu_1} = 100; i = \{1,2\}\). The truncated gain sharing mechanism coupled with either the guaranteed capacity investments, or with non-compliance penalties, ensure that the centralized optimal solution is also a decentralized equilibrium. 379

8.23 Example program with two tasks run by two under-investing firms in series: \(\Pi_0 = 200; Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25\). The truncated gain sharing mechanism coupled with guaranteed capacity investments. 380

8.24 Example program with two tasks run by two over-investing firms in parallel: \(c_1^K = 20, c_2^K = 10; \Pi_0 = 1000; \frac{\lambda_1}{\mu_1} = 100; i = \{1,2\}\). The truncated gain sharing mechanism coupled with non-compliance penalties, ensures the firms’ decentralized investments are in equilibrium and identical to the centralized solution. 384

8.25 Example program with two tasks run by two under-investing firms in series: \(\Pi_0 = 200; Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25\). The truncated gain sharing mechanism coupled with guaranteed capacity investments, or with non-compliance penalties. 385
9.1 A dynamic network with 6 locations and 3 epochs per location; arcs connecting epochs within the same location model inventory storage, while those connecting different locations model transport.

9.2 Different fulfillment strategies within a dynamic network.

9.3 Different fulfillment strategies within a dynamic network.

9.4 Example network with two arcs in series operated by independent firms: $c^K_1 = c^K_2 = 1 = c^O_1 = c^O_2 = 1; r = 10; d = 100; \beta = 1$. The firms overinvest w.r.t. to the network optimal capacity investment in arcs 1 & 2 resp.

9.5 Example network with two arcs in parallel operated by two independent firms: $c^K_1 = c^K_2 = 1 = c^O_1 = c^O_2 = 1; r_1 = r_2 = 10; d_1 = 200, d_2 = 100; \beta_1 = \beta_2 = 1$. The firms overinvest w.r.t. to the program optimal capacity investment in arcs 1 & 2 resp.

9.6 Example network with two arcs in series run by two over-investing firms: $c^K_1 = c^K_2 = 1 = c^O_1 = c^O_2 = 1; c^K_1 = 4; r = 10; d = 100; \beta = 1$. The truncated gain sharing mechanism coupled with either the guaranteed capacity investments, or with non-compliance penalties, ensures that the centralized optimal solution is also a decentralized equilibrium.

9.7 Example network with two arcs in parallel run by two over-investing firms: $c^K_1 = c^K_2 = c^O_1 = c^O_2 = 1; r_1 = r_2 = 10; d_1 = 200, d_2 = 100; \beta_1 = \beta_2 = 1$. The truncated gain sharing mechanism can be coupled with either the guaranteed capacity investment, or with non-compliance penalty policies.

10.1 The base series JIT case: $c^K_1 = c^K_2 = c^O_1 = c^O_2 = 1; r = 10; d = 100; \beta = 1$.

10.2 Increased demand in a Series JIT network: $d = 200$.

10.3 Reduced margins in a Series JIT network: $r = 5$.

10.4 Increased shortfall penalty in a Series JIT network: $\beta = 3$.

10.5 Increased cost of resource 1 in the Series JIT network: $c^K_1 = 4$.

10.6 Increased cost of resource 2 in the Series JIT network: $c^K_2 = 4$.

10.7 Increased flow cost on arc 1 in the Series JIT network: $c^O_1 = 4$.

10.8 Increased flow cost of arc 2 in the Series JIT network: $c^O_2 = 4$. 
10.9 Impact of uncertain loads in a Series JIT network: \( Pr(d = 100) = 0.25; Pr(d = 200) = 0.75 \). 462

10.10 Impact of uncertain loads in a Series JIT network: \( Pr(d = 100) = 0.25; Pr(d = 200) = 0.75; r = 5. \) 463

10.11 The base parallel JIT case: \( c_i^K = c_i^O = 1; r_i = 10; d_i = 100; \beta_i = 1; i = \{1, 2\} \). 465

10.12 Increased demand for route 1 in the parallel JIT network: \( d_1 = 200 \). 466

10.13 Increased demand for commodity 2 in the Parallel JIT network: \( d_2 = 200 \). 466

10.14 Reduced margins for commodity 1 in the Parallel JIT network: \( r_1 = 5 \). 467

10.15 Reduced margins for commodity 2 in the Parallel JIT network: \( r_2 = 5 \). 467

10.16 Increased shortfall penalty for commodity 1 in the Parallel JIT network: \( \beta_1 = 3 \). 468

10.17 Increased shortfall penalty for commodity 2 in the Parallel JIT network: \( \beta_2 = 3 \). 468

10.18 Increased cost of resource 1 in the Parallel JIT network: \( c_1^K = 4 \). 468

10.19 Increased cost of resource 2 in the Parallel JIT network: \( c_2^K = 4 \). 469

10.20 Increased flow cost on arc 1 in the Parallel JIT network: \( c_1^O = 4 \). 469

10.21 Increased flow cost of arc 2 in the Parallel JIT network: \( c_2^O = 4 \). 469

10.22 Impact of uncertain loads in a Parallel JIT network: \( Pr(d_1 = 100) = 0.25; Pr(d_1 = 200) = 0.75; Pr(d_2 = 100) = 0.75; Pr(d_2 = 200) = 0.25 \). 470

10.23 Impact of uncertain loads and reduced revenues in a Parallel JIT network: \( Pr(d_1 = 100) = 0.25; Pr(d_1 = 200) = 0.75; Pr(d_2 = 100) = 0.75; Pr(d_2 = 200) = 0.25; r_1 = 5 \). 470

10.24 Base model of Make-to-Stock system. 475

10.25 Make-to-Stock system with higher revenues. 476

10.26 MTS system with more expensive storage capacity. 479

10.27 MTS system with more expensive storage capacity, and higher revenues. 479

10.28 MTS system with higher holding costs. 480

xxix
10.29 MTS system with higher holding costs, and higher revenues. . . . . . 480
10.30 MTS system with higher shortfall penalties. . . . . . . . . . . . . 481
10.31 MTS system with higher shortfall penalties, and higher revenues. . 481
10.32 MTS system with higher transportation (or processing) capacity costs. 482
10.33 MTS system with higher transportation (or processing) capacity costs, and higher revenues. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 482
10.34 MTS system with higher variable (or operational) costs. . . . . . . 483
10.35 MTS system with higher variable (or operational) costs, and higher revenues. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 484
10.36 MTS system with higher variable costs, storage capacity costs. . . . 484
10.37 MTS system with higher variable costs, storage capacity costs, and higher revenues. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 485
10.38 MTS system with higher variable costs and holding costs. . . . . . . 485
10.39 MTS system with higher variable, holding costs, and higher revenues. 486
10.40 MTS system with higher shortfall penalties and capacity storage costs. 486
10.41 MTS system with higher shortfall penalties, capacity storage costs, and revenues. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 487
10.42 MTS system with higher shortfall penalties and variable costs. . . . . 487
10.43 MTS system with higher shortfall penalties, variable costs, and revenues. 488
10.44 MTS system with higher shortfall penalties, variable, and holding costs. 488
10.45 MTS system with higher shortfall penalties, variable, holding costs, and revenues . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 489
List of Abbreviations and Symbols

Symbols used in Chapters 6-8

\( T \)  Set of \( N \) tasks.

\( \lambda_n \)  Work content in task \( n \).

\( \lambda \)  \( N \times 1 \) vector of \( \lambda_n \).

\( F_T \)  Joint distribution of work content in \( T \).

\( F_n(\lambda_n) \)  Marginal distribution of work content in task \( n \).

\( \Psi \)  Matrix defining task relationships: e.g. predecessor or successor.

\( \psi_{kn} \)  Element of \( \Psi \).

\( \tau^s, \tau^f \)  \( N \times 1 \) vectors defining earliest start and latest end times for tasks.

\( \hat{\tau} \)  \( N \times 1 \) Vector defining the available time to complete the tasks.

\( R \)  Set of \( M \) resource pools.

\( K_m \)  Number of units of resource pool \( m \); or the resource pool capacity.

\( K \)  Vector of resource pool capacities.

\( c^K_m \)  Constant marginal capacity sizing cost of resource pool \( m \).

\( c^O_m \)  Constant marginal operating cost for resource pool \( m \).

\( c \)  Marginal capacity cost of a resource.

\( R_n \)  Set of \( M_n \) resource pools dedicated to task \( n \).

\( \hat{X} \)  \( M \times N \) matrix of resource pool capabilities w.r.t to the tasks.

\( \mu_m \)  Processing rate per unit of resource pool \( m \).
\( \mu \) \( M \times 1 \) vector of \( \mu_m \).

\( \theta_m \) Capacity scaling factor for resource pool \( m \); measures collaborative strength within the resource pool.

\( G_m(\mu_m) \) c.d.f. of processing rate \( \mu_m \).

\( G_m^{inv}\left(\frac{1}{\mu_m}\right) \) c.d.f. of processing time \( \frac{1}{\mu_m} \).

\( \tau^o_n \) Minimum processing time for task \( n \).

\( \tau_{mn} \) Total processing time when task \( n \) is assigned to resource pool \( m \).

\( A \) Set of \( L \) firms.

\( \alpha \) \( L \times M \) binary matrix of firm-resource ownership; element \( \alpha_{lm} \).

\( \hat{Y} \) \( L \times N \) matrix of firm capabilities w.r.t to the tasks.

\( \Delta \) \( N \times N \) Symmetric matrix defining collaborative requirements across tasks; elements \( \delta_{nk} \).

\( \mu^e_m \) Effective processing rate of resource \( m \) when assigned to its task \( n \).

\( X \) \( M \times N \) binary matrix specifying the assignment of tasks to resource groups; \( x_{mn} \) are its elements.

\( \tau^e_n \) Effective processing time of task \( n \) under an assignment \( X \).

\( \tau^c_k \) Cumulative processing time to complete all tasks leading up-to and including task \( n \).

\( \tau(X,K) \) Program completion time given assignment \( X \), and capacity \( K \).

\( D_n \) Delay in completion of of task \( n \) under an assignment \( X \).

\( D^c_k \) Cumulative delay in completing all tasks leading up-to and including task \( n \).

\( D(X,K) \) Cumulative program delay given assignment \( X \), and capacity \( K \).

\( c^P_m \) Constant marginal cost of delay for resource group \( m \).

\( C^{OR}_m, C^{OF}_l, C^O \) Total operating costs at the resource, firm, and program levels resp..

\( C^{PR}_m, C^{PF}_l, C^P \) Total penalty costs from delays at the resource, firm, and program levels resp..

xxxii
$C^R_m, C^F_l, C$  Total costs at the resource, firm, and program levels resp..

$\Pi_0$  Maximum gain (or revenue) available to the program.

$\beta$  Processing time or delay sensitivity of program gain.

$\Pi(X, K)$  Program gain given assignment $X$, and capacity $K$.

$\hat{\Pi}(., .)$  Program gain when it is delay sensitive.

$V(., .), \hat{V}(., .)$  Program value when gain is processing time and delay sensitive resp.

$V(., .), \hat{V}(., .)$  Program value when gain is processing time and delay sensitive resp.

$V_l(., .), \hat{V}_l(., .)$  Program value for firm $l$ when gain is processing time and delay sensitive resp.

$\gamma^I_{lRS}$  Investment risk share ratio for firm $l$.

$\gamma^W_{lWRS}$  Work and Investment risk share ratio for firm $l$.

$\gamma^C_{lCRS}$  Comprehensive risk share ratio for firm $l$.

$\gamma^R_m, \gamma^F_l$  Risk share ratio at the level of a resource $m$ or firm $l$, resp..

$\Omega_x(., .)$  Sample space of a random variable $x$.

$E(., .)$  Expectation of a random variable $x$.

$*$  Superscript or subscript denoting an optimal quantity.

$(\omega, \tau)$  Denotes lower and upper bounds of a quantity.

$\Upsilon, \hat{\Upsilon}$  Set of optimal capacity vectors for problem CL, CD resp..

$Z$  The set of feasible assignment matrices $X$.

$\gamma^R_l - L, \gamma^R_l - D$  Decentralized capacity investment problems for a firm $l$, under a particular risk sharing regime.

$m-$  The complement of resource $m$ in set $R$.

$\epsilon$  A small (often positive and real) number.

$\mathbb{R}^2_+$  Non-negative orthant.

$Pr(., .)$  Probability of an event.
Symbols used in Chapters 9-10

\( \mathbf{V} \) \hspace{1em} Set of \( V \) geographical locations.

\( \tau_v \) \hspace{1em} Set of \( T \) time-epochs at location \( v \).

\( \mathbf{N} \) \hspace{1em} Set of nodes obtained by the union of \( \tau_v \).

\( \mathbf{E} \) \hspace{1em} Set of \( M \) arcs (forward in time) connecting the nodes in \( \mathbf{N} \).

\( \Phi \) \hspace{1em} \( N \times M \) node-arc incidence matrix; elements \( \phi_{nm} \in \{-1, 0, +1\} \)

\( \Psi \) \hspace{1em} \( M \times M \) arc-arc adjacency matrix; elements \( \psi_{km} \in \{-1, 0, +1\} \)

\( K_e \) \hspace{1em} Capacity for flow of goods or products along arc \( e \).

\( c_e^K \) \hspace{1em} Constant marginal cost of capacity along arc \( e \).

\( \mathbf{A} \) \hspace{1em} A set of \( L \) firms.

\( \alpha \) \hspace{1em} \( L \times M \) binary matrix of firm-resource ownership; element \( \alpha_{le} \).

\( K_l^e \) \hspace{1em} Capacity contribution of firm \( l \) in arc \( e \); \( K^l \) is the resulting \( M \times 1 \) firm capacity vector.

\( \mathbf{K} \) \hspace{1em} \( M \times 1 \) total capacity vector.

\( \mathbf{Z} \) \hspace{1em} \( L \times M \) firm-arc capability matrix (binary).

\( \mathbf{D} \) \hspace{1em} \( N \times N \) demand matrix; elements \( D_{ij} \) represent demand for flow between nodes \( i,j \).

\( \mathbf{S} \) \hspace{1em} Set of \( S \) demand matrices \( \mathbf{D}^s \) corresponding to discrete scenarios.

\( F_{ij}(\cdot) \) \hspace{1em} Distribution of demand \( D^{ij} \) across nodes \( i,j \).

\( \delta^{ij} \) \hspace{1em} Portion of demand \( D^{ij} \) that goes unfulfilled.

\( r^{ij} \) \hspace{1em} Marginal revenue per unit demand between O-D pair or commodity \( (i,j) \).

\( \gamma_l \) \hspace{1em} Share of revenue (and/or penalties) that goes to firm \( l \).

\( x_e^{ij} \) \hspace{1em} Flow in the network of commodity \( (i,j) \) along arc \( e \).

\( I_j, O_j \) \hspace{1em} Set of incoming and outgoing arcs at node \( j \).

\( x^{ij} \) \hspace{1em} Total demand of type \( (i,j) \) that is satisfied.

\( R_{ij} \) \hspace{1em} Total revenues earned from \( x^{ij} \).
\( \beta_{ij} \)  Constant marginal shortfall penalty for a commodity.

\( c_{ij}^{e} \)  Flow costs for commodity \((i, j)\) along arc \(e\).

\( C_{ij}^{Q} \)  Total flow costs for commodity \((i, j)\).

\( C_{ij}^{P} \)  Total shortfall penalty costs for the commodity \((i, j)\).

\( \Pi_{ij} \)  Total profit earned for the commodity \((i, j)\).

\( B_{l} \)  Capacity investment budget for firm \(l\).

\( \gamma_{l}^{IRS}, \gamma_{l}^{CRS} \)  Investment and comprehensive risk share ratio, resp., for firm \(l\).

\( V_{l}^{g}(D, K, X) \)  Net value (profits minus investment costs) to firm \(l\) under risk share regime \(g \in \{IRS, CRS\}\).

\( \Upsilon \)  Set of optimal capacity vectors for problem CFP.

\( E(.) \)  Expectation of a random variable \(x\).

\( * \)  Superscript or subscript denoting an optimal quantity.

\( (\cdot, \cdot) \)  Denotes lower and upper bounds of a quantity.

Abbreviations

- PERT  Program Evaluation and Review Technique.
- CPM  Critical Path Method.
- JIT  Just-in-Time system.
- MTS  Make-to-Stock system.
- JIT-PUSH  Just-in-Time inbound supply and make-to-stock outbound deliveries.
- JIT-PULL  Just-in-Time inbound supply and on-demand outbound deliveries.
- PUSH-PULL  Inbound inventories and on-demand outbound deliveries.
- IRS  Investment risk sharing.
- WIRS  Work and investment risk sharing.
- CRS  Comprehensive risk sharing.
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIRS</td>
<td>Investment risk sharing.</td>
</tr>
<tr>
<td>TWIRS</td>
<td>Truncated Work and investment risk sharing.</td>
</tr>
<tr>
<td>TCRS</td>
<td>Truncated Comprehensive risk sharing.</td>
</tr>
<tr>
<td>TIRS-G</td>
<td>Investment risk sharing with guaranteed revenues (or capacity insurance).</td>
</tr>
<tr>
<td>TIRS-N</td>
<td>Investment risk sharing with non-compliance penalties.</td>
</tr>
<tr>
<td>CD, CL</td>
<td>Delay sensitive and processing time sensitive value maximization problems.</td>
</tr>
<tr>
<td>MINLP</td>
<td>Mixed integer non-linear program.</td>
</tr>
<tr>
<td>LP</td>
<td>Linear program.</td>
</tr>
<tr>
<td>NLP</td>
<td>Non-Linear program.</td>
</tr>
<tr>
<td>MIP</td>
<td>Mixed-integer linear program.</td>
</tr>
<tr>
<td>OA</td>
<td>Outer Approximation method for solving MINLP.</td>
</tr>
<tr>
<td>GBD</td>
<td>Generalized Benders Decomposition for solving MINLP.</td>
</tr>
<tr>
<td>O-D</td>
<td>Origin-Destination pair in a logistics network.</td>
</tr>
<tr>
<td>CSP</td>
<td>Centralized second stage multi-commodity flow problem.</td>
</tr>
<tr>
<td>CFP</td>
<td>Centralized first stage capacity investment problem.</td>
</tr>
<tr>
<td>DSP</td>
<td>Centralized second stage multi-commodity flow problem in a decentralized capacity investment regime.</td>
</tr>
<tr>
<td>DP</td>
<td>Centralized first stage capacity investment problem for a firm.</td>
</tr>
</tbody>
</table>
Acknowledgements

I thank my advisor Prof. Paul Zipkin, and my committee members, Prof. P. Majumder, Prof. L. Chen, and Prof. Reha Uzsoy, for agreeing to contribute their time and effort to this work. I'm grateful to the Ph.D. program at the Fuqua School of Business for fellowship funding, and to Prof. Zipkin for his research support. I'm grateful to UPS for discussions and support for research, several years ago, on various aspects of the ground and air logistics industry and of its Supply Chain Solutions division. Similarly, I have to thank Prof. Jeannette Song and Prof. Paul Zipkin who introduced me to Kuehne+Nagel: The underlying service and competitive model of lead logistics providers was explained during discussions with Dr. Rod Franklin, Dr. Juergen Rahtz, Mr. Dennis Williams, and many other managers and executives at K+N. The teaching opportunities at the University of Illinois, Urbana-Champaign, the Fuqua School (thanks to Prof. Fernando Bernstein), and previously at MIT Sloan (thanks to Prof. Gabriel Bitran), have helped me explain ideas more clearly than before. I also appreciate the faculty in the Business Administration department of the University of Illinois for their patience, as I worked on this thesis. I have been able to focus on this work in large part due to the support of my family, despite their suffering many years of unconscionable neglect. Without their – and most recently Kalyani’s – help and attention, I most certainly would not have survived my extended academic endeavors.
Introduction: Perspectives on Collaboration at Three Levels

1.1 Synopsis of Chapter 2: Supply Chain Collaboration And Our Economic Lives.

We begin Part I of this dissertation by outlining – in selective fashion – the deeper interconnections between recent macro-economic trends and the prevalent supply chain structures and decision making paradigms. The global economic downturn of the last two years has been swift and deep, sending organizations across a swathe of industries scrambling for new strategies. Many firms are searching for new sources of credit for financing their ventures, and new ways of running their operations that brings cost relief to the bottom-line. It is remarkable though that the recent economic crisis has made just a small dent – at least from what we have seen so far – in the overall trend and picture of how industries organize themselves. Yes, trade volumes are projected to suffer in the next few years, even as demand for goods and services has experienced a shock not witnessed since the 1980s. Admittedly there could even be a trend towards protectionism as governments bail out the institutions that are the engines of their own private sectors. However, extending product
markets regionally or globally is still a critical need for many firms looking beyond the downturn for growth opportunities. The cost-benefit analysis may look different now, but without question, companies and supply chains that are already operating globally are on better footing to emerge from the crisis in good health. This eventful setting allows us the opportunity to examine the need for supply chain collaboration in the context of the ongoing global and macro-economic crisis.

We first propose in Section 2.1 that both the underlying causes of this financial and economic crisis and the responses of firms to these events can be framed within an evolutionary perspective of supply chain (or industrial) organization and structure. It is an accepted evolutionary fact, for example, that the role of the corporation as an engine for the global economy has grown steadily since the industrial revolution. As testament to the enduring work of Peter Drucker, it is the management and stewardship of individual corporations that has finally brought every corner of the globe from out of the dark ages, all in the span of half a century. However, it appears that the reach and impact of the corporation as an economic unit already reached its peak towards the end of the previous century, and that going forward the supply chain or ecosystem is the engine that will be responsible for growth and prosperity during the 21st century. Whether we witness the same scale of growth as in the second half of the previous century, or whether we can sustain such growth far into the future, will depend on the effectiveness of this cohesion of firms that form the supply chain. We propose therefore that collaborative supply chain frameworks, whether formal or informal, will be the cornerstone of economic growth in this century. This simple observation – that could have enormous implications for businesses and societies alike – really motivates this dissertation. We make this statement despite the fact that the work contained within this thesis is far less ambitious, considerably more humble, and quite preliminary relative to the implied and potential scope of research in this
We also examine the structural and strategic responses of corporations – and their extended supply chains – to the unfolding economic crisis. As we parse through these scattered events and attempt to identify broad patterns in the responses of companies to a rapidly receding global economy, we uncover, for example, how the existing supply chain structures and strategies including vertical disintegration and decentralized decision-making may have indeed had a role in limiting the damage and spreading the economic risks across different regions of the globe. Yes, the entire world is caught up in this crisis and the Western world in particular may suffer an extensive phase of limited growth as many forecast, but I argue that such decentralized and disintegrated supply chains indeed may have saved us from another Great Depression, at least for now. Furthermore, this decentralized supply chain model may have set in motion the next great growth phase of the global economy which is centered around the developing countries in Asia and in Latin America. Some have argued that this is a fundamentally more balanced and equitable scenario compared to the currently Western-centric economic model. Nevertheless, this shift in the business landscape from that of dominance of multinational corporations based primarily in the West, to a model of a more equitable partnership of international agents provides the backdrop and motivation of the first several chapters of this dissertation. I therefore propose that while the multinational firm could be governed with corporate agendas and planning frameworks, the international supply chain requires more collaborative and participative decision structures.

The macro-economic successes and the failures are in turn the result of transactions and risks that are witnessed at the supply chain and operational levels, where firms have developed their operational dependence on their supply chain partners.
This is the fundamental link between the macro-economic events we recount, and the more conceptual and theoretical frameworks we present in later chapters. For example, the decentralized supply chain model offers scenarios where multiple firms and corporations combine their resources to produce goods and services. However it also exposes supply chains and the broader economies to risks, not only from the failure of one or more central agents, but also from the failure of firms to act cohesively in response to business problems and opportunities. Thus, while a vertically disintegrated (or leveraged) model has quickly become the norm in many industries based purely on efficiencies, cost, or profit considerations, many corporations are yet to realize that their collaborative capabilities, along with those of their partners, pose possibly the greatest risks to their business models and to the sustainability of their objectives. During this recent economic crisis, the corporations most at risk were dependent on their newly vulnerable financing agents to provide credit for capacity expansion or for inventory replenishment. The remarkable learning from the crisis and the ongoing recovery efforts therefore, is perhaps that it is no longer feasible to govern individual corporations as units independent from the extended supply chain. New frameworks are needed to manage the corporation with heed to the needs of, and the impact on the supply chain or ecosystem.

The current economic crisis has provided first evidence of a “rationalized” role for the corporation in relation to the growing importance of the specific roles played by corporations as agents within a broader supply chain. As we have recently witnessed, the failure of just one or two large corporations – say in the financial sector – can have a deep and devastating impact on the health of the entire global economy, which fact reveals the vulnerability and extent of the dependence of the average business on its key partnerships. It also reveals that firms that play a facilitative role in a supply chain of materials, information, or finances, can have dramatic impact on
the performance of the supply chain. Take the example of a firm such as American International Group, Inc. (AIG), which provides a range of insurance services to product and financial firms. Its services are indeed critical, but nevertheless play a supporting or peripheral to the operations of the vast majority of product and financial firms. During this current economic crisis, we saw how even a single firm such as AIG, through its size and potential failure can really impact global markets and the economy in a major way. Some will see this as evidence of the outsized role of individual firms.

However, I propose that this is really a functional re-alignment of a corporation’s value to its supply chain and to its industry, which has been wrought by evolving supply chain management practices. Corporations such as Lehman Brothers, Merrill Lynch, and their ilk, increasingly leveraged their operations with these risk aggregating agents, and hence entrusted their welfare, and that of their customers, to the continued validity of the risk aggregation model. They also simultaneously minimized their own relevance to their customers and to their value chain, and have therefore since paid the ultimate price in failure and bankruptcy. It is remarkable how similar the story is in other supply chains and industry sectors – one can look perhaps at the automotive or the electronics sectors – with the only difference being that it is the capacity investment that is leveraged by the product firms, while the risk aggregating agents are the manufacturing and transportation suppliers and service providers.

On the other hand, entire regions can become economic powerhouses within a decade because of an emerging cluster of such narrowly defined industrial or service suppliers. This is also an evolution that has contributed to the current malaise affecting the developed countries. As we propose and argue, the rise of the Asian
manufacturing and service cluster as competition to the Western labor force has been a temporarily weakening (or at best neutral) trend from the perspective of the Western economies, with the more immediate crisis precipitated by the operational failures within a cross-section of financial services in Wall Street and in Europe. This is not to say that Western and Japanese multi-national corporations have not benefited immensely in the past, or continue to do so, from the growing strength of far flung markets. However, the Western economies and their leading corporations that spearheaded 20th century growth now seem more vulnerable and at uncertain cross-roads in relation to the prospects for their average Asian counterparts. This argument leads to the idea that this global economic crisis is predominantly a shock re-balancing event, punctuating gradual shifts in comparative advantage coupled with evolving practices in supply and value chain management that have occurred over the last two decades.

This trend of divesting capacity or assets by many large Western corporations can also be traced to its roots in corporate philosophy where the short-term performance of corporate stewards such as executives and governing boards, is rewarded more than the effectiveness of their long term strategy or vision. This translates to less incentives for risky long term capacity investments that can build product and process capabilities, and tends to favor short life cycle innovations using external capacity (or assets) and process capabilities. Since publicly traded firms rely on professional managers who are often not beholden to the corporation beyond their contract duration, long term sustainability of strategy and policy decisions is no longer a feasible performance yardstick for managers. In contrast, it is interesting to observe the success of state-sponsored corporations in China, which have the necessary incentives to make long-term and risky investments in capacity and infrastructure that is now shunned by many Western firms. I propose in Part I of this dissertation that this type
of skewed incentives has led, perhaps unintentionally, to the development of what is now a truly global set of resources for firms and supply chains. The primary challenge for firms in this era – and with prevalent management philosophies – is therefore to harness and plan these global resources with diverse ownership structures, and furthermore, in a collaborative manner that ensures sustainable and efficient operations.

One contribution of Section 2.2 is an attempt to highlight distinct technological factors that have shaped the evolution of supply chain structure over the past few decades. For example, technology considerations (and implied costs) in many cases force a product development effort, or the manufacturing and logistics processes to be shared across many different firms. Better economies of scale with specialized firms, and diseconomies of scope within a firm, amplify both the opportunity costs as well as the risks for firms as they independently plan new ventures, and technology has thus played a major role in currently observed decentralized supply chain structures. We therefore strive to provide a concise outline of how prevalent supply chain structures and associated management strategies evolved from their roots in industrial capitalism as practiced in the previous century. *Section 2.2, while admittedly expansive in its scope, attempts to highlight the key connections between major technological or social developments of the last century, and the current capitalistic and decentralized model of global supply chains.*

In summary, the conclusion of the first two chapters is my view that supply chains across industry sectors need a roadmap that builds on the strengths and weaknesses acquired over the global economic expansion of the last six decades. *We also define four basic strategic imperatives for individual corporations within this new global but decentralized model.* For corporations and supply chains that strive to be at the forefront of their industries, these imperatives and goals could be the difference between
leadership and a more limiting marginal role. The fourth of these fundamental imperatives is the collaborative planning of the resources and activities of the extended supply chain, to reflect the common objectives of such a collection or partnership of firms. This last imperative will doubtless require several simultaneous streams of research and practical frameworks to address in any justifiable measure, but nevertheless provides the necessary motivation for the rest of this dissertation. These collaborative frameworks could be the conduit that can enable the smooth transition for individual corporations as they adapt to their rationalized supply chain roles, and help us avoid more economic shocks of the kind we are witnessing today. *Given this realization, we propose that companies today can deploy, at least in concept, the approaches and frameworks presented here, and expressly for the short run (the next decade or two) to help them work effectively with supply chain partners and emerge from this current global crisis.*

Here we recognize that the global economy is the interaction between individual, corporate, and collective (government) agents matching demand and supply for myriad goods and services. Better collaborative capabilities between individual firms and within cross-sections of supply chains can lead to greater efficiencies and better resiliency to external shocks and global crises. While macro-economists study shifts and trends in regional and global economies, we adopt the view that positive shifts require consistently greater throughput, productivity, and ultimately higher profitability at the level of individual firms and their extended supply chains. We propose, therefore, that sustainable growth and profitability in this era of economic multi-polarity and interconnectedness can only be achieved with the right frameworks for collaboration deployed within extended supply chains.
1.2 Synopsis of Chapter 3: Collaborative Planning As A Fundamental Capability Of The Supply Chain.

To start with, we propose in Chapter 3 that the critical limiting capability for most corporations today is collaboration with their partners in planning the extended supply chain or ecosystem. In that chapter, we show what the implications are of limited collaborative capabilities and inadequate management frameworks to support a collection of firms operating with common objectives and interests. Thus, in Chapter 3, we show that not only are collaborative planning frameworks a cornerstone for future economic growth, they are precisely so because they represent the critical barrier that individual firms must cross in order to be effective as a cohesion of units. Without such collaborative capabilities individual corporations cannot harness the potential of the extended supply chain: Indeed they will be starved of technological and human resources that are now necessarily distributed across the supply chain.

We take pains to emphasize this difference between collaborative capabilities in planning and in execution, where the latter has tended to be an infrastructure issue and one that has achieved great strides with advancing communications capabilities. Collaborative planning on the other hand is a critical managerial challenge, with deep issues to be resolved in concept and in practice. Further, successful planning requires a foundation or infrastructure for allowing execution, and we propose ideas in Chapter 3 for collaborative execution that can work in tandem with our collaborative planning frameworks. The main contribution of this chapter, however, is to outline in specific ways that planning as a corporate function suffers from the lack of collaborative capabilities or collaborative focus. In other words, this chapter helps us understand the pitfalls of applying corporate planning frameworks developed mostly for centralized and rigid hierarchies, to more decentralized environments that require
greater collaborative effort.

Chapter 3 discusses planning without either a functional bias (manufacturing vs. services) or a bias in terms of academic disciplines. Activities and decisions, regardless of their purpose or altitude in the corporate hierarchy, need to be defined and planned according to some logical framework that not only structures the activities, but also assigns them to agents and resources within the organization. In fact, all decisions and plans are subject to conflict for activities that involve multiple stake-holders with conflicting objectives. These conflicts are highlighted in a supply chain environment where the stakes for each firm are multiplied several-fold, but the pitfalls of planning in a supply chain context are equally applicable within the corporate boundary. We identify seven major categories of pitfalls of current day planning frameworks in a decentralized supply chain context. The discussion in this chapter is also a constructive criticism of the specific ways in which planning methods and frameworks have evolved from theory and practice based on how particular firms that have promoted or shaped them in practice, or based on how the academic field has found interest in these problems. Therefore, again here, we feel the need for an evolutionary perspective to arrive at a more meaningful criticism of various facets of popular corporate and resource planning approaches.

In response to this need we outline, albeit in a modest fashion, the underpinnings of such a framework for supply chain planning and collaboration. We rethink the basic constructs of how firms operating within a broader supply chain or ecosystem should go about their fundamental planning processes: that is, understanding market needs, laying out their investment bets, aligning their organizational capacities with market needs, and utilizing their resources effectively, in this 21st century. In particular, we propose a redefinition of corporate planning frameworks along five di-
dimensions that reflect current day supply chain realities: (i) distributed agents and diverse set of intersecting and many times conflicting objectives that are assigned to various roles in the supply chain, (ii) information sets specific to agent-task pairing, and controlled levels of access to such information for the distributed agents, (iii) decision-making hierarchies for subsets of independent agents, to manage task definitions, resource capacities and agent-task pairings, (iv) incentive frameworks that can reduce the friction in the decision-making processes, and finally (v) capabilities for the dynamic redefinition of tasks, agents, resources, information sets, and incentives to better react to supply chain circumstance.

Thus, in Chapter 3 we are describing – in a nutshell – the minimal requirements to be satisfied by supply chain planning as an exercise, in addition to the basic requirements from a more traditional corporate planning approach. The allusion and contrast here is to corporate planning that is rooted in an (enterprise) resource planning paradigm, which involves defining tasks, projects, programs, and processes, within the corporate boundaries. Collaborative supply chain planning, on the other hand, has to involve multiple firms (agents), objectives (interests), the allocation of decision-rights among supply chain agents, different incentives structures for different agents, and the capability to alter tasks, projects, programs, and processes at different corners of the supply chain as a response to evolving and uncertain circumstances. These are the significant hurdles to be crossed in collaborative supply chain planning as a paradigm, so that plans (and the planning process) themselves have sufficient credibility to survive the execution phase which ultimately yields the outcomes, rewards and losses for the supply chain and its agents. The five expanded dimensions to the planning exercise address precisely these limitations of corporate planning paradigms. It is possible to criticize this language of supply chain planning as not much different from the much broader and loosely defined terms such
as *management* or *control*. However, our view is that management is a term that is too broad to define and frame specific planning problems in most 21st century corporations and supply chains. *Management concerns itself with not only planning, but also strategy, execution, performance review, and organizational definition and control.* Our perspectives are limited to the structuring and planning of specific tasks within the supply chain’s operational environment. We therefore do not emphasize the debate and negotiation that is part of strategy formulation in collaborative environments, and also omit any serious discussion of technological and information infrastructure that is required for efficient execution in today’s supply chain environments.

Chapter 3 also aims to motivate multiple streams of academic literature that deals with the planning of various functional processes within the extended supply chain: for e.g. the planning of purchasing, product development, production, logistics, and distribution processes in a supply chain setting. In other words, when we propose a conceptual redefinition or alternative framework for collaborative supply chain planning, we simultaneously perceive the need for functional instances of planning that can demonstrate or validate one or more of the five expanded dimensions we propose: this is therefore the subject matter of Parts II and III.

It is useful to emphasize that Chapter 3 is intended as a guideline for the myriad approaches to modeling supply chain problems, both at a functional, as well as at a more abstract level. We intend the concepts presented there as a litmus test for whether specific modeling approaches would work in an evolving supply chain with diverse agents and objectives. If one or more of the expanded dimensions defined above are not incorporated, we propose that this represents a serious limitation of the modeling approach that would typically limit the applicability to collaborative
environments in some specific ways. In reality, we posit that it would indeed be extremely difficult, if not impossible to model supply chain problems and planning frameworks or solutions functionally in a way that would capture all of the five expanded dimensions. *In fact, the specific models presented in Part III also fail to capture at least one or more of these dimensions in an explicit fashion.* However, as always, there are trade-offs in modeling, between the level of detail, and the clarity of objectives and insights. Still, knowing which of the collaborative dimensions represents a vulnerability is useful in placing important restrictions and conditions on the model solutions and recommendations.

1.3 Synopsis for Part II: Supply Chain Programs as Vehicle and Platform for Collaboration

1.3.1 Chapter 4: Functional perspectives of collaborative planning: Innovation and the Project Environment.

The remaining two parts of the dissertation are devoted to very preliminary development of such collaborative *modeling* frameworks for specific functional settings and processes within a supply chain. However, the broad theme is to cast the collaborative planning and collaborative execution as the responsibility of a new organizational structure that transcends traditional corporate and firm boundaries. This umbrella organization together with its strategic motives (or directives), objectives, and resources is what we define as a *supply chain program*. While in many instances these programs encompass multiple functions such as product development, manufacturing, logistics, and services, in Part II of this dissertation, we deliberately adopt a more narrow functional view of the corporation and the extended supply chain. To clarify the workings of supply chain programs as we have defined them, we develop illustrative frameworks for two core functions that are now perhaps the selective fo-
cus (right or wrong is a debate for social scientists and macro-economists) for many large Western industrial corporations. We analyze product or process development activities within supply chain programs, representing the innovative capabilities of the program organization, and its ability to plan execute business critical projects. Secondly, we analyze processing or logistics activities, representing the execution or fulfillment functions within program – or more traditionally – supply chain environments. These frameworks are presented again at two levels: in Part II, we examine and describe the specific functional context in which collaboration is required and implemented. Somewhat independently, in Part III we present mathematical models of these environments that highlight the critical decision-problems that are at the heart of planning exercises that represent these two business functions.

For example, in Part II, we study capacity investment decisions that are part of development and innovation efforts in a supply chain context; such investment decisions really shape the collaborative efforts that are entailed, as they represent the primary and sometimes irreversible business decisions by partners towards the supply chain objectives. In order to understand such decision-making it is even more important to understand the context and the environment where such decisions must be effected. This is the task attempted by Chapter 4. For an illustration of collaborative frameworks in the project or innovation space, Chapter 4 defines a program in the context of a supply chain as a set of focused, interdependent projects and activities carried out by the partner firms with varying degrees of collaboration. It also describes the key operational elements of a program. We use a case analytic method in this chapter, and drawing from a number of different business environments of current and historical interest, we describe the features of collaborative projects – or supply chain programs – that define the scope and the parameters of a typical investment problem in such joint ventures. In fact, we also outline other operations
research and decision analysis problems that arise because of the peculiar constraints on decision-making in collaborative and decentralized projects. We pay particular attention to how planning in supply chain programs – even now – has features that reflect the five expanded dimensions defined in Part I. Therefore, we discuss the implications of task-agent pairings, information exchange rules and resulting decision hierarchies, the impact of risk in project efforts and outcomes, and finally the kinds of negotiation and incentive schemes that can lubricate the system that requires the continued participation of these collaborating agents.

1.3.2 Chapter 5: Functional perspectives of collaborative planning: Logistics processes.

In Chapter 5 of Part II, we attempt a similar task for the collaborative logistics environment or program. It is important to recognize here that unlike collaborative projects that often have a finite time-line, and where it may be harder to pin-point customer-server type of relationships, the vast majority of logistics environments are collaborative, at least in today’s world, because of choices made by manufacturers or buyers in a supply chain to share their manufacturing, transportation or distribution work (and /or capacity) with external agents. Thus, it is more common to find clear customer-server linkages in the logistics program in contrast to programs with a mission to innovate. The choices are typically business decisions influenced by the cost trade-offs presented by external agencies competing with internal resources or agents to provide the logistics services (as opposed to process technology or intellectual property constraints). As such, to understand collaboration in this environment is to really understand historical and prevalent approaches to the outsourcing of logistics work. The extent of collaboration is of course determined by the process environment, but in terms of business choices, is influenced by the scope of
the outsourcing to these external agents. Furthermore, outsourcing all of the logistics work often implies less collaborative effort, as opposed to a more piece-meal approach.

Whatever is the extent of collaborative effort, in Chapter 5, we study frameworks to manage emerging issues and modes of collaboration in logistics and transportation. We again adopt an evolutionary perspective to analyze how the deregulation, specialization, and the increased availability of third party and contract logistics firms, have influenced many firms to consider outsourcing their logistics needs to these external service providers. As a natural and related evolution, the capacity planning of procurement, production, and distribution activities within a supply chain now has the added dimension of determining the form and precise extent of the outsourcing of logistics requirements. Such third party capacity has to be reserved, incorporated, and utilized in ways that support the overall fulfillment strategy of the supply chain. In the first of a two part paper, we here develop a conceptual foundation for logistics capacity management in general supply chains, with an emphasis on the use of outsourced capacity. For this, we find it necessary to describe the antecedents of outsourcing in today's complex and global supply chains, and to describe the differentiated services that are now on offer to firms as they plan their logistics operations. Our attention to capacity decisions is motivated by the fact that planning collaborative effort requires firms to not only structure their tasks but again to lay out their resource investments and then allocate the resources to the assigned tasks. In this sense, there is a close parallel to the primary collaborative planning exercise in program management. However, it is also important to note that projects and programs require task structuring and assignment for the one-time completion of the project, while logistics processes have to be planned to be repeated indefinitely over a planning horizon or product life cycle. Moreover, the measures of performance in such infinitely repeated processes are different from the measures used to evaluate
projects. Chapter 5 also attempts to define some of the critical performance measures that will define success or failure for the collaborative environment.

Chapter 5 also highlights the pitfalls and opportunities presented by substituting internal assets with external logistics service providers. In particular, we develop some theory around how different product and supply chain architectures have corresponding and differentiated logistics requirements, and where and how firms may be best able to utilize external capacity and services. We also present similar ideas on how order management and fulfillment strategies – the critical operating policy backbone for any firm – can be used as a guiding framework for determining choices with regard to logistics capacity and service providers. We aim to highlight the transfer of capacity costs from a fixed to variable scale, where the variable cost component is provided by flexible options contracts. These are I believe important contributions of this Chapter that while independently important, also fit into the overall narrative of this dissertation that requires such deeper functional perspectives of collaboration and outsourcing.

1.4 Synopsis for Part III

1.4.1 Chapters 6 - 8: Task Assignment and Capacity Decisions in Collaborative Programs for Innovation.

Part III of this dissertation can be viewed as a selective illustration of how decision-making in collaborative environments can reflect one or more of the five expanded dimensions of planning that we propose in Chapter 3. The challenge is that our modeling efforts can incorporate more than one of these dimensions only at the expense of the clarity of insights derived from the modeling exercise. Furthermore, mathematical modeling – with a view to generating insights into systemic behavior –
typically has fundamentally different objectives than a programmatic or managerial framework. *Hence, there is a natural disconnect between the more expansive frameworks of Parts I and II, built on case analysis and historical perspectives, and Part III which relies on an axiomatic development of models to guide specific decision-making in program environments.* Parts I and II are thus examples of managerial and application frameworks that must exhibit complexity and scope in the range of issues discussed, while Part III is a more rigorous exercise that condenses the environment into the fundamental elements, and further analyzes the critical decisions in collaborative planning within the supply chain program setting.

Thus, in Chapters 6-8, I develop a model of collaborative supply chain programs through their fundamental elements. Important instances of collaborative programs include product development or innovation efforts involving multiple firms contributing technology, human, and financial resources. One fundamental element of collaborative programs, for example, is a task or a project that requires one of several heterogeneous resource groups owned and operated by the constituent firms. I assume that a resource pool is dedicated to a given task. Tasks, meanwhile, are assumed divisible to the extent that they can be shared between multiple resources within a candidate pool. I further allow for uncertainties in task work requirements and in supplier performance; and capture both synergies and inefficiencies implied by inter-firm collaboration. I examine first the centralized version of the problem, where a central program planner determines both the assignment of tasks to the candidate resource groups, and simultaneously determines the resource capacities in order to maximize the value of the program investment (or to minimize program costs). Under very general and plausible assumptions regarding the form of the task processing time functions, I develop an integrated mixed-integer non-linear programming model for the central planner’s task assignment and capacity investment problem. I develop
some simple rules for task assignment based on the internal collaborative capabilities within firms, their efficiency, and their marginal capacity costs.

My primary goal, however, is to formulate the associated decentralized version of the capacity problem, but where the central program planner determines the assignment of tasks to the resource groups first. The partner firms, as a response, independently determine their resource capacity investments, assuming prior knowledge of their partners’ cost structures. The decentralized capacity investment problem requires the specification of the gain or revenue sharing mechanism that allocates the program gains to each participating firm. For commonly used proportional gain sharing mechanisms, the decentralized capacity investment problem is shown to be non-convex in general. However, solving smaller scale instances of the decentralized problem helps identify the main advantages and disadvantages of proportional revenue sharing mechanism. These popular revenue sharing mechanisms cause firms to over-commit or under-commit capacity relative to the central planner’s needs. Furthermore, the decentralized equilibrium capacity investments could be vastly different from the centralized planner’s optimal capacity requirements. As a first remedy, I propose some basic modifications to the gain share mechanisms that coordinate - in some limited sense - the firms’ capacity investments, for any particular task assignment strategy. Coordination is defined here as the construction of an equilibrium capacity investment solution that is identical to the central planner’s capacity requirement. Using just bilateral incentive mechanisms between the central planner and each participating firm, I show how these modified gain share mechanisms can induce the partner firms to respond with capacity levels that optimize program level objectives. Not all of the proposed mechanisms are realistic, but they provide intuition on how to coordinate capacity investments in many such collaborative environments.
The main results and contributions of this work above are four-fold:

1. Firstly, our modeling approach is simple, highly flexible and modular: the primary elements of the collaborative project environment are tasks, resources, firms (alternatively, teams), and the market (opportunity). Each of these primary elements have attributes and associated parameters that are able to capture the complexities of the project environment: for example tasks are inter-related and may have variable work content, resource groups may have different capacities, and fundamentally different (or variable) speeds and collaborative capabilities, firms may have different cost structures and project objectives (with common measure such as expected profits), and markets may present different opportunities depending on project outcomes. This modeling framework can therefore accurately describe a vast range of project environments, and the results from the modeling are also therefore widely applicable.

2. Secondly, through our modeling, we demonstrate that the collaborative planning of capacity is fundamentally a difficult exercise for such project environments. In other words, practitioners should be aware that decentralized project environments are guaranteed to exhibit capacity (alternatively, investment) decision conflict and resulting costs (and chaos), unless managers examine the related coordination problem. This is true even when task assignment is centralized or is decoupled from the capacity decision. More importantly, this is true even with commonly used gain or cost sharing mechanisms in place to divide the effort equitably among the partners.

3. Thirdly, we demonstrate the existence of equilibrium capacity solutions to the decentralized capacity planning problem in general tree networks. But we also
show how even these equilibrium solutions, whenever they do exist, are likely to be inefficient as compared to the centralized optimal capacity investments.

4. Fourthly, we show how remarkably simply incentive mechanisms can potentially redress the capacity decision conflict, when the partners also agree to share the gains and losses, and therefore the risks. However, these incentive mechanisms can only work when there is an agreement between the partners to put in place restrictions on the gain sharing mechanism. For example, penalties for under-investment or over-investment, or alternatively, minimum payoff guarantees for the partners are shown to eliminate the inefficiencies in the decentralized system. The penalty system is robust and works in both “good” and “bad” markets, while minimum payoff guarantees only work in profitable markets. However, to achieve coordination, we show that we must limit the costs and gains accrued to any one firm and thus place restrictions on the cost or gain sharing formula. Thus, we show that the decentralized and collaborative environment can indeed be viable, but if and only if incentive mechanisms are put in place to mitigate the decision conflict.

5. Fifth, we demonstrate through our modeling efforts that there are a number of cost and gain sharing formulae that can be applied and that can potentially coordinate the decentralized project environment; in fact it is surprising that seemingly less “equitable” partnership structures can also be coordinated provided additional incentive mechanisms are in effect. Of course, partners are less likely to agree to the terms with weaker risk sharing mechanisms, but technically, we show that a range of risk sharing options are negotiable.

It is useful to point out here that there are several limitations of the supply chain modeling framework we present in these chapters that could be the basis for future work. Most importantly, we could consider programs where decisions regarding task
assignment and capacity investment or re-allocation could be made in a dynamic fashion based on evolving circumstance and outcomes. This would address the “dynamic redefinition” aspect of our planning frameworks in Part I. Secondly, we could incorporate more flexible information architectures whereby firms could have lateral knowledge of competitor cost structures rather than the more rigid schemes we impose for our modeling. Thirdly, we could explore different decision hierarchies where groups of firms have local decision-making power for a subset of tasks, while some others can have greater control over the program decisions. Finally, we could explore decentralized task assignment (or selection) where firms bid for certain tasks based on their capabilities, and where capacity investments and performance guarantees are part of the bidding process. All of these models and decision problems could potentially be constructed within the broader framework we have prescribed in Part III of this dissertation.

1.4.2 Chapters 9-10: Capacity Planning in Collaborative Logistics Programs.

While project-based efforts are becoming increasingly common platforms for collaborating agents within a supply chain, a process environment reveals several more issues in collaborative planning. Processes call for a distinct separation between planning and execution, perhaps more so than in project environments, and therefore the agents responsible for task execution typically are lower down in the decision-hierarchy. Hence while collaboration is designed into the processes, agents representing firms collaborate in execution with less decision authority. Hence the critical decisions impacting performance are really effected in the process planning phase, while collaboration in execution is programmed into the infrastructure or is procedural in nature. Thus while collaboration is observed largely between interacting agents belonging to distinct organizations, the critical decisions are really embedded in the planning process, where firms negotiate the task structure and the procedures
to be handed over to the agents responsible for execution.

It is precisely these environments that motivate the corporate and resource planning frameworks in use today. The broad based structuring and planning of tasks, and major decisions impacting payoffs are effected in centralized fashion by senior level managers (for example process and production planning), while execution is relegated to lower level agents who have process knowledge but whose decision authority is localized. In other words, decisions in the planning phase are where the extent and degree of collaborative effort is defined, and hence collaboration in planning is of paramount importance to the supply chain objectives. This collaboration occurs in different forms including negotiation: such negotiation can be based on exchange of feasible plans, or indeed through the coordination of decentralized decisions (via incentives) with respect to process parameters that have the most impact on individual or supply chain payoffs. Furthermore, recourse actions such as utilization of resources conditional on key planning decisions such as resource capacity, and based on evolving circumstances, are also sometimes best pre-determined collaboratively, or even in a centralized fashion. This takes decision-making away from local agents, and vests greater authority with agents who understand the process at a global level, and hence in theory, leading to more efficient or profitable supply chains. It is this type of process environment and decision hierarchy that we investigate in the concluding chapter of this dissertation.

In terms of contributions, firstly, we show through our modeling exercise how process flows are fundamentally different from project environments when it comes to effecting collaborative decisions such as task assignment and capacity investment. For instance, while projects and processes can both be represented by a network of interconnected tasks, projects typically require the completion of all tasks in the...
network. On the other hand process flows may utilize alternative paths through the network to varying extents. The network capacity would be determined in a collaborative fashion for both decentralized projects and decentralized processes, even as tasks in the network are assigned to one or more agents, but the flows in the network are subject to product demand or other evolving parameters such as costs or congestion in the network which would mean the capacities of different paths are utilized to differing extents.

This type of recourse (in a repetitive process environment) is the purview of critical tactical level decisions again centralized within corporate planning frameworks, but represent a significant challenge in current day collaborative environments. Thus, capacity decisions in collaborative process environments can be viewed as strategic decisions by the supply chain partners, while task assignment can be viewed as a recourse action that occurs at a tactical level. Execution can then be viewed as the performance of the tasks assigned selectively to one or more supply chain agents. In contrast, in collaborative project environments, capacity and task assignment are both strategic decisions, with limited flexibility in terms of recourse actions given the finite time-line of projects (at least relative to indefinitely repeated processes).

The above revelation implies that task assignment and capacity decisions in project environments are really coupled decisions (except perhaps in projects with dynamic allocation of resources and assignment of tasks) which must be effected simultaneously for maximum impact. For processes, on the other hand, task assignment as a decision problem is embedded into the (centralized or decentralized) capacity investment problem, with only firms investing any significant capacity capable of processing tasks assigned to them. Hence, it is possible to decompose the capacity investment and task assignment problems into sequential decisions: in the
second stage, conditional on the capacity decisions, we can determine the assignment of tasks to agents to maximize profits (or minimize costs), and based on the optimal second stage decisions, we can search for the best possible capacity decisions by the firms in the first stage, that will maximize the value of such capacity to the entire network. Alternatively, with a decentralized first stage problem, we can allow firms to independently determine their capacities that will maximize the network value. Thus, the program planning in collaborative environments can be cast into the framework of two-stage stochastic programming models. Limiting the state space in the second stage (via discrete scenarios, for example) can be made to curtail the computational requirements.

In Chapter 9, we construct precisely such a stochastic programming formulation for collaborative logistics programs. Using a narrative parallel to Chapters 6-8, we develop a realistic model of the collaborative program environment consisting of the fundamental elements: tasks, resources, firms, and market demand, and define the corresponding cost and reward structures. We condense the decision-making in such collaborative environments into two inter-related decisions: task (or flow assignment) that is determined in the second stage, and capacity investment in resources that is effected in the first stage. We then examine different decision-hierarchies within our model environment, including centralized and decentralized decision-making for each of the task assignment and the capacity investment decision. After a brief discussion regarding the tractability and feasible application of each of the four decision-hierarchies, we argue that a centralized decision-making paradigm would suit the second stage flow determination (or task assignment) problem better, since decentralized decision-making in the execution phase is really neither feasible nor easily analyzed from a supply chain management perspective. In contrast, the first stage capacity investment problem can be modeled either under both a centralized and
decentralized paradigm, and furthermore both are certainly feasible and applicable within real-world situations.

The main results and contributions of this modeling work are again four-fold:

1. Firstly, our modeling approach is again simple, highly flexible and modular: the primary elements of the collaborative logistics environment are tasks, resources, firms (alternatively, teams), and the market variables (demand or opportunity). We construct a finite-horizon dynamic multi-commodity network flow model and formulate a two-stage stochastic programming model to effect the key decisions in either stages. Tasks in this case are defined as product flows between location-time pairs in the dynamic network. Resources are again divisible, but dedicated to a specific location-time pair, while firms can share the ownership of resources along any arc connecting a location-time pair. Demand in the model occurs as a requirement to flow a possibly uncertain amount of product between one or more locations in the network over time. This modeling framework can therefore accurately describe a vast range of logistics environments, and the results from the modeling are also therefore widely applicable. We demonstrate such flexibility by showing how one could model a wide range of logistics/operating environments including Just-in-Time, Make-to-Stock, Build-to-Order, or Push-Pull systems.

2. Secondly, through our modeling, we again demonstrate that the collaborative planning of capacity is fundamentally a difficult exercise for such project environments. In other words, practitioners should be aware that decentralized project environments are guaranteed to exhibit capacity (alternatively, investment) decision conflict and resulting costs (and chaos), unless managers exam-
ine the related coordination problem. This is true even when task assignment is centralized or is decoupled from the capacity decision as with a two-stage stochastic programming formulation. More importantly, this is true even with commonly used gain or cost sharing mechanisms in place to divide effort, and share the rewards equitably among the partners.

3. Thirdly, we demonstrate the existence of equilibrium capacity solutions to the decentralized capacity planning problem in general tree networks. But we also show how even these equilibrium solutions, whenever they do exist, are likely to be inefficient as compared the centralized optimal capacity investments. We also demonstrate the equilibrium behavior of firms under different logistics strategy regimes such as Just-in-Time, and Make-to-Stock operations.

4. Fourthly, we again show how remarkably simply incentive mechanisms can potentially redress the capacity decision conflict, when the partners also agree to share the gains and losses, and therefore the risks. However, these incentive mechanisms can only work when there is an agreement between the partners to put in place restrictions on the gain sharing mechanism. For example, penalties for under-investment or over-investment, or alternatively, minimum payoff guarantees for the partners are shown to eliminate the inefficiencies in the decentralized system. However, to achieve coordination, we show that we must limit the costs and gains accrued to any one firm and thus place restrictions on the cost or gain sharing formula. Thus, we show that the decentralized and collaborative environment can indeed be viable, but if and only if incentive mechanisms are put in place to mitigate the decision conflict.

5. Finally we demonstrate through our modeling efforts that for specific assumptions regarding how the operating costs are shared in the network, only a subset
of possible cost and gain sharing formulae can be shown to achieve coordinate for the decentralized capacity problem. In particular, gain share mechanisms that share the profits in proportion to total costs including operating (or flow) costs cannot be guaranteed in all cases to coordinate the decentralized capacity decisions. Conversely, less “equitable” partnership structures that only consider the capacity contributions of the individual firms, for example, can be readily shown to coordinate the logistics program. This is in contrast to the results shown in Chapter 8, where a broader set of gain and cost share mechanisms was shown to coordinate the capacity investment decisions across firms.

In this fashion, in the final two chapters of this dissertation, we show how one or more of the five expanded dimensions that we propose in Chapter 3 can be incorporated in planning and the corresponding decision problems in collaborative environments. We show how one could model distributed agents (firms) and (decentralized) objectives within a traditional planning framework for collaborative projects or programs. We show in Chapter 9 how we can allow for the dynamic redefinition of tasks (and their assignment to agents) in a logistics network. Thirdly, we model specific information exchange rules within collaborative project or logistics environments, and discuss how the absence of such information set and exchange rules can really prevent decision-making at any level in either systems. Fourthly, we discuss for either environment centralized and decentralized versions of the key decision-problems as examples of the possible decision-making hierarchies: in Chapter 9 we discuss the feasibility/tractability of the permutations of decision-hierarchies for logistics programs, and show how certain decision-hierarchies can be eliminated from the feasible/tractable set based on simple criteria. Finally, we discuss incentive mechanisms at length through Part III, offering multiple models of cost and risk
sharing, and corresponding models of revenue or gain sharing in both collaborative project and logistics programs. In particular, we show how these incentive mechanisms can mitigate decision conflict and facilitate decentralized decision-making that matches centralized or global optimization, under some simplifying assumptions regarding the information exchange rules and regarding how the firms interact (bilateral vs. multilateral) within the program environment. For bilateral interactions, we show how simple incentive schemes can facilitate coordination of key decisions in the collaborative project or logistics environment.

For the remainder of this dissertation, we proceed by developing each Chapter in stand-alone fashion, while highlighting the interconnections with other chapters and essays only where necessary. This introductory chapter is intended to serve this purpose of developing those interconnections at a higher level where they can be more lucid. We start next with Chapter 2, defining and examining the impact of supply chain collaboration at the broadest possible macroeconomic level. Our objective is to convince the reader of the criticality of collaborative planning frameworks such as the ones we have developed in this dissertation – and however nascent in their maturity – to not only the performance of firms and supply chains, but also in the long run to our economic futures.
Supply Chain Collaboration and Our Present Economic Crossroads

2.1 The Global Economic Crisis of 2007–, and The Limits of Collaborative Planning

2.1.1 Introduction.

In this chapter, we examine the impact of the recent global economic crisis on the structure and organization of supply chains in a range of prominent and important industrial sectors. We also examine the responses by individual firms and their extended supply chains to the rapid unraveling of the financial sector, and to the ensuing demand shocks and altered economic outlook. A chronicling of these events provides insight into both the constraints and opportunities presented by a decentralized supply chain model. We question the extent to which firms have collaborated with their supply chain partners in their responses to the unfolding events, albeit within the narrow context of a few industry sectors. Finally and perhaps most importantly, we argue that both the roots of the ongoing crisis, as well as the ability of the global economy to limit the damage from assuming Depression-like proportions, lie within these peculiar decentralized and leveraged structures firms and supply chains
have adopted in the past two decades.

\hspace{1cm} 2.1.2 Macro-economic outlook and the current economic situation.

The International Monetary Fund (IMF) and the World Bank [175, 176, 177] have recently reported that the global economic output as a whole shrank by 0.6% in 2009 - the first contraction experienced since World War II. The developed economies experienced an average of -3.2% contraction in their output, which underscores the depth and the meaning of the current economic recession. The situation has been somewhat ameliorated by the brisk pace of growth and residual momentum of the emerging Asian economies of China and India. These countries are also expected to resume economic growth quicker and sooner than other large and significant regions. Still these two countries are an anomaly; the rest of the developing world, including Eastern Europe, Russia, and Latin America suffered major shocks to their local and regional economies.

As is commonly understood, the global economic crisis began with a series of heavy losses incurred in the financial sectors of the world economy – starting with major US investment banks laid low by the sub-prime mortgage crisis (see Zipkin [201]). Shocked by the mounting losses resulting from the sub-prime mortgage exposure, the large investment banking institutions stopped lending: not only to each other, but also to corporations and consumers. The banks were motivated to protect their own assets and interests and to minimize lending risks until the crisis blew over. This hesitancy on the part of major banks to lend even at minimum risk levels led to what is known as the credit freeze.

However, firms of every size need financing to sustain their operations and to
invest in future growth opportunities. This credit squeeze therefore resulted in a massive scaling back of short term economic output. Industrial production and employment rates suffered steep declines in the consumer economies of the United States, Western Europe, and Japan. Nervous consumers and industrial buyers alike contributed to a decline in demand for everyday products, services, and industrial goods. This negative feedback cycle prompted increasing unemployment and diminished hopes for any immediate recovery.

Many large and established giants who were performing poorly before the crisis – such as General Motors and Chrysler – filed for bankruptcy, and had no choice but to initiate asset sales or mergers with more solvent firms\(^1\). This same pattern was observed earlier, during the financial meltdown of 2008, with the vulnerable financial institutions. The governments of the major developed and emerging countries saw the depth of crisis and its unfolding impact, and have since announced unprecedented levels of loans and financial aid packages directly to the banks and to major corporations. The scale of the governmental response was directly motivated by fears of another crisis mirroring the Great Depression of the 1930s that had devastated the global economies for the two decades before the end of World War II. A significant majority of influential economists and government policy makers, who had previously thought there was little danger of the world ever encountering another economic shock like Great Depression, were now convinced that without the intervention of central banks, the existing financial and industrial structures would face imminent collapse.

Since the leading governments announced a concerted global intervention, many

large and prominent institutions\(^2\) have been bailed out by central banks and governments through the direct injection of debt and purchase of loss-making assets. While it is unclear whether these financial and industrial institutions were indeed capable of triggering another Great Depression just from their short term failures, much of the government intervention was also motivated by the need to calm down the stock market and the investing community. In support of the latter argument, there is evidence that the intervention strategy has worked in the short term, although at considerable cost to taxpayers and risk to the long term fiscal health of the developed countries. Indeed, with central bank interest rates close to zero, many of the large financial institutions are able to borrow from the central banks and lend to corporations and other banks with minimal risks and generate short term profits to bolster their balance sheets. Some of the large industrial corporations supported by government debt are also returning to profitability.

One serious issue remains largely unaddressed: countless smaller and medium sized firms have not been as lucky in securing either government debt or the support of their larger institutional partners. Statistics are still scarce as to the impact of the ongoing global economic crises on the smaller entities in various industry sectors, but the situation does not look positive for a vast majority of these smaller concerns. The fact that the surviving large and small financial institutions were able to reverse some of the losses in their balance sheets means they are in a good position to start lending to small and medium sized businesses and to individual consumers again. However, it is reported that in many cases they are still risk averse when compared to the unbridled years of growth before the crisis\(^3\). Consequently, many of these

\(^2\) Also called the “too big to fail” firms and institutions.

smaller businesses have had to look for new sources of lending – which happen in some cases to be their own bigger and more solvent supply chain partners.

Overall, on a broader scale, investors have been reassured to a certain degree of the stability of the world economy (at least assured that another Great Depression may have been averted) and have assumed cautiously optimistic positions, leading to a partial revival in the stock markets. A series of small but positive events – mostly to do with how well countries like China and India have weathered the crisis – have injected confidence into the average investor and consumer in the developed world; and they have increased their spending on basic and luxury goods and services. Warning signs still abound at every cautious up-turn, and vulnerable debt-laden economies in Europe such as Greece and Spain have dampened some of the optimism of a sustained recovery\(^4\). The spending incurred by the major governments to limit the damage from the financial and economic crises has also raised new fears of fiscal imbalance; and the specter looms of years of reduced spending, cuts in services, and high rates of unemployment, especially in the developed world\(^5\). In summary, a complete global recovery marked by robust all-round and reinforcing cycle of growth in both labor markets (that typically boosts consumer spending), and in industrial performance (that boosts profits, trade, and therefore the global labor markets) is thought by many prominent economists to be years away. The growth, according to some economists, may be jeopardized directly by fiscal imbalances; and according to some other schools of thought, indirectly, through reduced fiscal spending in response to such fears.


All said, this is the new economic landscape that firms and their supply chains have to adapt to in the short term. In the longer term, this crisis presents us with several opportunities to examine the strengths and failings of the current model of industrial organization and of existing supply chain structures. Are they robust enough to withstand another crisis of this magnitude in the future? Have the recent trends of outsourcing and globalization been an asset or a liability for the small and large players in such supply chains? Has the risk sharing been equitable, or have there been winners and losers based on how these supply chains have evolved or structured? Operationally, what has this crisis taught us in terms of the collaborative capabilities of firms in the bed-rock supply chains that support regional and the global economies? What does this crisis tell us of the major chinks in their armor: is it their planning capabilities or operational efficiencies or is it indeed their financial muscle and credit-worthiness that determines short term survival and long term success? These are some questions we pose and attempt to answer in a brief essay (spanning the next two chapters). Most definitely, we pose many more questions than we are able to answer, but it is my belief that these basic questions are as important to our future as are the answers – which may be a long time coming.

2.1.3 How have supply chains responded operationally to the crisis?

While we have described the ongoing global economic crisis in macro-economic terms, this dissertation is more concerned with the fundamental change we have witnessed in the structure of most businesses over the past two decades, and that we continue to observe even in today’s turbulent economic environment. It is largely an unquestioned thesis that the financial and economic crisis of the past few years has been remarkably global and interlinked in its causes, reach and its impact. In these next two chapters, it is my parallel thesis that the resulting impact on individual busi-
nesses, their performance, and their long term health also is also interlinked, where these linkages are best viewed in the context of their extended – and sometimes global – supply chains. Here, it is therefore my intent to connect the short and long term operational and supply chain dynamics, both within and across industries, to the macro-economic events of the past two years. Considering the reverse causality, I also argue that an operational perspective provides clues as to how these supply chains will evolve going forward, thereby influencing broader trends in regional and global economic outcomes.

As demand for products and services slowed, large and small firms alike have had to deal with the problem of over-capacity and excess inventory. For example, in the automotive industry, General Motors and Chrysler together canceled contracts with or closed thousands of dealership franchises in the United States alone. In some cases these relationships were decades in the making. Much of this capacity correction was of course, motivated by the steep decline in consumer demand that followed a rapidly shrinking labor force in the Western countries. In this particular case, it was also intended to restore profitability for these large firms over the long run: having fewer retail channels in a geographical area, it is hoped, would minimize competition among retailers and boost prices and margins for the manufacturers.

The capacity correction also came in the form of isolated shutdowns or sale of inefficient plants, or via the redistribution of productive capacity. The redistribution of productive capacity has for example included the elimination of entire supply chains supporting products with weakening demand, and consolidation of the residual ca-

---


pacity. Let us consider the automobile, electronics, pharmaceuticals, and retailing industries and illustrate these trends and phenomena.

Automobile demand had declined by more than 40% during the first quarter of 2009. The response by many firms has followed similar patterns. GM and Chrysler announced major product line and network restructuring plans to qualify for government loans intended to support their dwindling cash flows. GM also announced the phasing out of its long standing Pontiac, Hummer, and Saturn brands, effectively shutting down supplier, plants and dealership operations associated with these brands. Similarly, Ford recently announced the phasing out of its popular Mercury brand. Even Toyota Motors posted its first loss in over 70 years, as they (along with other foreign manufacturers) faced an inventory glut at US ports from rapidly declining demand for high end imports.

These plant shutdowns have been pervasive in the automotive sector (and more broadly in the manufacturing sector) for quite some time, especially in higher cost regions of the world, and the economic crisis has really just hastened these trends. Perhaps as a coup de grâce, the last surviving major automotive assembly plant (NUMMI) in California was recently shut down. Ironically this long standing General Motors plant represented a joint venture with Toyota started in the 1980s to revitalize American manufacturing prowess; the plant has since been purchased for a nominal amount by a much smaller electric automotive firm. NUMMI was sold by


11 Frank Russell, Tesla paying $42 million for Fremont’s NUMMI plant, May 27 2009, San Jose
Toyota citing unsustainable costs, with General Motors having abandoned the unit during its bankruptcy proceedings.

Beyond such capacity rationalization, for many supply chains, this economic crisis presented the first opportunity to react in a cohesive fashion, both through structural changes and via operational policy changes, to disruptive events of large magnitude. However, following the storyline of the few examples cited above, it appears at first blush that this opportunity to test cohesive and collaborative decision-making has gone unrealized, with the vast majority of firms relying on unilateral and independent decision making in response to the crisis.

This type of decentralized and unilateral behavior was most evident in the automotive sector. The large product or brand owning firms are in fact at the head of a vast supply base – spanning multiple continents – that provides materials, components, and a wide range of services to the industry as a whole. So, while much of the focus was on sustaining and bailing out the large assemblers, a great number of auto suppliers across the world have suffered not only from weakened demand, but also due to inadequate financing options. This is arguably due to the lack of requisite attention from central bankers and the economic stimuli that would have mitigated the fallout. At the other end of the spectrum, as the cash flows of the major OEM assemblers were impacted, suppliers faced ever longer windows of payment for their services, even as the credit squeeze limited their ability to boost working capital using external sources\textsuperscript{12}. Reports estimate that over 40 major US auto suppliers filed for bankruptcy in 2008, while doubts were being raised even about well-known Tier 1 companies such as Lear Corp., Visteon, and American Axle; the latter were at

\textsuperscript{12} Joann Muller, \textit{Detroit’s Other Crisis}, Forbes Magazine, March 8 2009.
risk of being de-listed from the New York Stock Exchange. Some suppliers became
dependent on the financial support of their big OEM customers, which were in turn
reliant on government relief funds for support.

It seems apparent that there was no clear response at the supply chain or in-
dustry level to the crisis. This was remarkable since the supply relationships in
the automotive industry, while often adversarial at the operational levels, have been
quite close for the past several decades within the leadership rungs. In support of
its key suppliers, GM indeed asked for special loans from the US government to pay
its suppliers on time and to ensure continued supply. However some of the hardest
hit firms were the small and medium sized suppliers, and the supplier bailout was
minuscule compared to the funds received by the major OEM firms such as GM and
Chrysler\textsuperscript{13}. Toyota Motors on its part, while not facing imminent collapse like GM
and Chrysler, had even begun to develop contingency plans to deal with key US
suppliers going out of business and causing supply disruptions\textsuperscript{14}.

In electronics manufacturing, it is estimated that outsourcing revenues fell by
11\% (to approximately $270B) just in 2009. The affected group includes companies
worldwide that provide design and manufacturing services (EMS/ODM) to brands
such as Apple, HP, Xerox and many other leading firms\textsuperscript{15}. The leading EMS/ODM
firms include Flextronics, Foxconn, and Jabil Circuits. Even Dell Computers, which
has been known for its refusal to outsource its assembly operations, has found it
harder to sustain its assets in this climate. By 2009, the company had announced

\textsuperscript{13} See David Shepardson and Gordon Trowbridge, \textit{Aid to suppliers signals support of auto industry: $5B assures firms to get paid for parts}, Detroit News, March 20 2009.

\textsuperscript{14} Joann Muller, \textit{Toyota Worried About Some U.S. Parts Suppliers}, Reuters, March 11 2009.

plans to shut down or sell many of its plants in the US and in Ireland to third party manufacturers\textsuperscript{16}.

There is also the silver lining in the personal computing and communications segments, as the vast majority of the world is still an untapped market for consumer products and applications. As a result, major computer makers, chip makers, and software developers have been quite successful in navigating the downturn. There is also hope that government stimulus spending on health-care, education, environmentally conscious technologies and logistics infrastructure will sustain and grow the pace of innovation in this industry\textsuperscript{17}. However, there are some troubling signs for the future: at the lowest tiers, the cost of wafer fabrication (lithography) processes in the computer processor chip industry has grown about fifteen fold in the last decade and a half, mainly to match rapidly changing and ever more demanding requirements of personal computing and communications applications. More broadly, and in comparison, the semiconductor industry has seen overall fabrication costs increase by 25\% since the early 1980s. The precision, performance, and miniaturization model of innovation has meant that this industry requires huge capital outlays for innovation. Without such investments at the lowest tiers, the prospects for spurring future value-adding growth are diminished in a diverse array of electronics applications. The credit squeeze and the scale of the losses in the financial sector have hit short term investments in technology, and this could have major implications for the electronics industry as a whole in the years to come.

The pharmaceutical industry has also witnessed considerable changes over the


\textsuperscript{17} Brian Fuller, \textit{Amid Economic Gloom, Semi Execs Chart Encouraging Future}, Electronics Design Strategy News, March 12 2009.
past two years in response to the economic climate. There were three major mergers and acquisitions among the top 20 firms in this industry within the first few weeks of 2009 alone: Pfizer acquired Wyeth in a record merger worth $68B\(^\text{18}\); followed by Merck’s reverse merger with Schering-Plough Corp valued at $41B\(^\text{19}\); and then swiftly by Switzerland based Roche Holding acquiring San Francisco based Genentech for $47B\(^\text{20}\). It is reported that these mergers and acquisitions were possible mainly because of government loans and credit lines to investment banks starting late 2008, as part of the economic stimulus. Still these mergers were affected by the credit crunch; so much so that they are reported to have come up with new approaches to financing share purchases— including selling corporate bonds directly to their current investors. Some experts argue that the gloomy economic forecast has been a driver for these mega-mergers, since the large pharmaceutical firms foresee expiring drug patents and fewer drugs in the pipeline in the coming few years; this coupled with increasing development costs of drugs and a more stringent regulatory environment\(^\text{21}\).

Another sector that witnessed significant supply chain or industry level changes was consumer retailing. By the first half of 2009, many big name retailers across a wide range of markets had announced bankruptcies, store closings, or scale-backs. Some of the names include Office Depot, Circuit City, Sprint/Nextel, Dell, GAP, Disney, KB Toys, Linens’n’Things, and CompUSA. A total of 27 large US retail chains had filed for bankruptcy just in year 2008. The impact of such store closings has been quite significant for manufacturers, wholesalers, and merchandisers.


(importers), along with massive employment losses at these companies. The employment losses in the US were especially painful, because mass retailing had become primarily a service business, with most of the manufacturing outsourced to low-cost countries, and much of the logistics outsourced to third-party logistics providers. It was estimated that a staggering 148,000 retail establishments would shut down in the US in 2009. It is perhaps not a coincidence that Chinese exports witnessed a steep decline in 2009, since imports from China for retail in the United States had been growing dramatically until 2008. In response, some large and dominant retail importers such as Wal-Mart and Kohl’s (see Gereffi [83]) have offered financing options to their key suppliers to ensure supply chain stability and continued operations.

There are emerging signals, however, that consumer spending may be recovering, and that could mean good things for retailers, their employees, and their supply chains.

There are many other factors though, that indicate that this global recession is unlike any other we have witnessed in the recent past, and that both consumers and business could take a long time to return to more stable consumption and production. This grim forecast creates the motivation for us to understand supply chain management methods specifically to address the challenges faced by firms as they try to survive the downturn. Going beyond, it is our belief that there are lessons in the way supply chains have been managed over the past few decades that this crisis allows us an opportunity to examine and build upon for better times. In the next few sections, I outline some of the successful planning approaches of the previous century, and comment on their effectiveness and suitability looking forward into the future. This will serve as the basis for the roadmap for supply chain planners and managers I propose in subsequent sections. First, however, I comment on the question of whether the great outsourcing trends and in general leveraged model of the

---

supply chain of the past decade has been a boon or a bane for the global economy. I argue that such trends have largely been a positive factor for the global economy, and that they may have forestalled an even deeper crisis.

At the same time, I propose that these trends have re-distributed productive capacity on a massive scale from wealthy and high-cost countries to the lower-cost economic regions. However, because this shift has occurred in a relatively short time period of a little more than a decade, it may have provided the seeds for the economic meltdown, by constraining the consumer economies of the United States and (to a lesser extent) Western Europe. Middle class consumers in the United States, primarily, have seen their incomes shrink in comparison to the top brackets. Meanwhile, corporate profits lay disproportionately in financial services. In the year leading up to the financial meltdown, approximately 40% of all corporate profits reported in the US were from financial services firms\(^{23}\). The housing market, which represented the predominant expenditure category for the middle income segments, while at the same time serving as well as a major economic and financial engine of the past decade, collapsed. Middle income housing – in turn being the foundation of the housing market – served as the catalyst for collapsing housing prices and fomented the mortgage default crisis, which, as most economists and observers agree, triggered the financial meltdown.

Independent of the underlying causes, the financial and economic crisis of 2007-2010 (and possibly beyond) seems to have been a missed opportunity for supply chain structures and policy making. We do not see a cohesive and coordinated approach in major industries to crisis management. Such opportunities were most evident in response strategies including inventory liquidation, capacity rationalization, and

operational financing to ensure stable operations. Nor have seen a coordinated approach to recovery, that incorporates the interests and concerns of the supply chain stake-holders on an equal footing with those of the firm’s shareholders. The predominant theme appears to have been the survival of individual firms, rather than continuity of products and services for either suppliers and customers, or the well-being of the supply or distribution channels.

This behavior of firms is not surprising since in the prevailing capitalistic structures, firms consider themselves responsible first to the shareholders, and then to their customers and to their partners. It is also perhaps too early to draw any conclusions on this decentralized approach to crisis management, but these events are likely to have a long lasting impact on supply chain partnerships and relationships. On the one hand, it is possible to question how some firms can claim leadership roles in the supply chain when the best interests of the coalitions are secondary to their own (shareholder) interests. On the other hand, as we argue later, it may indeed be that these decentralized and independent decision-structures are what allowed many firms to survive the toughest crisis they have faced in their operating histories. For the long run, such decentralized or even unilateral decision-making may have allowed impacted firms to recover and to emerge with new opportunities – not only for themselves, but also for their supply chain partners. The jury, figuratively speaking, will be out for quite some time to come, as academics and industry leaders take stock of these events, reactions, and impacts.
2.1.4 Has an outsourced & vertically disintegrated model of operations staved off a second Great Depression?

In the last two decades, many large industrial firms had moved to a variable cost, leveraged form of capacity development and production: including companies like GM, Ford, Chrysler in the automotive sector; HP, Apple, and Microsoft in the electronics sector; and recently Boeing and Airbus taking outsourcing to greater levels in the aerospace sector. On the other hand, some major Japanese conglomerates like Toshiba, NEC, and Mitsubishi have been reluctant to follow this leveraged model of operations. As a result of this outsourcing by the larger corporations, there has been significant growth and consolidation in the lower tiers of many supply chains, as countries in Asia and Latin America became the manufacturing base for the rest of the world’s supply chains. Tier 1 firms in the automotive sector such as Visteon and Magna, Flextronics and FoxConn in the electronics sector, and Li and Fung in the apparel industry have all grown dramatically as they captured more of the manufacturing and logistics business of the larger firms (Bitran et al., [16]).

It has therefore been interesting to observe how this leveraged form of operations has impacted the larger OEMs during this downturn. In the automotive sector, the Tier 1 firms such as Visteon and Delphi were in trouble even before the economic crisis. Delphi, spun off from GM in 2000 to become the largest US based auto-parts supplier, did not take long to suffer from the weight of residual liabilities from its GM years and filed for bankruptcy in 2005. It recently emerged from bankruptcy in 200924. The recent economic crisis finally downed Visteon which also filed for bankruptcy in 200925. These spin-offs had always been vulnerable because of their

24 Soyoung Kim and David Bailey, *Delphi exits bankruptcy after four years: Delphi completes sale of assets to lenders, GM*; Reuters, October 6 2009.

long term liabilities, high costs, and legacy labor contracts. So, one can argue that without divesting these assets, the parent OEMs would have been in even worse shape with the demand shocks of the recent economic downturn.

In the electronics sector, the larger firms such as HP, Apple, and Microsoft have been relatively stable during the economic crisis. A case in point is the consumer products firm Apple Computers, which having completely divested itself of its manufacturing assets has been remarkably nimble and has grown considerably over the last few years, with the downturn barely making a dent in its yearly growth patterns\(^{26}\). Similarly, other bell-weather firms such as HP and Microsoft have stayed profitable through the downturn. At the other end of the supply chain, the Tier 1 firms such as Flextronics have diversified their portfolios within the electronics industry to the extent that while revenues have fallen significantly as a result of a sharp decline in manufacturing orders, they are still viable and poised to make a strong recovery along with the broader economy. There has also been considerable consolidation at the level of Tier 1 suppliers, including the merger between Flextronics and Solectron, the two largest firms after Foxconn.

This kind of impact on upstream firms is not surprising. Margins are lower for the manufacturing processes, and the brand owners face declining demand will reduce manufacturing and service orders for the Tier 1 firms, who will in turn issue lowered forecasts to their own lower-tier suppliers. Since South-East Asia (and China in particular) has become the preferred base for electronics manufacturing, this cascading decline in manufacturing orders from the Western countries, and capacity reductions in South-East Asia, are partly responsible for steep drop in international trade and shipments in the year 2009. However, as evidence that this leveraged structure of

supply chain has protected some large OEMs from demand shocks, it is possible to observe that the Japanese electronics industry suffered massive losses at the trough of the recent downturn\textsuperscript{27}. This was in part because of their reliance on extensive manufacturing and production assets which they used to export products to the major world economies. The bulk of these losses came from the electronics giants such as Hitachi, Panasonic, Sony, Toshiba, Fujitsu, NEC, and Sharp.

For broader industry dynamics, as with all business cycles, we can expect over-correction of capacity to a certain degree; however in this case, the impact of such over-correction has been harder to predict. So far China has avoided a decline in its economic output, and so has India, and both are poised to resume close to double digit growth rates in year 2010 and beyond. Some economists believe that the aggressive fiscal stimulus enacted in China and the explosive consumer demand in the local economy have combined to stave off a crisis on the scale experienced by the Western countries. Similarly, India, with its conservative banking sector, and its large middle class consumer segment, has also minimized the local impact of the global crisis. However, some economists still express concern whether China and India’s growth rates are indeed sustainable\textsuperscript{28}.

Indeed, while the situation is dire, expecting an economic recovery at some point in the future is not too optimistic. In fact many of the firms mentioned above as being in trouble have already shown signs of pulling out of the slump. This includes the worst affected firms such as General Motors and Chrysler. By deploying mechanisms such as mergers, acquisitions, and sell-offs, many of these firms have re-

\textsuperscript{27} Junko Yoshida, \textit{Japan’s electronics industry shocked by $22 billion in losses}, EE Times, February 3 2009.

\textsuperscript{28} Shri Navrathnam, \textit{Asian Shares Lower; China Economy Concerns Weigh}, The Wall Street Journal, May 31 2010.
positioned themselves in their supply chains in order to weather the crisis, and have systematically planned to emerge from the crisis with a strong recovery. Examples include the afore-mentioned product line changes by General Motors and Ford, both of which have improved their performance in recent months; the Japanese electronics sector which through joint ventures and consolidation has once again returned to the growth track\textsuperscript{29}.

Thus, it is possible to observe both positive and negative risk factors arising from the choice of most Western corporations to adopt a vertically disintegrated or outsourced model of supply chain operations; see Figures 2.1.4-2.1.4 for a summary of these risk factors. At the firm level, they have seen cost efficiencies and lower investment risks, which has kept up the pace of innovation to compete with global players. The resulting higher profitability and cash flows of the boom years may

have sustained them through the worst of the current crisis. This strategy has also allowed the most vulnerable stages of the supply chain to diversify their process and product portfolios, and also with mass retailing, allowed these capital intensive manufacturers to mitigate their inventory related risks. This decoupling of inventory and manufacturing (capacity) risks and their dispersal up and down the supply chain appears to have played a critical role in allowing many large firms to survive a downturn of this magnitude.

Conversely, some corporations have lost the sustenance of their erstwhile benefactors and senior supply chain partners, and as a consequence they now have inadequate financing capabilities. The credit squeeze has hurt them the most. Operating cash flow constraints due to inflexible or incongruent contracts with suppliers and customers (also legacy liabilities) have led many firms to bankruptcy. It is possible that vertically integrated firms would have better been able to weather such crises.
On the supply chain level, the investment and capacity risk pooling that the independence of major manufacturers now provides, has prevented the capacity related liabilities of the major brand owners from overwhelming their balance sheets. It can certainly be argued that Ford, for example, has weathered the downturn (without government bailouts), at the expense of its former parts unit Visteon which has faced major financial troubles after their spin-off. However, the fact that Ford, and to a lesser degree General Motors and Chrysler, have managed to survive the crisis implies continuity of their global supply chains that rely on demand channeling from these OEMs. This continuity may have prevented a large scale shutdown of operations across and deep into the supply chains, and may have therefore saved millions of jobs worldwide, and entire regional economies that depend on such jobs.

Further, in a vertically disintegrated setting, retailers and brand owners have the flexibility to indeed cancel orders in the supply pipeline, and to issue lower supply forecasts without immediate capacity utilization concerns. In parallel, retailers also have the flexibility (at least in some cases) to liquidate inventory through price reductions, although resale price maintenance contracts with manufacturers can limit this flexibility. The US consumer price indices during 2009 actually fell for the first time since 1955 pointing perhaps to such increased pricing flexibility. This flexibility to respond to demand shocks may also have allowed these retail firms to shore up their balance sheets and continue their distribution operations that are vital to manufacturers in the long run. This type of crisis would have had a much more damaging and longer lasting impact – comparable to the Great Depression in scope, if not in depth – without the major shift in the last two decades by the large corporations to a vertically disintegrated model of supply chain.

30 See the CPI tables provided at www.bls.gov/cpi/#tables
On the last point, however, this increased order management flexibility leads to nervous buyers and retailers issuing drastic reductions in their order forecasts as they look to minimize their risk and deplete their current inventories before issuing replenishment orders. This phenomenon is of course the “bull-whip” or order variability effect (Lee et al. [126]). As is proposed in the literature, the bull-whip effect presents a real threat to supply chain cost control with demand uncertainty. Firms issue biased forecasts and orders in order to compensate for the uncertainty in the near term, and are therefore stuck with inventory related costs without adequate visibility up and down the supply chain. Following some of these lessons, firms in some supply chains today are better connected through information and communication systems, and can communicate demand and forecast information much faster and more accurately than was possible just a few years ago.

While this greater visibility allows for the mitigation of the bullwhip effect by minimizing inventory related risks and costs, demand shocks of the magnitude we have just witnessed (especially collapsing markets) can cause havoc upstream in the supply chain when firms issue sustained lowered forecasts. This is exactly what has happened in some industries. Retailers have responded to the crisis by altering their forecasts of orders well into the future, and this has led manufacturers to scale back on their production by eliminating capacity, rather than scaling back short term production or output. The greater uncertainty around the future economic climate also has led retailers to issue more uncertain forecasts. This uncertainty coupled with reduced orders, lead many firms to adopt a longer term response of scaling back capacity, as opposed to a short term response of reserving capacity but scaling back production.

These phenomena also led to the global reach and depth of the crisis in just a
short time span. Nervous consumers in the Western countries reduced their spending with retailers, the latter canceled orders and issued lower forecasts to their suppliers in Asia and Latin America, and in a short time span of a few months, entire economic regions and supply hubs suffered from the effects of scaled down capacity. Exports (and therefore trade volumes) and employment both were impacted negatively, and the economic contagion spread much quicker than perhaps in previous recessionary periods. What may have saved the day for these foreign suppliers is the resurgence of local demand for goods and services in countries like China and India, although doubts are being raised again whether these new customer segments in these emerging countries can sustain the global recovery on their own.

It is not a surprise therefore that so many smaller suppliers in the lower tiers of the supply chain have not been able to survive the downturn. In most vertically disintegrated supply chains, it is the norm to find a few dominant entities driving critical supply chain decisions, and as in the automotive sector, if these dominant firms are vulnerable, the risk of an incoherent and short term response to the crisis is that much higher. The case of thousands of dealership closings serves as a good example. The long term effects could be staggering: not only for the supply chain, but also the impacted regional economies. A related issue is the inability of the supply chain to exert any short to medium term control of critical customer service measures. For example, for jettisoned dealerships and product lines in the automotive sector, customers who recently purchased products from the closed dealerships face considerable costs as they work with alternative service providers. Moreover, as the vulnerable firms exit the supply chain, they take their functional capacities/capabilities with them, and this typically implies declining or unstable customer service measures (separate from inventory related measures). As standing products are jettisoned and features are rationalized, existing customers face the risk of poorer
service, and their defection could exacerbate the demand shocks for the entire supply chain.

Finally, and ironically, we again point out that the seeds for this global economic crisis may also be found in this rapid shift towards a vertically disintegrated model of supply chain. As the Western economies came to be driven by consumer spending, and as Western corporations outsourced labor intensive functions such as manufacturing to overseas partners, these countries simultaneously saw increased financial pressures on middle class segments, especially among the unskilled portion of the labor force. The inability of the middle class consumer, most remarkably in the United States, to sustain the consumption growth in these economies may have led to the mortgage and resulting financial crises that preceded the broader global economic crises. In other words some economists believe that there was a local economic crisis in the United States that caused the financial contagion and ultimately caused the weakening of the global economy\textsuperscript{31}.

These observations on the interplay between global economic trends, supply chain structure, and corporate strategy, lead us to explore the events and forces in 20\textsuperscript{th} industrial dynamics that have caused businesses and corporations to so fundamentally alter their structures and their relationships. This is the objective in Section 2.2. The motivation is also to understand the strategic imperatives of corporations in this past century, and to further uncover the pieces of this strategy that are in place, and other elements that represent gaps for most supply chains. As we shall see, gaps exist in the basic planning capabilities and approaches at a supply chain level, that are of critical import in times of economic and resource driven crises. Understanding these gaps is the subject matter of Chapter 3 of this dissertation.

\textsuperscript{31} See the writings of the Chicago school of “supply side” economists, notably Eugene Fama.
2.2 Unique Socio-Economic and Technological Markers Shaping Collaboration in 20th Century Supply Chains

2.2.1 The major technology breakthroughs.

To lay down the context for planning approaches that have been successful or prominent during the past century or so, it is useful to briefly comment on the historical drivers for today’s supply chains, in other words to understand how supply chains came to be structured the way we see them today. I argue that fundamental technology breakthroughs, in fuel and energy sources, agricultural methods, automotive and aerospace (including freight transportation), semiconductors, plastics, computing and information systems, military applications, media and communications including the internet, and finally in health-care systems have really been at the heart of how businesses have come to be organized and evolved in the last century. Of course, this is a rather grand picture of supply chain evolution, but both in terms of a chronology and in terms of impact, these are the handful of industry sectors that represent breakthrough events influencing global industry and corporate structure more significantly than other markers. Again, obviously, there are considerable interconnections and dependencies in these major developments, and identifying those interconnections would be an impossible task. At the same time it is still possible to observe in a few ways how these technological and scientific advances have served as the foundation for the way businesses are conceived and run in today’s global economy.

In parallel, we have witnessed remarkable socio-economic shifts in just the last two decades of the 20th century. Examples include the re-emergence of Asian economies, more democratic forms of government, mass consumerism and (arguably) better stan-
dards of living, free market style economies, and education and health care advances for the greater fraction of world population. Even political upheaval, which has been a constant for most of civilization, has been impacted greatly by these technological advances: the Cold War and the demise of the Soviet Union, the possibility of a long term European Union (after centuries of strife), and Japan and China’s emergence all serve as examples of how sophisticated communications technologies, and the resulting exposure to the Western cultural norms have wrought fundamental political shifts.

Here is one point: Now there are millions of new users of wireless cellular technology in India and China, who only had limited opportunity to use traditional land based trunk-lines for telephony. These users of information technology consume the same educational, business, and cultural media content (and disseminate their own) as an average user in the Western world. The personal computer – a contraption very much of the 1980s – and the internet – a phenomenon largely diffused in the 1990s – have together now made the most sophisticated and advanced technologies ever deployed, coupled them with valuable and real-time information, and made them available to the regions of the world with the lowest exposure to any kind of 20th century technology (even those that have not seen electricity for the duration). See Friedman, [76], and Prahalad, [152] for parallel ideas.

2.2.2 What technology has not solved.

Still, despite the technology advances and diffusion described above, we still possess only primitive understanding of the complex individual and collective human behavior and decision-making processes that determine our economic fate. Evidence of this can be found in the cyclical nature of businesses in most sectors and indeed of
the overall economy. The Great Depression followed the “Roaring Twenties” in ways that mirror the crisis of today (see Bernanke [14]). World War II which saw a contraction in the world economy was followed by the post war boom in most countries (see Rostow, [163]). The Asian “Tiger” economies were beaten back by the Asian crises of the 1990s and by Japan’s “lost decade”[113]. Similarly, deregulation to allow free market economies, and privatization in the socialist economies of Russia, Eastern Europe, and in South America managed to set off price shocks and economic turmoil in many of those countries during the 1990s (see Rajan et al. [155]). The latest global economic shock in turn was seeded by the excessive and risky sub-prime mortgage lending that fueled the US housing market boom of the past two decades. Few economists or business leaders had anticipated the sharpness of the current recession.

In other words, collectively, we have not yet managed to predict the upswings and downturns in our economies with any reasonable accuracy, despite claims to the contrary, and there is no reason to hope that we will, any time soon.

It is also important to observe that many of the technological advances of the past century have been commercialized and diffused primarily using cheap energy as a vehicle. This dependence on fossil fuels was in part engendered by increasing global trade. The movement of goods across continents was no doubt made possible by efficiencies achieved in shipping and logistics. However in turn it increased demand not only for imports and exports, but also made available an energy-rich Western lifestyle to the vast majority of the world’s population. This has increased by orders of magnitude the demand for fossil fuel as an energy source32. Not surprisingly, environmental and climate degradation and fossil fuel conservation has been a topic much debated in the supply chain world in the past few years.

However, the concern about the non-renewable nature of energy sources that sustain the vast majority of our supply chains was first brought to the fore by the energy crisis of 1973 – a shock driven by supply shortages and geopolitical risks. More recently in 2007-2008 the world experienced a speculative and demand driven price shock. Savvy investors and sellers have learned to trade energy as a commodity – oil and hydrocarbons in particular – based both on their current as well as future market values, and to repackage them into securities and financial instruments just as they did with the mortgage loans in the housing market. Rising energy consumption (until mid-2008) in China, India, and other rapidly developing countries, and the supposed “peaking” of energy supplies were cited as the main reasons for the record oil prices during 2008. However some fingers also point to how oil and energy based securities and derivatives became a focal point for global investors conducting high-frequency trading of these futures contracts (see Maugeri, [138]). While it is a matter of long standing economic debate as to how oil shocks can influence (or hinder) macro-economic growth (see Barksky, [12]), one can surely expect that this issue will again come to fore, when economies resume their growth pattern at some point in the future. At the very least, corporations and supply chains have to factor in questions of energy consumption and sustainability, and devise contingency plans for resource crises. In the long term, they may also have to re-engineer their processes, and evaluate alternative modes of supply and trade based on energy costs and related considerations.

2.2.3 Supply chain strategies to harness these breakthroughs.

What do these technological and socio-economic markers have to do with supply chain management, planning, execution, and analysis? It is my argument that these
markers have to be reflected in not only the organizational structure, but also the policy structure and operating practices of successful corporations, and increasingly of successful supply chains. A supply chain that does not take heed of these major changes outlined above, and adapt to the new environment of global access – and therefore competition – is not likely to survive very long or even be successful in the short term. For instance, the industrial giants of the United States – say an Alcoa or a DuPont – now operate in a much more fluid and competitive environment where they are vulnerable to events half way across the globe. Alcoa, for example, was under the threat of takeover by foreign players in Australia and Brazil\(^{33}\), who were not even direct competitors. Similarly, large investment banks such as Lehman Brothers (and Merrill Lynch), that took a century to build, were taken over and integrated in a matter of weeks by global players such as Barclays of England\(^{34}\) (and Bank of America\(^{35}\)).

Of utmost importance therefore, to any corporation or supply chain of any significance (or such aspirations) is to incorporate each of the following principles (see Figure 2.2.3) into their business and operating strategies:

1. Reach global markets, buyers, and customers: this requires distribution capacity that is global in its own reach;

2. Access resources (know-how, materials, suppliers, facilities, talent, and labor pools) on a global basis, and promote global diversity in its resource base;

3. Connect these resources, and connect them with the customers; using state-


The focus of this dissertation is on the last of the items in this proposed agenda for global corporations and supply chains. How can supply chains plan their resources, capacity, and operations with a global perspective? Now, these are by no means new principles: many prominent firms and institutions have adopted this global perspective (while incorporating local concerns) through the last few decades. This in part explains the explosive growth in international trade of the past century.
In particular, this strategy and pattern of globalization explains the emergence of non-Western countries from centuries of stagnation. Just in the span of two or three decades, they have overcome cumbersome cultural and political histories to participate in the new world economic order. Thus, resources are now being accessed and traded in truly international exchanges. The information technology revolution has managed to connect these countries and resources in real-time, and the supply and distribution systems have created trade routes that are relatively safe and reliable.

The remaining challenge of globalization is therefore to achieve cohesion and coordination in strategy, planning, and in action. It is my thesis that there is a need for new frameworks of global supply chain strategy, and corresponding frameworks of corporate strategy, that can facilitate and achieve such cohesion and coordination.

The first step, however, is to understand the evolution of supply chain structures, and strategies that have come to govern such decentralized structures. The resulting analysis also reveals the challenges, gaps, and the critical barriers for the formulation and successful implementation of such global supply chain strategies.

2.3 Structural Evolution Of Supply Chains And Collaborative Capabilities In The 20th Century.

2.3.1 Supply chain structures for mass manufacturing.

The technological advances of the early part of the 20th century provided the impetus for innovations related to mass manufacturing (Hounshell, [100]). As producers and marketers realized that economies of scale could be realized by reaching larger markets with the same product, they began reconfiguring their manufacturing processes for greater scale efficiencies. The predominant means to achieve these efficiencies was standardization, both in products as well as in processes. The resulting techniques
were labeled as “mass production”, and these techniques ranged from process and workforce specialization to inventory management and quality control. Companies at the time also realized that emerging technologies for automotive and aerospace applications were best delivered through vertical integration and consolidated operations. Once producers were able to produce in scale, there was a need for better and stronger distribution channels both on the inbound as well as on the outbound side. Rail and communications infrastructure also began to be utilized heavily for industrial purposes. An entire industry grew around advertising and marketing to gauge consumer preferences. This became the backbone for new product introductions, and also an important input for new technologies. I refer to Chandler and Hikino [34], and other influential work by Chandler describing the history of industrial production in the early part of the last century including [33]).

2.3.2 The growing importance of the individual consumer and product variety.

With growing middle class consumers and disposable incomes came consumer choice, and this made producers compete with variety in their offerings. Also, as products and tastes evolved, the fundamental manufacturing processes became more complex. This need to manage complexity drove many industries to adopt computer and information technologies (see Davenport [49]). Automation and information technologies also reduced the pressure on human labor to handle all of the process complexities and increasingly hazardous project work. Businesses soon discovered that the only limit to how information and computing technologies could improve productivity was their own imagination, and the computer’s basic capabilities [28]. The needs of businesses and technologists were met or exceeded at every stage with more advanced computing capabilities. Today’s professionals (especially in the op-

36 See Funke and Ruhwedel [78] for connections between product variety and the wealth of countries.
erations and supply chain management functions) and their workplace computers are virtually inseparable. Material handling techniques also evolved to support the growing needs of large scale logistics providers; sorters and routing equipment now enable web-retailing and merchandizing of millions of products at successful firms like Amazon.com, as well as most postal departments around the world [7]. The rate of innovation was also accelerated with computer aided design, prototyping, and manufacturing tools [65].

2.3.3 The multi-national corporation.

As mentioned earlier, supply chains also became more global, following resources, markets, and consumers across the world. The phenomenon of the multi-national firm or supply chain was in part created, and in part driven by, increasingly efficient shipping and logistics (Hanson et al. [96]). The International Maritime Organization helped create uniform shipping standards (containers, procedures, and regulations) for international trade in goods\(^{37}\). Larger carriers and more efficient transporters helped reduce costs dramatically for sea routes, while air freighting increased responsiveness in meeting changing demand in global markets. In the developing countries, industry consortiums worked with governments to develop vital local and regional infrastructure to facilitate the smooth movement of goods in-country (we refer to Chapter 5 for further details on advances in logistics practices).

International trade was also in large part influenced by political and economic trade agreements (called trade blocs) between countries. The World Trade Organization and various key agreements under its auspices have helped Western companies access previously closed markets, and in turn these emerging economies have ben-

\(^{37}\) See the International Maritime Organization website http://www.imo.org .
efited from better access to new technologies that helped them develop stronger manufacturing bases themselves [64][196]. With the opening up of previously closed markets, Western firms suddenly had choices of where they wished to innovate, manufacture or source. Labor cost differentials that had historically been growing but inaccessible to Western firms were now exploited to deliver even greater cost savings to consumers across the world. For example, the prominence of mass merchandisers such as Wal-Mart, and mass market chain clothing and apparel retailers is in part because of their ability to leverage these cost differentials, and to transfer the resulting savings almost entirely to the end consumers.

2.3.4 Air travel as a feasible mode for transport and freight.

The world aviation market grew starting in the 1950s, spurred first by US consumers traveling the breadth of the continent and traveling across the Atlantic for business and personal travel. The developed economies of Europe and Asia also took to commercial aviation in the period following the Second World War; the strength of the airline industry in a country was considered a signal of its economic development. However, the need for better management of the airline industry and its operations became a priority only after air travel had reached a broad enough market to become close to a public utility in the developed countries. Until that time, air travel was exclusive and a luxury service that worried less about operating costs, and instead focused on differentiation and customer service. Reduced competition along routes also suppressed the underlying inefficiencies in airline operations. The dominant US market, until 1977-1978, had operated under federally imposed regulations that allocated rights to flight routes to airline companies. The initial purpose was to help defray the fixed costs of operating a route, maintaining fleets of airplanes, and also control fares [144].
However, over time, there was considerable pressure politically to deregulate air travel to encourage greater competition, and to ensure better outcomes for customers. Consequently both US cargo and passenger aviation were deregulated in the late 1970s (see Chapter 5 for a more detailed discussion regarding transportation industry de-regulation). European deregulation took a different path: state run airline companies now established a broader network within the European region to share capacity and fly routes in other countries. Low cost airlines also sprouted in the US and in Europe in immediate response to deregulation. A more recent example of state run companies relinquishing control to private airlines is in India, which has witnessed explosive growth in air travel over the past decade. In virtually every case, such deregulation had an enormous structural impact on the industry. Almost overnight, efficiency and operational costs became a concern for airlines in a newly competitive market that allowed smaller and leaner entrants to compete with the larger and more established incumbents. Established players that were inefficient were forced to merge with competitors or exit the industry, while some, if not all, of the new entrants have found traction in the market.

As airline companies became more competitive, and as international trade has grown, businesses have also seen the benefits of air freighting their goods across countries, often to gain flexibility and responsiveness. The other major development in the aviation business has been the rise of the Asian, Latin American, and Eastern European economies. The past decade has seen demand for aircraft and aviation services grow considerably in these regions of the world. Boeing and Airbus, the two leading commercial aircraft manufacturers, both see the bulk of their demand over the next quarter of a century coming from these emerging economies; this was true at least until the current economic crisis that is severely impacting these emerging
2.3.5 *Telecom deregulation and the Information Age.*

Similar structural changes in another critical infrastructure industry, again on a global scale, were wrought with the deregulation of telecommunication and information services in the 1990s \[139\] \[174\]. Essentially, telecommunications worldwide have transitioned from being monolithic and monopolistic state regulated or even state run industries to privately owned and operated networks, composed from local and global competitors. The explosive rate of innovation and growth that followed in this sector has been phenomenal, and is often presented as a case study for successful deregulatory policies and practice. For example, the large scale investments from the private sector in long-range fiber-optic communications technologies is credited with creating the rise of off-shore information technology and service industries in countries like India (see again Friedman, \[76\]).

In particular, the deregulation act of 1996\[39\] was also intended to allow the entry of multiple carriers with similar service technologies (cable television or long distance telecom, for example) to compete in both regional and cross-state markets. There is considerable debate as whether indeed consumers have benefited from the potential for such competition, since the number of large telecommunications and mass media companies has been steadily declining over the years. However, from a purely technological standpoint, the removal of key state regulatory constraints and also state control of basic telecom services, has yielded massive benefits to businesses and individual consumers who have seen an unprecedented increase in the range of

\[38\] See Boeing literature www.boeing.com/commercial/industry\_info.html

Along with the use of the personal and business computing technologies, the rate of innovation and sophistication of communications technologies has opened new avenues for productivity growth and efficiencies that were not imagined to corporations and supply chains even as recently as the 1990s. This diffusion of communications technologies also paved the way for new dimensions of risk for corporations, however. Now, their actions and decisions are visible to stock markets in real time, whose reaction in turn can turn the fortunes of even the largest and most solvent of firms. Similarly, the rate at which business data becomes available to corporate leaders and decision-makers – about either their markets or their extended supply chains – is now much greater than before, and this blizzard of information keeps getting thicker by the day. The inability of firms to act upon key information is also now that much more visible to shareholders. This creates new challenges for firms to capture relevant information from the data through analysis, synthesize emerging scenarios from such information, and eventually to incorporate them into consistent decision-making frameworks.

2.3.6 From multi-national presence to off-shoring and outsourcing.

Once multinational firms helped established a base of resources, including professional and labor workforce in emerging countries, they began to see the opportunity to utilize more of these remote resources in ways that best suited the corporate and shareholder interests. This led to the off-shoring trends of the 1990s, as many firms took off-shore to cheaper locations their labor or local resource intensive operations. Initially, manufacturing and assembly operations that were less technologically intensive were taken to South and South East Asia, Latin America, and eventually to
Eastern Europe after the collapse of the communist regimes. Eventually, with the rise of the professional classes in those emerging nations, even high technology firms were able to exploit the skilled labor forces of countries like India, Romania, and Russia (at a fraction of the costs in more developed countries). Now, in addition to providing a base for low-cost manufacturing or assembly operations, these off-shore locations are also increasingly being used for high-value added activities including research and development.

Multi-national and technology oriented corporations such as IBM, Microsoft, SAP, Oracle, and Infosys, are now truly international in their scope and operations because every facet of their business model has a global footprint and reach. It is important to note that some of these very firms play vital roles in providing information and productivity enhancing services to the great majority of global supply chains in virtually every industry sector. As a result, such technology firms have had a huge role in enabling first the outsourcing, and then the off-shoring of information systems development and maintenance (see again Friedman[76]) that are critical to the smooth functioning of many large and complex global supply chains. While it is debatable whether these outsourcing and off-shoring trends have yielded guaranteed cost savings or have added significant value in terms of supply chain outcomes, it is without doubt that supply chains the world over have benefited from the collective experience of these information technology firms in developing best-in-class business process and management systems for a wide spectrum of global supply chains.

2.3.7 Cross-cultural fusion and convergence of consumers.

We have already commented on the structural evolution of other industries like the automotive, and the electronics and semiconductor, the pharmaceutical, and the
aerospace sectors. What is remarkable is that these industries which started out primarily in the Western European and US economies (and to a smaller extent in Japan) have now become truly global in their reach and scale. The developing countries of South America, Asia, and Africa, now have a critical role to play in these supply chains. This trend has generally been followed in most of the sectors of industry and business that were spawned in the 20\textsuperscript{th} century. In contrast, sectors like food and agriculture commodities – to some degree – and to a greater extent the arts and culture, still retain regional or local identities (as in the times before the 20\textsuperscript{th} century). This despite booming international trade and information exchange which has facilitated a more interconnected world with global commodities exchanges, multi-national food corporations, and a gradual fusion of cultures when it comes to consumer attitudes and personal consumption.

2.3.8 The process efficiency perspectives of a supply chain.

On the last point, innovations in designing and managing basic supply chain processes such as inventory and distribution management have followed the evolution of supply chains to progressively more distributed and decentralized business and decision structures. For example, new concepts such as postponement, delayed differentiation of products in the development/production, and distribution pipeline were introduced to derive supply chain efficiencies (see Feitzinger and Lee, [71]). Many of these innovations were driven by the need to manage the unprecedented variety in the products on offer to consumers. The mass retailing business model relied on providing this immense variety at discounted prices. With the advent of on-line retailing, companies could afford to derive the advantage of having multiple retailing channels. However, price competition became intense as the internet disseminated information about competing products on a global basis.
Retailers also have made great strides in how they manage sales information in real-time; point of sale systems and bar-code standards have helped big box retailers manage inventory and order replenishment functions. Companies like Wal-Mart and Proctor and Gamble, and scores of other retail supply chains have implemented collaborative planning, forecasting, and replenishment systems that aim to minimize order inaccuracies, inventory, and replenishment costs (see Fisher et al. [73]). The bigger retailers have also invested in warehouses and logistics infrastructure, and technology initiatives like RFID (see Zipkin [200]), to derive greater economies of scale in distribution, while at the same time achieving greater flexibility and responsiveness. Retailers have also played a major role in procuring from lower cost regions in Asia, South America, and Mexico; for example, Wal-Mart has opened a global procurement office in Shenzhen, China to procure from multiple country sources – mainly from low cost regions – and supply higher quality merchandise to their stores worldwide\textsuperscript{40}.

2.3.9 The services revolution.

Finally, services had been invented to support even the most primitive form of business that humans have engaged in during their evolutionary history. Examples of services that have existed for all of our collective history include the hospitality, health care, and education industries; in fact most of the critical services we consume today have always been a part of our history. For distant observers, it may be surprising therefore, that it was a long time into the 20\textsuperscript{th} century before “modern” planners realized the increasing significance of service businesses [158]. What motivated this new perspective may have been the relative decline of the manufacturing

\textsuperscript{40} See Wal-Mart’s press release http://walmartstores.com/pressroom/news/8437.aspx
and agricultural sectors in the developed economies.

With the diffusion of manufacturing technologies across the world and elimination of key trade barriers, Western (and Japanese) corporations were developing manufacturing operations in the lower cost regions of the world including South East Asia, South America, and Eastern Europe. With the 1980s and 1990s witnessing the deregulation of many socialist or communist economies, these regions not only were attractive as markets, but also as manufacturing bases for supply consumer goods to the Western economies. The role reversal has actually been quite dramatic; the workforce in the Western economies was transformed within two decades into a predominantly service workforce (manufacturing still remained a strong contributor for the economy), while mass manufacturing became a primary engine for the developing economies [109].

2.3.10 Structural evolution and the missing dimension of 21st century supply chain strategy and collaboration.

To summarize, we again draw attention to the implications of the structural evolution we have just described, for our proposed new principles of supply chain strategy. Quite simply, of the four items on the globalization agenda described in the previous subsection, all but the last of the items – supply chain planning with a global perspective – are readily and consistently achievable with 21st century technologies. This is both a technological as well as a more fundamental structural limitation based on how the Western capitalist society fashions its corporations and businesses (see Drucker’s landmark work, [59]).

Technologically, for instance, reaching new markets through state-of-the-art mar-
keting and distribution mechanisms; accessing global talent and resource pools; and connecting customers and resources up and down the supply chain with advanced communications and information tools are all feasible – and achievable – for the majority of well-founded corporations and supply chains. However, planning such global supply chains, so that resources and deployed and utilized most effectively and efficiently, and the right markets are served at the right times, is now a challenge that has grown several orders in magnitude. This is precisely because of the success achieved by many firms in implementing first three items on the agenda or road-map: namely global markets, global resource pools, and globally connected supply chains.

Now that firms have such global visibility into events and information in their markets and supply chains, it is an even greater imperative that the new breed of managers utilize this blizzard of information and data sources to avoid gigantic errors in supply chain planning and execution, and to generate the efficiencies that this global model has put in their sights. In the next chapter therefore, I outline some of the business planning approaches that have found favor in the context of the structural developments described above. While my immediate objective in doing so is to critique them, I build on such a critique to develop some basic but nevertheless foundational collaborative planning constructs that are aimed specifically at addressing the shortcomings of present day planning frameworks.
3

Planning Frameworks to Enable Collaboration in 21st Century Supply Chains

3.1 Overview and Organization

In the first few sections of this chapter, the goal is to present a broad-based critique and analysis of the effort to fit corporate planning frameworks that worked in more centralized and vertically integrated regimes to a more leveraged and decentralized supply chain organization. We present several observations and conclusions, some through examples, and some through argument, and point out the serious deficiencies in planning frameworks and implementation schemas used to manage many global supply chains. The proposition is that unless these deficiencies are addressed in both concept and in practice, supply chains of today, and of the future, are likely to be severely constrained in their ability to develop stable partnerships that are more effective and efficient than the erstwhile vertically integrated and centralized models. In order to reveal these deficiencies, we first define planning for the purpose of analysis in this and future chapters, and then chart the evolution of planning methods over the course of the previous century. We limit our historical overview to developments
in the past century, and even within that set, focus on the major branching points for planning frameworks and technologies. Specific choices and events during the evolution of the planning frameworks are highlighted for their role both in achieving significant efficiencies within the centralized and vertically integrated regimes, and in causing limitations for a leveraged supply chain model that require redress.

It is also useful to point out that several authors, notably Lee [125], have provided parallel descriptions of pitfalls associated with deploying centralized planning frameworks in decentralized and leveraged supply chain environments. This chapter is intended as a perspective on these issues that is derived from the first principles of planning, as opposed to a more direct functional/operational focus on issues such as inventory or capacity management within supply chains. The functional view is the matter of the subsequent chapters of this dissertation, where as a remedy to the pitfalls outlined here, we develop instances of such collaborative planning frameworks for the logistics and program management functions in decentralized supply chains.

In the later sections, we also provide a concise re-definition of planning as an exercise, and as a management approach, and attempt to provide specific remedies, in broad brush, to the pitfalls identified previously. The goal is to provide as concise a redefinition of the planning exercise in general, while identifying specific objectives and priorities that can guide not only the planning activity but also its interactions with its subjects and finally the presentation of its outcomes.

3.2 What Is Planning?

We first define planning in the broadest and simplest possible ways:
Planning is the (i) formulation of a schedule or sequence of inter-related activities at any chosen level of granularity, and over any chosen periods of time; (ii) with a corresponding assignment of responsibilities to one or more agents; (iii) with an allocation of resources required by those activities; (iv) to achieve a set of stated objectives or goals for the organization.

1. Therefore, in devising an organizational plan we do not differentiate between the types of activities; their scope and scale; their location; their timing; or the kinds or extent of resources they consume. The only requirement for a set of activities to be considered is that they be inter-related and of relevance to achieve the stated objectives.

2. Similarly, we do not distinguish between forms or shapes of an organization; in theory an organization could be the net collection of resources (both human and technological) that are required by the set of activities being planned. They could also span multiple firms and independent groups or organizations.

3. We do not distinguish between hierarchies of plans: in theory strategic planning in a boardroom; advertising budget planning in a sales department; capacity or production planning in a manufacturing facility; and workforce scheduling in a customer service call center are all viewed as instances of the same concept as defined above.

4. Finally, we do not differentiate between the time range of the planning exercise: short and long range planning are again illustrations of the same fundamental concept of planning.
Therefore in the rest of this chapter and during the next, we will discuss “planning” in these generic terms; it is implied that pitfalls and challenges attributed to the “generic” or loosely defined planning exercise will also apply to specific instances of planning regardless of the activity or process definitions, organizational scope, hierarchies, or of time horizons. Similarly, our proposed frameworks to augment 20th century approaches and to suit broader and evolving supply chain structures are intended to apply to most (if not all) instances of the fundamental planning concept defined above.

Before we move further, it is also important to differentiate between planning – either as a corporate function or as a business activity – and the much broader function of management. Management as a concept and construct is general encompasses the planning activity or function, and includes the development of a business strategy, organizational structures to support the business objectives, the structuring and definition of tasks, the planning of tasks and their assignment to organizational units, and finally the review of performance both at a business as well as at an organizational level to ensure all-round alignment with the business objectives. Thus, planning is one specific activity that is central to the management function, but the planning process still depends on the management function to define its inputs, its scope and parameters, and the audience and format for its outputs and recommendations. Hence, we caution the reader not to assume that these two activities and terms are too similar to warrant a separate definition of the planning process as we provide here.
3.3 Planning Approaches For The 20th Century Corporation And Supply Chain.

3.3.1 The measurement and division of labor; and management of sequences of tasks.

The earliest planning approaches started in fundamental ways to measure and plan small tasks for the purpose of efficient division and assignment of the tasks to the labor workforce. Taylor [185] started the “scientific management” movement which he argued was a better way to understand and communicate “work” both for the laborer as well as for the manager. His techniques however encompassed many managerial functions within the manufacturing environment including accounting systems to take inventory of work assigned and work completed, and production control systems to enable visibility for management into work progress. His most famous contribution, was in the form of time and motion studies that measured and published the standard times for simple or atomic tasks and activities within a factory environment. This was a first attempt at a modular division of human work so that it could be broken down and made interchangeable just as was done with industrial components in the preceding century.

Taylor’s disciples included well-known figures such as Gilbreth pair [85] and Henry Gantt [81][80]. Frank Gilbreth carried forward Taylor’s ideas in time and motion study; his work involved analyzing and making the simple motions of any human activity more efficient. Even today, the field of human factors in industrial engineering and bio-mechanics studies builds on these simple concepts, only in more sophisticated ways, and with more advanced tools for analysis and improvement. Gantt’s work carried forward Taylor’s ideas in planning and scheduling of activities. His main contribution was the development of several types of visual scheduling and planning aids. These relatively simple charts would help managers know how jobs were (to
be) assigned to workers with minimal interferences and conflicts, and whether they were ahead of or behind schedule in terms of due dates.

Gantt pioneered the concept of delivering a “production order list” of jobs, based on the schedule chart, to workers to complete for the day. Gantt also favored the concept of measuring activities in a standard fashion based on how much time any activity would consume on the schedule. Many of Taylor’s applications and projects however were for “job shop” environments, and found limited application in “higher-level” jobs and tasks that required greater skill, judgment, and innovation. Gantt’s work, however, was carried forward to activities of greater complexity including large scale project management, and innovation and development activities.

3.3.2 Computer aided approaches and the advent of operations research.

Quite naturally, managers soon realized the limits of these simple hand-made production scheduling and planning aids, and the role of computers to advance these methodologies was recognized. The field and practice of computer enabled planning simply carried forward these fundamental ideas for different contexts, and with different objectives. At some point after World-War II, synergies were realized with the advancements in the mathematical fields of operations research and decision science. Thenceforth, process and operational management methods grew as a branch of the applied sciences [1] (with the ongoing debate on what type or branch of science it actually is).

This coupling of operations research techniques with increasingly powerful computing applications in planning and scheduling proved to be a natural and highly successful pairing. In the 1950s, companies like DuPont implemented the critical
path methods (CPM) in project management [118], while the US Navy pioneered the Program Evaluation and Review Technique (PERT) where mathematical and network programming concepts were used as a foundation for the analysis and optimization of complex projects [165]. Companies like IBM (in parallel) developed complex applications in production planning and scheduling, again using relatively sophisticated database modeling and computer architectures. The 1960s-1980s witnessed the widespread acceptance of production planning methods such as Material and Manufacturing Requirements Planning and Distribution Requirements Planning. See Hermann [99] for a more elaborate history of production planning and scheduling methods).

At a strategic level, long range corporate planning systems first originated within large firms that were capable of the leadership and stability that such an effort required. In the period following World War II, management and corporate planning, as a field, also became more influenced by the mathematical/decision and the statistical sciences [143]. One of the corporate symbols of this movement was Robert McNamara, who served in the US military during the war, and then was recruited into Ford Motor Company. McNamara was trained and experienced in turning around inefficient and chaotic operations using an analytical approach that included more careful planning, and statistical methods to improve product and process quality. Other examples include General Motors under the leadership of Alfred P. Sloan, who allowed the noted scholar and thinker Peter Drucker to study GM’s operations and management structures [60][59], and who later pioneered systematic approaches to production and quality management. Similarly, IBM instituted its own corporate planning paradigm in the 1960s, and in what is possibly a testament to the success of such planning structures, the company is profitable and successful to this day [170].
Other prominent companies of the day who were instrumental in the formal or “scientific” management movement include Boeing, DuPont, General Electric, Procter and Gamble, and Westinghouse. Popular strategic planning and execution frameworks that were pioneered and experimented with at the time included “management by objectives” (Boeing), and product life cycle / brand management (P&G, GM). It was around this time in the 1950s and 1960s that companies also realized the intersecting roles of information gathering and processing; strategic decision making; and operational planning and execution. Separately the importance was recognized of organizational problem solving structures to the quality of product and processes.

3.3.3 Scaling the computer based planning approaches.

In response to this market for computer aided planning tools, companies like SAP and (the erstwhile) Baan based in Europe developed computer applications in the 1970s based on the concept of relational databases. They were designed to capture and connect large volumes of data within the organization and utilize them in the planning process. These software applications have also aimed to integrate the planning and scheduling function with other functions, such as accounting, human resources, marketing and sales within the firm. The 1990s saw a booming market for the implementation of these enterprise resource planning (ERP) systems. These systems were later scaled down from their intended use in large corporations to adapt better to the needs of planning in small and medium sized organizations. The end of the century saw more sophisticated and clever applications of the enterprise planning concept to supply chain settings, but often the benefits of these systems were limited to a single firm. This despite the fact that these systems often impact the suppliers and the customers of the firm in at times unpredictable and unintended ways (see a popular critique by Davenport [50]).
As a parallel development, in the airline industry, following de-regulation, demand management and efficiency in operations had become the twin pillars of success (indeed, survival) in this sector. Large companies like American and United Airlines invested millions of dollars in internal planning systems that determined capacity allocation to routes, organized fleet schedules, and managed demand through dynamic and discount pricing [184][140]. With the advent of internet applications, airlines also changed their sales processes to exploit this new medium of reaching customers. With increased competition, and with operations management being the key differentiator in success, airline companies have gone through multiple cycles of new entrants, consolidation, and mergers\(^1\). Over the years, aviation safety was also an issue that was addressed quite successfully using better planning and management approaches, including reliability and maintenance planning (see a stream of work by Barnett [11] on the safety implications of these approaches).

### 3.3.4 Incorporating safety and environment concerns in planning.

Environmental degradation and labor safety concerns from the excesses of capitalistic industrialization were the motivation for the formation of governmental and international regulatory agencies during much of the previous century. In the US, regulatory agencies such as the Environmental Protection Agency (EPA) and the Occupational Safety and Health Administration (OSHA), were both established by the Nixon administration in 1970. In Europe, there have been a series of country specific efforts and also regulation under the European Union (see Vogel [195]). More recently, the effort has expanded from just the regulation of the forward supply chain,

\(^1\) The September 11th terrorist attacks of 2001, precipitated the only break in the continued trend for growth for US (and world) airlines.
to include laws that regulate the management of industrial and consumer waste, and to create greater incentives (through penalties and credits) for resource and product re-cycling and re-use. Europe has again taken the lead in these environmental protection initiatives through the enactment by the European Union of two major pieces of regulation: The Waste Electrical and Electronic Equipment (WEEE) Directive of 2003, and the Restriction of Hazardous Substances (RoHS) Act of 2006. While these laws were enacted in this current decade, they were formulated and debated in the preceding years. The US as a whole has been slower to enact such industry specific laws, with the exception of individual states regulating subsets of products and processes.

Such regulations now serve as important constraints for planning and scheduling methods in a wide range of industries that require human labor to interact with hazardous work materials, and also those that potentially have a local or global environmental impact. At the strategic level, corporations and supply chains take into account country specific laws (including recently enacted laws in China) while developing a local manufacturing presence, and while developing global production and distribution plans. Similarly, labor hiring and firing and safety and health related operational practices by corporations are guided by country or region specific regulations, and this again impacts both strategic and operational planning. For example, plants operating in certain Western European countries may have to remain within bounds of utilization targets set by government regulations, as well as contracts with their labor force. Further, with legislation currently being debated on setting caps and prices on carbon impact and energy consumption of various business and manufacturing processes, corporations now have the added challenge of determining trade-offs on output levels at different stages or locations of their supply chains, and these trade-offs also have to factor into their planning function.
3.3.5 Incorporating service tasks and planning frameworks.

In the services sector, workforce or capacity scheduling and planning databases have now become routine and are even considered essential in diverse areas such as health care management, telephone customer service, and the hospitality industry. In the more technologically sophisticated sectors such as telecom and the internet, network capacity management is an area where planning approaches have worked relatively well. In fact, capacity planning requirements are really stringent in these sectors, where close to 100% availability and reliability is a common assumption and expectation in practice. Military projects and programs are another arena where operations research based planning approaches have been deployed with great success, to manage the logistics and support infrastructure, control inventory of equipment and fleets,
Figure 3.2: Evolution of select stochastic planning approaches in the 20th century.

and even to plan combat operations.

See Figure 3.3.4 for an approximate representation and time-line of the evolution of the planning approaches discussed so far. A parallel development took place in stochastic methods of work and resource planning, which has also led to major developments in manufacturing, telecommunications, information systems, networking, and computational systems that capture the underlying risk factors in the target environments. While we do not describe this evolution in greater detail, we refer to Figure 3.3.4 for such a depiction. Such stochastic methods trace their roots to probability theorists and applied statisticians; most notably the pioneering work of Erlang [69], in the very early part of the 20th century.
3.3.6  Perspectives on the planning function in the Information Age.

A majority of the planning approaches and applications discussed above are increasingly reliant on the availability, maintenance, and management of historical and real-time data, both within the firm and also across the supply chain. Advances in database management techniques and the ever decreasing costs of digital storage led to the development and use of large scale databases for the specific purpose of supporting the planning process. In particular, independently designed and managed databases supported different aspects of the planning process. Examples include customer relationship management (CRM) applications to support sales, forecasting, and order fulfillment; order management, scheduling, and resource databases; and supplier relationship management (SRM) systems to support requirements planning and coordination. Similarly warehouse management systems (WMS) are the choice for planners in the distribution and logistics domain. Planning approaches have become so successful in these individual operational domains that the main challenge for many corporations, large and small, is to integrate processes that span multiple functional domains.

In the next subsection, we aim to highlight what we see are the primary challenges to continuing with the planning approaches and system frameworks of the last century. Quite simply, these systems are designed to work within the confines of a single firm or a cohesive organization. When these internal planning systems meet the supply chain boundaries, the result is often chaos and unpredictable outcomes for both planners and the operations staff. The planning approaches of the past century lack critical collaborative capabilities that can allow these frameworks to work across the supply chain boundaries.
3.4 Pitfalls And Challenges Arising From 20th Century Approaches.

3.4.1 Broad limitations of existing planning approaches.

The previous discussion – and in particular on how planning systems and frameworks have evolved – points to the obvious and numerous advantages that such advancements have presented to businesses and corporations. It is arguable whether without these planning systems founded on the simple principles set forth by Taylor, the Gilbreths, Gantt, and Erlang all in the early 20th century, the economic growth witnessed since then would have been possible. Of course, many firms adopted these ideas and built on them in parallel; the profession of management science grew and contributed a series of advancements that motivated the wide-spread implementation and productive use of planning systems. Still, there are a number of fundamental (or conceptual) kinks, inherent pitfalls, and challenges in both the design and in the use of these familiar planning frameworks.

We do not differentiate in the design and in the use (at least here); since improper use is in some sense a design flaw. Systems need to be cognizant of their intended or unintended uses. Furthermore, in critiquing these systems, we are examining these frameworks from a supply chain manager’s perspective, whose responsibilities go beyond the boundaries of the firm and extend to the broader supply chain or ecosystem. This of course was not the intended purpose of today’s large scale planning systems; so our critique is more a reflection of the changing circumstances under which these planning systems operate. In other words, we examine the 20th century frameworks from the perspective of how we expect businesses and supply chains to evolve in this new millennium; hence there are bound to be shortfalls, and opportunities for advancement. Broadly, then, it is possible to categorize the pitfalls (Figure 3.4.1) and critical barriers in current day planning systems into seven fundamental categories.
1. Potential for ill-defined objectives, responsibilities, and roles.

2. Lack of standards or uniformity in design, development, and use.

3. The manufacturing versus service divide.

4. Developed for use within the firm or organizational boundary.

5. Numerous competing alternatives vying for resources and attention.

6. Limited exception and risk management capabilities.

7. Limited collaborative capabilities.
We focus on each of these pitfalls separately where we can. In the following section, we develop our road-map and vision for planning systems, and attempt in a very modest way to address these limitations. At the outset, though, we claim that each of the criticisms we offer are also the strengths of these planning systems viewed from another (possibly narrower) objective of generating greater efficiencies or productivity within the firm. Our overall purpose, again, is to show how supply chains as an ecosystem can design, deploy, and benefit from the use of well-structured planning frameworks similar to the ones that have been successful for a vast majority of stand-alone corporations. See Figure 3.4.1 for a synopsis of these pitfalls and their impacts on corporations and their supply chains.
3.4.2 Potential for ill-defined objectives, responsibilities, and roles.

Without doubt, we have come a long way from the simple visual planning charts to today’s planning systems that carry tera-bytes of data, compute at super-computing speeds, and manage corporations that involve thousands of people across the globe with billions of dollars at stake. However, the success of planning systems cannot always rest on its size, scope, or scale. On a fundamental level, a closer examination reveals several real pitfalls in implementing the most sophisticated of today’s planning systems; these challenges are magnified when we define and utilize these planning approaches within a supply chain context.

Hierarchical and cross-functional stake-holder conflict.

As would hopefully be evident from the discussion on the evolution of planning approaches, such systems have grown to serve multiple purposes, and to support a diverse range of business objectives. For example, many firms operate separate planning systems for every level of the decision hierarchy: strategic, tactical, and operational. Each of these hierarchies have different requirements and objectives, different inputs, and different kinds of decisions to support. While in the past hierarchical methods have been developed to address these concerns; in reality, each of the three levels of decision-making now have different sets of decision-support and planning tools. Different objectives, of course, may lead to conflicting directives to the organization based on the solutions recommended by the systems at each hierarchy; which one wins is still an issue decided by the operating culture and the organizational peculiarities within each corporation. Things get messier when one adopts a supply chain view to decision-making; here it is not only a matter of decision-hierarchy within a firm, but also decision authority across firms.
These approaches have now been refined for a given decision-hierarchy and supply chain context, and also tailored to work in specific environments. They are also fairly (locally) technical in their design and presentation; so that as a result, professionals and theorists both risk losing perspective and clarity on the fundamental nature and objectives of the planning process. Furthermore, there are often a great number of stake-holders within the firm, and across the supply chain, and each has varying degrees of control over the planning process. Each of these stake-holders is typically both self-interested, and furthermore sometimes has a limited understanding of the interests of the broader organization.

Uncertain scope of the planning exercise.

Furthermore, in many situations the planning process runs parallel to the process of setting the goals and objectives of the firm. When the firm’s strategic objectives compete with the immediate objectives of the plan, it could mean either that the firm’s strategy is not credible, or that the planners are unaware of the need, or indeed incapable of aligning the objectives of the plan with the firm’s goals. The latter situation is more common in supply chain settings. Quite simply every firm wishes to maximize its profits from its ventures, but often, adhering to a supply chain plan demands that individual participating firms make decisions that sacrifice immediate profits in favor of long term stability or success, or even survival of the supply chain. In such cases, the supply chain planning objectives and the firms’ own strategy may be at cross-purposes, and it calls for some resolution, either in the form of a correction in the strategy or with the firm re-evaluating its supply chain partnerships.
Reinforced bias.

Such conflict should theoretically lead to multiple objectives that can later be prioritized and reconciled; but often the stakeholder conflict in turn leads to a reinforced bias both in information exchanged, as well as in the preference over planning outcomes. This bias also leads to obfuscated priorities and ill-defined objectives. Justifying the chosen objectives is often a politically charged exercise within any organization (no matter its size), and therefore left well alone in the design and even development stage. Also, quite often these objectives are conflicting, which is but a natural outcome. However, planning systems may not yield solutions that satisfy every stake-holder, and therefore once again, there is less incentive to affirm the criteria that will guide the planning systems.

The other source of conflicting objectives becomes apparent when the planning process is cross-functional in nature (which is most often the case). Different organizational and process sub-groups have different demands as to the objectives. For example, the marketing and sales department would want a planning process to be responsive to the sales processes; the manufacturing and finance departments would be more cost conscious; while the accounting department would want the plan to honor the budget allocations and resulting constraints. The planner would have to find a way to honor the demands of each of these sub-groups; understandably in most cases goal setting becomes quite a controversial process. What is true for cross-functional planning is almost entirely true for supply chain planning; only in this case the conflicting objectives come from different sections and stages in the supply chain; each of which could represent an independent business. For example, the retailer of a product would want greater responsiveness to end-customer needs, the manufacturer would be more cost conscious and would worry about its profit
margins, and the component and materials suppliers would be concerned about how their investments in the product or service would be compensated for in the long run. Hence, a supply chain environment really compounds the challenges that come to the fore in a simpler cross-functional setting.

Responsibility for outcomes (or lack thereof).

For example, in many corporations, senior leadership takes accountability for strategic planning of products, investments, and organizational structure; middle management takes responsibility for tactical planning of organizational resource capacities, process planning, and project/program management; while the rest of the workforce takes responsibility for the ensuring that the tactical plans are executed to order. Even within the confines of the firm, each rung of the organization has different needs when it comes to the planning process and its execution. The workforce has responsibility for executing against the plan; but often encounters that plans set at the strategic and tactical levels are infeasible to different degrees. Such infeasibility in many cases is the result of the inherent uncertainties in the operating environment of the firm, which again many plans are incapable of recognizing.

Such situations also lead to a parallel universe where managers and the general workforce alike over-ride plans where possible, and individually hope for the best outcomes as they now are responsible for the consequences of stepping out of the plan. Consequently, the primary requirement of the workforce, in many cases, is that the plans and resources be quite flexible to allow for responsiveness to conditions on the ground. In other cases, where the workforce exhibits risk-averse behavior, they demand that the planning process take into account the uncertainties in the operational environment. Any width that the plan offers minimizes their own exposure to
risky operational decision-making. Such objectives could run counter to those of the middle and upper rungs of management, for whom such flexibility comes at considerable added costs and greater uncertainty in outcomes. In the case of such conflicts, how we define the incentives to cooperate with and implement the business plan – at the various levels of the organization – determines the success of the planning process, and also its long term credibility.

**Blurring of boundaries between planning and execution.**

A related issue in many organizations is that the planning process is often confused with the tasks entailed in its execution. This is apparent in organizations where there is no formal planning process sponsored or enforced by senior management. In such cases, the personnel responsible for task execution also make decisions concerning the structuring of tasks, and planning into the future. This is really a problem of procedures, communication and education within the organization; however, this issue takes on a more complex dimension in the supply chain setting. In a supply chain environment, the luxury of having procedures evolve over time may not be available, especially in dynamic environments where firms enter and exit the partnership at regular intervals. As a consequence, the different participating firms (and their personnel) often have to construct rudimentary and unstable plans, and also execute them according to the best of their ability. In many supply chain contexts, this happens solely due to lack of leadership and sponsorship from within the participating firms. Moreover supply chain planning require consensus in goal setting and defining objectives, which is often hard to come by; and therefore the planning process is conducted in an ad-hoc and real-time fashion.
Procedural versus a creative outlook.

Another common source of conflicting objectives is whether the planning process is intended to be a gradual participative process involving discussion, or whether it should be a “scientific” or mathematical procedure that is implemented in the form of a tool or aid. For example, in many companies the planning process is carried out as a “budgeting exercise”, where resources are accounted for, and supply is matched with the forecast demand for the firm’s output and services. In such case, planning (and its recommendations) is often cross-functional, supply chain driven, and a participative process that evolves over time. Within the same organization or supply chain, one could find other planning procedures that are more automated and tool driven. It is possible to derive a budgeting process from readily available enterprise resource planning software. These software are often monolithic and offer ready “solutions”, but lack the collaborative nature of human, more interactive planning processes. Hence sections of the organization that have to work with one or more of these alternative procedures are confused about whether to honor the plans and decisions originating from their counterparts within the organization.

Tradeoffs in decision-making are inherent and vital to any planning process or solution. However, when these trade-offs are not communicated in the proper or transparent fashion, it leads to further confusion and conflict within the organization. Lack of transparency in defining the objectives is often half the problem; the other half is due to the fact that sometimes plans are drawn up through complex discussions and sophisticated methods, communicating or disseminating the trade-offs within the organization is a risky undertaking: different personnel may have different levels of education and sophistication, and seldom is there a uniform perspective on the decisions made by planners or superiors. Matters are not helped, when in
some cases even decision makers are unaware of the precise trade-offs that they are effecting though their own plans.

**Planning versus execution authority.**

Assigning responsibilities and accountability within a planning framework is also a tricky proposition. In the extreme cases planners themselves have little authority within the organization. A lot more authority and power often rests with the line management that is responsible for execution with planning often relegated to being a support service. This is a recipe for disaster since in such cases, planning is almost always a bureaucratic exercise that has little credibility with the line management that is responsible for execution. At the other extreme, planners come from senior management, but little authority is provided to the line management responsible for task execution or adapting the plan for ground conditions. Again this creates problems as line management often has more understanding of the ground level realities of the business, but they either have to work within the rigid confines of the plan – doomed to mediocre performance, or they have to take undue risks to over-ride the plan.

In some cases, critical operational planning responsibilities are assigned to relatively low skilled personnel or those with little line authority; these planners take on the role of “schedulers” or “program managers” coordinating the tasks of other members in the organization; this is again a situation that invariably leads to poor outcomes both for the planner as well as for the organization. Burn-out rates are quite high in such positions, as these individuals have to deal with crises on a daily basis, because they are now responsible not only for the planning, but also for the execution of the tasks. Within a supply chain context, such program managers have
even harder roles; now they have to deal with line managers across different firms but have little authority except in the form of informal feedback and evaluation processes. The one exception is when the program managers or schedulers deal with suppliers in which case they wield some power because they can impact the performance appraisals of the suppliers.

3.4.3 Lack of standards and uniformity in design, development, and usage.

*Organic growth of planning systems.*

As discussed in the previous section, many planning frameworks originated within large firms; the resulting approaches were heavily biased towards the environment within which they were used. Later on, as professionals became mobile, and as the ideas and concepts spread through academic and industry work, many different industries began adopting such planning systems and associated methodologies. In the past two decades, these systems have been formalized to a certain extent by the prominent software providers; this trend has propagated to the extent that now many planning systems are part of the information and communication systems adopted by firms. Hence, the peculiar attributes of these information systems are then carried over to the planning processes adopted by these firms; these planning processes in many cases are peculiar to each firm and are a by-product of the organizational histories and the experiences of both the firm and its personnel.

As a consequence, the planning processes in different firms often have only limited commonalities in practice, even though the underlying tasks that are being planned are virtually identical. For example, Intel Corp., and Advanced Micro Devices, both operating in the semiconductor market may have fundamentally different approaches to product planning (a critical strategic planning activity in the industry) based on
their leadership and corporate cultures. The impact of organizational culture on the planning process is further compounded by the fact that there are few, if any, standards or widely accepted frameworks to organize and implement the planning exercise – at any level or for any function. For example, what one firm may refer to as “budgeting” would mean “capacity and production planning” for another. Furthermore, even different business units within a firm may be using different planning methods; the reasons could be several: they may have different sets of customers or be operating in different markets, or they may be units integrated from the merger with another firm, or they could be aligning themselves to the needs of different supply chains.

Disparities in standards within a supply chain.

There are some standards being developed for the supply chain context; the Supply-Chain Operations Reference-model (SCOR), which is a “process reference model” for supply chain management, and “has been developed to describe the business activities associated with all phases of satisfying a customer’s demand”. While these frameworks are successful in providing guidelines to organizations on the supply chain planning process, in reality, many firms still retain their planning and operational peculiarities even while following some of these guidelines.

For some of the more well-structured problems, such as capacity or production planning, there are a number of possible algorithmic approaches depending on problem modeling convenience and choice. There is some consensus in academia and in the planning literature as to which of these approaches work best in some versus environments versus others. However, in practice, there is often limited knowledge of the underlying algorithms, or even about the best way to structure and communicate
the planning problems in a fundamental and correct fashion. The end result is again
the lack of unified standards for developing planning algorithms, for implementing
them through the appropriate processes, or for understanding and communicating
the solutions.

Rate of innovation in computing and database platforms.

A related problem is that the delivery vehicles or platforms for implementing plan-
ning algorithms, along with the software to enable implementation, have witnessed
rapid change over the past couple of decades. This makes it hard for firms to keep
pace with the best-in-class algorithms, delivery platforms, and the software. The
fact that such software tools are quite complex and require considerable training and
investment to learn, adopt, does not make the task of standardization any easier.

In addition, given the limited training in different pockets of a typical organiza-
tion, there is a great deal of effort required in adopting new and improved systems.
Also, because of a lack of continuity in standards, many system upgrades are exer-
cises in unlearning the old processes in their entirety and learning new ones. Most
new planning system implementations are therefore disruptive and can be quite risky
to the core business of fulfilling customer orders, if not carried out with care.

Reputation effects on adoption rates.

Once implemented, these new systems are evaluated using the subjective criteria of
its diverse set of users. A planning system is only as good as its reputation with its
user base, and therefore even if a system is functioning as per its original specifica-
tions, it could be considered a failure based on opinions of key personnel. On the
other hand, a new planning system considered to be successfully implemented can wreak havoc with a firm’s critical operational processes, and in extreme cases, even result in a firm’s failure.

3.4.4 Corporate versus Supply Chain Focus

*Scope limited within firm boundaries.*

The strategic imperatives for most corporations have already been laid out for them: they need to access markets in multiple geographies, access resources beyond their firm boundaries, and connect them so as to enable value-added transactions. From a supply chain perspective, the lack of standards results in major “inter-operability” issues in developing and implementing supply chain planning processes. In other words inter-operability issues between individual corporate planning systems can impede the very transactions that the strategic imperatives call for. Just as with cross-functional conflict in objectives, different firms may have different processes they are familiar with, different software they have invested in, and different objectives and trade-offs they need to consider. These factors contribute to the general instability of the supply chain planning process and to the unpredictable performance of firms when assigned to various constituent tasks.

Many of the biases we discussed across stake-holders, hierarchies, and across functions, carry over to the supply chain setting. Only in this case, all of these biases are compounded by the inter-firm boundaries and the ease of communication and transaction capabilities. More critically, the notion of a supply chain level strategic, tactical, or operational plan does not exist in management parlance. What does it mean to have a supply chain capacity plan, or a supply chain production plan? Who decides the parameters of the planning process across corporate boundaries; who
provides the inputs; who is accountable for adherence to the planning recommendations? These are all still concepts at a very nascent stage in their development, and we are indeed far from making planning processes seamless across supply chain partners.

*Lack of visibility into supply chain level information, material, and financial flows.*

At the very least, a planning process must have a clear definition of the tasks and resources to be planned and the performance or objective criteria that will guide the execution. In a supply chain setting, in most cases, no single firm has visibility into all of the tasks or processes required for manufacturing a product or delivering a service. Without this clarity in task definition, a cross-firm planning process remains an oxymoron. Secondary data such as resource requirements or financial commitments are even less credible without clear task or process definitions. Since no one party has access to all of this information (one can think of private versus public information architectures), it is therefore an unresolved question as to who will be responsible for conducting the planning exercise.

The only alternative is to publicly announce and exchange internally developed plans and then resolve conflicts in a group setting. However the logistics of such a planning exercise are daunting given the combinatorial implications.

*Lack of flexibility to react to supply chain events.*

Within a corporate setting, plans can in theory be revised to account for unforeseen events and exceptions that occur in real-time. Either plans can be over-ruled or worked around so as to minimize impact on overall outcomes. For example equip-
ment failures in one facility within a firm can warrant the deployment of back-up resources and extra work hours so as to minimize the impact on the planned work flows. However, in a supply chain setting, such unforeseen events can have impact outside the firm boundary; in places and at firms that may have no authority or ability to enforce contingency measures. Worse, the information related to unforeseen events may be kept private or may not reach other firms in time for them to react. There are numerous examples of such lack of reactive capability or flexibility in a supply chain setting.

3.4.5 Competing alternatives.

Automation (procedural) versus manual (creative or responsive) approaches.

The primary tension in most organizations when it comes to sophisticated or pre-conceived planning systems is that it minimizes the human element and impact both in the planning as well as in the execution phase. A good example is an automated scheduling system that has limited visibility or access to real-time information. The human agent can adapt better to this uncertainty through social and informal (uncoded) interactions, and therefore can sometimes access real-time information and therefore be more efficient and productive in scheduling tasks. The problem is that human beings are not consistent or reliable as procedures and computer based systems, and are further more error-prone. This sets up a classic tension between the more procedural and consistent algorithmic approach to corporate or supply chain planning, and the more flexible human approach that is three-dimensional and incorporates more subtle information exchanges. Thus the primary competitor for planning systems and approaches is the human alternative which is to plan less and be more responsive to events as they unfold.
Low-technology versus unproven high-technology alternatives.

Another frequent refrain is that planners introduce and replace technology just for the sake of incorporating the latest technological innovations, as opposed to a real need in terms of the efficacy of the planning system. High-technology systems are also more capital and set-up intensive and are bulkier and have more redundancy for most planning environments. This leads to a choice between more humble low-technology systems that have a smaller scope and depth but are proven to be effective in those smaller domains, and more sophisticated frameworks that are expansive but risky and error-prone, at least in the initial stages of their deployment. A famous example, of course, the conflict between systems that follow the lean manufacturing or Toyota model of production and operations, and the information system driven approach advocated by the large software firms; actually this also serves as a good example of the conflict between people oriented versus database driven automated systems.

Specialized legacy applications versus integrative frameworks.

Specialization to a process also offers problems for general planning systems which are designed to be more expansive and encompassing in their reach, but as a trade-off have to be selective in terms of their specialization in any one function or process within the organization. The lack of such depth in broad-based corporate planning systems leads to a localized rebellion among process experts and savvy users who assert their special needs by incubating and growing their own legacy frameworks and systems that serve their function well. Often these specialized systems are well-justified; without them key functions would suffer in their performance and the overall organization could be impacted. Still, the competition between local renegade systems and corporate umbrella systems is another source of conflict to be aware of.
3.4.6 Separate emphasis on manufacturing, services, and on information technologies.

Floundering in the Information Age.

In the context of the last century, the world economy and consequently the majority of industries and corporations that have grown in that time period have done so on the strength of the manufacturing function. The exception to this rule has been the last two decades following the invention of the personal computer and the diffusion of internet technologies. It is no surprise therefore that planning systems and processes were first invented to manage these manufacturing systems. The primary need was for greater efficiency and productivity in a manufacturing concern’s activities, and hence many planning frameworks even today reflect the piecemeal, discrete, and staged structures of manufacturing systems of that time. The questions for planners centered around how to minimize effort and cost in manufacturing tasks; how to increase the productivity of the workforce and the equipment; and how to satisfy the demands of manufacturing firm’s customers in the best way possible. In later years, planning systems were also responsible for the quality of the manufactured product, and for organizing the exchanges of process information within the organization.

The Information Age has changed the center of gravity from monolithic planning systems to more individualized and localized, but networked organizations and supply chains. The range and depth of information available a few years ago only to centralized and physically intimidating mainframe databases, is now potentially available to every employee in the extended supply chain. The dramatic increase in the access to business relevant information without a similar step change in decision structures has posed special challenges to planners and decision makers in large and complex organizations. Essentially, distributed agents in the supply chain having greater access to real-time information also increases the burden on planners to jus-
tify their plans and decisions in the face of such information parity; there is also the risk of plans becoming more short term and reactive in their emphasis, as the information onslaught reduces incentives to adhere to long term strategies.

*Dealing with human behavior and psychology.*

With the advent of the service economies in the West, academics and practitioners now had an incentive to intensify their efforts in analyzing and planning service systems. This view however is not without contention; in reality it is possible to argue that all business activities in some ways provide service to the consumers, and therefore even manufacturing firms can be classified in some ways as service organizations. However, what is fundamentally different about the management of services is the relatively greater emphasis on human psychology and behavior. Moreover, planners have always been challenged by the inability of the workforce to conform to the sometimes rigid requirements of their frameworks and solutions approaches. For all of the 20th century, and even today, planners have been confounded by the failures of their best-laid structured and engineered frameworks for various inexplicable reasons related to organizational and individual behavior. I propose that this incongruence between the sophistication of the planning process as it applies to structured but inanimate processes, and the relatively primitive ways we still conduct ourselves in an organizational setting is at the heart of many of the problems that exist in the field of management; corporate and supply chain planning being just one instance of this overall phenomenon.
A creative service perspective of the supply chain management function.

These gaps have now come to the fore as we try and apply our techniques and methods learnt in structured processes of manufacturing to more unstructured transactions and activities we witness in the service context. From a supply chain perspective, these gaps are critical. Processes in a supply chain can encompass both highly structured (say manufacturing) tasks; as well as more chaotic creative tasks such as product design or customer service and support. Thus, supply chain management tasks could typically be classified either as service or manufacturing activities. However some tasks are harder to categorize, involving creative services that support and enable the critical information, financial, and material flows. Without encompassing such loosely defined service or transaction enabling categories, plans drawn up for a supply chain would not carry much credibility. Planning only a subset of structured activities in a process while ignoring other value-adding but less structured tasks, would constitute an incomplete set of instructions for the organization.

Modeling and planning service transactions.

Another shortcoming related to service tasks, is that services can be understood as being delivered and consumed in real-time within a transaction. For example services cannot be inventoried, and the performance and quality of service transactions is closely linked with the quality of the service provider. Therefore, planning in the service context has to be more refined in nature and also be more flexible and adaptive to the environment as experienced by the service provider. Furthermore, the planning is also required to be cognizant of not only the resource providing the service, but also the intended recipient or consumer in the transaction. For example, for planning a set of financial transactions and activities in a supply chain the planner needs to understand the personnel involved in the transaction, their level of training,
the type of transaction that is involved; but the planner also needs to understand the perspective, expectations, and limitations of the service recipient. This subtle point is often overlooked in planning solutions for the service industry; in most cases firms conduct resource planning exercises in the service industry, but the psychology, expectations, and possible heterogeneity (or variation) in the providers and consumers of the transaction are ignored.

*Modeling and planning creative and innovative operations with complex information flows.*

For effective operational planning within a firm, and even more so across a supply chain, it is quite critical that both structured (“manufacturing” or “productive”) tasks, as well as loosely defined (“service”) tasks be incorporated. Of course, a key unresolved question here is whether loosely defined tasks can be planned at all. Still it is our contention that understanding and mapping such loosely defined transactions can lead to effective planning approaches for these tasks. The main issue rather seems to be that organizations do not attempt to map or identify these tasks into their processes. A good example is found in the field of innovation and product development: there are several structured and well understood tasks in the development process including design, prototyping, development, and commercialization. However, technology and associated services are key enablers of the development cycle, and many firms today use advanced information systems and specialized software to enable collaboration, communication of ideas, and to maintain records.

For a real world example encapsulating these various issues: in the case of Airbus and its recently introduced A380 commercial airliner, the company invested billions of dollars in its development. While the company and its suppliers invested a great
deal of effort and resources in the detailed planning of the development cycle, the project still suffered delays and incurred lost revenues in the billions of dollars as a result. It turned out the different organizations (Airbus and its key component and sub-system suppliers) did not ensure that the information technologies they deployed were compatible: the end result was a problem in some critical sub-systems that led to program delays. Thus, a foregone investment in a key information technology service and enabler – one that would have cost only in the hundreds of thousands – led to losses in the billions of dollars for the supply chain.

3.4.7 Limited collaborative capabilities.

*Centralized versus Decentralized Supply Chain Approaches.*

From the preceding discussion, it should be abundantly clear that planning as a management concept is mostly reserved for application within a firm or an organization. When the organization is defined more broadly to include multiple firms, the rules for planning are changed altogether, and the prevailing concepts are no longer consistent with the needs of the extended environment. Firstly, most planning frameworks take the perspective of a central planner or a unique entity that has access to all of the information relevant to the plan, has access to all of the data on the activities, understands all of the inherent trade-offs, and finally knows the best objectives for a given environment. For many environments, it may well be true that a central planner’s perspective is the best possible one given the alternatives. However, for situations that demand the collaboration of multiple distinct firms that exhibit self-interest and competing objectives, a central planner’s approach may not be the best fit. Obviously, a supply chain is an example of such a decentralized environment. The issue is not the proposed or possible benefits offered by central planning; rather the issue here is the effectiveness and suitability of central planning to the diverse
ecosystem that is represented by many supply chains.

**Controlled access to firm level data.**

Sometimes, we observe that limiting collaboration may actually be beneficial to the central planner; too many conflicting objectives uncovered through a collaborative or consensus approach, could result in a much weaker prescription. A common example of such choices is in databases designed to actually preclude or limit collaboration to varying degrees, and allow access to a few select users within the organization. Only a few users within the organization have the authority to alter the inputs or parameters of the planning exercise; it is understandable therefore that external partners have even more limited access. Without access to the information being used to plan key supply chain processes that span multiple firms, there is not much scope for joint discussion, and even less scope for a collaborative planning.

**Confusing information exchange with collaborative decision-making.**

The one exception is the recent emergence of web-based user interfaces and database portals have enabled greater visibility into planning data across organizations. However, while technology exists that can allow real-time exchange of data; access can be limited by policies set by a firm. Furthermore, even current day technologies are not equipped to enable inter-firm collaborative decision making. In fact, most of the technological advances of the past quarter of a century, from Electronic Data Interchange to more complex XML based data transfer protocols have been designed to enhance the bandwidth and quality of transactions within supply chains. However, there has been limited advancement in technologies to advance collaborative planning and decision-making. This could a contentious claim, since some would argue
that real-time communication capabilities are now improved several fold from when planning frameworks were first adopted. However, the reality is that effective communication is a necessary, but not sufficient step to achieving collaborative planning capabilities.

*Clout and financial strength determine planning hierarchies.*

In fact, planning even within a firm is collaborative only if the culture and leadership allows; it is common to observe in firms that some functions have greater clout and bargaining power than others. In certain firms, the financial planner becomes the proxy for the central planner, and controls a lot of attributes or parameters (scope) of the process. In other firms or industries, marketing and sales executives are the kings and queens. Invariably, one or more key stake-holders are overlooked, and the effectiveness of the planning process is diminished.

Such skewed relationships become even more apparent and accumulate destructive potential in supply chain settings. It is always a debatable and contentious question in a typical supply chain partnership as to who exactly plays the role of the customer or leader. These roles could also change and reverse over time, as the financial or market stock of different players (and their capabilities) crest and trough at different stages of the relationship. Planning hierarchies can therefore be subject to the prevailing winds of market forces and specific supply chain needs.
3.4.8 Limited risk management capabilities.

*Plans offer an ex-ante view.*

Needless to say, but worth repeating. The earliest planning approaches were cognizant of the uncertainties in the environment. Since they were not reliant on sophisticated (and expensive) technologies, plans were altered based on the needs of the month or day. Moreover, the scope of planning was limited to small groups or organizations, unlike today where potentially hundreds of thousands of activities can be under the purview of managers. While in theory and in the literature planning under uncertainty is a topic that has been studied for decades, in practice uncertainties are often harder to deal with. This is partly a behavioral issue: under uncertainty, planners need an *ex ante* objective whereby they consider the expected costs of system operation or even (in sophisticated settings) the worst case or variance in performance. However, those responsible for execution observe and evaluate the plan *ex post*; as a consequence the planners almost always look bad from the point of view of those implementing the plans!

*Performance versus flexibility.*

In response, many planning approaches have special exception management provisions; which allow plans and activities to be altered based on contingencies or exigencies. Better exception management features lends a planning framework more flexibility; and also greater credibility from the perspective of users. However, from the perspective of the planners, such exceptions are really troublesome to deal with ex ante. But risk management is a broader concept that goes beyond exceptions and just laying out contingencies. Risk management is also a collaborative process that involves negotiations and bargaining between the planners and the workforce over the possible range of outcomes associated with plans.
Risk is harder to manage across firms.

Risk management is more in line with the central ideas of planning than most frameworks we observe today reveal. Moreover risk management is a fundamentally collaborative process that involves liberal exchange of information: both historical and future estimates. Mapping possible outcomes of a plan and understanding the consequence of each potential outcome is a primary step in mitigating risks. In a supply chain setting this requirement translates to greater understanding of the impact and consequences of a given plan on the participating firms and their operations. Thus, effective risk management requires that firms collaborate and share critical information about how they believe the plan will perform in their own environment.

Perverse incentives for bias and distortion.

One could expect considerable controversy, bias, and conflict in such an exchange and bargaining process. Partners may have the perverse incentive to bias and distort information that are inputs to the risk mitigation process in order to ensure more beneficial outcomes in expectation. For example, in production planning, buyers may inflate their material requirements forecast that they share with key suppliers, while suppliers may distort information on available capacity. The buyers fear material shortages, while suppliers may wish to reserve capacity for more profitable customers. Ex post, the effectiveness and performance of the plan depends on how the different stake-holders collaborate on dealing with exceptions and contingencies. If the partners do not share information on outcomes (adverse or beneficial) in a timely fashion, they could limit the ability of parties within their own organization, along with the broader supply chain, to respond to the contingency. For example,
if buyers do not share real-time demand information with the suppliers, then the suppliers will not be utilizing their own capacities in an optimal fashion, leading to greater costs in the supply chain. Conversely, if suppliers do not share information on potentially disruptive outcomes, buyers may not have enough time to react, and again the outcomes for the supply chain as a whole could suffer; thus transparency is the key issue here.

In summary, we have discussed seven critical challenges to planning as it applies to supply chains today. It is our belief that current day planning frameworks are limited in their scope and ability to address these issues that impact performance at the supply chain level. These are not a comprehensive set of pitfalls or challenges, and there are many other aspects of planning that can be critiqued (constructively) for a given setting or application. But it is my assertion that supply chain managers and planners would benefit the most by addressing these potential shortcomings. In the next chapter, we provide a fundamental view of how planning frameworks can begin to address some of these challenges in a modest but still realistic fashion.

3.5 Expanding The Planning Concept For 21st Century Supply Chains.

As defined in the previous chapter, planning involves developing a pre-defined schedule of activities assigned to one or more responsible agents, along with an allocation of resources to those activities, to meet certain organizational objectives. The central thesis of the previous chapter is that instances or applications of the fundamental planning concept have been successfully developed for vertically integrated corporations and cohesive organizations. The set of circumstances and priorities entailed by globalization and the vertically disintegrated model of supply chains now calls for a different set of instances and applications to be created to work for a global
Therefore, we do not need a complete redefinition of the concept of planning. We only need to expand it to work within a supply chain setting. We propose that this expansion be carried out along the following five basic dimensions (see Figure 3.5 for a summary illustration).

3.5.1 Distributed and decentralized agents and objectives.

Few supply chains today are accurately represented by an easily identifiable collection of firms operating in a confined region, let alone within a single organization. As a consequence national and organizational boundaries have to be crossed many times before a product or service is within reach of, or adds value to the intended customer. In order for planning to be a meaningful, credible, and relevant exercise
in a supply chain context, we need to create credible instances (or applications) of the core planning concept for such diffuse organizations.

We first need account for the simple fact that both the activities and the resources deployed are really spread across the one or more firms participating in the supply chain. For example, the assembly of an automobile or electronics product typically involves a number of different firms who supply the subsystems, aside from the firm responsible for taking orders and completing the assembly process. Similarly, a financial service transaction requires the availability of resources and technologies belonging to several firms. This map of how key resources are shared in a supply chain context really defines the organization for the supply chain; and this is the ecosystem that supply chain managers will be responsible for.

The agents can also be decentralized, and a planning process needs to account for this fact. Decentralization implies that groups of agents may be independent in setting their objectives for the planning exercise. They may also have different sets of resources at their disposal and may contribute them towards the plan at their own discretion. The planning process also needs to be able to recognize multiple conflicting objectives among the agents.

For similar ideas, that are applicable in the different domain of artificial intelligence and planning, we refer to Cammarata et al. [31], DesJardins et al. [57], and Durfee [62]. The objective in artificial intelligence approaches is to provide a computational decision support system for the organization to manage distributed agents and tasks. In this dissertation we are concerned not only with the decision-support system but the organizational structure that will best support planning at the supply chain level.
3.5.2 Dynamic redefinition of activities, resources, agents, and responsibilities.

This redefinition is aimed at resolving the issues of risk and contingency management identified in the previous chapter. The first dynamic redefinition deals with how (supply chain) activities are allowed to change over time. These are contingency alternatives that allow the supply chain activities outlined by a planning document to be altered (at any level). Without clear bounds or limitations on contingency plans, the supply chain plan could be vulnerable to events occurring in its different corners. Without such contingencies in place, however, a planning framework really does not make any sense; moreover, the resulting plans could be too rigid, and cause the supply chain to be less flexible or responsive to environmental changes. Building such contingencies lends credibility to the plans, and also provides some measure of assurance to the parties that obligations, roles, and responsibilities will be honored over time. At the same time, it also provides some downside hedge to concerned parties, as they look to minimize the impact of events that could lead to potential losses.

For example, it is common in the electronics manufacturing industry for the product firms (like Apple, HP, or Microsoft) to issue advance material purchase authorization notes to their third party manufacturers. The reason for this is that material shortages especially in critical phases of the product life cycle could imply great losses for the supply chain. Here, contingency plans define what happens when due to demand shortfalls or customer order cancelations, materials already purchased go unused and are piled up as inventory. In some cases, based on prior agreements the product owner assumes responsibility for the inventory costs; but in other cases, the third party manufacturer would use the materials purchased for other ongoing programs with other firms. Such inventory risk management plans allow the part-
ners to make critical decisions in the interest of the supply chain without individual exposure to debilitating risks. For a less sanguine example, we can remind ourselves about the inventory write-offs in the technology sector amounting to several billions of dollars at the end of the dot-com era in early 2001.  

Secondly, the ownership of these resources, and therefore the authority to allocate these resources, can change over the life of the product, service, or program, thereby causing differences over time of the availability, price or value of these resources to the supply chain. For example, many large electronics manufacturers have sold or leased a significant number of their manufacturing facilities to the third party manufacturing firms such as Flextronics. Now that Flextronics owns or operates these factories and resources, with certain confidentiality limitations, it has the ability to ration their availability, to price them differently, and to distribute them across their entire customer portfolio, and not just the parent marketing firm.

3.5.3 Task specific information sets and agent access.

For both planning and execution, the who/where/when/how set of instructions for each task or activity need to be encoded in an information set specific to each task. Since the tasks can be assigned to one more decentralized agents, it may be necessary to provide selective access to each agent responsible for even a single task. The same principles apply to inter-connected or coupled tasks. When interconnected tasks require information sharing among responsible agents, they must again have an instruction set that describes how much information is to be shared between groups of agents, and how much access to the information set is to be provided.

An example is when tasks in secure organization groups are outsourced to subcontractors; the information relevant to the completion of the task is shared based on the security clearance or similar certification obtained by the sub-contractor. Higher levels of access generally imply greater flexibility and performance by such external agents, while lower levels of access can ensure security but typically at the expense of task performance. However, these information sets and subsets and access levels are a pre-requisite to the planning process.

In some situations, these information sets and agent access levels can also change over time. For example retailers may share more of their forecasts with their suppliers, as the quality of the forecasts improves closer to the selling season. Such dynamic updating of information sets can impact task performance positively. In other situations, dynamic updates can hurt – for instance, one could think of retailers who issue frequent forecast and order updates without penalties, leading to increased risks and increased costs for suppliers as they plan future deliveries to accommodate retailers’ constantly changing forecasts. Hence, the richness and scope of the information set, the extent of information sharing, and their rate of change can all be parameters for the planning process, and can be adjusted and tailored for particular environments. Nevertheless, they are important attributes of tasks and their interconnections, and this dimension is critical for supply chain planners.

Another important reason for maintaining task specific information sets is to allow for agents responsible for tasks to resolve uncertainty in the task environment. In this sense, the greater the uncertainty in a task environment, the greater the need for information to resolve this uncertainty. For an excellent overview of the role of information processing in the context of organizational design, we refer to Galbraith [79], and Tushman and Nadler [191]. For other knowledge-based representations of
firms and their value-adding processes, we refer to Grant [88].

3.5.4 Hierarchies of decision-making authority

Fourthly, we need to identify decision-making authority among the distributed and decentralized agents. We need to define which agents get to decide the task-agent-resource assignments, which agents determine the schedule of tasks, and which agents determine the alteration of such assignments and schedules for contingency purposes. In other cases, we may also need to assign authority among agents for bargaining and conflict resolution. Sometimes the decision-making authority could also include the ability to determine incentives for agents to collaborate and achieve common supply chain objectives. More simply, we need to create a decision-hierarchy among the distributed supply chain agents, with decision roles and boundaries similar to how we set up similar structures within any organization. We note that this decision-hierarchy is different from the analytical decision-hierarchy proposed by Saaty [164], which calls for a hierarchy of decision criteria (as opposed to authorities). Here we propose for a hierarchy of decision-making authority allocated among subsets of the distributed agents.

For example, as we will see in Chapter 4, supply chain agents may need to collaborate to bring a new product or a process to market. In this case, as part of the supply chain planning definition, we need to assign control of certain aspects of decision-making to one or more key agents from within the supply chain. In many industries, supply chain partners may nominate agents from within their organizations who have such broad authority to allocate decision rights, define information sets and sharing rules, sometimes even define the objectives of the planning exercise for the rest of the agents in the supply chain. These super-agents constitute a “pro-
gram” command center, and their oversight is important to ensure that decentralized agents collaborate and devise plans that serve the stated objectives of the group.

Even for detailed planning, if a task requires resources contributed from several firms, then decisions need to be made with respect to the availability, utilization, and allocation of these resources; decisions may also need to be made to alter these plans based on contingencies. For a planning process to be successful in a supply chain, these decision rights must be *a priori* defined and allocated among the agents. One may end up with a hierarchy of decision rights, where decisions made by one set of agents may be either supported or over-ruled by agents assigned a higher level of authority, but such allocation rules need to be clarified before-hand for this decision-hierarchy to be effective.

Ideas similar to the ones above have been articulated in the context of broader organizational theory by Simon [171]. Huber and McDaniel [102] argue that complex and turbulent organizations (similar to the supply chains of our interest) are best structured or designed based on the needs of the critical decision-making processes that impact their performance; similar to our approach here, the authors also connect ideas in organizational decision-making, the information processing view, and technology based decision-support systems. There is another view of the importance of decision-making hierarchies to the functioning of modern organizations. Chandler [32] credits a top-down hierarchy with enabling agents leading large and complex organizations to perform higher level planning functions (such as strategy and investments), without being directly involved in the evolving minutiae of tasks and processes assigned to agents deeper down in the decision-hierarchy.

However, in the supply chain setting, unlike the modern corporation, we cannot
always expect that agents at the top of the hierarchy to be composed of the major firms (by size) in the supply chain. Rather we need a more balanced allocation of decision rights based on the criticality of agents and their task responsibilities to the overall supply chain objectives. For example, in the case of mass retailing as espoused by firms such as Wal-Mart, the retail and last-mile distribution processes only constitute a small subset of the activities in the extended product supply chains that may include raw materials production, component development, product assembly, and international distribution. However, companies such as Wal-Mart exert tremendous influence over the most critical and highest level of decision-making in the supply chains for the products they carry. In a more vertically integrated organization, the opposite would be true, as the weightiest part of the supply chain operations would have the most say in strategic planning and decision-making.

3.5.5 An incentive framework for agents.

As a fifth dimension, we need to incorporate – relative to the decision-making roles – the bargaining processes that determine how the costs and rewards associated with various activities and resources are apportioned between the supply chain parties. This is really an incentive framework without with many supply chain activities in a partnership run the risk of failure. This is most frequently seen in cost and risk intensive activities that make decentralized agents nervous about outcomes and implications down the road. For any supply chain planning exercise to be effective, at any level of the decision-hierarchy, there must be scope for bargaining and negotiation among the distributed agents; and especially among those agents that have considerable resources and investments at stake. The parameters of the bargaining and negotiation steps in the planning exercise are typically defined within incentive frameworks, as we outline below.
First, for an industry example discussed in detail within Chapter 4, we can consider the Boeing Dreamliner development plan conceived in 2003. Boeing’s suppliers had agreed to foot the non-recurring costs of component development in return for intellectual property rights and minimum supply volumes. The idea was to enable Boeing to outlay the necessary multi-billion dollar investments for the program, while sharing the risks associated with the program suppliers. This was not an isolated incentive framework: within the same industry, a different product development plan initiated by Airbus Industries (of Europe) called for Airbus’ suppliers to assume 25% of the non-recurring costs incurred during the development of the A380 family of commercial airplanes (see Chapter 4).

Another example from a different industry is the case of Dell’s assembly operations which are supported by vendor managed inventory. Dell is able to achieve negative cash to cash cycle time, whereas suppliers who agree to the production and payment plan stand to benefit from long term or preferential supply contracts. Similarly, in the service industry it is common to see cost and revenue sharing incentive contracts in franchising arrangements between either retailers and manufacturers or brand owners and franchisees. These are all examples of incentive frameworks that enable more effective planning of supply chain activities. These incentive frameworks are typically juxtaposed with the conflicting objectives of the set of decentralized agents. The incentives serve to mitigate the extent of the conflict.

In summary, there is a critical need for traditional business planning frameworks to be reworked and re-defined, in light of what we have learnt in the past two decades about supply chains. Furthermore it is possible to outline five dimensions to redefine and expand the general planning process for a supply chain context: an inter-
organizational map of the key shared resources, availability of the shared resources over time, decision-making roles and structures with respect to the plan, contingency planning and risk management, and bargaining and negotiation processes.

3.6 Addressing 20th Century Pitfalls Through The Expanded Planning Dimensions.

Here in this section we begin addressing some of the particular challenges we identified in the application of 20th planning frameworks, by deploying the fundamental ideas above of how basic planning concepts can be expanded for supply chain environments. We examine three of the pitfalls in detail, and show how a combination of one or more of the five expanded dimensions can be developed in conjunction with existing approaches. We also provide examples of how these five expanded dimensions can be realized and implemented in practice. Figure 3.6 summarizes the dimensions of supply chain planning that are most relevant to addressing some of the major pitfalls.

3.6.1 Clarifying the objectives, roles, and scope of planning.

The idea here is to provide some resolution for the stake-holder conflict that hinders planning in a decentralized supply chain context. This can be achieved primarily by three mechanisms:

1. Defining distributed and decentralized agents and capturing their individual goals and objectives where possible. Some of this may be private information for these agents, which agents may be reluctant to share with the planners; for which we will need the next mechanism:

2. Incentives for individual agents to share their short and long term objectives from the planning exercise;
3. Decision-making hierarchies so that objectives can be prioritized and a preference order (or weighting) of these priorities can be constructed

Let us consider a production planning example for a supply chain. The plan typically needs to specify the production levels at several facilities owned by one or more supply chain partners, for several time periods into the future. Now, each of the partners may have different objectives for the planning exercise: the lead firm may have an interest less in growing profits but to increase market share of the product beings manufactured. The suppliers on the other hand may be interested in short term profits, and may therefore want to produce less than the lead firm wishes to sell in the market. For each partner, different facilities in their ownership may similarly exhibit different objectives: some may want to maintain high utilization while some
others may be concerned about the safety or environmental concerns from increased production volumes.

Having distributed agents - one for each owner-facility pairing - allows us to capture the objectives of individual partners w.r.t. to the facilities they own (either jointly or in whole). Attributing potentially independent objectives to these decentralized agents allows the planner to capture the diversity of interests in the supply chain organization. It is also possible to weigh these different considerations according to negotiated rules.

Secondly, an incentive framework aims to reward these decentralized agents for moving towards the needs of the collective, or to penalize them for deviating from the collective goals. In the production planning example, these incentives can come in the form of higher margins for suppliers who wish to scale down production relative to the centralized solution; and lower margins or higher costs for agents that wish to produce more than the plan recommendations. These are simple ideas, but worth emphasizing; moreover they are a foundation for a collaborative supply chain planning.

Now, to enable planners to determine the best outcomes for the collection of agents, one needs to devise a decision-making hierarchy and assign decision authority to subsets of the decentralized agents. For instance, the lead firm and the Tier 1 suppliers could constitute the highest level of decision-making authority. A given tier 1 supplier and their own Tier 2 suppliers could constitute a grouping within second rung of the hierarchy, and so on. The idea here is to rationalize the individual objectives of the decentralized agents and to create representative goals for a particular sub-set of agents. While nested sub-groups of agents (and their objectives) would be
easier to manage, we do not require these nested structures (see Figure 3.6.1 for an illustration of the concept). Importantly, in contrast to popular approaches to the mathematical solution of the planning problem, we do not require a single isolated decision-maker at the top of the hierarchy. In other words, it could be a collection of agents who make decisions (rule by committee for example) with the higher authority. This in some ways models the natural human approach to planning large organizations.

To clarify the roles and scope of the supply chain planning exercise, we require the other two expanded dimensions.

1. Roles are clarified through information sets associated with tasks; these would
specify to the best extent possible the interconnections with other related tasks, the objectives of the task and its particular requirements, its assigned owner and the status of the task, among other attributes. Allowing selective access to this information on a task to one or more agents who are impacted by the task gives more clarity to the agents of their relative role in the supply chain organization, and also the scope of the planning exercise.

2. A dynamic redefinition of the agents (and their objectives), tasks, information sets, could in certain situations highlight the changed objectives of the organization, and also transition tasks from the planning to the execution phase where contingency mechanisms would need to be deployed for risk mitigation. Here access to the evolving information sets for the distributed agents determines their ability to respond to different evolving scenarios.

3.6.2 Bridging the manufacturing vs. service divide for the Information Age.

It is arguable whether there was a particular need to maintain separate functional terminology such as manufacturing or services to describe basic business processes. However this is the legacy of the management and planning approaches of the past century, and we nevertheless have to bridge this gap, if we are to provide a seamless planning function not only for broader supply chains, but also for individual corporations. It can be argued that manufacturing and services are both fundamental business activities created by human beings for economic gain; so there must be sufficient similarity in their structure and organization, at least so they can be planned using a common framework. What we claim here is that three of the five expanded dimensions of planning introduced above, can address most of the gaps dividing manufacturing and service functions.
A common definition for tasks.

Manufacturing (or technology) and service functions are typically differentiated by
the resource groups required and their respective skill sets. The popular view of the
manufacturing or technology intensive tasks is that they are more of an engineering
or applied science, and therefore the parameters of the activities are governed by
the hard criteria of scientific reasoning and logic. Conversely, service functions in-
volve transactions at a human level within or across the organizational boundaries,
and therefore require a different set of skills and resources in the form of trained
professionals. The parameters of the tasks are thus heavily dependent on the types
of resources utilized, and these parameters are also more diffuse in their definition.
However, it is our contention that there are always service aspects to manufacturing
roles, while many service tasks have elements that are engineered or are part of a
broader process that is rigidly structured.

For example, product development is a function that involves the less structured
human interactions, where designers try to capture customer needs and preferences,
and then to engineer products sometimes utilizing scientific methods and technolo-
gically sophisticated resources. Information technology is another function that
involves both understanding and delivery services to the user community, while en-
abling business transactions through structured and logical data/information manip-
ulation methods. It is also possible to classify creative or knowledge work (such as
tasks assigned to marketing or strategy professionals) as a service performed for the
organization or its customers by the individuals playing those roles; but where such
knowledge work is founded on certain scientific principles or utilizes technologically
advanced resources (although in some cases, the knowledge work rests entirely on
human creativity).
The question we address here is whether one can provide a common definitions for service and technological (or production) tasks, which could then serve as a foundation for a common framework to plan such tasks across a supply chain. We argue that this framework is indeed achievable, and propose an axiomatic definition for elemental tasks in general organizations or supply chains.

We define tasks based on four sets of attributes:

1. **Requirements or Specifications**: outlining the purpose, the means, and the timing of the task requirements;

2. **Information sets** relevant to fulfilling the requirements; such information would encompass the instruction set for performing the task, and would also explain the connections with other related tasks in the organization; information sets could also include feedback from other agents or tasks;

3. **Resources** that are needed to complete the task: their availability and deployment mechanisms can also be included in the task definition; resources could encompass raw materials, personnel, supervision, equipment, or other inputs to the task.

4. **Performance criteria**: these are the criteria to examine a range of outcomes for the task including its degrees of completion, qualifying conditions for failure or success, or in general defining measures that capture the impact of the task on the organizational or supply chain performance.

For example, consider again the production planning example that involves planning the quantity and timing of production of a certain stamped sheet metal item
over some time period into the future. We are interested in first defining the sheet metal stamping operation. Using the above axiomatic approach, we could either define the task at its most basic elemental level (stamping a single item), or at a more aggregate higher level where we not only specify the elemental stamping task, but also the quantity and timing of the batch production.

1. Using the latter approach, we would define the requirements to be the stamping of a pre-specified number of sheet metal items, at a particular plant location, over a period of one week. The requirements would also specify in broad terms the mechanism or the methods for stamping such as the die specification, the weight of the press, the surface temperature of the die, and rate of the lubrication, among many other process parameters.

2. The equipment that can be used for the stamping operation would be specified, along with any preferences. Further the supply of sheet metal, fixtures or dies, and other resources such as lubrication, inspection gauges, material handling equipment, would also be detailed. Personnel for carrying out the stamping operation at the station, along with supervisory or quality control agents would also be considered task resources.

3. The information set for the task would include the specification of the materials, real-time status of supply of the raw materials or potentially of orders/demand during the task window; feedback on the quality of the items in the batch are also typically provided.

4. Criteria for success or failure at the task would typically include the timeliness of order completion, the inventory costs both on the supply and the output side, or the rate and severity of defects. Other criteria could be applied such as the safety record of the personnel or of the equipment, and adherence to
maintenance schedules.

Now, let us consider using the same attributes to completely specify a service task, again in the same sheet metal stamping facility. The quality control agent in such operations is usually on call, and attends to the stamping operator to examine sheet metal, die, or stamping process defects and helps adjust the parameters to eliminate or minimize defects.

1. The role of the quality or inspection agent would be specified: examine sheet metal samples, record defects, categorize them, and problem solve the root causes, and provide feedback to the stamping operators.

2. The equipment that can be used for the inspection would be specified, along with other resources such as manuals, and reference literature, and databases for enabling problem solving.

3. The information set for the task would include the process parameters and settings that produced the sample, customer requirements and tolerances, and updated rules for determining the appropriate certificate to issue for the sample (pass or fail, for example).

4. Criteria for success or failure at the inspection task are typically the rate of defective items being produced, the likelihoods associated with the inspection such as the probability of a false positive or a false negative, etcetera. Other service criteria would include response time to defects, and problem solving lead times.

While the above example was constructed from a manufacturing setting, we could provide identical characterizations for a fast food restaurant process, or for a call cen-
ter or a bank’s customer service operation. Finally, as is evident from the production planning example, tasks are not required to be atomic: they may represent a collection of tasks at a more granular level. In operations terminology, a higher level collection or sequence of tasks with identifiable set of specifications, resource groups, information sets, and success criteria would constitute a project.

A service oriented and networked organization.

Once tasks have been constructed axiomatically as above, we can begin to develop a representation of the supply chain as a network of tasks that are interconnected by service transactions. In order to define these linkages, we need to assign task responsibilities among the supply chain agents; with this each task is assigned at least one owner. Each task is also assigned at least one customer agent. The customers for a task may be other tasks, and by translation the agents responsible for those customer tasks; however, customers may also be agents not assigned to a task.

Agents responsible for a task play the role of servers in the service transaction, and respond to the task requirements laid out by the customers. The organization (or the supply chain agents as a collection) provide the necessary resources to the servers, along with the information infrastructure that can generate the information sets needed for task completion. The criteria for success are again typically defined by the customers, but may also be determined by the organization. These linkages of server and customer agents may cross firm boundaries, and therefore create a networked supply chain organization with accountability and requirements defined at the atomic task levels.

A good example of such cross-firm linkages is in retailer-manufacturer relation-
ships. The retailers are typically the buying agents and would be the customers for the manufacturers who supply one or more items for retailing. The retailers would be responsible for providing the demand forecasts and firm orders to the manufacturers, along with feedback on end-customer preferences and quality of the products. The criteria for success would be the inventory related service levels, or the lead time for order delivery achieved by the manufacturer, along with the quality measures for the product. In this example, the task and the service transaction have been defined at an aggregate level: the production of individual items would constitute lower level tasks, and so would the warehousing and transportation of the finished goods to the retailer.

Finally, all the linkages throughout the supply chain organization would possess attributes of general service transactions, including varying degrees of server or resource heterogeneity, the real time nature of service interactions, and the intangible nature of certain service tasks. For more rigidly structured “manufacturing tasks”, resources such as equipment would exhibit less heterogeneity, and the output of such tasks (such as production batches) could be inventoried, while the quality of the service could be measured along tangible dimensions. For some others, including creative or knowledge intensive tasks, the service attributes would be more skewed towards the less tangible, heterogeneous and more interactive. However, the key idea is to create a common framework of tasks and customer-server linkages that can encompass a broad range of fundamental task types found in a typical organization, and further utilize this common framework to create a supply chain of task and agent linkages.
Figure 3.8: A service oriented architecture of the organization and supply network.

Managing knowledge and information processing tasks.

As a final step towards bridging the divide between manufacturing and service oriented processes in a supply chain, we need to address the increasing importance of the third category of tasks in any general organization: namely tasks that either contribute to or leverage the knowledge and information base of the organization, or those that support critical decision making. In many organizations, such knowledge work is categorized as a service or administrative activity that enables the core tasks and processes. Here we argue for a more central role of information processing tasks agents in the organization and supply chain. We propose a multi-dimensional role for knowledge tasks. We further suggest that the primary objective of such information processing and synthesis tasks should be to augment and enrich the core value-adding processes and constituent tasks created from the service oriented architecture of the organization.
Information processing tasks have nearly identical definitions and attributes as any general task. The major differences lie in the following:

1. The requirements which are mostly guidelines or objectives determined at higher levels of the decision-hierarchy; for example the objectives may be to feed real-time information to individual tasks (or responsible agents) on the status of related tasks, or on the external factors that impact task processing. In other cases, the information may be in the form of feedback on task performance, while in some others the objective may be to update the definitions (along any of the attribute dimensions) of tasks to reflect a changing environment.

2. The resources used are increasingly a combination of highly skilled or trained professionals and sophisticated computing resources. These resources are also
frequently shared across a wide range of tasks and higher level processes; sometimes for efficiency reasons, and other times for the purpose of serving/connecting various tasks and distributed agents through common information channels.

3. The knowledge or information tasks are also different in the volumes of data from the external environment they synthesize and share with the distributed agents. The information sets local to a task are therefore often subsets of the information generated by these knowledge tasks.

As mentioned above, one of the key objectives for such knowledge or information processing tasks is to effect a dynamic or real-time redefinition capability for task definitions, including changing specifications and performance criteria, updated information sets, or indeed to augment or reallocate task resources. These redefinitions may also include agent-task assignments, based on updated information or objectives determined at higher levels of the decision-hierarchy.

To the question of how these information or knowledge processing tasks bridge the manufacturing versus service divide in most planning approaches, consider the fact that most corporations today capture and store vast amounts of data and spend billions of dollars processing them to enable their core operations and processes. Currently, these information technology divisions are structured either as R&D organizations, or as administrative services that support the core manufacturing or service functions; indeed in the majority of cases the critical manufacturing and service processes are both administered and supported by common information technology and knowledge processing agents. These information processing tasks therefore constitute the nervous system of the organization, as they bind the service and man-
ufacturing tasks to accomplish the organizational objectives.

In a supply chain setting, again such knowledge of information processing tasks are the interfaces between tasks that lie within different organizational boundaries. Agents from two or more organizations are assigned the responsibility to communicate information across the firm boundaries, and to negotiate and determine common responses to problems or circumstances that impact multiple firms. These are also the agents who relay task related data across the firm boundary, and synthesize reciprocated external information to reconfigure internal tasks. As such we believe these knowledge or information processing tasks play the role of central nervous system for a service oriented network or supply chain.

To summarize our development of the augmented planning framework so far, we have defined five dimensions along which we propose to extend the planning concept of the 20th century: distributed agents and objectives, a decision hierarchy among the agents, information sets and selective access rights, an incentive framework to guide decision making within such hierarchies, and finally mechanisms for the dynamic updating of agents, tasks, and other definitions for the organization. We propose using these five augmented dimensions to expand the scope and objectives of a typical planning exercise from a corporate to a supply chain level. In order to bridge the gaps between structured activities, unstructured service tasks, and more chaotic and unpredictable information or knowledge tasks, we further propose a common definition for such tasks within the extended organization and that are within the scope of the planning exercise. The tasks can be defined completely with four sets of attributes: requirements or specifications, resources, performance criteria, and information sets. In typical situations, the specifications and performance criteria could be expected to be subsets of the information set for a given task; however, this infor-
information set frequently incorporates data external to the task, and also that generated by the task itself. Finally, we propose a service oriented architecture or definition for the organizational processes being planned: the idea is to couple tasks (and their responsible agents) as customer-server dyads. These dyads chained together define the service oriented process that is the target planning environment.

3.6.3 Building collaborative planning capabilities for the service oriented architecture.

The natural question therefore is how we can utilize such a service oriented process network to create a collaborative supply chain, where the activities along individual dyads, or even the performance of individual tasks are in correspondence with the objectives and needs of the extended organization. Using constructs we have already defined, we will proceed to illustrate how such collaborative planning capabilities can be developed. Note that we distinguish between collaborative capabilities in execution, and collaboration in the planning exercise. Collaboration in execution is achieved primarily through mechanisms that allow agents responsible for coupled or linked tasks to interact and to create visibility across the information sets for the agents. Collaboration in execution is also achieved when agents can learn and implement joint decision-making and joint problem-solving at the task level.

Collaboration in planning on the other hand requires commitment by the distributed agents to the supply chain planning concept at two levels:

- To devise and accept specific configurations of the five expanded dimensions of distributed and joint objectives, task definitions and information sets, decision-hierarchies, incentive frameworks, and rules for dynamic updating of such configurations; in other words agree to the parameters of planning at the supply
chain level.

- Secondly, these agents must actually plan the tasks and processes within the service oriented architecture. That is, they must define tasks based on requirements, available or potential resources, performance criteria, and other information. They must also assign task responsibilities among agents, resolve conflicts through the incentive frameworks, and create managerial or knowledge tasks to enable the emerging processes.

Thus, collaboration in *execution* is partly an issue of infrastructure (to allow real-time interaction and communication among agents), while collaborative *planning* is primarily a conceptual and managerial challenge. How do we get the distributed and decentralized agents to commit to the planning concept itself, and secondly to actually conduct the planning exercise in a joint fashion? In this sense, collaboration becomes synonymous with coordination, where the decentralized decision making somehow achieves the same outcomes as centralized planning. We therefore propose that collaboration in execution is better handled by technology and infrastructure development, while we here we develop mechanisms that allow for collaborative and coordinated planning in supply chain settings. The execution challenge of joint decision making at the task level however is one that requires us to deploy decision-making hierarchies we defined earlier and that allow for such independent problem solving at the task level, which we also address briefly below.

*Decision-making hierarchies that enable collaboration and coordination.*

Here we pose the fundamental question: what are the characteristics of decision hierarchies that enable collaboration among distributed and decentralized agents in
planning and execution? Furthermore, it is of critical importance that the collaboration be effective and achieve the goals of the planning exercise. We argue that decision hierarchies must satisfy all of the following criteria for this purpose; these are the necessary conditions to achieve collaborative planning, but by no means the sufficient ones.

1. The first proposition is that decision-hierarchy must be exhaustive in terms of representing the distributed agents in the supply chain and also exhaustive in terms of the tasks defined within the service oriented architecture. In other words, every agent-task pairing in the must have some decision-rights in the hierarchy. Otherwise, the decision-hierarchy does not capture all of the interests and objectives in the organization, either from a task performance perspective or from an individual or group interest perspective.

2. The second claim is that decision-hierarchies must be composed of distinct subsets of agent-task pairings that are contained within the grand set making up the supply chain organization. In other words, there cannot be an agent-task pairing at any level of the hierarchy that is not considered part of the supply chain, and that does not participate in the service oriented organizational network. Thus, every agent-task pairing must either be a “customer” or be serving some other such pairing in the service oriented architecture, and therefore must have well-defined commitments and interests in the planning exercise.

3. Thirdly, every subset of agent-task pairings must be provided independence in setting objectives, and further in decision-making over the range of tasks associated with this hierarchy. Without this independence in decision-making for the selective sub-domain of the service network, we risk lack of control in task performance for the responsible agents.
4. Fourthly, we propose that decision-hierarchies assign differential authority to subsets of agent-task pairings so that it is possible to create an ordering of these subsets to determine the rank or level of any subset in the hierarchy. This ordering need not have the usual transitivity, completeness and reflexivity properties of the rational preferences, but it must be transparent and evident to all agents in the network. Differentiating authority for the subsets allows us to create an ordering so that conflicts in the extended organization can be minimized and resolved to the fullest extent possible; since lack of differentiation between subsets of agents may yield decision-hierarchies that provide incoherent resolution of conflicting objectives in the organization.

5. As a fifth characteristic, we propose that this differential authority across the subsets of agent-task pairings can be achieved through parent-child relationships. The properties of this relationship must allow the objectives of all the child subsets to transfer in whole to the parent subset, while the decisions are enforced in the reverse order. Every subset, except one, must have exactly one parent subset of agent-task pairings. This special subset at the top of the hierarchy that has no parents represents the objectives of the entire service oriented network (or extended organization), while it also carries the burdens for enforcing the decisions for the entire organization. Note that we do require that this highest ranking authority be composed of only one agent; indeed we could have situations where this subset of pairings with the greatest authority represents multiple agents from within the supply chain.

6. Finally, we propose that sharing of agent-task information sets within a particular subset, and up and down the hierarchy be declared upfront. In some cases, these information sets could be kept private by the agents, while in other cases the agents could share the information sets to realize the best outcomes.
Thus, the degree of visibility into another agent-task pair's information set is to be determined as part of the rules for collaboration.

We now discuss how these characteristics in a decision-hierarchy can enable collaborative or coordinated planning. The first such enabler is the construction of the decision-hierarchy from elemental agent-task pairings. Doing so not only allows subsets of agents to define their objectives and rules of collaboration in terms of impacted tasks but allows the organization to see the linkage between agent decisions and impact on task performance, but also conversely allows the agents to see the impact of task performance on their objectives. The parent-child relationship further allows us to relate task performance at any lower level in the hierarchy to the objectives and decisions at progressively higher level in the hierarchy. Independence in objective-setting and decision-making within a subset of agent-task pairings allows complex and large planning problems involving larger subsets of agent-task pairings to be broken down into successively smaller (at least in scope) planning sub-problems: the decisions at lower levels in the hierarchy are communicated up through its parents, where they may be considered as inputs to the planning problem at the higher levels of the hierarchy.

Next, we argue that a decision-hierarchy with properties as defined above requires an accompanying and tailored incentive framework to achieve collaboration. Without such an incentive structure, the decision-hierarchy would not represent true collaborative planning. The reason is quite simple: in the hierarchy as we have defined above, it is possible that the decisions at the highest level of the hierarchy (that are typically in the best interests of the supply chain may not meet the criteria for good or efficient decisions at a lower level. If the lower level decisions are decentralized,
then this is an unsustainable scenario. Another rationale for incentives is in collaborative decision-making within a particular subset of agent-task pairs. Even within a particular subset, different agents may again have to arrive at a compromise or even an optimal decision, which may again be in conflict with the optimal sub-problems for each agent-task pairing.

The problem here is to ensure incentives (or disincentives, as the case may be), in order to avoid such conflicts among agents in decision and the intended outcomes for the distributed agents. These incentives can come in the following forms:

1. A re-distribution of costs or gains so that each agent within a subset always prefers to arrive at the decision of a parent or even a partner agent (in that subset), rather than persisting with a decentralized decision.

2. A sharing of risks from decisions so that on some measure defined over the uncertain outcomes, again, each agent prefers the decision favored by the collective.

3. In some situations, the incentives could be directed at the higher level agents in order to accommodate and resolve conflicts in favor of the lower level or marginal agents in the supply chain.

4. There may also be a need for incentives to share greater information among agents within a decision group.

We will see examples of the first two types of incentives in Chapter 7. The third type of incentive frameworks are more common in decision hierarchies where the higher level subsets within a tree are composed of only one agent; in other words, one firm has control over the decision-making of all the child subsets of agent-task
pairs. In these situations, there is a need for incentives for the single agent to accommodate the interests of the child subsets. This is seen for example when a lead firm partners with a contract manufacturer who in turn has decision-control over component suppliers; in the case of a few mission critical components, the lead firm may provide some incentives to the contract manufacturer to accommodate the special needs or objectives of this smaller supplier.
4.1 Introduction To Supply Chain Programs For Innovation.

4.1.1 What is collaborative program management?

In many cases, firms require the help of a number of supply chain partners in the development and deployments of new products, processes, and improvement initiatives. Not surprisingly, the terms of the partnership have a pronounced impact on supply chain performance. Going beyond a product focused effort, suppliers are critical to launching successive generations of a product, technology, or service, and for delivering and managing product and service variety. Some recent and popular examples can be found in diverse sectors of industry including consumer electronics; continuous improvement initiatives in the fashion of Toyota Motors, and large scale technology initiatives such as the RFID implementation drive currently led by Walmart corporation. Figures 4.1 and 4.2 depict frameworks implemented by Boeing and IBM to govern critical programs that were
part of their business portfolio during the last decade.

We define a program as a collection of interrelated projects and activities, focused on the common strategic objectives of one or more organizations. A project is often a largely self-contained subset of the those activities, often achieving the tactical objectives, while tasks could be defined as activities that have an operational view. We also see cross-firm and cross-functional programs as the driving force of numerous supply chains. Many authors in the field of management, engineering, and systems development view program management, separate from the supply chain context, in similar light; and see it as a framework to manage multiple related and concurrent projects [6][149] [187][188]. Even in practice, programs are frequently defined in terms of the design, control, and management of the constituent projects and tasks.
The literature on the topic when viewed within the supply chain context, however, is surprisingly sparse given its immense importance for most businesses and corporations. This could possibly be attributed more to semantics than to an oversight on the part of management scholars; in essence the concepts from the widely studied field of project management port over quite usefully to many program environments [119][120].

However, it is our belief that in the past two or three decades since supply chain management has gained prominence, program management both as a requirement, and as a managerial framework has assumed greater importance for most businesses. Some firms now use program management principles as an underlying framework or foundation that supports all of their supply chain related efforts. This trend has
been accelerated in the past decade or so with firms out-sourcing their work to their partners at unprecedented levels. At the minimum, there is now widespread recognition of the important role of programs in supply chain management and control; and in some cases to the extent that the two terms are synonymous. This last phenomenon is evident in some industry sectors such as high technology development and manufacturing, especially in the defense sector. Some firms in the technology sector also use program management to structure much of their account management and customer related activities.

Programs are executed by interdisciplinary and cross-functional teams with complementary skills, coalesced from multiple firms who partner to accomplish the specific program objectives [129] [173]. To break this concept down in parts: typically each firm brings its own unique capabilities to the program, although these capabilities could be common to more than just one firm. The program organization is frequently composed from the groups or teams contributed by the participating firms; what makes supply chain programs special is that often these teams and their members operate with varying degrees of autonomy from their own firms. In many cases these teams or groups are instead responsible to the program leadership for achieving the goals and objectives set by the program. What makes this organizational setup sustainable is that quite often, programs are conceived to support the strategic imperatives of the collection of firms. In turn these goals and objectives require and are the result of complex negotiations at the strategic, tactical and operational levels of the partner firms. This common understanding of the goals and objectives legitimizes the cross-functional teams or groups to mobilize resources from different divisions of the partner firms, plan and coordinate the infrastructure development, plan the material logistics, and then execute the activities entailed by the program.
For a more detailed example, consider the case of the new Boeing 787 Dreamliner program - intended originally to design, develop, and build a medium size, long range commercial aircraft by year 2008, that provides 20% greater fuel efficiency and greater flying comfort than comparable offerings [107] [157] [186]. These innovations come at significant cost - to the tune of $13B – hence in an effort to minimize its own exposure to program risks, Boeing instituted a “build-to-performance” design and development approach. A key supplier would be given broad specifications for its sub-system or components, and the latter would collaborate with other second tier or laterally positioned suppliers at its own cost [146]. However with this new approach, there have been major delays in the program since, with Boeing announcing several changes to the schedule [90]. Many of these delays have been attributed by Boeing’s own leadership to inadequate supply chain and program planning, and to a flawed supply chain strategy [150].

One consistent theme in the examples we cite throughout this chapter is that the structure of the program environment has great influence on whether and how the supply chain reaches its objectives and performance targets. We define program structure through the set of firms that participate, the terms of their engagement; by the manner in which the program assign roles and responsibilities to these firms, and the planning of the entailed projects and tasks.

We argue that programs are distinguished from other organizational forms in supply chains by the following:

1. Programs have a clearly stated purpose and mission that drives all of the constituent projects and activities;
2. They must have committed and strong leadership, and the authority to formulate or at least influence strategy at the participating firms, and also to resolve operational problems;

3. There needs to be a clear understanding of which firms are partnering in the program, to what extent, and in what roles;

4. They must have a financial, human, and technological infrastructure and resource pools to plan and support its activities;

5. Finally, they must have clearly stated and commonly agreed upon rules of coordination and governance.

See Figures 4.3-4.7 for a more detailed understanding of what we term program structure, and also how the notion applies in different program environments. In companion chapters 6 and 7, we use a decision-theoretic framework to address the third, fourth, and fifth aspects of the program structure above; that in turn aim to resolve questions of partner selection, resource assignment, capacity determination, and coordination of program activities.

Apart from developing a decision-support tool, an important objective of this research is also to highlight the pervasive influence of program risk on all aspects of its management. We argue that risks in the program environment often arise from a small but fundamental set of sources:

1. Market or customer related risks; where the effort expended by the program is closely aligned with the needs of some process, a set of customers, or that of a market segment which in turn are not completely predictable or knowable in advance of program planning;
2. Internal resource efficiency or productivity risks, when resources or groups devoted to the program operate at efficiency levels not foreseen by program planners or marketers;

3. Complex task and project level interdependencies that make one firm vulnerable to the choices, performance, and outcomes of other firms working on related or common tasks;

4. Risk from collaboration and cross-firm interactions, whereby there could be inefficiencies, for example work requirements escalation, from the inability of firms to communicate or work together seamlessly (perhaps in turn due to inter-operability issues); and

5. Program level interdependencies, where local decisions such as capacity commitments by a firm, or local task assignment strategies on the part of the program planner, can have program wide consequences.

Subsequently in Chapter 6, we develop a unified modeling framework that attempts to captures all of these different sources of program risk, and incorporate them into the key decision processes. It is useful to comment here that risk entails both an upside and a downside to outcomes for firms; and in a program setting, increased risk for one particular firm, while undesirable to that entity, may work towards lowering the risk for the collection of firms at the program level. This is captured by the notion of risk sharing in supply chain programs; we try to understand the basic need for risk sharing in programs, and derive some related cost sharing mechanisms that effect risk sharing in a collaborative setting.
4.1.2 Collaborative programs as the foundation for supply chains.

Many programs have a lead firm or central organization that requires the participation, investment, and sustained effort of a number of different supplier firms towards a focused objective; such as a new product or service process introduction, environment-friendly technology shifts, business process re-engineering, or cost reduction efforts in retail or assembly operations. The success of the supply chain venture therefore depends crucially on the success of the constituent supplier projects - failure of even one supplier to complete their share of the work or to achieve their program related objectives could result in failure of the supply chain in some specific ways. Hence, the supply chain as an entity, is in some ways conditional on the successful execution of the program objectives.

Going beyond the operational (or execution) details, it is also our belief that many supply chains, large or small in scale, are founded in the form of collaborative programs. As such the firm or partnership structure of supply chains is best under-
stood from a program management perspective. Questions such as which firms get to participate, to what extent, and the ways they contribute their expertise and capabilities in the supply chain, are thus negotiated and resolved largely at a program level. Thus, in some cases at least, the structure of a supply chain and its founding program are indistinguishable. Hence, this perspective allows us to cast important supply chain questions such as partner selection, supplier capacity determination, resource flexibility and task assignment in terms of their impact on program objectives.

For example, Boeing and its suppliers currently have the well-defined program objective of developing the Dreamliner, and delivering against their backlog of orders profitably to their airline customers [55]. For these objectives, Boeing has gone...
through an extended phase of identifying the key partners and their specific roles in the development process, pilot and volume manufacturing, and beyond through the program lifetime. The partners, individually, and as a group have determined the capacity and investment required for the different phases of the development, testing, and volume manufacturing. Together, or interpolated with the capacity decision was also the exercise of sharing the resources (or plans to do so) and allocating the small and large tasks among the partner firms. Similarly, on the revenue side, the firms have collectively decided, with leadership from Boeing, the terms of the revenue sharing and also the intellectual rights to the technology that is developed during the design phase. In other words, the entire supply chain structure (possibly for the duration of the product life cycle) has been shaped and determined in a program context, and in ways that are most responsive to the program objectives.

**Figure 4.5: New product development programs.**
For another prominent example, Wal-Mart and the US Department of Defense have independent rolling initiatives with their key suppliers to implement RFID enabled information systems in their logistics chain [136] [160] [161]. They have also undertaken a partner selection step, the technology or process design phase, a planning phase to determine the best way to roll out the initiative at the partner firms, and have determined the roles and responsibilities for the partner firms once the new RFID enabled logistics processes are put in place. Again, the supply chain structure here is shaped, if not founded, based on the firms’ decisions on how best to achieve the program objectives.
In the health care domain, the World Health Organization together with the governments of industrialized countries has a program currently underway, to enable vaccine manufacturers in those countries to mass-produce and distribute the vaccine globally to protect against a pandemic spread of the H1N1 influenza virus [132]. The program objectives helped guide the selection of the primary vaccine producers, the determination of capacity requirements at the suppliers for production and for delivery, contractual terms such as cost and risk sharing, and wholesale pricing of the vaccine. Here again, the structure of the entire vaccine supply chain, at least for this specific product, was shaped by the broader program goals.

It is therefore an aim of this and the next several chapters to formalize this translation of program objectives to supply chain structure; as we will discuss shortly, we do this by framing the structure questions through a unified decision-theoretic
model. What is interesting about this decision model is that it is a participative model (as indeed required for most program environments), as opposed to an exclusionary one where one firm makes the decisions for the entire supply chain. As such there are elements of game theory and coordination in the formulation of the program planning problem.

4.2 Organizational Structures, Decision Processes, And Their Interactions.

4.2.1 Firm roles in program management.

Often, one main function of the lead firm or of some other responsible agent(s) at the program level, is to allocate broadly defined projects or even individual tasks falling under the program umbrella, among the set of firms or resource groups willing or able to do such work. The firms participating in the program could have different ca-
abilities, may exhibit different cost structures, and therefore may commit different levels of capacity to the same task; as such they may also have different expectations for payoffs. Given the characteristics of the participating firms, and the relative differences in their capabilities, the program planner or the lead firm has the onerous job to determine how to allocate work arising from task requirements among the partner firms.

Firms participating in a program model typically independently decide the resource capacity and investment to contribute towards the program. In other words capacity investment is a firm level decision and the program manager cannot directly dictate the level of capacity for a resource owned by a firm and assigned to the program. This follows naturally from the way many firms are structured today; their primary responsibility is to their own shareholders. Customers and supply chain partners are valued only in the sense that interactions with them contribute value to the shareholders. That being said, there are instances where the business relationship and contractual obligations may allow the program manager to dictate capacity levels of resources contributed by the partner firms. In the companion chapter 7 we will consider environments where the capacity investment decision is a decentralized one; in that only the firms owning the resources can ultimately determine their capacity levels. However, we also formulate a centralized capacity problem (assuming the necessary information transparency); this is so that we can highlight the discrepancy between the centralized and decentralized solutions in typical collaborative programs.

The capacity decision – being relegated to the firm level – is therefore be guided primarily by firm level considerations, as opposed to overall program level value. Individual firms often look to maximize their own value or return on investment in the program, and will set their capacity commitments according to those firm level
objectives (except perhaps if the firm in question is also a program leader); and firm level decisions do not directly maximize program value nor do they minimize program level costs). Rather firms will set their capacity levels and investment in response to the value they can expect to receive from participating in the program. However, the program planners do have some leverage in the capacity decisions: the central decision-maker typically has the authority to determine the task assignment, and secondly may in some cases the sole authority to write contracts with all of the other partner firms. In this way, the program planner can at least ensure the basic conditions for a firm to determine its capacity in ways that will support the overall program value. This is the decentralized decision-hierarchy with some leverage for the central planners or program managers that we will seek to model in the companion chapters 6 and 7.

In those chapters, we will only encounter situations where the value derived or cost incurred by an individual firm is proportional to the program level value or cost objectives. However, for a given level of capacity commitment by a firm, the share of program value for a given firm in turn depends on the capacity investment decisions of the other participating firms. Of course, the task assignment decision influences the capacity investment: firms that are not assigned a particular task do not have any incentive to invest in capacity towards that task.

4.2.2 The task or work assignment mechanism.

In determining the work assignment in this manner, the program manager has to balance a number of different competing factors, both hard accounting measures, as well as more subtle business relationship concerns. From an operational perspective, the assignment of tasks to the different firms and resource groups should in theory
take into account the available work processing capacities of these resource groups, the cost structure at these firms and their relative differences, and the reliability and efficiency of the firms in completing their share of the work. The program manager would typically be guided in the assignment decision by the cost objectives for the program, constraints on the program budget, the value (or revenue) derived for the program under one assignment strategy versus another, and also constraints on the work-share of different firms. The last is often with a view to limiting program exposure to performance risk at a certain firm, or conversely to ensure minimum business volumes for certain others.

From a business strategy and relationship perspective, there could be other factors to consider— for example, the length of history between firms, geographical and international trade considerations, future competitive threats and intellectual property issues, among others. If a firm that is a partner for the current program is also a competitor in other markets or programs, then firm with more to lose from any competitive challenge would try to limit the collaboration to the extent economical. Firms also receive work-share in supply chain programs because of regional market considerations. Boeing has historically allocated program work-share on its flagship programs based on international marketing considerations; for example, the Japanese airline market is a significant customer base for Boeing, so many of its key suppliers happen to be from Japan, market geography playing an important role apart from the technical qualifications and competence of those suppliers [153]. This is despite emerging threats of entry by these key suppliers in Boeing’s lucrative commercial aviation market.

On the other hand, firms may require non-compete or other contractual clauses to allocate major work to its partners who may be competitors in other spheres.
For instance, on the well-known Microsoft Xbox360 program, Microsoft collaborated with IBM in the early 2000s, for developing the design and for the fabrication of the CPU for its game console product. Interestingly, Microsoft decided to retain intellectual property rights to the chip design, and certain aspects of the fabrication process. This was motivated by contractual issues it faced working with Intel on the CPU for the previous generation of the Xbox product. Intel, it turned out, was not as flexible in reducing chip prices in concert with Microsoft’s market penetration plans, and held the rights to both the chip design, and the fabrication process of the “Pentium” line of CPUs. Thus, with the current generation of the Xbox360 CPUs developed by IBM, Microsoft can in theory allocate the manufacturing work to the most flexible and responsive fabrication partner in the program, and does not have to rely on IBM alone. Interestingly, IBM was also a major partner in the development of the CPU for the competing Sony Playstation3 program [4][181][182].

4.2.3 The coupling of the assignment and capacity decisions: The role of negotiation.

A work assignment strategy or mechanism is not predicated on the available capacity alone. Rather for many supply chain programs, the assignment of tasks and their work requirements depends more directly on the offered, tendered, or committed capacity. For example, a firm bidding for a certain work-share may not have all of the offered capacity online, but based on the assignment strategy would ply fresh investments in capacity to support its share of the program work. As such, we view capacity devoted by a firm or resource to a program as an investment towards the program objectives. In this view, firms that have some pre-existing capacity to devote to the program have lower overall cost of program capacity than firms who need to develop all of the capacity tendered or committed to the program.
To revisit the Microsoft case: for the first generation of the Xbox program the company contracted with Intel to supply off-the-shelf CPUs over competitor Advanced Micro Devices (AMD); the negotiations for the CPU supply were based less on technological or design/development concerns (the Graphics Processing Unit—or GPU—is considered more central to the game console product design), and instead emphasized operational and pricing concerns. Microsoft needed a supplier who could provide chip-making capacity for the intended high volume manufacturing of the Xbox, and one who could do so at reasonable prices off the block. Microsoft also need a supplier who could provide ramp capacity within the relatively short time-frame of launch. Intel was able to promise expensive chip manufacturing capacity, both for ramp-up, and mature supply phases to Microsoft at no additional fixed cost; Intel had sufficient capacity already in place for its existing Pentium line of CPUs, to guarantee the required volumes to Microsoft.

However this capacity flexibility came at a higher unit marginal cost to Microsoft, and with less leverage in the relationship and lower operational flexibility over the long term in terms of cost reduction capabilities. AMD on the other hand (not being as large and dominant as Intel) had capacity limitations; the chip-maker wanted Microsoft not only to invest in additional capacity, but to guarantee break-even volumes as a hedge for its own capacity investment. It turned out that AMD had a successful product launch during the year of the negotiations with Microsoft, and this product had consumed most of AMD’s excess capacity[180].

In the models we analyze, we assume a strict separation in the decision responsibilities of the program agents; the firms only decide on capacity, while the program manager determines the program optimal task assignment mechanism. However,
since the program value depends on the firms willing to commit capacity at a levels acceptable or even optimal for the program, the task assignment mechanism must take into account the value derived by the individual firms from participating. Conversely, the firms take into account the task assignment mechanism while offering their capacity levels.

Hence, even though modeling the negotiation process is hard, we nevertheless capture this aspect of program environment by positioning the decisions of the program planner and the firms in such a way as to give the “first mover” advantage to the program planner. It must be stated, however, that this decision hierarchy is possible only if we assume perfect information exchange between the program planner and the individual firms; any private information breaks our decision model and makes the planning exercise, let alone the coordination issues, that much more challenging.

4.2.4 Program risks from collaboration: Their causes and their implications.

In Part III of this dissertation, one of our objectives is to understand how uncertainty and risk in the program environment impacts task assignment and capacity investment decisions. To this end, we try to model the different dimensions of program risk faced by the participating firms. Some of these risk factors are internal to the firm, but some are external in the sense that firms cannot directly control these risks, but can only react to them through their respective capacity investment decisions. The internal risks typically arise because of inefficiencies of firms in processing their share of a task; the net effect of these inefficiencies is to increase the processing time of the work.

1. Firstly, there could be some variability in the task demands during execution; this is best seen in design or development or even creative work; independent of
the cost advantage at a firm, there could be some escalation in the work content attributed to “unknown unknowns”. While this may not seem acceptable from a work conservation point of view, this reflects reality to the extent that work requirements could escalate after the task assignment step. It is not uncommon for firms to experience requirements escalation when product designs or process steps are revised during task execution for very legitimate reasons.

2. Secondly, there could be capacity related risk, in the sense that the committed or offered capacity may not actually be realized, and the effective capacity could either greater or lower depending on uncontrollable events. For example, sudden capacity disruptions such as plant shutdowns or even personnel turnover can cause firms to default on their capacity commitment.

We present models in Part III that capture both of these types of internal risks. Next, program risks that arise from network interdependencies are considerably more complex, and can occur at multiple levels of the program, and from multiple sources, as we discuss below.

- One form of network risk exists simply because there is an underlying correlation in the work requirements of two or more related tasks. However, this correlation in task work requirements can exist for a number of different reasons which makes this type of risk hard to measure and control:

1. In the common situation where projects or tasks are split across firms and resource groups, a simple and direct correlation can be observed in that increasing the work-share of one firm, implies that the other firms will
together be responsible for less of the task, and so on. We will refer to
this type as competitive or substitutive risk.

2. There could be an inherent connection or natural interfacing between
tasks, and therefore correlation in their work requirements, independent
of whether they are split or shared between resource groups. In prod-
uct development programs, the development of interfacing components
or modules could imply positive or negative correlation in work require-
ments, depending on the type of interface. Similarly, in quality improve-
ment programs (such as Six Sigma initiatives), process improvements in
linked stages of the supply chain require proportional effort, otherwise,
overall product or process quality improvements cannot be guaranteed.
We refer to this dimension as process or design risk.

3. The third possibility is for there to be a correlation among task work
requirements based on which firms are assigned tasks that require collabor-
ation. Even when the tasks are not shared or common between two
or more firms, a set of tasks could require that the responsible firms or
resource groups collaborate in some specific ways to accomplish the task
objectives. In development programs, teams and resource groups working
on interfacing components are required to collaborate to a certain extent
to achieve the performance and quality requirements of the overall mod-
ule. If the firms are handicapped to varying extents in their collaborative
capabilities, then the collection of tasks may be impacted in terms of
their total work content. For example, if a component that has multiple
interfaces with other components is assigned to a firm with limited collabor-
orative skills, then the work content of all the other related development
tasks could escalate as firms resolve more problems arising from the im-
paired collaboration. This type of task demand correlation is the result of firm characteristics; the demand covariations are caused by the collaborative characteristics of the set of firms involved, and hence we term this as collaborative risk.

4. The fourth type of task demand correlation arises because of task scheduling in a time constrained program environment. Tasks often have predefined precedence and succession relationships with other tasks; so it is possible to consider situations when preceding tasks are delayed or consume more resources than planned, leaving less time and capacity to complete a given task. We refer to this type as schedule risk.

- A second form of network or program risk arises from the capacity decisions made by the partner firms on both related or unrelated tasks. In the case of related or interfacing tasks, a firm under-investing in capacity for a given task could cause delays, work escalations, and therefore even higher costs for firms assigned to the related tasks. We refer to this aspect of program risk as network capacity risk.

- Finally, a third and encompassing form of network risk is experienced when the overall program progress is impacted by sub-par performance on one or more key tasks assigned to the various firms. Assuming that firms derive value in proportion to the overall program outcomes it is possible to conceive scenarios where every participating firm experiences losses because of adverse program level outcomes. The case of the Boeing Dreamliner is a good example, where the stock price of key suppliers to the program, including that of Boeing has dropped significantly after successive announcements of program delays. In many cases, firms also experience more direct operating losses as revenues or
gains are deferred and cash flows are depleted as they await program completion. We will refer to this type as the cumulative program risk.

The process or design risk factors described above can be framed or pitched to firms as the technological costs of the program, and can be better understood and explained in the context of the program task structure and work-flows. The caveat here is that with poorly laid out task structures or with excessively complex work-flows, the risks could well become unmanageable; and the program outcomes highly uncertain and more likely to be adverse given such lack of clarity or predictability. A strong technical foundation for program management and careful data gathering can highlight the key sources of process or design risk.

Competitive or substitutive risk can be cast as the result of the relative differences in the resource capabilities and the technological advantages of firms. The reason this is a type of risk is that in many cases these technological advantages are not evident or guaranteed upfront either to the firms or to the program managers, but are inferred from the historical performance of the firms.

Network capacity and cumulative program risks, on the other hand, are actually more of a concern from the point of view of supply chain governance. In essence, supply chain programs could impose costs for constituent firms from exposure to delays or risks in unrelated tasks or externally performed work which could hold up the entire program. As we'll see later, this poses a challenge to the coordination of the program, where the program manager seeks to induce capacity decisions by the firms that will lead to program optimal outcomes.
4.2.5 *Information rules and decision hierarchies.*

The major determinant of whether firms can control for types of network risks is the set of rules that govern the exchange of information and the level of transparency in the program management. For example, with limited knowledge of the partner firms’ capacity investments into related or unrelated tasks, or with limited understanding of how the overall work in the program is shared, firms would essentially be making decisions and operating in isolation and with greater uncertainty about their network interactions and therefore greater exposure to network risks.

In the models we consider in subsequent chapters in Part III, we assume that firms have perfect knowledge of the program planner’s task assignment mechanism, as well as of the capacity decisions and cost structures of the partner firms. However, in our models this knowledge of the partner firms’ capacity decisions is assumed available only on an ex-post basis; that is firms do not have lateral communication with other firms before they determine their capacity. Rather the communication happens bilaterally between the program planner and the individual firms. For each possible capacity decision on the part of a firm the program planner communicates to that firm the overall impact of that decision on program outcomes, and also communicates the payoffs to that firm at that level of capacity. This information structure precludes the possibility of lateral gaming, and formation of coalitions in the capacity decisions of two or more firms. We leave the modeling of imperfect or private information, along with lateral gaming between firms to future research.

In terms of the decision hierarchies, the program management environment is often game-theoretic in nature; whether the interaction is cooperative or non-cooperative depends on the specific environment and on the details of a particular cross-section
of the program (and sometimes also based on the program manager’s pre-defined rules of coordination). Unless the lead firm and the partner firm are competing via their resource pools for the assignment of a given task, in our models, the interactions between the program planner and the partner firms are mostly co-operative in nature. Even when the lead firm and the partner firm do compete for the same task, we assume that the program planner is completely neutral and the task assignment mechanism is not biased in any special way to the lead firm unless the lead firm’s advantages are measured directly in terms of cost structure, efficiencies, or capacity investments relative to the other firms. Also, in the models presented here, the assignment of a task to the candidate resource pools is a binary decision, and as such firms do not receive partial or shared responsibilities for a task (although such a scenario can also be handled within our broader modeling framework).

Figure 4.8 illustrates a much broader hierarchy of decisions from the perspective of different stake-holders within a typical program. We only model a select few of these decisions and issues through the frameworks presented in Part III.

4.2.6 Resource interactions in collaborative programs.

Together with the game-theoretic interaction between the program manager and individual firms, there is also a lateral interaction between firms as they bid for work-share and look to maximize their own value. The interaction could be of three types based on the resource groups operated by the respective firms:

- Competitive and substitutive interaction, where the resource groups from different firms compete for a share of a common task. For example, in the case of the Microsoft Xbox360, Microsoft contracted first with the electronics manufacturing firm Flextronics to manufacture and assemble the product, but later
on added two more assemblers: Taiwan based Wistron, and Canada based Celestica to the program. While overall volumes increased over time, the assembly work share of Flextronics declined even as Wistron increased its share of the shipments to more than 50%. In this case, the three assembly plants could be understood to have a competitive interaction as they compete for a share of a fixed demand from the assignment mechanism.

- Complementary or cooperative interaction, where resource groups, independent of ownership, do not have any tasks in common. In such circumstances, the resources are not vying for task share, and in fact may contribute value to each other in light of the network capacity risks described earlier. Many arms length suppliers to a program interact in this fashion; these suppliers often work on modular or separable aspects of the program, and the interfaces between these modules are codified well enough to allow the firms to make program related decisions independently from each other.

- Competitive and complementary interaction, where the resource groups from different firms have responsibility for multiple tasks not all common to the resource capabilities. In such circumstances, the resource groups may compete for the share of the common tasks, while they complement each other on the other tasks.

- Cooperative and substitutive interaction, where the resource groups belonging to the same firm may be capable of performing the same task. In such circumstances the resource groups may each garner a share of the task and are therefore substitutive, but in fact they are also cooperative since they each contribute value to the common ownership.

- Collaborative interaction, where resource sharing a common task have to in-
teract with each other to complete their respective portions of the work. In reality, tasks in complex programs could require varying degrees of collaboration, and the extent of the collaborative needs, and therefore the impact on the demands for the interacting resource groups could in fact be determined based on the split in the workshare. For example, in the design and development of the Boeing Dreamliner, Boeing outsourced a greater portion of the design of the aircraft to its key suppliers than in previous programs. As a result, the level of cross-firm collaboration required also increased, and according to some reports, to an extent that became unmanageable for the program. Boeing subsequently insourced key steps in the process from those suppliers who could not cope with the increased collaborative requirements.

- Alternative and substitutive interaction, where the resource groups, not necessarily of the firm may be capable of performing the same task, but where one resource is given a task or work share in preference over the other, but where in the event of capacity shortages or resource failures, the alternative resource could be brought online to process the workshare instead. This happens frequently in the case of schedule or service constraints where backup or contingency capacity whenever available, is deployed to protect against resource or supply failures.

In our models, we will capture all of the above interactions except the last, and we reserve some comments in the future research section on issues related to alternative or contingent capacity.
4.3 Program Costs And Accounting Principles.

4.3.1 Accounting for costs in a collaborative environment

Every activity carried out under the program umbrella has to incur some cost; it is typically a matter of accounting to determine the reason for the activity, the utilization and consumption of various program resource groups that were involved, and then measure and record these costs for the purposes of attribution and performance evaluation. Naturally, there are a number of possible techniques that can be deployed towards cost accounting in programs, based on the program environment and its unique requirements.

For example, traditional methods such as standard cost accounting, where the activities are costed based on the incurred fixed and variable costs, would be justified in manufacturing programs where a lead firm and a supplier may have an arms length relationship and where there is little collaboration required by the supplier in fulfilling its program responsibilities. More advanced techniques based on the standard cost accounting principles of breaking down cost into direct and indirect sources, such as marginal costing or throughput based costing can be used for more complex operations at the firms supporting a supply chain program.

However, when the relationship between the firms is no longer arms-length, but when tasks and projects fit into complex work flows spanning the various firms, traditional accounting principles break down fairly quickly. In fact they have the potential to seriously undermine the effective governance and functioning of the supply chain program, especially when there are niggling disputes over cost attribution and budget planning. In such cases, a better alternative for most program environments is the activity based cost accounting method.
The activity based costing method relies primarily on the ability of an organization to list and categorize the major activities carried out within a program or a project, map these activities to resource groups operated by the organization, and to attribute them to various value adding objects such as projects, products or other “services”. The total cost of the activity can be measured by mapping each activity to the resource groups consumed, and by maintaining and updating a “pricing rate or sheet” of the resources used by the activity. A given activity may support a number of such projects or services, and by measuring the frequency of attribution to each of them, the total cost of the activity can then be spread across the products or services for performance evaluation. In this way the cost of operating the resources is eventually attributed to value adding products and services and a profit and loss balance statement can be drawn for each of the products and services. We refer to [27] for further details.

This method of attributing activity costs to the products or projects being serviced fits naturally within a program environment. Resources, both human and technological, are typically shared by a number of different tasks, while these tasks in turn may support a number of different program objectives and at different stages or phases in the program. Many program environments are also subject to a high degree of operational uncertainty, and it is therefore difficult to predict exactly which activities will be entailed in the completion of a given project (“unknown unknowns”) and the rate at which resources will be depleted by these activities. As long as activities can still be categorized into accounting buckets, the method will still be valid; in certain situations, activities could be defined and entered into the accounting system as the program evolves. The same idea applies to program resources as well; over time a program will add some resources and retire some others, but as long as the
program has the means to inventory these resources and price their consumption, it would still be possible to cost the services rendered to the program.

The issue of resource pricing is important from the perspective of firms owning the resource groups. These resource groups would typically be listed as assets owned by the firm, and therefore the firm’s shareholders would demand that these assets yield “returns” or add value to the firm. When the firm contributes or offers these resource groups to the program, it then tries to price the consumption of these resource groups based on time or service units; typically this resource consumption pricing offered would be somewhat independent of the level of the demand, but one could cite examples where considerations of economies or dis-economies of scale come into play. Often the programs are also charged, in part or in whole, for the fixed cost of the resource when the asset is purchased specially for the program. In such cases, the participating firm is essentially an investor in the program, and here considerations of return on investment, and asset yield, are ever more prominent.

This raises the issue of how to price the resource groups and activities based on the opportunity cost of capital and effort to the firm; activity based costing has been enhanced using concepts from corporate finance for this purpose. Essentially, firms now measure the profit and loss statements from their involvement in supply chain programs using more sophisticated measures including return on capital, return on assets, economic value added, discounting of investments and returns, among others. In Chapters 6 and 7 we will focus on a select few measures for deriving insights into firms’ behavior in collaborative program environments, as opposed to a more holistic balanced scorecard approach.

At the firm level, we will consider primarily on the activity costs resulting from
work assignment from the program to its resource groups, and the capital costs in developing and contributing these resource groups to the program. The processing costs for a resource are assumed proportional to the allocated demand. The capital costs are assumed to include both the capital expenditure as well as the opportunity cost of investment incurred by committing resource capacity for the program duration. Furthermore, we assume that the program accounting system is efficient in capturing the true costs of the activities and the capacity investments and that it is transparent to the program planner (if not to all of the firms). In other words, we assume that the program planner has perfect information and knowledge of the cost structure at every other firm participating in the program. While this is a stringent assumption, we need it nevertheless to make the analysis tractable, and for deriving useful insights into program decision-making.

Another distinction in program costing methods is based on whether the activities are recurring, i.e. repetitive work to be conducted over the life of the program, or non-recurring, where the activities are not expected to be repeated over the program duration. A natural way to categorize these costs, at least in certain situations, is to assign recurring costs as variable costs – that depend on the volume or frequency of those activities – while accruing non-recurring costs as fixed charges. This is not logical in all program situations, however, since it is possible to have recurring overhead costs that are better assigned as fixed costs. Hence, from our perspective, and in our models, we designate only the capital or set up costs as non-recurring, while we assume that recurring costs are incurred in the processing of the tasks.

In reality, viewed within a broader firm-level accounting system, many supply chain programs themselves are viewed as non-recurring projects, and firms may categorize their program related costs as fixed or non-recurring costs. However, we
believe our method of designating capital expenditure as non-recurring and task processing costs as recurring has its advantages in designing and analyzing program cost and value sharing mechanisms, as we detail below.

4.3.2 Cost and value sharing mechanisms for collaboration.

Given a strong accounting foundation, programs can begin to measure and allocate costs in an efficient manner to the participating firms. It is rare that a program is its own corporate entity for financial purposes; hence in most situations all costs incurred towards a program must find a way back to the participating firms. Which in turn leads to the issue of how the costs are shared between the firms, and this is a critical piece of the overall program strategy. The natural way to do this would be to just allocate the costs directly to the source, and let firms bear the true costs of the program as they are realized; this is also almost always an inequitable assignment of program costs. The simple reason being that it is in practice extremely difficult, if not impossible, to correlate the value derived from the program separately with each of its cost centers. In other words, activities and costs incurred therein, contribute in a synergistic fashion to the program objectives.

For example, in the case of the Boeing Dreamliner development program, while it is possible to breakdown the cost of the airplane to its components and modules and their developers and manufacturers, the tangible value of the program is composite in the sense that its customers buy the aircraft for its advanced features and relative advantages over competing products. In this way, while the costs can be attributed directly to the firms, it would be a contentious process to even begin to attribute the tangible value derived by the program to the costs incurred, let alone derive the value contribution of individual activities. The same is true for many other pro-
grams, independent of scale and complexity.

The way firms and program managers deal with this is often via established precedent and by setting cost targets; this is partly to avoid being bogged down by accounting details. For instance, the program and the individual firms may agree on some estimates (derived from historical precedent) on the work involved in groups of activities, and agree to remain within pre-specified bounds of a cost target for those activities. The same concept would apply to the capacity investments. In the event that the costs deviate from the bounds, the matter is treated as an exception and dealt with on a case by case basis. These cost targets taken together and categorized by firm at the program level would constitute the cost share of the firm in the program.

In fact the same principle is sometimes applied to how the value derived from the program is shared between the firms. Only in this case the historical precedent is in determining the profit margins for the firms with respect to either the cost targets or at the broader measure of cost share. For example, in the electronics industry, it would not be uncommon for the assemblers and manufacturers of the game console (such as Flextronics or Celestica) to be paid by the program management (such as Microsoft) a flat margin (say 3-5%) over and above the estimated cost targets for the product assembly. This margin would be set in a way so as to compensate the assembler not only for the recurring costs of assembly in high volume, but also for the fixed cost of setting and ramping up the assembly operations. In the sales channel, Microsoft charges a royalty fee in percentage terms to game publishers who use the Xbox360 platform to develop and sell their games; this is again an example of sharing the revenue or market value of products for a program context.
Costs and value sharing over time, i.e. over the program life cycle, can by addressed by evaluating the non-recurring and recurring costs incurred by the firms over the same horizon. For example, when the lead firm assumes a portion of a partner firm’s non-recurring costs, it is essentially works as a “front-loaded” cost sharing mechanism that can induce greater levels of capacity investment by the firm that is in the interests of the overall program. Conversely, when the firms agree to split the recurring costs in some manner, the cost sharing mechanism allocates costs in a way that balances revenue inflows and cash outflows for the participating firms over the program life cycle. The challenge is to set these cost and value sharing parameters in a way that is acceptable (or even optimal) for all key participants, and in a way that also beneficial to the program. In the subsequent chapters of Part III, we will explore such mechanisms that can maintain a stable coalition of partners in the supply chain over the program duration.

The consequences of poorly designed cost and value sharing mechanisms can be disastrous (or at least enormously expensive) for programs, as in the case of the Microsoft Xbox program. The lead firm, Microsoft, lost nearly $5B over the life cycle of the first generation of the product, as it was locked into inflexible component and service pricing regimes with its suppliers like Intel and NVidia, that did not support the overall program objectives of penetrating the game console market through aggressive cuts in end-product pricing [4] [181] [182]. The same is also true for Boeing, as it currently grapples with considerable delays and unhappy customers with order backlogs; Boeing adopted a strategy where it assumed less of a portion, than in previous programs, of the non-recurring development costs, given the $13B price tag of the development effort. The company decided to out-source a lot of the key steps in the development cycle, promising its suppliers a greater share of the long term program revenues; this strategy has been cited by its own executives as one of
the major reasons for the repeated delays and snags in the development, leading to substantial losses and deferred revenues for the entire supply chain [106][150].

4.4 A Preview Of Related Chapters 6 and 7.

It has been our intent through this broad essay on collaborative programs to highlight the defining or critical elements (namely tasks, resources, firms, and program objectives), attributes and characteristics of program environments. Our broader goal is to set up a more precise and compact modeling exercise that incorporates the key elements above, and also the dynamics of a typical program planning exercise. In the following chapter devoted to collaborative innovation programs, we extend this qualitative discussion of the program environment a bit further, in order to outline the special challenges that arise from the complex interactions between the above critical program elements. We define and develop a modeling framework in Chapter 6 for the purpose of analyzing and evaluating key decisions in program planning and execution. Section 4.2 helps us identify which elements are critical from a modeling perspective, and how to create a compact representation of the program environment and decision processes. To same end, Section 6.1 explores some of the technical literature on related topics in supply chain management, product development, project and program management; here we look for ideas that complement our own, but also for differing points of view.

In particular we restrict our attention to setting up two interrelated decisions: the capacity of resources contributed by the partner firms, and the assignment of the program tasks to these firms. The purpose of Chapter 6 is to arrive at models that are amenable to the solution of the capacity problem given the task assignment decisions, and vice versa. It turns out that the problem of jointly determining the
assignment and the capacity levels, given our specific modeling framework, is not as tractable from a computational point of view; especially for large scale problems. Although we show that a globally optimal solution exists within our framework, the optimal solution could be computationally expensive. However, in our belief, various modifications in our modeling framework could lead to a computationally feasible joint optimization framework for large scale problems; we leave this particular issue for Chapters 6 and 7. Suffice to say that we make several assumptions in our models to simplify the typical program environment: enough to construct solvable problems, while at the same time retaining the richness and complexity of the decision problem.

In Chapter 7, we develop some fundamental insights into the workings of our mathematical model. In particular we present some important convexity and monotonicity results that guarantee that our model can be solved to optimality for the program planner’s objectives. However, the optimality of the central planner’s problem raises the question as to whether the program optimal solution is indeed optimal for, and therefore palatable to each participating firm. For example, the task assignment mechanism can maximize program value (or conversely minimize program costs) but will it be acceptable to all of the participating firms, and over all of their individual decision spaces? With the discomfort posed by these issues in the background, we formalize the program planner’s centralized problem in Chapter 7.

A key assumption that underlies the centralized problem formulation is that the cost structure at the individual firms is completely transparent to the central planner. Further, along with the knowledge of the solution proposed by the centralized problem; the individual firms also hold information on the other firms’ cost structures. Despite this liberal information exchange, lateral gaming behavior with regard to the capacity investment decision is not allowed, even in the decentralized model.
Rather, we will assume a special decision hierarchy whereby individual firms only respond to incentives defined by the central planner, even for their decentralized capacity decisions. These key assumptions help us set up the decentralized capacity decision problem in the same chapter; where firms have to independently determine their capacity contribution, in response to the program planner’s task assignment decision, given the other firms’ capacity decisions, and in response to the incentive schemes defined by the central agent. In the absence of such restrictive assumptions, the decentralized problem becomes a multi-player competitive game, which is significantly harder to analyze (we leave the analysis of such scenarios for future research).

We illuminate the critical issue of coordination through Chapter 7, wherein we show how the program manager can develop and modify several value or cost sharing mechanisms to induce the firms to invest at capacity levels that are in alignment with the program objectives. We propose several mechanisms that are guaranteed to achieve the coordination in the supply chain capacity decision. However, not all may work equally well for a given setting, so we highlight the pluses and minuses of each class of coordination mechanisms and discuss their applicability. We also provide some intuition on when and why it is possible to coordinate the capacity decisions for a given program context.

Our subsequent chapter, however, is devoted to a broad-based description, very much similar to the content here, of collaborative programs for logistics, or indeed those that have a process structure, as opposed to a project structure or innovation agenda. Similar to the discussion here, we describe the evolution of the program concept in logistics and process management, while examining many issues through the prism of “outsourcing”. Outsourcing of logistics and manufacturing processes has been much analyzed in the recent literature on operations management and strategy.
We examine the opportunities and challenges presented in the development and incorporation of outpoured logistics capacity, since the development and incorporation (utilization) of outsourced capacity calls for the same principles of collaboration and coordination as outlined here.
Collaborative Programs for Fulfillment: Issues in Outsourcing Capacity

5.1 Overview and Organization

We start by describing the antecedents of outsourced logistics services starting with the deregulation of the transportation and freight industry in the 1970s. We then describe potential logistics and transportation activities that have been candidates for third party sourcing, while commenting on the specific path of evolution of the maturity and growth in the role of these service providers. With this broad understanding of the elements of outsourced logistics, we attempt to characterize the value proposition in such outsourced arrangements both from the perspective of the “shippers” – logistics agents internal to the firm – and the service providers. We also comment on the pitfalls and drawbacks to the outsourcing of logistics providers. These are often the questions that the third party service providers must answer, both before winning the logistics business, and post-contract, to defend their practices and performance. These are also hurdles that shippers must pass before they hand over chunks of their order fulfillment processes to third parties at considerable
risk to their own performance and reputation. Throughout, we examine the prevalent issues from the perspective of logistics capacity requirements and the planning and staging of such capacity.

One natural extension of the discussion presented in this chapter is a framework for modeling logistics programs that captures (or has the flexibility required thereof) their most important and performance defining elements and their interactions. This chapter is intended therefore as the motivation for such models, and indeed we develop Chapter 9 based on the extensive understanding of logistics programs we develop here. One purpose of this material here is to describe the environment that is the target application for the models, and to place our subsequent modeling efforts in the context of a broader approach to managing logistics and transportation related activities in a business world that demands collaboration and partnerships for success. This chapter can also be read and viewed as a conceptual foundation for designing logistics processes that support an overall product or supply chain fulfillment strategy.

5.2 The Antecedents Of Outsourcing And Inter-Firm Collaboration In Logistics.

Figure 5.1 summarizes the major developments in the logistics and transportation world that have been relevant as contributing factors, and also parallel to the trend towards greater outsourcing and utilization of third party service providers. In the following discussion, we highlight each of these developments, and relate them to our own broader research objectives.
5.2.1 **Deregulation.**

Most important business regions in the world have witnessed a gradual deregulation of logistics and transport activities over the past three decades. The exceptions to the general trend of deregulation are the renewed security related regulations that have been imposed after the terrorist attacks of September 11 2001. Much of the regulation we will discuss in this chapter has to do with how transport and other logistics capacity on an industry level is made available to shippers and consigners within a regional, national or indeed the global economy.

In the United States, for example, it is argued that the trucking regulation in the United States – starting with the Motor Carrier Act of 1935 – was never intended to limit carrier capacity for a given product or route. Rather, the capacity limitations were the unintended consequences of price control. By letting bureaus

---

1 See Boyer [24].
of trucking carriers and truckers unions determine prices for carriage, the demand for carriers was artificially dampened, leading carriers to need fewer equipment and by lowering the overall capacity available for shippers and consigners. Secondly, the trucking and railroad regulations limited the ability of carriers to transport a wider range of products, or over a wider range of territories, leading to a reduced ability to pool demand from different shippers. The lower utilization of transport capacity again increased the average cost of shipments for a given route, thereby inflating the price of carriage relative the optimum loads. The result of such trucking regulations imposed several constraints on the routing of loads; including potentially circuitous routes to comply with carriage authority (or lack there-of).

The Motor Carrier Act of 1980, significantly altered the regulatory mechanisms for the pricing of for-hire carrier services. The main thrust of the new legislation was to eliminate price-setting bureaus and to allow shippers to negotiate with individual carriers directly on prices for various services. Economists and industrial organization experts widely heralded the deregulation as a step that would generate greater efficiencies, both for shippers as well as for the carriers. Evidence suggests that average prices for interstate transport followed a downward trend; even as efficiencies were generated through the improved location of inventory, and in the development of a hub-and-spoke model of fulfillment. Surprisingly, in the years following deregulation, the for-hire trucking industry saw a concentration (or shrinkage) in the number of large carriers, possibly because the downward price path for their services was unsustainable in the short run.

The relaxation of tonnage restrictions on trucking, especially on the Less-than-Truckload (LTL) segments also served to increase available capacity within particular

\[^2\text{see for e.g. a stream of papers by Nancy Rose and co-authors, [111][162].}\]
routes. The consequent lowering of the cost of LTL carriers also boosted the share of LTL shipments in relation to truckload shipments. The net impact on both shippers – who could now ship more efficiently in smaller batches – and on carriers who could solicit demand from a wider spectrum of shippers, has been recorded as being beneficial. Examining industry structure, the Motor Carrier Act of 1980 has had a major impact on the trucking industry. Several econometric studies\(^3\) have shown that the reform act of 1980 facilitated the entry of a large number of common carriers into the industry, especially in the smaller range of trucks (for e.g. vans). In the years following deregulation, the trucking industry in fact exhibited an increasing returns to scale; as carriers added more capacity, they were able to generate more efficiencies through the solicitation of more customers, leading to more efficient utilization of the overall capacity.

Surprisingly, deregulation did not yield the anticipated large scale entry of for-hire providers into the industry; rather the for-hire industry became more concentrated in the years following deregulation. It has been suggested\(^4\) that while there were scale benefits for efficiency and risk pooling from concentration, the artificial limits imposed by the license and authority regimes before 1980 had prevented firms from growing organically by adding more routes and customers, or from acquiring other firms operating in other spaces. Also surprisingly, in the decade following deregulation, the anticipated shift from private to for-hire (or outsourced) providers did not materialize as such at an overall industry level. Rather industry structure shifts were observed within the LTL segments; many firms could now develop their own specialized and full truckload fleets, while delegating LTL shipments to for-hire providers. Deregulation of trucking within the European Union has been much more gradual

\(^3\) See [24], [116], [199], and [141].

\(^4\) See Boyer, [24].
and recent. Recent research\(^5\) has documented similar and positive benefits for both the shippers as well as for the carriers. The difference being that Europe has witnessed a shift towards for-hire or outsourced carriers as a result of regulation.

The railroad industry has reported a different experience than the trucking industry when it comes to deregulation. The Staggers act of 1980 was the counterpart in the railroad industry to the Motor Carrier Act that impacted trucking channels. In fact part of the motivation for railroad freight reform was to make rail transport more competitive in relation to trucking by lowering costs and breaking down the price-setting bureaus. Research in the years following deregulation\(^6\) did not reveal the anticipated shift of volumes from truck in favor of rail transport. However, the decline in prices has been more gradual in this industry, and subsequent research has reported that in the ten year period following deregulation, industry costs have reduced considerably while industry profits have climbed through a series of mergers and consolidation of capacity\(^7\).

In ocean shipping, the US reform act of 1998 has also fostered a number of changes in industry structure and in productivity and performance\(^8\). The need, model, and consequences of the ocean shipping deregulation has been similar to that of the trucking and railroad industries. In essence, shippers could write contracts confidentially and directly with carriers and service providers without the influence of price-setting conferences. In fact, with the advent of internet technologies for communication and price negotiations, ocean shipping has witnessed a surge (up to a 200% increase in

\(^{5}\) See Lafontaine and Valeri, [123].

\(^{6}\) See Boyer, [23].

\(^{7}\) See Ivaldi et al., [105].

\(^{8}\) See a 2001 Federal Maritime Commission Report, [41].
the number of service contracts just in the two years after the reform act) in service contracting when measured in the number of contracts, and in the volume of cargo shipped.

While these seem to be industry level transformations, the implications for shippers planning for logistics and distribution capacity are immense. Removing artificial constraints on route and product capacity, shippers can now access capacity from a potentially unrestricted pool of service providers. Logistics capacity planning is now therefore more of a requirements based analysis, rather than a search for the most efficient and cost-effective providers or the best routes given regulatory constraints. The latter issues are still important execution problems; however, firms can now focus on determining their capacity requirements independent of artificial constraints on execution. Our models we develop later in Chapter 9 therefore represent this higher level in the capacity planning hierarchy.

5.2.2 Containerization.

One trend that has streamlined intermodal transport (between trucks, rail, and ocean shipping) has been the containerization of freight. In contrast to the break-bulk method, where truckloads are reduced to smaller components and then re-assembled for either rail or ocean shipments, containerization simply involves using the same container across all modes of transport. Worldwide adoption and use of standard containers has resulted in lower turnaround times for all modes of transport involved\(^9\), and has greatly improved the speed of freight and service levels experienced by shippers.

\(^9\) See Crainic and Kim, [46].
Containerization has also made capacity more flexible and fungible for both shippers and LSPs, even though shipping volumes would appear to have been discretized with standard containers. Shippers in many sectors can now reserve “space” on truck routes, on railroad, and ocean liners, where the available space is neatly measured in the number of containers that can be efficiently loaded, stored, and unloaded from the transport equipment. Previously, the predominant measure of capacity, especially on international and long-haul routes was tonnage and equipment based (these measures are still applicable, based on the nature of the goods being shipped). However, for a vast majority of products, such as in retail and packaged goods, and in electronics, containerization has meant that logistics capacity is now planned in standardized container volume (truckload vs. LTL, for example).

5.2.3 Information and communication technologies.

The adoption and use of Electronic Data Interchange, Satellite communications, Internet based protocols, and more recent technologies such as RFID has fundamentally changed the operating practices of logistics service providers. With these technologies, it is now possible for logistics and transportation firms to be more in tune with the order management systems of the shippers, and therefore to be potentially more responsive to their needs and service requirements.

Broadly speaking, the literature has grouped information technology improvements into two categories: information processing or management information systems (MIS), and communication systems. Information processing systems include popular enterprise applications such ERP (enterprise wide resource planning), DRP (Distribution requirements planning), WMS (Warehouse management systems), TSP

---

10 See for e.g. Delfmann et al. [56], Bowersox and Daugherty [21], and Closs et al. [39].
(transportation systems planning, routing, and scheduling), and E-commerce and procurement exchanges, to name just a few of the sprawling application space. Electronic Data Interchange, Satellite communications systems, Global positioning, Mobile Computing and Networking (for e.g. Qualcomm’s solutions for the trucking industry), and (RFID) Radio Frequency Identification Devices are examples of advancements in communications technologies for the logistics service providers.

The productivity increases from the relatively rapid and large scale deployment of such technologies are yet to be completely understood and measured. However, the scale of adoption of both the MIS and the communication systems technologies by logistics service providers over the past two decades seems to point to significant benefits for both the providers as well as for the shippers (of course, another interpretation is that the technologies are well-marketed!). Recent research\textsuperscript{11} has pointed to the changes in industry structure that have come about from the adoption of these information technologies. Trucking companies and their resources, for example, are now as much managers of information, and effective communication channels, as they are handlers of the goods that are being shipped. Thus, shippers expect that the logistics service providers will not only interface with their order management systems for seamless order handoff, but also communicate with the end-customer to provide information on the location and timeliness of the orders that are being shipped.

Another consequence of the rise of information technologies is that logistics has now become a more specialized field. Achieving superior performance in consistent service levels, responsiveness, and cost, especially with the added fixed costs of advanced technologies requires expertise and experience in some or all of the recent

\textsuperscript{11} See for e.g. Nagarajan et al. [145] on the trucking industry.
technologies mentioned above. It is therefore much more difficult for shippers to invest and maintain their own fleet of transportation and logistics assets in the same performance band as specialized and focused for-hire providers.

While not based on rigorous empirical studies, a large number of surveys sponsored by industry consultants and service providers has indicated that firms are now outsourcing more of their logistics fleets and capacity requirements to third party for-hire providers than before. In the last decade, a number of studies\textsuperscript{12} have suggested that more firms have outsourced their logistics function to specialized for-hire providers than ever before. While many firms cite service quality as the predominant reason for hiring and retaining the service providers, and information technologies much lower on the supplier scorecards, it is also true that the majority of outsourced services are transactional in nature; precisely those that benefit from the improved communication and transactional information systems that are on offer from the specialized providers. At the same time, paradoxically, shippers appear to be impatient with respect to the IT capabilities offered by the logistics providers.

Perhaps for this reason, and with the increased sophistication of best-in-class logistics planning systems, firms are now looking to systems integrators or lead logistics providers who can oversee the management of the individual LSPs by providing industry knowledge and expertise, MIS and other software systems, and procurement best practices to the shippers\textsuperscript{13}. These systems integrators have overseen a more dramatic shift towards capacity outsourcing – sometimes as neutral brokers – from independent for-hire logistics providers.

\textsuperscript{12} See for e.g. a GeorgiaTech Logistics Institute Report [124], and a survey by Lieb and Bentz [128].

\textsuperscript{13} See Bitran et al. [16].
5.2.4 Realignment of inventory, manufacturing, and logistics management principles.

In parallel with the advent of information technologies have come a series of process and operations innovations in the logistics arena. Industrial engineering and operations research methods in execution have witnessed remarkable advances over the past two to three decades. The business process re-engineering and the core-competence movements guiding business thinking in the 1980s and 1990s led many large manufacturing firms to rethink the role that manufacturing, and logistics as a sub-component, played in their overall strategy. The re-engineering of manufacturing processes (proposed by Hammer, [95]) was followed by cost cutting initiatives including the outright sale of manufacturing and logistics facilities to strategic partners\(^{14}\). With manufacturing and/or inventory management outsourced, it became an important question for firms whether to hold on to, or to outsource their logistics processes. Some firms held on to their logistics divisions, believing that customer order management systems were integral to their business success, while some others saw logistics outsourcing as a logical extension of the divestment of manufacturing capacity\(^{15}\). Today, the prominent contract manufacturers, especially in the electronics sector, have their own logistics divisions that compete for services with the in-house logistics capabilities of the shippers\(^{16}\).

Along with these industry structure changes, we have also witnessed remarkable breakthroughs in accepted logistics and warehousing practices, that change the definition of capacity for logistics planners. Retail firms such as Wal-Mart grew on

\(^{14}\) See for e.g. Bowersox [19], and a report on Flextronics [135].

\(^{15}\) See for e.g. Nortel’s divestiture deal with Kuehne+Nagel, [86].

\(^{16}\) See again Bitran et al. [16].
innovations including cross-docking, implemented within their private logistics fleet and assets. For another example, practices such as automated sorting and clever order splitting, merging, and routing established by the online retailer Amazon.com at its warehouses enabled the scaling of warehouse operations to process millions of customer orders in both peak and off-peak seasons. Dell, with its vendor managed inventory, and its principles of “high-velocity” fulfillment, provides an example of such innovation in fast-paced high volume manufacturing; similar changes in logistics practices, some following the Just-in-Time principles advocated by Toyota Motors, have been implemented through the co-location of tier 1 suppliers in automotive, and in electronics. There is now a closer and more tightly managed integration between customers who place orders, and the back-end logistical systems that fulfill orders. This is true not only of business to customer retail, but also of business to business order transactions.

What does this tight-knit configuration of procurement, manufacturing, and distribution imply for logistics capacity planning? Well, for one, capacity planning for logistics can no longer be easily distinguished from the planning of other resources in the supply chain. Conversely, because much of the capacity in logistics is now being outsourced, enterprise (or firm level) planning systems now have to incorporate external capacity, in all of its forms – fixed or variable, reserved or flexible – when determining the requirements of outstanding or forecast orders from customers. Thus, on the one hand, most environments today require that capacity planning be

17 See Stalk, [178].
18 See research by Xu and Graves, [198], for a description of such an order management system.
19 See a report on Dell’s global production-distribution network, [121].
20 See for a marketing and supply chain perspective of JIT, [75].
21 See again Bitran et al. [16].
22 See again Lieb and Bentz, [128].
conducted in a seamless and integrated fashion across the procurement, manufacturing, and distribution functions; however, existing capacity planning algorithms and systems are restricted to internal or firm-level resources, and do not extend readily to include out-sourced capacity. Chapter 9 aims to bridge this gap by extending logistics capacity planning to outsourced and often the flexible form of capacity, for a broader supply chain context that involves both the shipping firm as well as its logistics service providers in the planning exercise.

5.2.5 Specialization and commoditization of logistics services.

One of the paradoxical trends in the logistics world is that individual processes that make up the logistics function have become simultaneously specialized and commoditized (the same argument can be made for many manufacturing processes today). The specialization has occurred due to the tight-knit integration between procurement, in-bound logistics, manufacturing, and distribution. For example, the logistics requirements for the electronics assembly, or for that matter disassembly and repair, are highly specific to the manufacturing process, and in some cases to the product. The material handling, storage and transport of the components and sub-assemblies prior to, and after assembly, are highly customized to the specific electronics manufacturing process, and to the product design$^{23}$. Similar observations can be made for current practices in automotive assembly$^{24}$, or in the case of current and future aircraft assembly processes$^{25}$.

The key factors that engender specialization are of course, the product or pro-

$^{23}$ See again a GeorgiaTech Logistics Institute Report, [10].
$^{24}$ See a trade magazine article, [97].
$^{25}$ See for e.g. the case of the Boeing 787 Dreamliner development, [107].
cess specific trends in different industry sectors. This can be observed, for example, in modular product assembly operations; different sub-systems of the product are manufactured off-site and then brought together for final assembly by the logistics service providers according to a just-in-time, or just-in-sequence model of delivery. The interfaces between the different subsystems dictate the degree of specialization of the logistics processes. We view an interface between the subsystems as all of the activities that need to be completed before they can be assembled; this is a more comprehensive process view of modularity, in addition to the product architecture view proposed by Ulrich, [192].

If the interfaces are standard and not component specific, then the specialization required is low; this for example is true for a bulk of discrete widget type auto parts. However, for other subsystems, the specialization or customization required could be extensive. For example, in the automotive assembly operation, special conveyers and fixtures are designed for the storage and transport of some sub-systems like seats, fabricated or welded parts, and drive-train components. The logistics providers – whether internal and external – have the responsibility to not only operate these material handling systems, but also in some cases to invest in and develop the handling equipment according the process specifications. The same concept holds in other manufacturing and distribution systems; for example, distribution of retail and consumer or pharmaceutical products may involve the design and development of specialized palettes and transport equipment that satisfy one or more unique requirements of the product such as temperature control, spoilage or damage avoidance, and reverse logistics.

The paradox here is that while logistics may have become more specialized and product focused, they have also become more commoditized in the sense that the
skills, tools, and information systems needed for such specialization are now widely accessible, and transferable across different domains and logistics providers. In short, logistics service providers have “learned to learn”, and to innovate in order to provide competitive contract logistics services to shippers and manufacturers\textsuperscript{26}. In the manufacturing sector, especially in the United States, some of this commoditization has been accelerated because firms wanted to divest their more expensive labor and resource commitments to their logistics partners; given the rate of change in products and process technologies, this made sense. The reuse of these technologies was much more likely, when in the hands of LSPs with multiple clients in the industry, than for a single firm with a limited product portfolio. The economies of scale to develop logistics capacity were in favor of LSPs serving the industry as a whole.

In this latter sense, the specialization of logistics is the result of niche processes (both in manufacturing as well as in packaging and distribution); whereas the commoditization is being driven by third party LSPs and their partnerships with many different OEMs and product owners. Thus, one could think of specialization and customization as an innovation that is diffused and commoditized through an industry network facilitated by the third party service providers. The commoditization occurs gradually with the diffusion of technologies and best practices within and across industries. For this reason, logistics for electronics manufacturing are likely to follow similar broad patterns across the major manufacturing centers in the world\textsuperscript{27}, and competition is fierce within the commoditized segments. Individual variations (or specialization) in the processes are caused by the shippers demanding customized service for their own peculiar environments.

\textsuperscript{26} “Commoditization” being a term used in a Harvard Business Review article by Davenport,[51].

\textsuperscript{27} Such as SouthEast China, Eastern Europe, India, Mexico, or Singapore.
The capacity management implications of these trends also stand in contrast: specialization implies closer and longer term relationships between a firm and its logistics service provider. It also implies greater investments in logistics technology and equipment on the part of the service provider; hence specialization is a high risk and high reward proposition for the service provider. It is not surprising that firms, and OEMs in particular, wish to divest these functions to external providers, since it reduces the fixed costs for the OEM and therefore reduces breakeven volumes for profitability. The value proposition for the service provider is in the long term nature of the relationship; it moves the service provider up in the chain of suppliers. It also allows the service provider to negotiate higher margins, as the service becomes more customized and the length of contract is extended. A third and long term value for the service provider is in the technology of specialization itself; with such projects, the LSP can successfully compete for projects and market share both within that market segment, as well as transfer such new technologies to other domains where possible. With specialization, capacity is not as easily transferable, and is also not readily interchangeable unless the alternative capacity is designed with the exact specifications.

Commoditization, on the other hand, implies replaceable and transferable capacity; it also implies capacity that can be augmented with some restrictions. In financial terms, commoditized capacity has a greater variable than a fixed or upfront component, in relation to the more rigid specialized capacity that may require significant investment. Also, commoditized capacity may be cheaper than the specialized capacity, when measured in dollars per unit item produced or shipped; this is mainly due the interchangeable nature of such capacity. Often such commoditized capacity can be procured or reserved on spot markets and may be available from a wider pool of service providers than specialized capacity.
In Chapter 9, for example, we will consider a continuum of specialization in capacity; in particular, we consider both an upfront (or fixed) as well as a variable component to capacity costs. That said, we will assume that all forms of capacity considered have to be reserved ahead of their use.

5.2.6 Growth in sea and air freight; demand for expedited services.

At the turn of the millennium, it was estimated that a staggering 40% of global trade by value was in the form of air freight. According to the World Trade Organization statistics, as a whole the first five years of international trade during this decade have seen between 15 and 20% increases in both air and freight transport. The US has tended to dominate air freight in its export of transportation services, while the EU is a powerhouse in sea-freight. Between 2000-2007, China, has also come to dominate the growth statistics in sea freight; with a seven fold increase in its import of sea-freight services, and reaching an astonishing $20B worth of sea-freight exports in 2007 from a virtually insignificant level in 2000. Equally impressive is China’s growth in its sea freight infrastructure, with over 20% of the container vessel traffic of the world flowing through China.

It is also true that these growth statistics have been dented by the global recession of 2008-2009, although the high growth areas such as China and other parts of Asia have started rebounding in 2009 and 2010. Still, as a trend, it is possible to observe that the growth in international trade has followed the broader trends in globalization and development of the emerging economies of South America, Eastern Europe,

---

28 See a chapter by Humphreys [104].

India, and China. The emergence of low cost countries as the new manufacturing centers for products ranging from apparel to electronics, and to automobiles, has no doubt fueled the demand for sea and air transportation services. As a consequence of such unprecedented growth, shippers became concerned about the availability of sea and air freight capacity to support the needs of burgeoning international trade. Shipping and jet fuel prices in 2007 and 2008 also reached historical peaks, with speculators foreseeing increasing demand in relation to the supply of fuel for meeting transportation needs.

Shippers reacted to the forecast capacity shortages by writing longer term contracts, and reserving sea and air freight capacity in advance of use, and for longer durations. For example, in consumer goods and electronics, it was a common practice for freight forwarders to reserve sea and air freight capacity for their customers (typically retailers and OEMs) to manage order and demand during the peak holiday season in the US and Europe. These long term contracts not only reserved capacity (options) but also locked in freight rates or pricing, so as to hedge against the risk of rate increases that could alter the economics of off-shore production and transport for these firms. The dramatic decline of 2009 has no doubt hurt some of these firms, as demand for their goods experienced shocks never seen in half a century.

Some econometric analysis\(^{30}\) has presented evidence that in the period until 2000, air freight rates had declined relative to sea freight rates. However, given that a great portion of freight moves under contract (i.e. prices are negotiated confidentially between the shipper and the carrier), a true understanding of the behavior of

freight rates is hard to measure\textsuperscript{31}. What determines capacity pricing from a value perspective? As with any other market, prices are determined by the basic services such as transportation, documentation, and other critical services, but also by premium factors such as reputation, and value-added services such as expediting and short-term availability.

It is for the value-added expediting component that air-freight is such a profitable business for freight-forwarders\textsuperscript{32}. It is widely reported that air-freight traffic has grown much faster than the world economic output over the last four decades\textsuperscript{33}. As manufacturing centers and their end markets became more geographically dispersed, air freight reduced the distance on the time dimension, albeit at a premium over sea freight. The trade-off between the additional transport costs and the lower cost of manufacturing has in the case of certain products consistently tipped in favor of air freighting from distant production centers. For example, highly perishable products whose production is restricted to certain regions geographically distant from the target markets – such as flower bulbs from the Netherlands, high value personal electronics, or emergency shipments of repair components – have been ideal candidates for air freight. As firms have become more savvy about managing perishability, and have learned to make better trade-offs between transportation and opportunity costs, air freight as a component of world trade has grown correspondingly, at least in our view.

The capacity value chain in the air and sea freight world exhibits some similarities in structure: Carriers are agents who own the transportation assets in the form of

\textsuperscript{31} Containerized Freight Rate Trends, The Center for Supply Chain Research Smeal College of Business Administration, Pennsylvania State University, Draft, April 15, 2008.

\textsuperscript{32} Personal communication with Dr. Rod Franklin, VP of R&D, Kuehne+Nagel.

\textsuperscript{33} See Bowen and Leinbach [18].
either cargo ships, or cargo airplanes. Carriers announce a schedule of their capacity over time, and solicit bids from shippers and freight-forwarders for utilizing capacity on the transport vessels. The forwarders then reserve capacity in bulk on the vessels in anticipation of business from shippers; the forwarders serve as consolidators of shipments from one or more shippers for the carriers. In addition, the freight forwarders also provide ancillary services such as customs documentation, brokerage, and facilitate letters of credit or other forms of financing for the cargo if it involves international destinations.

Industry practitioners estimate that more than 60% of all domestic air freight in the US, and 90% of international air freight is managed by freight forwarders working with the carrier airlines\(^\text{34}\). At the same time, there is intense competition for cargo booking between different carriers, and indeed between the carriers and the freight forwarders. Forwarders are after all brokers who provide value added services to their shipping customers; so recently there is a trend among airlines and shipping firms to reach out directly to large and profitable customers for their cargo business. Heated competition along certain routes has been blamed for lowered freight rates both in the air and sea segments, although empirical evidence remains scarce.

While the industry as a whole cycles between periods of capacity shortage and excess\(^\text{35}\), forwarders and shippers hedge against both capacity and price risks by booking cargo capacity in advance. The management science literature has also studied this issue recently from the perspective of airlines maximizing their revenue potential by pricing their cargo capacity based on customer segmentation (see the

\(^{34}\) See Coppersmith, [43], [44].

stream of recent work by Gupta and co-authors, (3)(92)). Freight forwarders, for example, may pay a capacity reservation fee in advance of utilization to book cargo capacity, and the rates they are charged by the carriers depend on the cargo volume, and market considerations such as demand and competition among carriers for the forwarder’s business. Any capacity that is not pre-booked is sold on an ad-hoc basis to direct shippers (those that do not go through forwarder intermediaries), or to forwarders looking for last-minute capacity for their shipments. This sets up a challenging revenue management problem for the carriers: How much capacity, for instance, should they allocate to forwarders, and how much should they retain for ad-hoc use; in this case, unlike traditional revenue management, ad hoc capacity may not always be priced higher than pre-booked capacity.

The preceding discussion is partly to highlight the importance of capacity reservation in the world of shippers, forwarders, and carriers. Similar to production capacity which often has a positive lead time for installation, transportation and freight capacity, especially of the out-sourced kind, also can have positive lag, and has to be reserved ahead of time. In contrast to the revenue management literature, the models presented in Chapter 9 consider the view of the shipper or the manufacturer in the supply chain (as a central planner or agent); given known data on the reservation and utilization (or exercise) prices, we attempt to solve for the optimal capacity reservation quantities.

5.3 A Service (Provider) Based Segmentation Of Logistics Processes.

5.3.1 A capacitated view.

In this chapter, the implicit focus is on those aspects of capacity that indeed need to be “planned” and installed in place in advance of fulfillment. However, as can be
expected, not all logistics activities require advance planning; moreover it is typically only those activities that utilize fixed capacity or involve capital investments that require advance reservation. In order to understand this distinction, it is useful to develop an outline of typical processes in supply chains that fall under the logistics umbrella, and further to understand which ones are more prone to be outsourced. The categorization is not exhaustive (see Figure 5.2), but only serves to provide a physical description of logistics activities that we will repeatedly abstract in much of our following discussion. As can be seen, the basic processes involve the physical movement or handling of goods, but there are a number of ancillary activities surrounding the material movement, that take on special significance in a supply chain context.

We are mainly concerned with measuring requirements for those activities that are subject to capacity limitations. For example, direct transport, shipment consolidation (or merging), and cross-docking, and light manufacturing/assembly, are activities that one can surmise would typically be constrained by capacity limitations. The capacity limitation itself can come in many forms including labor, equipment, schedule, storage space, and fuel, among others. However, we will not be as concerned with the dimensions of capacity for an individual process, as with the limitation placed on the output of the process relative to its demand.

On the other hand it is often the case for a service provider that information systems (including call centers), or for that matter professional services are relatively easy to scale, by expanding the consulting staff or working overtime, or by adding more server capacity. The capacity of transport equipment and (controlled) storage space, especially those that are highly capital intensive; or the capacity of a repair operations line, are not as easily altered, and require careful planning and installation
in advance of their utilization. As such in this chapter, we will focus on the latter set of processes.

5.3.2 Lead logistics providers.

We differentiate lead logistics providers from third-party or for-hire logistics providers as follows: Lead logistics providers provide a comprehensive logistics solution that encompasses inbound supply, warehousing, inventory staging and kitting, and finished goods delivery to the customer. In providing such an “end-to-end” solution, they sometimes invest in and deploy their own logistics assets, while in other cases they may engage multiple third party logistics providers for developing a comprehensive solution for a firm. The key difference between the third-party model and the lead logistics model is that the latter through its investments and transaction costs, shares the risk of developing a logistics solution with the outsourcing firm, while also sometimes demanding a portion of the cost savings or even revenues that
are incurred. Such risk and gain share contracts are designed to provide incentives to the lead logistics firms to not only reduce operating costs but also to maintain or improve service levels for a firm’s customers. For this last reason, these contracts tend to be multi-year renewable contracts, since the performance impact of capacity changes and policy redesign may not become evident in the short term.

Lead logistics providers therefore provide a comprehensive logistics service. Rather than having a transactional relationship with the firm where the firm dictates the utilization of the services, these firms provide a more performance-based service. They focus on the aggregate cost and service metrics of critical importance to a target firm, and evaluate the current state of the logistics system for potential changes that can reduce costs and improve end-customer service levels. The relationship with a firm or even a supply chain is more collaborative rather than transactional, since to achieve the stated performance objectives requires all parties to work together and change the working practices of the supply chain. The need for this model of service increases with the need for greater responsiveness from the logistics processes; as depicted and explained later in Figure 5.10. As different parts of the supply chain are required to be more responsive with lower inventory buffers, the need for a lead logistics provider increases correspondingly to ensure that capacity is available to support the entire order fulfillment cycle, and that the capacity allocation does lead to local bottlenecks while having a lot of excess capacity at the supply chain level.

5.3.3 Systems integrators.

Systems integrators are agents that perform one added level of service to their clients: that of interacting with suppliers (or even end-customers) to ensure compliance with the supply chain fulfillment processes and order performance management. The col-
Laboration objective is central to the value proposition of the systems integrators, and in fact their service proposition is to shield their clients from the complexity of coordinating logistics with multiple suppliers or end-customers. This type of service has typically found more traction on the in-bound side, because firms are more nervous about giving up control of their customer-facing logistics.

Logistics systems integrators often have in-depth specialization in one industry; often they are extensions of contract manufacturers, or spin-offs from manufacturing and assembly divisions from large manufacturing firms. For example, companies like Magna and Lear in the automotive industry; and Flextronics or FoxConn in the electronics industry. Li and Fung in the apparel industry performs a systems integration service for its clients who depended on the firm to navigate the quota system for imports into the United States, by staging production and transport of apparel from several Asian countries (Bitran et al. [16]). Often, these firms perform services in addition to a logistics systems integration, and provide investment capital to suppliers on the one hand, and design and prototyping services for their retail clients. In most cases, they are the face of the client for the network of suppliers, and therefore have significant control of these lower tier segments of the supply chain. This simplifies the coordination tasks for the client, and they can have a single point of access to all suppliers, and have a single channel for resolving supply issues and to manage supply risk.

5.3.4 Tiers and pricing of logistics services.

Figure 5.3 differentiates lead logistics providers from systems integrators, and their comprehensive services from those of piecemeal providers. While the industry structure could change over time, it is still possible to distinguish the providers based
on the range of services rendered. The highest level of service is the most complex and value-adding service that can be performed. This is the synchronization of the multiple logistics activities (such as the ones discussed through Figure 5.2) so that all of the myriad transactions, material flows, and communication channels support the fundamental order fulfillment processes of the supply chain. This synchronization not only needs to happen at the schedule or operational level, but also at the asset capacity or tactical level, and at the strategic level where risks and gains are shared among all the logistics providers that are engaged.

This level of service can only be provided either by an established and dominant firm in the supply chain, or by a long term player that has considerable expertise in the specific industry. Without such credibility and trust, planning, coordination,
and execution are all very difficult tasks, and a uniform logistics or supply chain solution (or at least a semblance of such cross-firm coordination) cannot be guaranteed. It is for this reason that in manufacturing industries this role is reserved for long-term contract manufacturers, agents who have in-depth knowledge of the supplier networks, and also of the logistics processes. These providers can be found today in a given industry vertical, but very rarely does their expertise cross-over between sectors. The services are priced based on a combination of a flat management fee and a variable component based on the volume of transactions.

The next level of service is today provided by the lead logistics and the so called fourth-party logistics providers. The primary value proposition is the aggregation of logistics capacity that is required for an “end-to-end” solution that supports the order fulfillment objectives. This aggregation of capacity is made possible either because these agents are asset based (i.e. own and operate their own fleets and warehouses), or have access to a wide range of third party providers because of a strong logistics reputation, and wide ranging experience in many different industry verticals. In today’s market, the large logistics firms such as USPS, DHL, UPS, and FedEx are also the most successful lead logistics providers. However, there is strong competition from firms with a long history providing either asset-heavy third party services, or asset-light freight forwarding solutions. Through their business model, the latter have deep knowledge of multi-modal logistics providers, and are therefore able to offer lead logistics services to their clients. In many cases, these lead logistics providers compete with the in-house logistics divisions of their larger and more established clients.

Fourth party logistics providers are often those who do not invest in their own assets, but rather are aggregations in these sense that they offer brokerage services
that can create capacity across the fulfillment channel. Thus, a fourth party logistics provider would work based on a brokerage or management fee, as opposed to a gain share or risk share model of service pricing. Lead logistics providers on the other hand often invest or pool their own assets to provide aggregation, but also act as a neutral broker to procure external capacity for the supply chain needs. Depending on the range of services tendered, there is a fixed management fee component, and a gain share component; the gain share component is negotiated based on the risks and opportunities perceived by both the provider and the supply chain.

In the third rung, are the basic service providers who offer their assets for utilization either directly by the firm, or through the lead logistics or system integrator intermediaries. These providers range from the large and well-equipped firms like UPS, FedEx, and Maersk. These firms tend to have strengths in one particular mode of transport, or within a particular line of services, and compete based on reputation and focus on those specific areas. The pricing is based on the transaction volumes, and heavily discounted based on the market capacity and availability of services at any point in time. Many of these providers have recently developed revenue management systems that dynamically price their asset capacity over time given demand and supply in the market, and based on the volume of transactions (see Gupta, [92]).

5.4 Value Proposition And Pitfalls In Outsourcing Logistics.

Before we proceed to take for granted this model of utilizing third party LSPs, we discuss here the economic advantages of outsourced arrangements. The arguments for or against third party logistics are not very different from those debated in manufacturing and for broader service sectors.
5.4.1 *The advantages presented by LSPs to lead firms.*

The primary incentive for manufacturers or shippers is the reduction in fixed liabilities in the form of transportation assets. Not only are there fixed costs but also recurring costs in maintaining and upgrading these assets. The presence of these fixed costs typically increases the break-even production volumes for the firm or the supply chain. So, for sure, there is a downside risk in investing in fixed logistics assets; however in the absence of outsourcing contracts, there is also an upside risk in that excess demand can go unfulfilled for the lack of logistics capacity both on the inbound and outbound side. Another drawback of fixed transportation capacity that is not deployed outside the firm is the cost of running empty trips on the return legs.

Divestment of fixed assets such as warehouses is also presented as an attractive cash flow opportunity to firms. As product life cycles shrink and as there is uncertainty around future demand for products in specific regions, firms can not only eliminate risk, but also accrue short term cash infusions by selling their facilities and assets to third party service providers. This trend was also observed in the manufacturing sector as large OEMs in the automotive, and electronics business sold their factories and other fixed assets to the contract manufacturers and service providers. In some cases, such as Nortel’s divestment agreement with Kuehne+Nagel, there was also a transfer of personnel, and therefore the skills and knowhow in logistics to Kuehne+Nagel. However, this transfer of know-how was not one-sided; Nortel also gained the expertise of Kuehne+Nagel in international distribution and logistics, especially in regions such as South America, and the Asia-Pacific, where Nortel had limited prior experience\footnote{Personal communications with Kuehne+Nagel Lead Logistics Services, RDU Office.}. 

\footnote{Personal communications with Kuehne+Nagel Lead Logistics Services, RDU Office.}
There has also been a trend recently of small and medium sized firms contracting with third party LSPs not only for their transportation needs, but also for inventory control, and also inventory financing. For example, ownership of inventory is assumed only after delivery, and the LSP assumes liability for the inventory in the pipeline; in such cases the financing is a value-added service, and the LSP charges a premium over the basic freight rates. For the smaller sized firms, this frees up cash flow, and also frees them from non-value added tasks such as obtaining letters of credit from banking intermediaries, where their smaller size could sometimes be a barrier. The arrangement works on the opposite end as well; the seller may contract with the LSP to obtain payment upon transfer of the goods, and does not have to wait until final delivery to the buyer, typically in far out markets. For inventory control, some firms – small or large – may rely on the LSP to provide state-of-the-art systems and methodologies for inventory management; especially those that can improve their customer service levels, and/or drive down inventory costs. The caveat here is that focusing just on driving down inventory costs without heed to the impact on service levels is not advisable as a long term strategy!

5.4.2 The risks posed by the outsourcing of logistics.

As mentioned earlier, the pitfalls are common to outsourcing internal processes of any variety; there is a long term dependence on outside firms for critical activities in the supply chain. Whether or not they are considered “core” to the firm’s business model, it is still a fact that logistics and transportation activities are critical to customer fulfillment flow-paths; the long term cost of customers unhappy with the service levels or service quality is hard to measure. In the case of Nortel, after two contract renewals with Kuehne+Nagel, it was decided that having direct con-
Figure 5.4: Value Proposition in Engaging Logistics Service Providers.

tact with customers in the fulfillment process was more valuable than the significant efficiencies and cost savings achieved with outsourcing logistics.

Another risk from depending on external capacity is that the variable cost of transport may be subject to market forces and to its availability to match with manufacturing and delivery schedules. If the transportation and storage routes are busy and used by many firms (as is currently the case with SouthEastern China), it could be hard to predict capacity costs, especially during peak seasons when there is shortage of capacity in those routes. Further, upon realizing such pitfalls, even if the firm decides to rebuild internal capacity, it could have to contend with the loss of expertise in the form of personnel and through the cost of relearning the logistics
work-flows which may not be trivial.

One of the major challenges faced, both by the firm and the LSP is the hand-off period when the latter assumes responsibility for certain processes for the first time. The learning curve may be steep, especially if it is accompanied by new systems or ways of working that are foreign to either firm. The resulting disruption to the firm’s customers may negate the cost savings accrued; firms may sometimes realize these pitfalls only post-fact, and the real trade-offs may not be evident upfront when the decision to outsource is made. It is also a challenge to measure the benefits of outsourcing with any clarity especially in the midst of product and process changes accompanying the hand-off, and the initial pain from the change in working practices is often a hurdle for managers to overcome.
5.5 The Linkage Between Product, Process, And Supply Chain Architectures.

In this chapter we restrict our attention to physical products, as opposed to service products; although service products also require logistics and transportation planning (think Disneyworld!). For a physical product, it is our fundamental belief (not unjustified) that the product architecture, or the attributes of the product, determine not only the supply chain structure, but all of the constituent processes including logistics and transportation activities. While logistics activities are no doubt governed by supply chain structure and material management policies, all said and done, there is a particular product or component that needs to be handled and shipped, and therefore the product architecture and design is central and determines every facet of the logistics function.

Thus if we are concerned with the shipping of the end-product, the dimensions, weight, density, intended use, fragility, customer need, and several other attributes are important when considering shipping alternatives, and designing customer delivery processes. On the other hand, if we are tasked with shipping the components that make up the product, the number, size, and state (pre-assembled, or separate) of those components when delivered, along with other relevant attributes need to be taken into account. Depending on the product architecture and manufacturing process, it could be that the transportation and shipping activities dealing with the finished product are dwarfed by the activities on the inbound side. The opposite could be true in other product/process combinations; all of these phenomena are driven by the product architecture: that is, what makes up the product, and how so?
5.5.1 Advanced technologies and turnkey programs.

In many cases, the process used to manufacture the product assumes equal importance; we define a process as the set and sequence of activities that lead to a finished product. In fact, in some situations, the process is indistinguishable from the product itself; the set of processing activities becomes unique to the product. Examples include highly customized products including infrastructure, manufacturing equipment, luxury and specialty automobiles, and aerospace products such as satellites and aircraft prototypes. The logistics and transportation involved in such product supply chains is also highly customized and not easily codified; this is not surprising given that logistics activities are a subset of the overall development and manufacturing process. We use the terminology “integral product” (“integral process”) to describe such products (processes); for such products (processes), the interfaces between the sub-components (sub-processes) are not so easily codified (see Ulrich, [192]).

When the interfaces between the sub-processes are not easily codified (for the sake of repetition), it is more difficult to split the responsibility of these activities among many firms. Outsourcing of sub-processes becomes that much more challenging, and requires a more tight-knit relationship between the partner firms (see an excellent conceptualization by Gereffi et al. [84]). This of course, has implications for the outsourcing of logistics; in turnkey programs such as new product development and launches, it is therefore common to observe a systems integrator, or a lead logistics firm, interacting closely with the lead firm to coordinate the many individual sub-processes involved in the program logistics. Examples include the Microsoft Xbox development and launch programs\footnote{See [94] for a case study of the Xbox program.}.
Thus there is a correspondence, of sorts, between product or process modularity, and supply chain modularity; the implication being that modular supply chains have multiple firms interacting in formal and arms-length relationships, each having responsibility for a chunk of the product or the process. Figure 5.6 provides a concise, albeit simplified, depiction of combinations of degrees of modularity in product or a process, and the degree of modularity in the supply chain.

5.5.2 Integral products or processes with modular supply chains.

As Figure 5.6 indicates, it is possible to observe modular supply chains working on integral products with closely coupled manufacturing and logistics. Although in this case, the modularity is despite the greater communication and coordination costs, and despite the risks arising from greater complexity. Rather the main driver of such modularizing decisions is the need to distribute the risks associated with the fixed costs of developing and marketing the product. The Boeing Dreamliner servers as a prime example of such a modularized supply chain; the decision to split the responsibility for design, development, and logistics was despite the complexity of the new product development program. There have been considerable delays in the program since its inception; although it is possible that the risk-sharing strategy may pay off for Boeing in the long run, especially since there is a sizeable backlog of customer orders for the new airplane.

5.5.3 Modular products with integral supply chains.

On the other end of the spectrum are highly modular products with sometimes well-codified (standardized), and modular processes. So, why would the supply chain
be close-knit and not involve multiple providers who can drive down costs? For example, consider the introduction of new consumer electronics products, such as the Microsoft Xbox, or the Apple iPad. While, from a design point of view, these products could be considered integral and complex (imagine the circuit design, for e.g.), from a mass manufacturing perspective they are indeed modular products and are built from standardized components and processes. Indeed, these products are not manufactured by Microsoft or Apple, but exclusively by contract manufacturers such as Singapore based Flextronics for the Xbox, and Taiwan based Foxconn for the iPad. The reasons, particularly in manufacturing and logistics, have to do with product learning, and with economies of scale. In the initial phases of the product, it is vulnerable to quality and process defects, which are best ironed out with a smaller set of suppliers; also the cost of the product and its logistics are higher in the early stages which calls for greater economies of scale and scope.
5.5.4 Stable technologies and supply chain processes.

Finally, there is the more familiar category of stable products and processes that are well understood and codified, so that multiple contract service providers could be engaged for not only various chunks of the product/process, but also for the same task or activity (to foster competitive pricing). Food and other consumer products that are staple, and raw materials and components for established products and technologies fit into this category. The logistics processes are well learned and understood within the industry for them to be commodity like, and therefore there are a number of suppliers vying for the same business. Familiar examples in the logistics world include small and medium sized package shipment, and similar order fulfillment processes. Amazon.com for example, contracts with both USPS and UPS for its distribution and fulfillment processes, and individual consumers may have had experience receiving shipments from either carriers.

5.5.5 Logistics in turnkey projects.

What does all of this taxonomy mean for logistics capacity management? Well, there are several implications based on the particular configuration of modularity in supply chain partnerships, and product/process modularity. Firstly, logistics capacity requirements are hard to measure within integral processes; this is because the processes are so tightly coupled, that the effective capacity is measured only by the weakest link, which is hard to predict in development projects. It is hard to measure which activities will entail resource/capacity consumption or utilization, and furthermore, the extent to which the capacity will be consumed for various processes. Thus, a crane used in construction, may be tasked with material movements in various parts of the site, and the extent of it utilization is not easily predicted in advance. Similar examples could be made for transporting components to build a
product prototype (such as an airplane); the carrier may have to transport the same component multiple times depending on rework and re-design of the component.

In these environments, the shortage costs are also very significant, as lack of logistics capacity can really hurt the customer delivery schedules leading to potential losses for customers and/or loss of goodwill. As a result, lead firms typically prefer to invest in their own logistics capacity, especially if the share of logistics cost is small relative to overall program costs; the exceptions are typically when the lead firm does not possess the technology or expertise to carry out certain logistics tasks, or when there are reliable partners who can provide the required turnkey solutions. Finally, economies of scope are more important in such programs, as opposed to economies of scale; there is more of a need for highly flexible logistics capacity as opposed to a need for pure efficiency.

5.5.6 Modular logistics with integral supply chains.

When the technology is unproven and exhibits a high degree of customization, logistics capacity may still not be any easier to forecast, even when the supply chain processes are modular; however, since modular processes have well-defined boundaries, the capacity requirements may be easier to measure and communicate. For example, with the introduction of a new product – say a line of apparel products – logistics capacity requirements are readily measured in terms of volume, units of shipment, and the anticipated schedule. However, since the demand for the product is relatively unstable before the start of the selling season, the logistics capacity requirements are hard to forecast too far in advance. Once the selling season begins, demand forecasts can be issued with more accuracy, and a schedule of logistics capacity requirements can be provided to the LSPs. Shortage costs tend to be high in
these situations, since new technologies and products are usually sold at a premium. Also, in this case, efficiency is as much a concern as flexibility, since demand could be unpredictable and logistics capacity needs to scale well to meet demand swings. Firms typically outsource logistics both for technology and (fixed) cost reasons; however even in this case, given that the precise logistics needs are difficult to forecast, firms prefer to work with a lead logistics firm to coordinate needed capacity quickly in order to respond profitably to growing demand.

5.5.7 **Tightly coupled logistics with collaborative partnerships.**

Similar to the previous discussion on supply chain architecture, it is possible to observe extensively and/or widely sourced logistics capacity even when the processes
are tightly coupled. In many such cases, the lead firm does not view logistics as a core capability or aspect of its business model. Examples can be found in the services, and advanced technology sectors; for example a firm building a new plant, or a hotel chain developing a new location, or a government infrastructure development program. In these cases, these organizations are best served by sourcing logistics capacity from LSPs that have similar experience with other firms; such expert resources come at a significant premium, but there is no in-house capacity as a feasible alternative. There are also considerable coordination costs that come with managing a diverse set of LSPs, and a systems integrator or a lead logistics provider is of great importance to such projects and programs.

5.5.8 Managing modular logistics with collaborative arrangements.

Of course, the biggest and most competitive market for LSPs are supply chains with stable technologies and processes. Given the diffusion of process knowledge, many LSPs have the capacity to support such supply chains, which drives down the costs, and also places an emphasis on service levels. Capacity is also interchangeable and augmentable, so arms length contracts are more the norm in this segment. Process modularity implies reasonably well understood capacity requirements, and hence contracts are written based on forecasts developed well in advance. Capacity for some specialized processes may be procured through reverse auctions. Given the arms length relationships, 3PL or contract LSPs are ideas candidates for such supply chains. Lead logistics providers or 4PL providers are only needed when there are a sufficient number of 3PL contracts to warrant a neutral agent who can negotiate better rates and also provide value added services such as shipment coordination. Again these systems integrators are typically of value only when the lead firm has limited expertise in the logistics domain, or the performance of 3PLs is of concern
relative to the needs of the firms’ customers.

5.6 Differentiating Supply Chain Fulfillment Strategies.

It goes without question that the manufacturing system structure and policies governing inventory and material movement determine choices that shippers make w.r.t. logistics. A one-for-one replenishment scheme within a base-stock model of inventory control will typically involve more transportation and logistics capacity than perhaps a batch replenishment policy. For another example, the choice and preference statements between different modes of transport (air vs. sea freight, for e.g.) is typically the result of the needs of the supply chain, and of course of the product at its various stages of development and manufacturing. In this sense, it is the vagaries of the supply system that drives true demand for logistics services. The end-customer forecast, demand, or delivery schedules are often altered and/or distorted by events in the manufacturing process; both on the inbound as well as on the outbound side. Hence it is indeed possible to differentiate such needs based on the inbound versus outbound (supply vs. fulfillment) perspective, and based on how removed the logistics activity is from the end-customer of the product.

It is also an important objective of this chapter to clearly outline how supply chain structure (whether by design or happenstance) and the specific inventory and order management policies arising thereof, determine logistics capacity requirements and planning. Researchers have attempted to develop taxonomies with which to describe and differentiate different supply chain structures and their environments; and their efforts have been quite successful in shaping both the theory and practice of supply chain management (see for example Bowersox et al. [20], Chopra et al. [37], Simchi-Levi et al. [169], and Zipkin [202] for excellent references). In this chapter,
since we are concerned with capacity management of logistics activities, it is the order management aspect of supply chain strategy that is of primary interest to us.

Here we describe, and contrast, four fundamental fulfillment strategies that are remarkably able to capture in broad strokes the vast majority of order management and inventory policies in use. These are indeed broad classifications, and therefore do not claim to represent the inventory control and order management policies in detail, but rather they are the guiding frameworks around which specific control policies are designed and implemented. They are also important given the driving philosophy behind operational improvements and policy making over the past few decades: lean operations and supply chain management. The foundations of lean management can be found in the need for greater efficiency; however, most manufacturing and supply chain operations are measured not only their cost control, but also by their service levels, responsiveness and reliability. These four fulfillment strategies essentially describe the trade-off between these competing pressures on supply chains. Any fulfillment strategy that is carefully devised aims to balance the need for greater efficiency (lean) against the need for greater service performance (service) as measured by the customers.

Efficiency is typically an internal performance criterion, and ensures that the firm is able to fulfill customer demand profitably; whereas service levels ensure that customers are satisfied or that the firm meets its contractual obligations with the customers. On both the efficiency and service dimension, we have an overly simplified classification of supply chains based on whether they fulfill customer demand from finished goods inventory, or whether customer orders are fulfilled without any such buffer stock. Thus, consider a firm that produces to replenish finished good inventory from which customer demand is fulfilled; in theory this firm should be able
to satisfy all of the customer demand without delay, and therefore has perfect service levels (of course, reality is a world of exceptions, where firms may not have inventory of the right product, and inventory may not be sufficient to cover all of the demand, etc.). However, this firm would be operating at less than 100% efficiency since it builds products ahead of demand materializing, and is also assuming inventory risk. On the other end of the spectrum, consider a firm that only produces and delivers products after a customer order has been received; in this case, the firm is efficient in terms of incurring more inventory related costs; although it would be accurate to say that the service level is less than 100%, if indeed fill rate is a component of service performance.

Next from a logistics perspective, we divide the firm’s supply chain into two
stages: the inbound and the outbound sides. There are logistics activities and or-
der management systems designed to supply materials to this echelon in a broader
supply chain. The item is assumed manufactured or processed in this echelon, and
upon completion, there are logistics activities that deliver the processed materials
to the customers of this firm (see Figure 5.8 for a depiction of this supply chain
cross-section). The fulfillment strategy for this cross-section of the supply chain can
be represented compactly by the following elements:

1. Finished goods inventory; if items are available in this buffer stock, customer
demand would fulfilled immediately.

2. In the absence of finished goods inventory, orders from customers of this unit/firm;
how they are accepted, and registered with the production/logistics system.

3. Ownership of the finished goods inventory; if the finished goods inventory is
owned by the firm, then in lieu of orders, the firm necessarily produces the
items based on a forecast of customer orders, typically generated by the firm
from independent sources; conversely if the customer owns the finished goods
inventory, the replenishment signals are in the form of orders to the firm’s pro-
duction/logistics system. (Note that any inventory that is part of the supply
chain shown in Figure 5.8 is because it is owned by the firm.)

4. Component or raw material (WIP) inventory; again if components are available
in stock, the firm can build according to either a forecast or an order bank by
consuming items from the WIP buffer.
5. In the absence of WIP buffer, the firm would have to place orders for components and materials with its suppliers, based on requirements it identifies either from its forecast of customer orders, or from its customer order bank.

6. Ownership of the WIP or component inventory; if the firm owns the WIP buffer, the inventory at the buffer can be assumed replenished to forecast; otherwise, if the WIP buffer is owned by the supplier, then any requirements arising from production of the end-item are necessarily issued as orders to that supplier.

Based on these elements of a fulfillment strategy, we can then define four configurations of order management in the supply chains. These configurations implement either an order based, or a forecast based fulfillment at the customer and the supplier ends. Before doing so, we discuss whether inbound logistics are really any different from outbound logistics from a planning and capacity perspective.

5.6.1 Inbound versus outbound logistics.

If an electronics firm runs out of finished goods inventory, or is behind schedule on a customer order, it could expedite the shipment to the customer by air; sometimes at a significant premium over the alternative of land or sea shipment\textsuperscript{38}. For inbound distortions, one could think of suppliers who are late in staging components for pick-up, or for that matter when multiple supplier schedules have to be coordinated/consolidated for delivery onto an assembly system. In every such case, the LSP may

\textsuperscript{38} Personal communications with Dr. Juergen Rahtz, VP, Kuehne+Nagel, Lead Logistics Solutions.
indeed make choices w.r.t. the mode of transport by trading off the value of time-to-delivery and the differential cost of the transport alternatives. This is a consistent decision model for modal choice regardless of the situation, and stage (lower or upper echelon) in the supply chain. However, the exact nature of the trade-offs and the set of alternatives before the logistics manager is what is dictated by the manufacturing process, and is conditional on the particular stage in the supply chain where the shipment or handling occurs.

One could surmise that the value of time-to-delivery would be greater at stages closer to the customer, especially when there are penalties involved for late deliveries. On the inbound side, the value of time-to-delivery is correlated with the cost of disruptions to the preset manufacturing schedule. The set of alternatives may also be different; it could be possible to expedite delivery of relatively small parts for automotive assembly by air\textsuperscript{39}, however, it is highly unlikely that a finished car or light truck will be delivered by air to a dealer or a customer. Interestingly, therefore, the true (opportunity) cost of supply disruptions is frequently dependent on the set of transport alternatives as well!

To summarize, depending on the product and customer in question, the shortage or delay penalty costs could indeed be greater at stages closer to the customer. Exceptions to the rule almost always involve hard scheduling constraints at intermediate stages, where shortage of component or materials is highly disruptive. However, what complicates the capacity planning exercise is that the transportation cost differentials between the set of transport or handling alternatives could also be very different for each stage in the supply chain. This could happen both due to prod-

\textsuperscript{39} Personal communications with manager of Alcoa Automotive (Forging) Plant, Hawesville, Kentucky.
uct characteristics (for example, the finished product may be really cumbersome to handle and expedite), and due to supply chain considerations (for example, the end-customer may be on another continent, while suppliers may be local). Hence, it is very difficult to develop thumb rules for logistics capacity planning based on process or supply chain considerations alone. The same is true for attempts to identify the merits and demerits of capacity outsourcing based solely on considerations of supply chain echelon or inventory control policies alone.

Even though the specific choices with regard to logistics capacity and service providers is hard to pin down, it is possible to develop a general theory around the needs and requirements of different supply chain structures, operating strategies, and control policies; this is what we aim to describe next.

5.6.2 Make-to-stock systems.

When the firm owns the FGI and the WIP buffers, this implies that the inventory replenishment through production at the firm (or echelon in this cross section of the supply chain) is triggered by a forecast of future customer orders; we then have a pure make-to-stock operation on both ends of the supply chain. We refer to Figure 5.9 for some attributes of products and systems that we find frequently deploying a make-to-stock model of fulfillment.

Some popular examples include automotive manufacturing in the pre-Toyota years, consumer staple products, and a wide range of consumer durable goods including electronics and electrical goods. Indeed it is safe to assume that the vast majority of manufactured goods still follow a make-to-stock fulfillment strategy for their supply chains, even though there has been a strong focus to achieve inventory
reductions without sacrificing customer service levels. Make-to-stock supply chains (or at least cross sections following this strategy) tend to be vertically integrated, with limited collaboration with suppliers, and arms length relationships with distributors. Inflexible production processes tend to be partly behind this strategy, since the constraints of the production technology require inventories to smooth the flow of orders through the supply chain. Inventory costs tend to be de-prioritized relative to other concerns such as customer response times or fill rate based service levels, even though there may be attempts to control these costs. Procurement lead times are also sometimes a factor that requires an inventory buffer to protect manufacturers from demand variability.

The incentive for a sales driven or push strategy to selling comes from a vulnerable market position; these products tend to be commodity or staple products, so customers could substitute to competing products. Thus inventory serves as a competitive mechanism for firms. What may sometimes offset the cost of FGI inventory is the low product variety, although it is possible to mention prominent exceptions such as Proctor and Gamble selling 28 different variants of toothpaste. Batch distribution or transportation also is a factor in make-to-stock systems; for example it would be very inefficient (and dangerous!) to make hourly trips from a fuel depot to replenish a gas station. Similarly on the back end crude oil from the Middle East is shipped in large vessels and stored in refinery depots and government reserves. These process constraints are often the underlying causes why a supply chain cannot operate without significant inventories.
5.6.3 JIT→Push system.

When the firm owns an FGI buffer that satisfies all customer demand, but does not own any WIP inventory, we have what we call a JIT→Push system. In this system the finished product is built to a forecast, but the components are ordered as needed from the suppliers. In fact, it is quite possible that there is actually WIP inventory that supplies the firm’s component or material needs, but that it is being held and owned by the supplier; such supply arrangements are of course called Vendor Management Inventory (VMI) contracts. In our definition, the supply system is Just-in-time from the perspective of the firm or echelon in question, because the WIP inventory is not recognized by the accounting system of the firm.

Examples of such systems are found in many automotive firms of today including Ford and Toyota; dealers issue a forecast of demand for cars and trucks, and the firm produces against such forecast schedules (although they are called “order
banks”, they are strictly speaking an aggregation of multiple forecasts from dealers). Exceptions are when large dealerships actually place firm orders that are financed sometimes by the automotive firm’s own financial arm, or when a large set of orders arrives from corporate and government customers. Other examples include popular computer or electronics manufacturing with vendor managed inventory that is supplied in small batches just before assembly.

The JIT→Push fulfillment strategy is frequently deployed when the lead time for manufacturing and customer distribution is high relative to the lead time for component supply; which happens when the manufacturing and demand centers are geographically distant. It also is prevalent when the firm has disproportionate bargaining power relative to its suppliers, and can essentially mandate a vendor management inventory system. There are few product variants with a high degree of substitutability, but typically a large number of components that are expensive to carry. Demand is also highly variable which calls for holding as little component inventory as possible, while market share concerns lead to a push or sales model of distribution. The illusion of continuous flow in operations is achieved through the push model; this helps to smooth material flows and balance the utilization of expensive production equipment.

In order to implement the JIT system on the back-end, either the suppliers must be willing to do VMI, or there must be tight integration between them and the firm. In return, the front end push system guarantees smooth flow and in some cases guaranteed volumes for the suppliers; in essence the firm takes on the risk of FGI inventory, and smoothes demand variability at the component level. The back-end processes tend to modular; otherwise JIT tends to be very hard to coordinate among multiple suppliers.
5.6.4 *Push→Pull system.*

For the reverse arrangement, when the firm maintains WIP inventories, but forgoes FGI stocks and builds the product only against “firm” customer orders, we have what is known as a *Push→Pull system* (see Simchi-Levi et al. [169]). In this system, the WIP buffer is replenished based on a forecast, and is owned by the firm. This strategy is frequently deployed to protect the customer from long replenishment lead times for components.

For example, the lead time for most semiconductor chip components is at least a few weeks, whereas computer manufacturers who label themselves as “build-to-order” cannot allow customers to wait that long to receive a shipment against an accepted order. Another popular example is found in fast food restaurants. This strategy is also deployed when the price of the component inventory fluctuates over time; so firms prefer to hedge against future price increases by writing supply contracts well into the future; examples include fuel distributors and depots, and aluminum or other metal fabricators; or indeed any industry with customized products made from commodity materials.

The *Push→Pull* strategy is ideally suited for firms with large variety in their product offerings, but which are all made from a few platforms or modular components; the products are differentiated very close to the customer. Perhaps the most famous exception is Amazon.com, which executes a near perfect *Push→Pull* fulfillment strategy without differentiation even in the back-end. Amazon.com is able to achieve this through an on-line retail or Pull channel, but where the inventories of millions of different items are heavily centralized into giant warehouse operations.
In addition, the pull system on the customer end is designed to deal with very choosy customers who do not substitute easily within the product portfolio; customers who demand variety as a pre-condition to purchase. Examples include fashion products that are highly perishable (where a sales driven or push customer fulfillment strategy very expensive), and of course an on-line retailer. The pull strategy also relies on a highly flexible and responsive front end; without fast logistics and scalable and highly responsible production, the pull strategy is difficult to implement with customers. It therefore helps to operate in a seller’s market, where the product (or at least the sales process and pull model) is well-established and customer demand at an aggregate level can be forecast with some accuracy.

Having a pull system requires a very high degree of trust between the firm and its customers; customers have to believe that products will be delivered within a reasonable time limit; whereas the firm has to believe that the purchase contract will be valid at the time of delivery. This is typically not a good strategy for new and profitable products whose demand can be volatile; the push strategy works better in such situations during the early stages of the product life cycle. The exception is when the product is heavily marketed and minimum demand volumes can be guaranteed; in which case the firm can wait for customers to signal their orders before stocking the shelves (one could give examples such as pre-orders for Harry Potter books, or retail stores ordering initial stocks of the new Apple iPhone or iPod products). Again, buffer stocks of components protect the firm from long lead times or unreliable suppliers. In typical processes, the back-end tends to be less modular than the front end, customer facing processes.

In fact many purported build-to-order systems are actually Push→Pull, systems
because if suppliers with long lead times resist VMI, the only option is to hold some
inventory of those components (think Dell and Intel: Intel cannot be mandated to
do VMI in all cases); in fact we describe the build-to-order strategy next.

5.6.5 Build-to-Order system.

When the firm owns no FGI or buffer stocks, and production and fulfillment is trig-
gered by live customer orders, we have a pure build-to-order system. Build-to-order
systems do not have any redundant inventory in the system; all of the inventory can
be traced to outstanding customer orders. The strategy is borne out of high cost
and/or redundancy of inventory, and a high degree of customization (i.e., product
variety is very high). Products tend to be modular, but this is not always true; when
the product is so unique and tailored that all components and their interfaces are
customized to the needs of the order. In relative terms, achieving responsiveness
through inventory buffers is not feasible, even if in some situations, the lead times
for assembly and distribution are high. In many cases in fact, the short lead times
for assembly, and procurement enable the build to order system; procurement lead
times can be shortened by vendor managed inventory systems. We term these as
VMI→Pull strategies. For example, companies like Dell Computers or Bose Sound
Systems require that component suppliers own stocks of their materials, and the
end-product is priced according to their availability; assembly lead times are rela-
tively short compared to the component production and transportation lead times.
A build-to-order system also relies on tight integration between suppliers, manufac-
turers and distributors; Dell calls this the virtual integration model\(^{40}\).

\(^{40}\) see a Harvard Business Review with Michael Dell, CEO of Dell Computers, [133].
ple is construction or infrastructure development (like highways and railroads). The processes are highly flexible to accommodate for mass customization with sometimes highly customized components; another objective to investing in flexible processes is responsiveness. The products and components could either be modular or integral (think of housing projects with uniform modules).

5.7 Planning Logistics To Support The Supply Chain Fulfillment Strategy.

5.7.1 Logistics capacity for make-to-stock systems.

Make-to-stock supply chains are more schedule driven; this attribute extends to the supporting logistics also. Capacity requirements are therefore easier to estimate. Inventory replenishment is the primary driver of material handling and logistics; as such the performance of logistics is measured by the inventory related service measures. It is not uncommon for transportation and warehousing managers to be measured by the fill-rate for customers, or by incidence of stock-outs. By definition, make-to-stock systems foster a batch mode of operation (one-for-one replenishment is typically expensive and infeasible in these scenarios), and logistics capacity is also consumed in batch quantities. Since inflexible technologies are common underlying factors, logistics processes are also similarly constrained along with the manufacturing. Capacity requirements are easier to measure and forecast; the forecast horizon is tied to the production planning horizon. Which in turn implies that logistics capacity can be staged with alternative options over such a time frame.

Logistics services tend to be more commoditized in this sector; stable demand implies long histories of operations, and technology diffusion guarantees intense competition in the logistics areas. The resulting scale efficiencies imply that logistics ser-
vices within such systems are more conducive for outsourcing. The inventory buffers serve to protect against stock-outs (implying higher stock-out costs), but only as long as replenishment happens according to schedule, and this makes consistency and reliability key attributes of successful LSPs. However, collaboration with suppliers is not as much a priority as with other systems, and arms length relationships and cost plus type contracts are much more common.

The exception is when material handling and logistics are technologically coupled with the manufacturing processes; for e.g. when items need special jigs and fixtures for transportation. Even in such cases, over time, such technical relationships can be well-codified and supply alternatives can be developed successfully. As a result, an important characteristic of logistics processes is that they are modular with respect to the other supply chain processes. Hence, there is a correspondence, of sorts, between make-to-stock fulfillment strategies, and modular supply chains, independent of whether the product itself is integral or modular. The modularization, or decoupling, of the supply chain processes is made possible by the inventory buffers between the different stages. Hence, attributes and desired features of logistics for modular supply chains are also common to those supporting a make-to-stock model of operations.

The inventory buffers decouple the back-end and front-end logistics depending, of course, on the size of the buffer relative to the demand (or supply). Hence, there is less of a need to coordinate or synchronize in-bound and out-bound shipments as we grow the inventory buffers, which is precisely the intent in many applications; such coordination could be more expensive than the added costs of inventory. The capacity requirements of the front and back-end logistics operations are by design more balanced, since both the inbound and outbound shipments are coordinated to
the same forecast; see again Figure 5.8. This coupled with more predictable shipment schedules leads to reduced need to have excess logistics capacity relative to the schedule; even though the overall capacity requirements and costs may indeed be greater than other pull-based strategies.

5.7.2 Build-to-order systems.

There is a similar correspondence between build-to-order systems and the advanced technologies and turnkey programs summarized in Figures 5.6 and 5.7. Build-to-order is generally the fulfillment mode of choice for firms operating in that space of product or process modularity and with the highly integral supply chain structures. As such logistics tend to be very different those needed to support make-to-stock strategies. Thus, demand for logistics capacity may be highly erratic, and very specialized to the various tasks in the project; forecasting is again a challenge, so flexible capacity is at a premium.

The primary performance requirements are responsiveness and adaptability, and the ability to work closely with the program managers. The needs could be one-off requiring highly skilled personnel. Capacity in such environments needs to be committed for production and/or assembly; even though utilization may not be uniform. Hence, pricing of this type of logistics capacity and service is often at a high mark-up. For example, cranes and other heavy equipment need to be deployed at a construction site for the entire duration of the project. Similarly, the Boeing 787 Dreamliner has dedicated and refitted 747 wide-bodied jets to ship parts from different manufacturing centers of the world to the assembly site in Everett, Washington.

Again, logistics services are highly integral to the build-to-order system, and
there is very little scope for decoupling activities. This requires careful synchronization between the front and back end logistics activities. For example the capacity of logistics resources on the inbound and outbound side must be synchronized to guarantee the required service levels. Operationally, this means coordinating the schedule of resources needed on the inbound and outbound sides. More strategically, LSPs need to provide long term partnerships and relationships to be successful at providing such turnkey services; the well-justified alternative for many firms is to just buy logistics capacity and operate a separate internal logistics division. Thus, the imperative for collaboration is strong. Pricing is therefore service driven and value based, as opposed to purely a cost based approach; any cost premiums for LSPs can be justified against the reduced opportunity cost of lower quality service.

Figure 5.10: Attributes of logistics capacity for various fulfillment strategies.
5.7.3 JIT→Push strategies.

The ready interpretation is that outbound logistics are similar to those deployed for make-to-stock systems, and inbound processes have attributes that are desirable in a build-to-order fulfillment mode. It is indeed true that the outbound logistics for a Push strategy is schedule driven, capacity easier to forecast, and shares many of the same attributes with logistics in the make-to-stock model.

This analogy goes only so far. The reason being that JIT and build-to-order are not necessarily identical operating environments. JIT systems are typically geared for uniform or balanced flow. Also recall that this does not necessarily imply a build-to-order system because inventories could still decouple the supply and manufacturing process; it is only a question of who owns the inventory and how it is staged. Thus for some JIT systems, the back-end is truly a build-to-order system; in such cases the logistics processes and capacity requirements share the same attributes as those of a pure build-to-order system. Capacity must be flexible, tightly integrated with the manufacturing capacity and processes (both on the supplier side as well as on the firm’s side), and highly responsive to standing orders. LSPs also need to be highly attuned to the operations managers, as supply orders may change priorities, or as the production schedule changes.

However, when the JIT strategy is effected through transfer of inventory ownership or shifting inventory locations up the supply chain, there are a completely different set of criteria in use. In such situations, there is a whole new space for specialized transportation and staging services, and a new set of opportunities for third party LSPs. The reason being that shifting inventory upstream has an important reason; namely variability for demand at the component level which causes a
need for expensive inventory which in turn cannot be sustained by the firm. In some situations, this variability is caused by the interaction between multiple product components with complementary demand; in other situations the components are indeed substitutes based on end-customer needs. In the complementary scenario, the inventory costs compound with risky demand, whereas in the substitutable case, customers may require a large number of variants which is again expensive. The solution is to simply outsource the inventory stocking to the suppliers, which goes by the term VMI.

The challenge in the context of VMI is to marshal the different components to supply an order kit corresponding to the production schedule; which is the new space for LSPs. Since the VMI system involves both batch supply to the stocking location, and coordination and kit-assembly for every production order, the overall logistics capacity requirements could be higher; this is in fact the main trade-off in implementing JIT systems. Firms respond to these added costs by co-locating the component suppliers and the logistics service providers in areas proximate to the assembly system. The co-location of supplier inventory requires tailored warehouse capacity, and the kit assembly and transportation requires specialized skills and transport. As such, this is a high-value added sector for third party LSPs. Although in many cases, the firm may choose to invest in its own logistics capacity to achieve greater responsiveness to the front-end, and for performance control. For example, Ford (or their key suppliers) partners with third party logistics providers for implementing a JIT-Push system with VMI\textsuperscript{41}, whereas Wal-Mart invests in its own (cross-docking) warehouses and logistics capacity\textsuperscript{42}.


\textsuperscript{42} See again Stalk, [178].
The capacity needs on the inbound side are synchronized to the production schedule; hence some level of capacity synchronization is built in with this strategy. Furthermore, because the front-end logistics capacity is planned against a forecast, the requirements are forecast with greater accuracy, and the back-end JIT operation also benefits from this stability. Hence, even without the decoupling of the make-to-stock system, the supply chain can still smooth the logistics capacity requirements. The situation further improves with VMI contracts, since in this case, the risk schedule disruptions due to component stock-outs is reduced, and this simplifies the logistics capacity planning exercise for the firm even further.

5.7.4 Push→Pull strategies.

Finally, a Push-Pull system may have its own unique set of logistics processes, and associated capacity requirements. Since variety and differentiation are the key drivers, a hub-and-spoke model of logistics is often a good fit to implement the design of the fulfillment strategy. This requires scalable warehousing capacity and systems coupled with a very responsive front end, similar to that in build-to-order systems. Similar to the JIT→Push model where kitting and assembly logistics plays a critical role, here the logistics of differentiation is the critical barrier.

This differentiation could be achieved in different ways in different supply chains: For example in the case of Amazon.com, the differentiation is achieved via automated sorting systems that pick, sort and route inventory into thousands of individual customer orders that could be for multiple and very different items. Out of the millions of SKUs that Amazon.com offers, customers could combine one of several items into a single order, with the option to have them shipped either together, or in separate
deliveries. If the customer requests one single delivery, the effective order variety on offer is unlimited. The problem is compounded when items are stored in separate warehouses; in this case, the logistics provider has to merge these items from the same order at a common location and then ship them out to the customer (see for example, Xu and Graves, [198]).

Another popular example of such differentiation that is logistics or transportation intensive, is the case of the Hewlett Packard printer manufacturing system. Given the relatively long lead times for component supply, and given the location of the manufacturing centers, production has to be planned well in advance, based on demand forecasts in several regions of the world. The challenge of course, was that different regions of the world have different power supply, and hence with different printer models and different country specifications, HP was selling over a hundred different variants of the printers. The solution was to make vanilla or common products until a certain stage (the so called “Push-Pull” boundary), and then differentiate the product only after much of the demand uncertainty was resolved. This was done within the European markets by having a single distribution center in the Netherlands, where the bulk of the region’s inventory was carried, and perform the final country specific configurations subsequent to receiving retailer or distributor orders. The final configuration involved light-assembly, as well as packaging\textsuperscript{43}.

Similar to the JIT→Push strategy, there is a need for specialized services at the boundary between the stocking and the fast-fulfillment stages. The logistics activities on the front end are closely coupled, whereas the back-end logistics are required to be highly efficient with cost and inventory control a major priority. Given the large number of variants involved the front-end capacity is required to be flexible, \textsuperscript{43} For more such case examples, see Rietze, [159].

241
but also scalable depending on the demand volumes. The need for synchronization (of schedules) between inbound and outbound logistics is reduced in this case; again as the size of the inventory buffer grows, there is a degree of freedom available to schedule back-end capacity, while still matching long-run average capacities. It is also true that being directly exposed to the variation in end-customer demand, without FGI buffers, makes front end capacity planning very hard; while at the same time, depending on the size of the inventory buffer, the back-end may also be face some of the effects of such variability. However, there could be some benefits in the back-end, especially if there is some risk pooling available from demand aggregation across the variants. Figure 5.10 attempts to compare and contrast the important attributes of logistics capacity across different fulfillment strategies.

5.8 Capacity Measurement, Investment, And The Outsourcing Option.

5.8.1 Towards a consistent measure of freight capacity.

The diversity of logistics processes listed in Figure 5.2 leads to the conclusion that it is futile to create a common measure encompassing all of the processes. To start with, we could restrict our attention to those activities that are capacity constrained, and whose capacity we are interested in determining; those that do not suffer from capacity limitations, we can safely ignore from this narrow perspective.

Since all of the processes exist to support an end-customer order, one possible approach is to understand how many orders a given service activity or process could sustain over a common time interval. This is a first step towards consistency. Thus, one could measure capacity of a process by the number of end-customer orders that can delivered over a specific time window within the planning horizon. Stretching
this time interval, we can eventually measure the capacity of an individual process over the entire planning horizon. This would indeed provide a consistent measure for capacity across all of the myriad logistics processes that support an order. Given a small enough (but still uniform) window, this measure of capacity can also be understood as throughput (or output rate) of that process in end-customer orders.

Let us examine some complexities involved in this approach. Firstly, the requirements could vary over time. However, this throughput measure of individual process capacity could easily be pegged to those evolving requirements. Secondly, the same process (say transportation between two fixed locations) can have different throughput based on the variant of the customer order being fulfilled. To resolve this, for the case of a process serving multiple order types, we can assume a weighted average requirement for such processes. Hence, because a process could be specific to a subset of the customer order types, there may be uneven capacity requirements observed across the set of processes, but the measure is still a consistent one. Thirdly, the demand within each order class may not be known in advance, and could be subject to randomness, which implies that our order throughput measure of process capacity requirements will also therefore be uncertain. For a work-around, we may then wish to split and redefine the processes so they are specific to an order type, and so on...

See Figure 5.11 for an illustration of these ideas; in particular, the two processes depicted are loaded differently when measured in demand for Order Type 2. For the sake of abstraction, we could assume that the two processes do not involve any common resources from Service Provider X. Note also that there is also an issue of how best to allocate process capacity requirements to each service provider; this sourcing split could depend on the pricing and availability of the two service providers during the planning horizon. The figure shows how different order types can utilize
different amounts of process capacity, and therefore one needs a common measure of capacity for such systems interconnected by process flexibility and service provider capabilities (in this case throughput of Order Type B). Finally, observe that in this illustration, capacity is measured at the process level, and not at the resource level; one could imagine situations where the processes share resources.

A major drawback of the order throughput measure is also revealed through the example. While the requirements are translated to the service providers through the sourcing split decision (2:3), it is not clear why 40% of Process 2 orders are sourced from Service Provider X. The second issue raised is when, say, the inbound and outbound processes share common resources (such as a fleet of trucks); in this case we need to incorporate another hierarchy of requirements: at the level of the resources supporting the two processes. The level of granularity is a choice for analysts making capacity related decisions; but for this concept paper, we have abstracted the requirements to be measured at the process level. Resource level analysis requires delving into different categorizations of resources; the obvious challenge here is that the diversity of resources typically in use, and their individual measures of capacity could become even harder to translate into a common or consistent measure. This, furthermore, is also beyond the detail we seek to present here.

A final point is that this approach to measuring capacity requirements is well suited to a Push strategy where a pre-defined schedule of deliveries is available. In a pull-based system, control of the internal logistics is handed over to the end-customers, who in effect specify when and how they intend to utilize the logistics capacity; this leads to greater variations in arrival (and hence throughput) of orders, and ultimately in the process capacity requirements. An order throughput based measure works well when we have greater control over the schedule of end-customer
order shipments. For such Pull systems, we need to define a more schedule and delivery window based measures of capacity, as we discuss next.

5.8.2 **Lead time and schedule based measures of capacity.**

Unlike the stylized depiction of Figure 5.11 processes may not always be decoupled from a schedule or time viewpoint. For example, the processes could be constrained to lie within specific time intervals in the fulfillment or delivery schedule. Thus, outbound delivery of an order will have to wait until the assembly or manufacturing steps have been performed, and manufacturing can commence only after all components are made available. In a Pull environment, customers place an order and wait for it to be delivered. Hence, the true measure of service is not how many orders we delivered, which naturally leads to a throughput based measure, but rather the
timeliness of the order delivery, and how many orders we deliver within a target service time. For a Pull system, we therefore follow a single order class through all of the logistics processes that need to be completed for delivery.

Consider Figure 5.12 for another illustrative example. Order Types A, B, and C have customer required delivery times of 5, 7, and 5 days respectively. There are two service options (and providers) for the inbound component delivery, and only a single option for the outbound delivery. The air freight option for inbound delivery takes 1 day irrespective of order type, while the rail option takes longer. The assembly lead time is also 2 days irrespective of order class. Given the customer delivery time window, we see that Order Type C is not a candidate for Just-in-Time delivery, and hence the firm needs to either develop an air freight capability through Service Provider X, or arrange for some inventory buffer on the inbound side (VMI or internal) that will reduce the response time to less than 5 days; currently there is not sufficient delivery capacity to promise that level of customer service. Similarly, Service Provider Y is not a feasible provider for inbound deliveries towards Order Type A. On the other hand, assuming that air freight is more expensive than rail (which it almost always is), we see that Service Provider Y is the best option for inbound deliveries for Order Type B.

While the example above is very simple, it does point to the main benefits of a lead time based, and schedule compliance based measure of capacity. Even without a resource view, this type of measure tells us where the potential bottlenecks are in the fulfillment process, and what alternative capacity could be engaged to alleviate the bottleneck.

To make this rather simple example a relatively sophisticated problem, one only
needs to consider lead time variability (or uncertainty) for each of the four processes depicted. Say for example, the inbound rail transport lead time for Order Type B is now uniformly distributed between 2.5-3.5 days (assuming congestion and other rail schedule concerns in the winter season). In this case, there is a 50% probability that the rail option would cause the delivery lead time for this order to exceed the customer target. Hence, if the cost of expediting is less than the delay penalty cost, the firm would seek to source capacity from Service Provider X for this order type. Exactly what fraction of the orders would use the fast channel would depend on the relative cost of the two modes; but the point is that lead time variability alone is sufficient to warrant a contract with both service providers.

There are again several drawbacks to this approach to determining capacity requirements. Firstly, this approach does not make clear the link between resource investment and the capacity of the dependent process measured in delivery lead time. Rather the investment is a more binary or discrete decision: for example,
the question is whether to contract with Service Provider X, or Y? Not how much capacity to procure from either provider. Secondly, as pointed out earlier, it is not immediately clear how to divide the order bank between the service providers; and thus shipping volumes, which are a critical component of any service contract, are not so easily specified. The order throughput based measure is clearly better able to capture aggregate measures of capacity.

5.8.3 *Combining order size and lead times into a throughout measure.*

On closer inspection, superimposing the order size measure of capacity onto the lead time measure of capacity resolves a key problem revealed by either of the illustrative examples. The issue of the sourcing split is decided, at least in concept, by considerations of lead time uncertainty, while the issue of volume of orders through either channel is decided, again in concept, by minimizing the sum of the expected expediting and delay penalty costs.

For example, one would then measure process capacity through the maximum number of orders of a certain type that could be delivered over the process lead time. In the case of a rail option, this would be as many orders as could be delivered via several containers. In the case of air shipments, it would be as many orders as could be shipped via several pallets. This also leads to a more consistent measure of capacity across different processes. Thus, if the number of orders that could be shipped by a single air shipment is much smaller than the rail option, then despite the shorter response time, the throughput (orders in a shipment/lead time) could indeed be greater for the rail option. Then if it is cheaper to hold several rail cars worth of shipments as inventory rather than expediting smaller quantities by air, one has a different cost-benefit model for the sourcing split.
5.8.4 Resource level measures of capacity.

In this chapter, we have taken a higher level process view of logistics networks and their capacity requirements. Implicit in such a view is an assumption – highly erroneous for many supply chains – that every process has exactly one unique resource assigned to it, whose capacity is being measured in accordance with the customer needs. In reality, there could be multiple resources required to complete a process, and furthermore, these resources could be consumed at different levels. For example, two different trucking routes within a country could potentially use the same fleet of trucks leased from a third party provider. In this case, the capacity requirements are registered at the process level, but must be translated to the resource level capacity measure so that the fleet size could be planned. However, translating order throughput measures into resource level requirements is just a logical extension of how we translated customer demand to into process level requirements. The translation process could be highly technical and very specific to the resource, but it is the same idea taken one level down the planning hierarchy.

In fact, if data were available on the intended utilization rates of different resource types, we could even measure resource requirements in order throughput rates. But from a capacity planning perspective (and also an execution or capacity implementation standpoint) we are ultimately investing in the resources, so the resource requirements are better expressed in the amount or number of resources needed. Even better, at the resource level, if we know the true cost of resource acquisition, we could express the resource investment in monetary terms. As mentioned earlier, we will forgo the resource level view for this concept paper, except when we discuss aggregate capacity pricing of outsourcing alternatives. The resource level view becomes more
appealing in modeling and algorithmic contexts, where the capacity sizing problem is solved at the resource level.

5.8.5 Inventory storage capacity measures in push systems.

The bias in this chapter has been towards measuring and analyzing capacity requirements for transport and freight related processes. However, inventory storage capacity can also limit the responsiveness of the system that is designed to carry inventory. To understand this, consider the fulfillment of Order Type C in the lead time measure example of Figure 5.12. Since, it is not possible to use a build-to-order approach to satisfy orders of this type (the target response time is less than the raw transport time), the only other alternative is for the system to carry some inventory, either at the finished good stage or at the inbound stage to supply stock to assembly. Suppose carrying inbound stock is cheaper; also suppose that following a classical EOQ model that the number of orders of Type C was fixed 20 orders per day, and that components were supplied every day at that same rate to the inbound stock center, before assembly starts for the day.

Since the supply happens on a daily basis, we can then make sure there is a minimum stock of components to build 20 orders. Will the system work if there is only warehouse space for 10 orders? The answer is yes, because we will only have supply for half-a-day’s worth of orders, and the remaining 10 orders can only be built from the next day’s deliveries. Thus 50% of the orders will have a lead time of 4 days, and the other 50% will have a response time of 5 days which is still feasible. If the warehouse capacity is smaller than 10 orders, the system is no longer capable of guaranteeing 100% service levels. One could extend the same logic to determining the capacity of not only the warehouse, but also of the entire factory for building the
orders.

It is therefore evident that warehouse or inventory storage capacity is interlinked with the throughput of the supply system, and the demand rate for that inventory. So, what is a consistent measure of warehousing capacity? In most situations it is just volume of space either in containers that transport the inventory, or in warehouse locations. The capacity question is really how much space we need during or after transportation; and this means measuring the inventory requirements.

Little’s Law provides an answer: Given the average demand (or supply) rate, and the lead time of supply, the average inventory in the system is the product of the demand rate and the average waiting time for the orders (see [130]) before they can be used in assembly or shipped off to the customer; furthermore this inventory requirement is measured in orders. Hence, we have a naturally consistent way to measure the inventory requirements. Of course, this measure may be hard to implement, especially when thousands of individual components make up an order, but in concept the order measure of inventory is consistent and can be used when there are multiple order types, and multiple storage locations, etc. Of course, inventory can also be measured in monetary terms, but that is not a useful way to compute warehousing space requirements, which is our main objective.

The obvious drawback to the above approach to actually determining capacity requirements, is that demand or supply could variable (random); then the warehouse may need to accommodate more the average number of orders. Then determining capacity requires an understanding of the variations (distribution) in the demand or supply. That is not an easy problem to analyze, especially for general networks with multiple inventory storage points. The way firms deal with such excessive require-
ments is to rent or lease the excess capacity needed from outside providers. Hence, the warehouse sizing problem is also one where the translation of fixed to variable capacity cost, and the flexible options contracts could be important design problems.

The one reason to not group inventory storage decisions together with transportation capacity decisions is when storage costs (either in its fixed or variable forms) are typically small relative to fixed and variable costs of transportation. In fact this is true in a vast majority of applications. The inventory storage itself may be expensive, given holding, and handling costs, but the storage capacity is typically measured in volume of space required, and for most supply chains, warehouse space itself is not a serious limitation. The rationale behind methodologies such as cross-docking is usually to avoid costs from standing inventory, rather than to minimize warehouse space\footnote{Facility location models explore the impact of spacing and distance on throughput measures, so too much space may also be a problem}.

5.8.6 The fundamental trade-offs in logistics capacity planning.

An individual process.

Here we sketch the basic trade-offs in capacity planning for an individual or stand-alone logistics process. In doing so, we ignore the inventory storage capacity problem, and just focus on the freight capacity measure. The objective is typically to maximize profits or to minimize capacity shortage and excess costs. We will assume a consistent order throughput measure of capacity for all processes under consideration. Since we do not consider any inventory storage, this is a pure build-to-order system. We could also suppose that there is a fixed cost to building capacity, as well as a variable cost component that depends on the scale.
The shortage of capacity implies that the order throughput of a particular process is below the required level dictated by the end customer. Since the measure of capacity is throughput, the shortage costs are realized through penalties for delays in fulfilling orders. In the case of lost sales, the penalties are rather severe, and the entire revenue for those orders is registered as the penalty cost. In other situations, the capacity shortage will result only in penalties that result from contingency actions such as expediting. For either case, we will assume that the penalties are again measured in dollars per unit of capacity shortage.

Capacity is typically a perishable asset; and this is true for our simple example. By having the capacity to output orders at greater than the rate demanded by customers, we have a redundancy because there is no scope to build inventory and meet future demands. The cost of this redundancy is whatever premium we paid to have access to this excess capacity.

In a deterministic environment, where demand and supply rates are known for a time window under consideration, there is not much of a planning problem. We simply attempt to invest in capacity that is very close to the actual demand rate for orders (assuming that all orders are profitable). The real problem of course is when either supply or demand is time-variant, or random. For the time variant case, if capacity decisions (in the short term) are irreversible, then the trade-off is between incurring capacity shortages during some heavy demand phases, and incurring redundancies for other low demand phases. If the demand rates for a given time window are random, then the trade-off is between the expected cost of capacity shortage and the expected costs of capacity (not unlike a news-vendor or newsboy problem in inventory planning).
Serial and parallel processes.

With parallel processes – for example the supply of components before an assembly processes – the shortage and redundancy is not only relative to the demand, but also relative to the parallel process capacities, depending on the type of coupling. It would seem intuitive to have equal capacity for parallel processes in an assembly type system, and this would be true when the demand and supply rates are deterministic and constant over the planning horizon. However, in the case of time varying or random demand rates, one could easily provide examples where different shortage or acquisition costs could lead to different capacity choices. Another reason for redundancies across parallel processes could be for the sake of synchronizing schedules. Inventory buffers could also be used to synchronize schedules and to manage redundancies, just because of lot size limitations of different transportation processes; for example a railroad shipment would deliver a bigger lot size in less frequent intervals than a truckload shipment.

The interesting feature of serial systems is that end-customer orders would be fulfilled by the lowest echelon, and hence we could determine the capacity of this stage based on the order arrival process. Similarly, each of the upstream stages has its capacity pegged to the immediately down-stream stage requirements. This means the last link capacity is pegged to the customer order process; the second echelon is pegged to the first, and so on. Inventories can help smooth the flow of orders, but their cost must be weighed against the cost of just matching the throughput capacity of the interlinked processes. The justification of redundancies in capacity for any individual arc are not as clear in this case; except when there are technological constraints such as lot size or schedule conflicts.
Short and long term stability.

An interesting observation that holds true regardless of the network structure is that inventory and logistics capacity (measured in order throughput rate) are in effect short-run substitutes. One could hold more inventory in the short term in order to avoid expanding the capacity of a feeder route, or without suffering shortages relative to end-customer demand. However, in the long run, inventory and logistics capacity (or velocity) are not substitutes. In the long run, the logistics process capacity must match the material supply rates for parity, and furthermore, unless it is profitable to forgo some customer orders over time in favor of lower logistics capacity, the long run throughput rate must equal the arrival rate. Essentially, if all of the customer orders are admitted into the system, we must have long run average capacity to at least match with the order rate, to prevent instability. This is nothing but the well-known result (or pre-condition even) for stability in queueing systems.

5.8.7 The broader network design problem.

The goal here, if not already apparent, is to provide some guidelines and insights on capacity investment or expansion for various processes in the logistics network. In a perfect world, where orders can be predicted in their quantity and timing well ahead of their arrival, we could then use these requirements for individual process capacity, and then construct a logistics network – without consideration of a budget or availability of capacity – to output precisely the orders required, at precisely the rates demanded by customers. In such a deterministic scenario, this consistent measure of process capacity tells us how much of the network resources have been consumed by each order, and therefore tells us the true logistics costs of fulfilling each order.
Capacity investment constraints and service level trade-offs.

If however, the available investment budget is limited, one has to decide which processes are more critical based on a priority scheme, or based on the revenue stream from different order classes. In this case, an interconnected and flexible network will not fulfill every order as demanded, and therefore we have to optimize the capacity allocation within the network based on such criteria. A related problem is to actually measure, given a capacity configuration, the fraction of orders we are able to fulfill from every class. This is also called the service level for an order class, and sets up a classical design problem where we must determine the capacity to be invested for each process subject to minimum service level requirements.

Network operating policies and impact on capacity.

In most real world networks, the routing of the orders, in a dynamic sense is also a factor that impacts the capacity. For example in a congested network, just routing orders through the same congested routes is guaranteed to reduce effective system capacity. This is a very rich field of study not only in logistics and transportation networks but also telecommunications and internet traffic systems. We refer to Bramel and Simchi-Levi, [25], for some interesting policy analysis and network design problems. Indeed, a vast array of fundamental problems in network design and policy making have been studied in great rigor. These include facility location, fleet sizing, measuring effective fulfillment capacity and delivery service performance, measuring delays in processing due to network congestion, the routing of orders within a congested network, the optimal packing of containers, to mention just a few. We classify them as important operational approaches that are necessary for detailed performance analysis and for an integrated approach to the capacity planning problem. One challenge is that of synthesizing more elaborate and interconnected models
of network capacity planning that are still relevant to high level sourcing decisions.

Far from such an ambitious exercise, our objective in this paper has been to highlight a select few trade-offs in the network design problem, that result directly from a firm’s choice in fulfillment strategy. In the following discussion, we explore how best to choose a contract type (fixed price or an options contract) and the range of services offered from a menu that might be available to the firm. The choice in contract type is an indicator of whether a particular group of services should be outsourced in the first place, since an internal provider that can match the fixed price and also the range of services of the external provider, is obviously also a candidate.

5.8.8 Comparing different capacity cost or pricing structures.

Fixed price and options contracts.

As discussed in Section 5.4, with outsourcing, instead of a fixed, one time investment in internal logistics capacity, the supply chain could utilize variable amounts of external capacity over time without incurring costs that result from excess capacity relative to the demand requirements. In many situations, this reduces the exposure to business risk for the supply chain, and reduces break-even sales volumes for profitability. At the same time, with an outsourcing alternative, the firm can potentially scale up the requirements as needed.

In addition to the flexibility in procuring variable capacity levels over time, the pricing (or cost to the firm) of the capacity could also flexible. Instead of paying a bulk amount for the capacity procured as a fixed investment, the service provider could offer procurement options at a reservation fee, which could then be exercised at a pre-determined strike price when needed by the supply chain. Thus, a fixed in-
vestment in the form of a reservation fee for a certain capacity, not only ensures that the firm will have all of that capacity available for use in the case of strong demand, but also the firm will not incur any downside risks related to redundant capacity. This provides another level of transfer of fixed to variable capacity costs, but more importantly this allows the firm to hedge against both upside and downside demand risks.

Of course in most cases, the service provider structures the offering so that the capacity options if fully utilized will be more expensive than the fixed price (without options) contract. Otherwise, there is no incentive for the firm to purchase a fixed price contract, and the firm can just hedge against downside risk by purchasing the cheaper options at the reservation price.

*Aggregate capacity contracts.*

With the higher tier lead logistics providers, the pricing of capacity also reflects the extent of aggregation in the capacity of interlinked logistics processes. Thus, many service providers would offer a single capacity procurement contract for several routes or items on the inbound or several order classes on the outbound side. In some cases, the service provider could offer to provide both the inbound and outbound services. Some of these advantages have been discussed earlier, when we described the role of lead logistics providers and systems integrators. The benefits are often three-fold (at least as envisioned by the service provider).

The first benefit would be better synchronization of the capacity of various coupled processes and needed resources, so that the overall capacity requirements would be lower and the processes more efficient. In reality, the impact of such synchronized
capacity could be enormous; by staging resource availability to correspond with the process requirements, and by pooling flexible resources that can cater to multiple processes, the lead logistics firms can provide substantial reductions in capacity requirements, when measured at the resource level. Such synchronization would not be possible if processes are delegated to many different service providers, in a way that limits resource commonality, and also limits the scope for coordinating and staging resources across different processes.

Paradoxically, it is the synchronization of capacity that is a major barrier to outsourcing; internal logistics practices evolve over many years of operating to provide such capabilities. This is what constitutes expertise in a given industry sector or supply chain process. However, the fixed to variable cost translation is often valuable enough to a firm dealing with uncertain demand or supply environment, so that it forgoes such synchronization capabilities developed in-house.

Separate from the synchronization benefits, are also the economies of scale achieved by pooling common resources for use across multiple processes. This typically helps in resource procurement, as a large batch of resources can not only be procured cheaper, but also their maintenance costs over time could also be lower. This is one main reason why larger service providers such as UPS and FedEx typically have lower total operating and maintenance costs for their fleets of trucks and aircraft, relative to smaller providers. These savings are eventually passed on to their clients through more competitive pricing, volume discounts, and greater reliability of service.

Similar ideas apply to the administrative overheads such as the cost of running a complex and responsive IT and communications systems; with the more established lead logistics providers, not only are these systems more advanced that with
the average competitor, but since these providers serve multiple customers, they can spread out the cost of development and maintenance of these IT systems across their client base. The impact of such IT systems on capacity management is not trivial; with better support systems capacity can in theory be synchronized better from a scheduling and staging perspective.

A third benefit from aggregation of course, is that flexible resources can yield not only economies of scale in capacity procurement, but also potentially risk pooling benefits if process requirements are uncertain and are possibly negatively correlated. To use the example of UPS or FedEx again, a standardized fleet within well-defined process segments (that is grouped by loads, range of distance) makes any truck or a plane within a segment interchangeable and flexible to serve any customers requiring such loads not just on specific routes, but for all routes that fit within the distance range. Contracting with multiple vendors for different routes and investing in capacity separately for each route may not provide the same level of risk pooling benefits.

While detailed econometric studies have not been reported, to our knowledge, on the predominant motives for firms partnering with lead logistics providers, it is our conjecture that the synchronization, scale, and risk pooling benefits should explain much of the cost savings realized by their clients.

5.8.9 Fit between contract type and the fulfillment strategy.

As depicted in Figure 5.10, fulfillment strategies have differentiated logistics requirements, and it is possible to outline the attributes of corresponding logistics capacity requirements. Similarly, based on such requirements, it is also possible to provide some guidelines as to how these requirements could either be procured or developed;
which is our objective here in this final section. Quite naturally, the procurement contract is dependent on the type of service that is engaged and the scope of the service provider’s relationship with the supply chain. Note that we are here concerned with the fixed and variable cost of the capacity, as opposed to the fixed or variable cost of the order fulfillment. The difference is that any component of capacity that can be converted from fixed to variable, potentially adds to the variable cost of order fulfillment, but reduces the fixed cost exposure of the firm from capacity investment.

1. If the service provider is an internal logistics division, one could characterize the contract as a fixed price contract; the firm incurs in most cases an administrative overhead, a fixed cost for the initial investment in logistics capacity, and a variable cost component for the utilization of the capacity.

2. With a third party or contract logistics arrangement, the firm could write either a fixed price or an options contract. The options contract would reduce the fixed cost component; there would still be a fixed upfront investment in the form of a reservation fee per unit of capacity reserved, and a subsequent variable cost for every unit utilized.

3. With a lead logistics contract, there is an opportunity to aggregate the capacity requirements, and design the same kind of contract as with any other outsourcing service provider. The aggregation could happen at three levels: an aggregation of inbound logistics activities, one for the outbound deliveries, and a third level of aggregation combining inbound and outbound capacity requirements into a single contract.
4. We treat the systems integrator and a lead logistics provider identically with respect to their ability to aggregate the capacity requirements; an additional impact of systems integrators as discussed earlier is that they could make more efficient use of the available capacity, and lower the capacity requirements in aggregate.

Figure 5.13 attempts to explain broadly the fit between contract type and the role of the logistics processes within the order fulfillment strategy. There are three broad categorizations depicted: based on whether the logistics process support a Push or a Pull, based on whether it is an inbound or an outbound (customer facing process), and third the level of aggregation involved.
Push systems.

In general, fixed price contracts are beneficial in a Push system, because the transportation or storage largely follows a schedule and therefore the requirements can be forecast in advance with relatively few errors. Of course, uncertainties in the supply, transport, or demand processes could lead to fluctuations in capacity requirements, but because our definition of a push system is one that builds to forecast, and draws from inventory there is some degree of protection from demand or supply variations. Of course, if the options contract is not too expensive in relation to the fixed price, it could indeed be the best approach; but here we are describing it as a tendency. Aggregation typically reduces the variations in requirement even further, so if it is beneficial to write a fixed price contract for an individual process in a Push environment, it would typically also be the case for an aggregate contract.

Build-to-Order systems.

The Build-to-Order systems, on the other hand, typically are well served by options contracts. The primary reason being that we do not have the inventory buffers and forecast based schedules that protect logistics processes from supply and demand variations. Options contracts are indeed utilized as mechanisms that provide contingency capacity to system managers to transport components and orders that deviate significantly from the estimates. The reservation fee is typically lower than the alternative fixed price, which allows the firm to reserve more capacity in advance than they would otherwise. While the exercise or strike price for capacity utilization can be a significant expense, in situations where the opportunity cost of delayed orders is of more concern, this creates contingency capacity for a wider range of demand and supply scenarios. Another main benefit of an options contract over a fixed price contract is that the entire requirement is not committed in advance, and if require-
ments are below estimates, the firm saves by not carrying redundant capacity.

Aggregation is even more powerful in these systems. It can potentially reduce the margin of error in the requirements forecast, and the fixed costs in the form of the reservation fees can be determined with greater accuracy. Aggregation provides even greater risk pooling benefits when the (resource) capacity is substitutable across processes being aggregated.

*Hybrid systems.*

In hybrid systems, the front end operating policy typically has greater influence relative to the back-end policy; the customer facing end sets the beat for the rest of the supply chain. If the front end is a pull system, while intermediate inventory can protect the back end from the extremities of demand, there will still be greater variability in the back-end operation. Hence, it is typically advantageous to write options contracts for processes on either end of the supply chain. However, aggregation can still minimize the impact of such demand variations to the extent that fixed price contracts could work for a Push→Pull systems. This is possible because the back-end push system can help reduce the impact of supply variations, leaving only demand variations to counter. All said and done a fixed price contract would be cheaper for the same capacity utilized as an options contract.

When the front-end operates according to a Push system, but the back-end delivers to orders (JIT), the front end processes are now vulnerable to supply variations, and the need to expedite or down-scale capacity requirements makes options contracts valuable. On the outbound side though, we typically have inventory buffers that serve the end customer, so the need is to expedite orders when the inventory
buffers are depleted because of short term supply disruptions; here again options contracts could be useful. Unless inventory costs or storage capacity concerns arise, the need to down-scale front-end logistics capacity would not be as strong.

5.9 Summary And Further Work.

In this chapter, we have attempted to explain using fundamental operations concepts, the differences between a logistics strategy that relies on internal firm resources, and an outsourced strategy that engages external capacity for various logistics and transport activities in the supply chain. Our approach in this paper was independent of the product or industry sector, and was rather founded on a general concepts from operations and supply chain management that are universally applicable. We focus on the capacity sourcing and planning aspects, as opposed to the execution and utilization decisions.

The chapter is aimed at both an academic as well as practitioner audience. This paper contributes to the literature on logistics capacity planning, and is also applicable to general models of capacity sourcing and investment encompassing both logistics and manufacturing processes. Our first (admittedly far from being ground-breaking!) contribution is a broadly synthesized but still concise view of antecedents of outsourcing in the logistics sector; this is nevertheless important as a motivation for understanding how best to incorporate external capacity in a heavily outsourced world. We also provide a service based segmentation of typical logistics processes and provide brief descriptions of the different tiers of service providers and their basic value proposition to firms or supply chains; this helps us understand evolving industry structure and illustrates the range of procurement options w.r.t. logistics capacity.
We then present a framework for linking logistics process attributes to the product and supply chain architectures. Different product environments clearly need different logistics solutions, but our framework is novel in that we explain how different degrees of modularity or interconnectedness in the product can actually create supply chain environments that are more or less conducive to the outsourcing of logistics processes. We believe this is not only a contribution to the literature, but also an important framework that can guide (if not effect) decision-making in practice.

The third contribution is a similar and compact framework that outlines how different supply chain fulfillment strategies and choices firms make in this regard, can lead to differentiated logistics capacity requirements. The broad division is between systems that build inventories in response to forecasts, and those that build directly to customer orders. We consider representative supply chains implementing variations of these fulfillment strategies, and explain how the corresponding logistics requirements could be differentiated. Furthermore, we also present a framework that indicates which service providers could be best suited for executing a chosen fulfillment strategy. These concepts are not necessarily new, especially to practitioners who witness the industry evolution first hand, but the main contribution, we believe, is in the synthesis and in the representation.

We then present several consistent measures of logistics capacity; the consistency is critical from an integrated capacity planning perspective that spans multiple service categories and resource types within a supply chain. These consistent measures allow us to explain the fundamental trade-offs in logistics capacity planning, and allows us to compare different sourcing and contracting options using a common language. We conclude with a framework, also a contribution to the literature on
capacity sourcing contracts. This framework explains how supply chain fulfillment strategies could guide the choice between cheaper, more uniform, and predictable fixed price contracts; and the more flexible, and typically more expensive options contracts for logistics sourcing.

This chapter is intended as a pre-cursor to the more rigorous logistics capacity planning model of Chapters 9 and 10, where we aim to understand in greater detail the economic incentives that spur outsourcing decisions and develop some guidelines to frame the resulting procurement contracts. That work can therefore be considered a companion piece, where we provide more rigorous justification for the guiding frameworks and concepts presented here. Further modeling, beyond Chapters 9 and 10, in order to clarify and even challenge the conceptual frameworks presented here is certainly an avenue and future stream of research. The other avenue is to gather empirical evidence that reveals whether supply chain managers indeed make sourcing decisions in accordance with these guidelines, and if not, the constraining factors thereof. There is also scope to look at industry verticals and observe variations on these frameworks that have either been adopted with great success or that refute them as counterexamples. Finally, we have not discussed some emerging issues of interest to the logistics industry such as green supply chains (for e.g. accounting for carbon impact of logistics), and reverse and take-back logistics. Such issues also have a bearing on decisions to procure and utilize external capacity.
6.1 Modeling literature review.

As we have seen so far, the issues of interest to program managers are wide-ranging and multi-faceted, as they are challenging. These issues are compounded when dealing with multiple contributing firms, as it introduces a game-theoretic dimension to program decision making. From an analytical perspective, therefore, it is our task in this paper to pare the problem down to a few key aspects of decision-making in a program environment, and the environment itself to a manageable set of assumptions and model descriptors. Our aim in this section is to highlight literature that has guided our modeling choices.

Before embarking on our own modeling exercise, we surveyed the literature for quantitative models used in academia and in practice to describe program environments. Some early empirical surveys in the R&D sector [127] reported that quantitative techniques such as the Program Evaluation and Review Technique (PERT),
Critical Path Method (CPM), first developed and adopted in the 1950s and 1960s, had found considerable traction in a wide range of industries. (See Elmaghraby [66] for an excellent reference on these network modeling approaches). From an engineering and product development standpoint, a stream of work by Eppinger and co-authors describe applications of model based project management techniques in a variety of industries including automotive and consumer electronics [67] [68] [147] [190] [193]. In terms of maturity of project management practices, variations have been reported across industry sectors; with the defense, and oil and gas sectors leading in terms of their capabilities and sophistication [42]. The pharmaceutical, telecom, financial and construction industries also display considerable maturity. In the electronics manufacturing sector, contract manufacturing firms such as Flextronics and original equipment manufacturers such as Cisco Systems have been regarded for their successful models of what we refer to as program management with their key supply chain partners [53] [54].

However, the traditional focus of the quantitative methods in project management, especially in the network models has been on structuring, scheduling and sequencing of the tasks. From the point of view of evaluation, these network models reveal the critical paths in the project work flow and compute measures such as the earliest completion time of the project. The major reason they are not directly applicable to environments we consider is that incorporating multiple firms and determining the best assignment of tasks to firms embeds the assignment decision into the evaluation model, and this does not yield very well to clear analysis. Also, the network model is governed by the different cost or value objectives of the firm, and again having a single network model to address all of these objectives at every firm is not feasible. For example, decentralized capacity decisions on the part of firms are not so easy to incorporate even in deterministic environments. There are how-
ever many models in the literature on the assignment problem in project contexts, and similar to our own work, they interpret the assignment problem as capacitated resource assignment [52] [58] [148] [154] [183]. All of these models rely on mathematical programming approaches; but not all fit into the PERT/CPM framework. Our models on the other hand, while still relying on mathematical programming as a foundation, attempt to incorporate some of the key modeling elements of the PERT/CPM framework.

One of the central trade-offs addressed in some of the network based approaches is that between additional investment of resource groups to certain tasks (entailing greater project costs), and the degree of lateness of the project; in other words the trade-off between the cost and time dimensions; the seminal work done by Fulkerson [77] has been later been expanded in the literature. Our models capture this trade-off, but only indirectly and with significant differences; we do examine the issue of how much to invest in the program resource groups (and by translation, on the assigned tasks), but balance the additional costs incurred with the penalty costs incurred as a result of the processing delays. In our models, the penalty costs are linear in the time delays, but there is scope to expand this to any convex function of processing delays.

The last issue presents an important distinction of our work from the traditional network based models: our models are not restricted to measuring time or completion metrics in project management. Hence we do not aim to provide detailed time-based measures such as the distribution, or for example the second moment of program completion time. Rather our interest is in understanding how the time measures impact the corresponding costs to the program, and further impact key decisions such as task assignment and capacity investment. For this, we assume a
known (and implicit) relationship between the tardiness of a program, and program level measures such as revenues and costs. Hence, there is a trade-off in our model between investment, time, and cost measures, but this happens on a more aggregate basis than in the traditional network models.

We believe our approach above is well-justified, since time-to-market concerns in programs are largely driven by the time-sensitive nature of program revenue \([40]\). Typically the tardier the program or the longer its lead time, the lower its value to the firms involved. Furthermore, effecting multiple trade-offs requires us to translate all impacted performance measures to a single measure of program performance; in our case, the profits or value derived by the firms from the program. Thus the central trade-offs in our model are expressed purely in terms of costs of resource capacity investment, task assignment, and delay penalty costs; we believe this makes the decisions and trade-offs more clear.

Our models also try to capture uncertainty and risk in the program environment; as discussed earlier, the risk in our program models can arise from three sources: task work requirements, realized resource capacity, and program value. The network modeling literature explores risk in projects in similar fashion, but there, the uncertainty is in the time-duration of the network activities \([36]\) \([35]\) \([108]\). There also exist models and literature, more in the flavor of strategic decision-making and economics of product portfolio selection, that consider the impact of project risk on the investment decisions by firms \([103]\) \([131]\) \([151]\). There are however few significant papers, to our knowledge, in the management science literature on how to perform task or resource assignment, together with decisions on resource capacity in an uncertain program environment. As such our paper presents an important contribution to the literature in this regard.
There is however a vast literature in the area of capacity investment and management in general. We were influenced by the stream of work reported by van Mieghem et al. [98][194], and Fine and Freund [72], and Jordan and Graves [110]. We incorporate ideas and modeling techniques from these works, even though our models are not identical in construction or in context to any one of these works in particular. The issue of decentralized decision-making by firms leads to a game theoretic formulation of the capacity investment problem, which again to our knowledge, has not been explored in great depth in the project or program management literature. The related problem of actually coordinating capacity investments by the partner firms for the purposes of achieving project or program objectives also deserves further study. Our paper is the first, we believe, to meld together these critical and performance defining elements of a general program environment, and through a model that addresses both strategic as well as tactical questions.

With regard to task assignment, there are again very few models in the management science literature with specific application to program environments. However, in a broader task scheduling environment, this issue has been explored to a certain extent in both general as well as in specific applied contexts [87] [168] [189]. There is also scattered work on resource or task assignment decisions under uncertainty, but much of the work concerns capital budgeting and financial commitments to competing project or product portfolios. In this paper, we assume that task assignment is a centralized program level decision-process; this has implications in that we exclude a firm’s choice in tasks given how related tasks are allocated to other firms. Even then, our models represent a significant advance to the literature on program and project management.
Similarly, with regard to coordination mechanisms, such as work, cost, and value or revenue sharing, we again find that the literature as it relates to project and program management is rather limited, especially in the quantitative and decision-theoretic domains. The existing works are more in the flavor of economic theory, and are descriptive rather than normative in their intent [15] [172]. That said, although their model is restricted to the domain of collaborative product development, the recent work of Bhaskaran and Krishnan [15] is the first to analyze work, cost and revenue sharing in the same vein and in the same model. There is also a relatively mature body of work on coordination of supply chains through contract mechanisms (see for example, the work of Cachon et al [29] [30]), but the predominant emphasis is on vertical chains and material transactions in serialized supply chains, as opposed to lateral or collaborative work.

For contracts in decentralized projects, there is the recent work by Kwon et. al. [122]. The authors consider a manufacturer involved in a project with several suppliers and having the option of either paying each supplier upon completion of their individual task, or waiting until both tasks have been completed to pay the supplier. With exponential task completion times, they show that delayed payment can be more profitable for the manufacturer’s profits (in an equilibrium where the suppliers determine their own work-rates), when the program revenues are low. When revenues are greater than some threshold, the delayed payment contract is shown to be detrimental to the manufacturer. This contract is also not beneficial with an increasing number of suppliers. Our model captures the externality effect of one supplier on the others in the network, through the manner in which we model the program processing time (delay) as the cumulative time (delay) it takes to complete each level of tasks in the program. Further, since the program profits are decreasing in completion time (delay), and since the profits are shared according to proportional
revenue sharing mechanisms, the impact of individual supplier’s work rate on the profits is captured, both at the program level, and at the level of each firm.

In summary, our modeling and analytical approach in this paper has been influenced by the literature in several areas of management science. We have selectively adopted elements of the literature and existing models, and have constructed our own models that are as a result significantly more applicable to the environments of interest to us; and are therefore able to interpret program management environments in a more realistic fashion. We do lose some detail in other areas: most importantly, we do not address decision-making over time. Also we view program decisions as being made without time considerations or considerations of program evolution. In reality firms may retain the flexibility to revisit their decisions as the program and market evolves, and also stage their key decisions over time in order to utilize new information as it becomes available and to incorporate feedback from previous decisions.

We also forgo much of the detail and richness, especially in operational and inventory dynamics, offered by several quantitative supply chain models. But our main concern is not material flows, but rather the elements of supply chain structure as discussed earlier; so our modeling is at a more aggregate level. Similarly, we do not incorporate market needs and demand behavior or dynamics that often drives supply chain program objectives. However, we believe there is room to explore some or all of these issues in any further work that expands our own modeling framework.
6.2 The Fundamental Elements of Supply Chain Programs

6.2.1 Program Management Models: Tasks.

Consider a supply chain program with a set of constituent tasks (alternatively, projects, or activities) given by the set \( T = \{1, ..., N\} \). Tasks are discrete, but entail measurable and divisible base work demands or requirements \( \lambda_n \), forming an \( N \)-vector of demands \( \lambda \). The base demand vector \( \lambda \) could be subject to uncertainty and the individual demands \( \lambda_n \) could be correlated. Let \( F_n(\lambda_n) \) denote the marginal c.d.f. of the work requirement or demand distribution for task \( n \), and \( F_T(\lambda) \) the joint distribution of base demands. In particular, we assume that the first and second moments of \( F_T(\lambda) \) are well-defined. There are first some comments necessary as to how the task requirements are measured; that is, what are the dimensions of \( \lambda_n \)? There are several possible approaches or definitions, depending again on the program environment; we model the two possibilities below.

In the first case, \( \lambda_n \) could represent the number of times task \( n \) is repeated, or reworked, until it is considered complete. In this way, \( \lambda_n \) represents the number of distinct requests placed by the task on the resource groups involved. Also, with this approach, \( \lambda_n \) would be dimensionless. For example geometrically distributed \( \lambda_n \) could then model the number of trials required before task \( n \) is completed. We could also model continuous distributions (based, for example, on the geometric distribution) where repeat trials of task \( n \) are not as intensive as the first attempt at the task, but where future trials are equivalent to some positive fraction of the first one (in terms of their work requirement).

In the second case, \( \lambda_n \) could denote the number of discrete job requests for resources serving the task. This models situations where the task involves repeating
a basic activity a number of different times, even when there is no rework. For example, the task could be the development of several prototypes of a product component; then \( \lambda_n \) would represent the number of prototypes of the component that are required.

In our models, we allow for both approaches; in particular our modeling framework is indifferent to which of the above definitions is used for \( \lambda_n \). Moving further, all tasks need to be completed for the program to finish; task \( N \) signifies program completion, while task \( 1 \) denotes its commencement. We then model the program network as a tree defined using tasks as its nodes, and edges representing the transition between tasks. To enable this structure, the sequence, i.e. the precedence and succession relations between tasks are assumed fixed and known. Task \( N \) is at the root of the tree and succeeds all other tasks in the program. The task ordering relations are then encoded using a matrix \( \Psi \), as follows. We define \( \psi_{kN} = +1 \) for all tasks \( k \) that immediately precede task \( N \); correspondingly, we define \( \psi_{NK} = -1 \). Similarly, we define \( \psi_{nk} = +1 \), and \( \psi_{nk} = -1 \) for all tasks \( n \) that immediately precede \( k \); \( \psi_{nk} = 0 \) for all other pairings \((k, n)\) of tasks. Finally, we assume that there is at least one task that immediately precedes \( N \), while every other task \( k \neq N \) has at exactly one successor.

Within this tree structure, tasks could have pre-defined earliest start, and latest end times. Let the \( N \times 1 \) vectors \( \tau^s \) and \( \tau^f \geq \tau^s \) denote the earliest start and latest end times associated with the tasks. Then, we use a vector \( \hat{\tau} = \tau^f - \tau^s \) to denote the available time to process and complete task \( n \). However, note that \( \psi_{kn} = +1 \implies \tau^s_n \geq \tau^f_k \). This completes the definition of the tree structure for the program.
Figure 6.1: Illustrative tree structure for program: Task $N$ is at the root of the tree; every other task has exactly one successor. Note that $\psi_{12} = \psi_{28} = \psi_{N-3,N-1} = \psi_{N-5,N-1} = +1$, while $\psi_{21} = \psi_{28} = \psi_{N-1,N-3} = \psi_{N-1,N-5} = -1$. While task $N - 8$ precedes task $N$, we still define $\psi_{N-8,N} = 0$, since it is not an immediate predecessor.

See Figure 6.1 for an illustration of an arborescent, or a tree structure for a program with $N$ tasks; this paper only deals with programs that exhibit this structure. The tree structure, while very general and flexible enough to accommodate many different types of environments (serial and parallel being extreme cases), is still a specialized program structure. In reality, programs may have a more complex structure that do not resemble trees; we leave the analysis and optimization of such non-tree like structures for future research.
6.2.2 Program Management Models: Resources.

A set $\mathbf{R} = \{1, ..., M\}$ of resource groups is assumed devoted to the program; in other words all of the capacity of any resource in a given pool $m$ is available to process program work, if necessary. The capacity or size for resource group $m$ is denoted by $K_m$, implying there are $K_m$ units of the identical resource groups in that group; however in our models we will assume that $K_m$ is a continuous decision variable. We will further assume that in a decentralized setting, the capacity $K_m$ for resource $m$ is a decision variable only for the firm owning that resource (for other firms $K_m$ becomes a parameter). Hence we have an $M$-vector of resource group capacities $\mathbf{K} = \{K_1, ..., K_M\}$. The cost of investing $K_m$ units of capacity is assumed to be a convex increasing function in $K_m$. For example, with linear capacity sizing cost, the cost of capacity at level $K_m$ would just be $c^K_m K_m$.

In our models, tasks are assigned to, completed, and delivered by resource groups; but it is not necessary that tasks be assigned in whole. Rather the task demand can be split and allocated in fractions to the capable resource groups. Each resource group, on the other hand is assumed to be specialized and devoted to one particular task. We denote the subset of resource groups dedicated to task $n$ by $\mathbf{R}_n$, where $||\mathbf{R}_n|| = M_n$. Also, we use an $M \times N$ binary matrix $\hat{\mathbf{X}}$ to denote the task capabilities of the resource groups; every row of this matrix has exactly one positive (unit) element, while a column can have multiple unit entries.

The resource groups in our models are also heterogenous in multiple respects as discussed below:

1. Some resource groups are either inherently, or by design, more productive or efficient than others, even when they have similar task capabilities; the most
plausible explanation for this phenomenon is technological or process differences among resource groups. For example, two logistics providers who use different routes for the same package delivery, based on internal routing considerations, will entail different distances traveled, and therefore different time spent for the same package shipment; it is still possible that the overall cost of the shipment is lower for the firm that makes the package travel greater distance before delivery.

2. In our models, we assume that a unit of resource group $m$ processes work assigned to it at rate $\mu_m > 0$. As a modeling choice we decouple $\mu_m$ from the set of tasks, so $\mu_m$ is not conditional on any of the $\lambda_n$. Importantly, $\mu_m$ is also independent from $\mu_j$ for any other resource group $j \neq m$. We denote the c.d.f. of the processing rate $\mu_m$ for resource group $m$ by $G_m(\mu_m)$, and the c.d.f. of its inverse (the processing time) $\frac{1}{\mu_m}$ by $G_m^{inv}(\mu_m)$; this models resource, and therefore supply risk for the program.

3. Resources could exhibit different marginal cost rates of task processing; we assume that the marginal processing, or operational cost of work at resource group $m$ is given by $c^O_m$ per unit of work.

4. Finally, the processing rate of a resource group $m$ could exhibit scale effects with capacity $K_m$; this could result from the collaborative inefficiencies (or synergies) of the resources in the group, as they work on the task that they are assigned. Thus, given capacity $K_m$ the effective processing rate of the pool of resources is modeled as $K_m^{\theta_m} \mu_m$, where $\theta_m \geq 0$. In this way, it can be verified that the processing rate benefits are diminishing in capacity, and are therefore concave in $K_m$ when $\theta_m \in [0,1]$, and conversely the resource group processing rate is increasing convex when $\theta_m \geq 1$; thus modeling both inefficiencies and
synergies from collaboration on a given task. It should be noted that $\theta_m$ is a measure of intra-firm (or within task) collaboration, as opposed to cross-task (and potentially inter-firm) collaboration to be discussed later.

Taken together, all of these resource attributes imply the following fundamental processing time function. Given a requirement or demand of $\lambda_n$ arising from task $n$, and with capacity $K_m$, resource group $m$ has processing or activity time of:

$$\tau_{mn} = \frac{\lambda_n}{K^\theta_m \mu_m}.$$  \hfill (6.1)

In some situations, tasks could take some minimal time to complete, independent of the capacity of the resource group; in which case we denote this minimum time by $\tau^o_n$, and write the fundamental processing time function as:

$$\tau_{mn} = \tau^o_n + \frac{\lambda_n}{K^\theta_m \mu_m}.$$  \hfill (6.2)

However, as can be observed throughout our analysis, the minimum processing time has limited impact on the decision model, and hence in our analysis, we will use Equation 6.1 to define the processing time for a task.

Empirical support for such a processing time function comes from Graves [89]. The topic of discussion in much of the time-cost curve literature is really the nature of the cost function as we compress the completion time, but Equation 6.1 is consistent with this behavior, and is actually implied by the convexity of the cost curve. Graves points out that the literature on product development or research and innovation projects generally assume that the fundamental time-cost trade-off curve is convex. The greater the investment (or cost input) into a task, or a project, the lower the completion time. Moreover, models in the product development literature generally assume that the returns on investment into time-compression are
diminishing. Brooks [26], for example cites examples from software engineering and development where incremental additions to project staff are not as productive as the original team given learning curves and inefficiencies in knowledge transfer.

This agrees well with our basic model where the processing time $\tau_{mn}$ is convex in $K_m$ but where the slope of the time-cost curve depends on the scaling parameter $\theta_m$. If $\theta_m < 1$, then incremental capacity additions are not as productive in compressing the processing time. Following the example in Graves [89]: if a single task $n$ is shared by a team of $K_m$ the pair-wise communication effort among a team of $K$ persons increases in the order of $\frac{K(K-1)}{2}$, a steep increase that would diminish the impact of expanding the resource pool.

Mansfield [134] shows evidence – albeit from a few decades ago when communications technology was rudimentary compared to today’s standards – of how the processing time function for a sample of observed projects indeed followed the convex curve. Graves [89] also demonstrates that the convexity of the project completion time in the cost inputs is unchanged even for general project networks (such as PERT), using well-known sensitivity analysis concepts in linear programming. Finally, Scherer [166] shows how the convexity of the processing time curve is still preserved in uncertain environments. In our models, we will allow a number of variables in Equation 6.1 to be subject to uncertainty including the load $\lambda_n$ and the processing rate $\mu_m$, all the while preserve the convex nature of the completion time and therefore program payoffs in the resource investment $K_m$. 

281
6.2.3 Program Management Models: Firms.

The supply chain program (or organization) is defined from a set \( A = \{1, \ldots, L\} \) of firms willing to participate. We could designate Firm 1 as the lead firm. The firms take a stake in the program by owning a share of one or more resource groups; we use an \( L \times M \) matrix \( \alpha \) to represent the firm-resource ownership structure. Generally, an element \( \alpha_{lm} \) of this matrix satisfies \( \alpha_{lm} \in [0, 1], \forall l = 1, \ldots, L; m = 1, \ldots, M; \sum_{l=1}^{L} \alpha_{lm} = 1; \forall m = 1, \ldots, M \).

However, in the analysis presented here in this paper, we will only consider the special case where each resource can be owned by exactly one firm. This may not be as stringent an assumption as it seems; split ownership of resource groups can be dealt with by splitting them into multiple sub-groups each owned by only one firm. In other words, since ownership cannot be shared at the resource level we have \( \alpha_{lm} \in \{0, 1\}, \forall l = 1, \ldots, L; m = 1, \ldots, M; \sum_{l=1}^{L} \alpha_{lm} = 1; \forall m = 1, \ldots, M \). This distinction has important implications for the equilibrium behavior of firms as they independently determine their resource capacity investments.

Given the ownership structure of the resource groups, we can derive \( L \times N \) matrix \( \hat{Y} = \hat{\alpha} \times \hat{X} \) that represents the acquired task capabilities of the firms, where \( \hat{y}_{ln} > 0, \iff \text{ firm } l \text{ is capable of performing task } n; \hat{y}_{ln} = 0, \text{ otherwise for } l = 1, \ldots, L, n = 1, \ldots, N \). Note that \( \hat{y}_{ln} > 1 \) whenever a firm \( l \) owns two or more resource groups, each capable of performing task \( n \).

A central agent, called a program planner, determines the assignment of tasks to the firms. The role of the program planner can be taken up by the lead firm, but in all cases, we assume that the program planner determines the assignment as a neu-
6.2.4 Program Management Models: Task-Resource Assignment.

In the models we describe in this paper, the tasks are discrete entities, and can be assigned to only resource; in particular only those resource groups \( m \in R_n \) are candidates for the assignment of task \( n \). We use an associated matrix \( M \times N \) matrix \( X(\leq \hat{X}) \) to represent the actual task assignment; a binary element \( x_{mn} \in \{0, 1\} \) represents the decision to assign task \( n \) to resource group \( m \). The matrix then exhibits the following natural properties:

1. \( x_{mn} \in \{0, 1\}; \forall m = 1, ..., M; \forall n = 1, ..., N. \)
2. \( x_{mn} = 1 \iff n \) is assigned to resource group \( m \in R_n; x_{mn} = 0 \) otherwise.
3. \( \sum_{m \in R_n} x_{mn} = 1; \forall n = 1, ..., N. \)

A given assignment strategy \( X \) implies an assignment \( Y \) of task demands to firms where \( Y = \alpha \times X \); an element \( y_{ln} \) satisfies a set of properties analogous to those of \( X \).

1. \( y_{ln} \in \{0, 1\}; \forall l = 1, ..., L; \forall n = 1, ..., N. \)
2. \( y_{ln} = 1 \iff x_{mn} = 1 \) for some \( m \) such that \( \alpha_{lm} = 1; y_{ln} = 0 \) otherwise.
3. \( \sum_{l=1}^{L} y_{ln} = 1; \forall n = 1, ..., N. \)

Figure 6.2 is an illustration of how the tasks, resource groups (and their component resources), and the firms are related in the program context. As mentioned earlier, we do not allow for firms to have shared ownership of resource groups. Further we assume that a resource group is devoted entirely to a task; in other words
resources are not shared across tasks (so this removes from consideration issues such as overlapping resource usage by tasks). However when there are multiple candidate resources, we select only one resource group for a given task as part of the overall assignment decision.

6.3 Measures for Task Requirements and Performance

6.3.1 Measuring collaborative requirements across tasks.

We measure the collaborative requirements across and within tasks indirectly in terms of their impact on task processing rates. In turn the impact on the task processing rate has two components: the first is the within task component where resources within a group have to collaborate for completing a given task, while the
second deals with how the task relationships impact the processing rate. Since the
tasks are necessarily performed by different resource groups, and potentially by dif-
ferent firms, the second component captures the dimension of inter-resource and
inter-firm collaboration. The first component measures how the effective processing
rate of a resource group scales with its capacity $K_m$; recall that $\theta_m$ is the parameter
for the resource group that governs this relationship.

For the second component of inter-resource group collaboration, we use the fol-
lowing approach. Given a tree structure (or precedence relationship) $\Psi$, we associate
a symmetric multiplicative factor (or parameter) $\delta_{nk}$ between a pair of tasks $n$ and
$k$ that that are immediate successors and predecessors, i.e. $\psi_{nk} = +1$ or $\psi_{nk} = -1$.
Going beyond we associate this parameter $\delta_{nk}$ between tasks that satisfy the follow-
ing precedence relation: $\psi_{nt} = \psi_{kt} = +1$ for some task $t \neq k, n$; in other words,
tasks $n$ and $k$ have a common successor; the interpretation is that these tasks while
not successors or predecessors of each other, are in some sense concurrent. As we
define below, this parameter $\delta_{nk}$ serves to either speed up or slow down the resources
assigned to the respective tasks. When the $\delta_{nk} > 0$, the processing rate is positively
impacted by the collaboration, and the resource groups assigned to either task are
sped up as a result of collaborative synergies; conversely when $\delta_{nk} < 0$, there are in-
efficiencies from cross-task collaboration, and both resource groups are slowed down
as a consequence.

Significantly, we assume that pairs of tasks that do not satisfy these precedence
relations above do not have any collaborative relationship, i.e. we define $\delta_{nk} = 0$ for
all other pairs of tasks. Note that we do not require $\delta_{nk} = \delta_{kn}$.

Let us then denote by $\Delta$, the $N \times N$ matrix of $\delta_{nk}$, for all $n, k \in T$. Now, we will
explicitly define how these collaborative requirements impact the task processing. The net impact of the collaborative requirements is to alter the processing rate $\mu_m$ of any resource assigned to task $n$ (or for that matter $k$). We define the effective processing rate $\mu_m^e$ of resource $m$ capable of performing task $n$ as:

$$\mu_m^e = \mu_m \left(1 + \sum_{k \in T} \delta_{nk}\right); \forall m \in R_n.$$  \hspace{1cm} (6.3)

For the above definition to be meaningful, we will require that $\sum_{k \in T} \delta_{nk} > -1; \forall n$. Note again that synergies from collaboration between tasks are modeled by $\delta_{nk} > 0$, while inefficiencies are modeled by $\delta_{nk} < 0$.

This may seem a somewhat arbitrary definition of which tasks require collaboration; in theory we could extend our models to cases where potentially any pair of tasks could require collaboration. However, note that there is a subtle distinction here between collaboration between tasks, and the correlation of the demands they place on resources. Thus, correlation is a broader or more pervasive phenomenon that could impact tasks that do not necessarily precede or succeed each other in the immediate sense. However, in many program environments, it is more often the case that tasks that are connected in the sense of immediate precedence relationships require active collaboration. Conversely, tasks that are a step or two removed from each other are less likely to require any active collaboration; if they did, then the task precedence $\Psi$ would indeed have to be altered to imply closer connectivity between those tasks.

To reiterate the assumptions regarding $\Delta$:

1. $\delta_{nk} \neq 0 \implies (i) \psi_{nk} = +1$ or $\psi_{nk} = -1$; i.e. task $n$ either precedes or succeeds task $k$, or (ii) $\psi_{nt} = \psi_{kt} = +1$ for some task $t \neq k,n$; in other words, tasks $n$
and $k$ have a common successor.

2. $\psi_{nk} = 0 \implies \delta_{nk} = 0$.

3. $\sum_{k \in T} \delta_{nk} > -1; \forall n$.

Finally, it is useful to point out the differences between the resource pool collaboration parameter $\theta_m$, and the cross-task collaborative parameter $\delta_{nk}$. In our models $\theta_m$ serves as a characteristic of the resource pool for a given task $n$, while the matrix $\Delta$ is a characteristic of the network of tasks. This is a deliberate modeling choice; we wish to highlight that variances in collaborative synergies or inefficiencies can be explained by inherent task requirements, as well as the fact that some resource groups are better collaborators than others. Since resource groups in our model are dedicated to a task, their collaborative capabilities have a local impact on individual task processing. On the other hand, the network-wide collaborative requirements also have a local impact on the resources processing a specific task, since it could either slow them down or speed them up depending on the net impact of the collaborative requirements. In this way, we model collaboration as both a resource and network characteristic, and explore both its local task-level, and network-wide impact. The organizational studies literature (see Dailey [47]) also attempts to characterize the performance of groups at innovation efforts through two independent concepts: cohesion, and collaboration. In this language, $\theta_m$ would be the parameter that would reflect group cohesion, while $\delta_{nk}$ would measure the collaborative content between tasks.

This model above for measuring the impact of collaboration appears consistent with the literature on R&D, product development, and innovation: we do find some parallel views of how collaboration among tasks (really among resource groups) im-
pacts task processing. Dailey [48] presents some empirical evidence in the organizational studies literature of how collaboration among resource groups impacts task productivity. Interestingly, the study finds that collaborative relationships have a positive impact on resource productivity, especially when there is greater uncertainty regarding the workload content in the task, and when the inter-dependence between tasks was not too high. Our model allows for both phenomena to be captured. The parameter $\delta_{nk}$ can be lowered to reflect greater uncertainty regarding either $\lambda_n$ or $\lambda_k$, thereby calibrating the processing rate of resource groups associated with either task. Similarly, greater inter-dependence between tasks has been shown empirically to have a negative impact on productivity: this is clearly the main purpose of describing the impact of collaboration on processing rates (and therefore productivity) in a pair-wise fashion.

Other than these works cited above, we do not find detailed empirical studies (other than anecdotal evidence) that analyze collaborative content within a network representation of the project or program environment. As such, this model explaining the impact of collaboration on the fundamental time-cost trade-offs (that are part of the classical project management theory and practice) could be an important addition to that literature.

6.3.2 Effective processing times for tasks.

Given our preceding discussion on how the processing rates of resources can be impacted by task collaborative requirements, we can write the effective processing time from a potential assignment of task $n$ to resource $m \in R_n$ (assuming $K_m > 0$) as:
\[ \tau_{mn}(X, K_m) = \frac{\lambda_n x_{mn}}{K_{m}^{\theta_m} \mu_m} \]  
\[ = \frac{\lambda_n x_{mn}}{K_{m}^{\theta_m} \mu_m (1 + \sum_{k \in T} \delta_{nk})}. \] 

In Equation 6.4, the collaborative requirements of task \( n \) impact the processing time for the task through the effective processing rate \( \mu^e_m \). If the task exhibits overall synergies through collaboration with other related tasks; i.e. if \( \sum_{k \in T} \delta_{nk} > 0 \) then the effective processing time for the task will be reduced, and conversely if \( \sum_{k \in T} \delta_{nk} > 0 \) implying net inefficiencies through collaboration, the effective processing time will be amplified. Secondly, observe that the overall impact of collaboration is independent of the assignment strategy. Essentially the impact of collaborative requirements on task processing times is proportional to \( \frac{1}{1 + \sum_{k \in T} \delta_{nk}} \), independent of which resource is performing the task.

6.3.3 Time and delay measures for tasks.

Given this background, we define the effective processing time for task \( n \) under assignment \( X \) as:

\[ \tau_n^e(X, \{K_{m \in R_n}\}) = \sum_{m \in R_n} \tau_{mn}(X, K_m). \]  
\[ = \sum_{m \in R_n} \frac{\lambda_n x_{mn}}{K_{m}^{\theta_m} \mu_m (1 + \sum_{k \in T} \delta_{nk})}. \]

Given an assignment \( X \), the effective processing time \( \tau_n^e(X, \{K_{m \in R_n}\}) \) may exceed the time \( \hat{\tau}_n \) available for completion of task \( n \). Then, given capacity \( \{K_{m \in R_n}\} \), we can derive the delay experienced by task \( n \) as follows:
\[ D_n(X, \{K_m_{\in \mathbb{R}_n}\}) = [\tau^e_n(X, \{K_m_{\in \mathbb{R}_n}\}) - \hat{\tau}_n]^+ \] (6.6)

6.3.4 Time and delay measures for programs.

As discussed earlier, we have defined the program as a tree structure (or network) with the tasks representing the nodes of the tree. This greatly expands the possible range of program structures that can be modeled, including cases where the tasks could be arranged serially, or conversely cases where the majority of tasks (save N) are arranged in parallel.

The cumulative processing time for the program under assignment \(X\), and for a fixed capacity vector \(K\) is then computed recursively as follows:

1. For the leaves of the tree, i.e. those tasks \(k\) such that \(\psi_{nk} = 0; \forall n\), the (cumulative) processing time to achieve or complete task \(k\) is simply:
   \[ \tau^c_k(X, K) = \tau^e_k(X, \{K_m_{\in \mathbb{R}_k}\}). \] (6.7)

2. For all other nodes, we compute the cumulative processing time to complete that task recursively as:
   \[ \tau^c_k(X, K) = \tau^e_k(X, \{K_m_{\in \mathbb{R}_k}\}) + \max_{t: \psi_{tk}=+1} \{\tau^c_t(X, K)\}. \] (6.8)

For simplicity of notation, we denote the program completion time as:
   \[ \tau(X, K) = \tau^c_N(X, K). \] (6.9)

Next, in similar fashion, the cumulative delays time for the program under assignment \(X\), and for a fixed capacity vector \(K\) are also computed recursively:
1. For the leaves of the tree, i.e. those tasks $k$ such that $\psi_{nk} = 0; \forall n$, the cumulative delay experienced before completing task $k$ is:

$$D_c^k(X, K) = D_k(X, \{K_{m \in R_k}\}). \quad (6.10)$$

2. For all other nodes, the cumulative delays incurred before completing that task are computed recursively as:

$$D_c^k(X, K) = D_k(X, \{K_{m \in R_k}\}) + \max_{t: \psi_{tk} = +1} \{D_c^t(X, K)\}. \quad (6.11)$$

Again, for clarity of notation, we define the cumulative program delay as:

$$D(X, K) = D_N^c(X, K). \quad (6.12)$$

6.3.5 Cost measures.

The cost measures that are defined here will be incorporated into the program level and the firm or resource level objective functions that will in turn be optimized to determine the optimal assignment and capacity investment vector. The optimization problems themselves are defined in sections 7.4 and 8.2.

Costs incurred in our models are of two types:

1. Processing or operating costs that are incurred by resource groups for completing their share of the task; these costs are independent of the capacity deployed towards task completion, and also independent of the processing rate of the resource groups; but for a given task, they can be specific to the resource group. We assume the operating costs are linear in the task demand placed on the resource; and denote by $c^O_m$ the marginal cost per unit demand placed by task $n$ on resource group $m$. 

291
2. Penalty costs for delays in task processing; these costs are incurred by resource groups assigned to the task which experiences delay in completion. We assume again that the penalty costs are linear in the delay $D_n(X, \{K_m \in R_n\})$, but that the per unit delay costs could be different for each resource group in $R_n$; thus we denote the delay penalty cost for resource group $m$ as $c^P_m$.

With the above definitions, it is possible to write the processing and delay penalty costs at the resource, firm, and program levels, as follows:

$$C^O_{mR}(X) = c^O_m \lambda_{mn}(X), \forall m \in R_n, n = 1, ..., N; \quad (6.13a)$$

$$C^O_{iF}(X) = \sum_{m=1}^{M} \alpha_{lm} C^O_{mR}(X); \quad (6.13b)$$

$$C^O(X) = \sum_{m=1}^{M} C^O_{mR}(X). \quad (6.13c)$$

Similarly, the delay penalty costs at the resource, firm, and program levels are defined as:

$$C^P_{mR}(X, K) = c^P_m D_n(X, \{K_m \in R_n\}), \forall m \in R_n, n = 1, ..., N; \quad (6.14a)$$

$$C^P_{iF}(X, K) = \sum_{m=1}^{M} \alpha_{lm} C^P_{mR}(X, K); \quad (6.14b)$$

$$C^P(X, K) = \sum_{m=1}^{M} C^P_{mR}(X, K). \quad (6.14c)$$

Taken together the costs at the program level can then be written as:

$$C^O_{mR}(X, K) = C^O_{mR}(X) + C^P_{mR}(X, K) + c^K_m K_m; \quad (6.15a)$$

$$C^P_{iF}(X, K) = C^O_{iF}(X, K) + C^P_{iF}(X, K) + \sum_{m=1}^{M} \alpha_{lm} c^K_m K_m; \quad (6.15b)$$

$$C(X, K) = C^O(X, K) + C^P(X, K) + c^K/K. \quad (6.15c)$$
6.4 The Value of Innovation and Resource Investments

6.4.1 Discounting of program value or revenue functions.

In the models we consider, the overall program gain or revenue could be influenced by the program completion time or indeed by the program delays. Regardless of the modeling choice, as a sensible relationship, the greater the cumulative or maximum delay, the lower is the program value or revenue accrued. In doing so we attempt to capture in some fashion the time sensitivity of program value.

We use a very simple linear discounting model of program gain as follows. The maximum gain \( \Pi_0(\geq 0) \) is achieved when the program completes at time 0, or in an alternative model discounting with delays, when the program incurs no delays. With program completion time \( \tau(X, K) \), or alternatively delay of \( D(X, K) \), the gain for the program is defined as:

\[
\Pi(X, K) = \Pi_0 - \beta \tau(X, K) = \Pi_0 - \beta \tau_N(X, K); \quad (6.16a)
\]

\[
\hat{\Pi}(X, K) = \Pi_0 - \beta D(X, K) = \Pi_0 - \beta D_N(X, K). \quad (6.16b)
\]

Here, the parameter \( \beta(\geq 0) \) is used to model the time sensitivity of the program to time-to-completion, or to the delays incurred. It can also be interpreted, roughly, as the marginal value of time for the program.

The net value accrued to the program is the gain minus the total costs incurred towards the program:

\[
V(X, K) = \Pi(X, K) - C(X, K); \quad (6.17a)
\]

\[
\hat{V}(X, K) = \hat{\Pi}(X, K) - C(X, K). \quad (6.17b)
\]

There are several properties of the linear discounting model that are useful to our modeling context.
• Firstly, the linear model, like all discounting approaches, makes the program gain sensitive to delays; in particular, the revenues or gains are non-increasing in the delay $D(X, K)$ or in the activity time $\tau(X, K)$.

• Secondly, the linear model is time consistent; and thus assumes that the gain or revenue is no more or less sensitive to the delay function in the shorter range of delays, than in the longer range. As an argument against this property, however, the growing consensus among behavioral economists and neurobiologists is that individual and collective human behavior is better modeled by time-inconsistent discounting models that discount more in the shorter range of delays; examples include the hyperbolic family of discount models. Unfortunately, the hyperbolic, or even the commonly used exponential family of discounting models are not easily amenable to analysis in our modeling framework, and lead to non-convex problem formulations.

• Thirdly, the linear gain model is a concave composition of the delay function $D(X, K)$ and of the processing time function $\tau(X, K)$, and therefore any convexity properties of the processing time or delay functions extend, via linearity, to the program gain function. This enables us to incorporate the discounted gain in the program level and firm level objectives, and obtain globally optimal assignment and capacity decisions.

6.4.2 Proportional gain and cost sharing mechanisms for firms and resources.

We consider three fundamental types of cost and gain sharing mechanisms; in all cases, the overall program gain is shared between the firms in proportion to their share in some measure of program cost. As we will see later, all three mechanisms can be used, when adequately restructured, to coordinate decentralized investments
in program capacity. While we will define the notion of coordination more formally in the following sections, in the context of our models, it implies that the program planner is able to induce the participating firms to invest in capacity at levels that maximize the overall program value.

1. In the first type we call the investment risk sharing (IRS) mechanism the firms just share the gains accrued in proportion to their share of the program capacity costs, i.e:

\[ \gamma_{IRS}^l = \sum_{m=1}^{M} \alpha_{lm} \frac{c_m^K K_m}{c^K K} \]  

2. In the second type we call the work and investment risk sharing (WIRS) mechanism, the overall program gain \( \hat{\Pi}(X, K) \) or \( \Pi(X, K) \) is shared between the firms (or resource groups) participating in the program in proportion to their share in program operating plus capacity costs, i.e. \( C^{O}(X, K) + c^K K \). In other words, firm \( l \) obtains a proportion \( \gamma_{WIRS}^l \) of the program revenues, defined by:

\[ \gamma_{WIRS}^l = \sum_{m=1}^{M} \alpha_{lm} \left( \frac{C^{OR}(X) + c^K K_m}{C^{O}(X) + c^K K} \right) \]  

3. In the third class we call comprehensive risk sharing (CRS), the gain \( \hat{\Pi}(X, K) \) or \( \Pi(X, K) \) is shared between the firms in proportion to their share in total program related costs; i.e. their share of \( C(X, K) \). Thus:

\[ \gamma_{CRS}^l = \sum_{m=1}^{M} \alpha_{lm} \frac{C^{OR}(X) + c^{PR}(X, K) + c^K K_m}{C(X, K)} \]  

That we write the share of the firm costs in terms of the sum of the resource costs in the expressions above is for a specific reason: we will aim to coordinate the supply chain investments componentwise via \( K_m \), as opposed to coordinating firm-level
investments. The difference in the two approaches has much impact on our ability to achieve coordination of program investments.

Thus, the net value for firm \( l \) under the comprehensive risk share regime (we have analogous expressions for the net value for the firms under the other gain sharing regimes) is equal to:

\[
V_l(X, K) = \gamma_{lCRS} \Pi(X, K) - C^F_l(X, K); \quad (6.21a)
\]

\[
\hat{V}_l(X, K) = \gamma_{lCRS} \hat{\Pi}(X, K) - C^F_l(X, K). \quad (6.21b)
\]

6.4.3 Impact of risk in the demand and supply variables.

Equation 6.13 represents two fundamental sources of risk: the base task demand \( \lambda_n \) and the resource processing time \( \left( \frac{1}{\mu_m} \right) \). \( \lambda_n \) in some sense represents “demand” risk from the perspective of the resource groups, and \( \left( \frac{1}{\mu_m} \right) \) represents the supply risk from the task point of view. We assume that the base demands could be correlated. Significantly, we assume that the base demand \( \lambda_n \) for task \( n \), and processing rate \( \mu_m \) for resource group \( m \) are both strictly positive under all realizations of the random variables. Let us then define \( \Omega_\lambda = \Omega_{\lambda_1} \times \ldots \times \Omega_{\lambda_N} \) and \( \Omega_\mu = \Omega_{\mu_1} \times \ldots \times \Omega_{\mu_M} \), where \( \Omega(\cdot) \) represents the sample space of the random variable argument.

Since, the two sources of risk are independent, we can write the expected cost measures from an assignment of work from task \( n \) to resource group \( m \) as:

\[
E[C^R_m(X, K_m)] = \int_{\Omega_{\lambda_n}} \int_{\Omega_{\mu_m}} C^R_m(X, \{K_{m\in R_n}\})dG_m dF_n \quad (6.22)
\]

Proceeding with the same logic, the total expected resource, firm, and program
level operating costs from assignment $X$ are given by:

$$E[C^F_l(X, K)] = \int_{\Omega_{\lambda_1}} \ldots \int_{\Omega_{\lambda_N}} \int_{\Omega_\mu} C^F_l(X, K) dG_1 \ldots dG_M dF; \quad (6.23a)$$

$$E[C(X, K)] = \int_{\Omega_{\lambda_1}} \ldots \int_{\Omega_{\lambda_N}} \int_{\Omega_\mu} C(X, K) dG_1 \ldots dG_M dF. \quad (6.23b)$$

Similar expressions can be provided for the expected operating and processing costs at the resource, firm, and program levels.

Another dimension of uncertainty in our models is the maximum program gain $\Pi_0$; let the distribution $H(.)$ of the gain parameter be defined over sample space $\Omega_{\Pi_0}$. Then, the expected value derived from the program is given by:

$$E[V(X, K)] = \int_{\Omega_{\Pi_0}} \int_{\Omega_{\lambda_1}} \ldots \int_{\Omega_{\lambda_N}} \int_{\Omega_\mu} V(X, K) dG_1 \ldots dG_M dFdH; \quad (6.24a)$$

$$E[\hat{V}(X, K)] = \int_{\Omega_{\Pi_0}} \int_{\Omega_{\lambda_1}} \ldots \int_{\Omega_{\lambda_N}} \int_{\Omega_\mu} \hat{V}(X, K) dG_1 \ldots dG_M dFdH; \quad (6.24b)$$

To summarize the computations so far, we started out with a fixed assignment or assignment $X$, and wrote the basic expressions for the effective task demands and operating or processing costs. Then, assuming a fixed capacity investment vector $K$, we wrote expressions for the processing times, task delays and associated costs (conditional again on the assignment); at each resource, each firm, and at the program level. Then we incorporated uncertainty in the base demands on the resources, in resource processing rates, and in the maximum program gain, and gave expressions for the corresponding expected costs and value measures. In doing so, we made use of the assumption that the resource processing rates are independent of each other and from the base task demands.
In Section 7.2 we describe some fundamental convexity and monotonicity properties of the cost functions derived above. These properties allow us to set up, in Section 7.4, a mathematical programming model at the program level to determine the optimal task demand assignment strategy given the capacity investment, and further to determine the program optimal resource investment levels. The same properties also allow us to define optimization problems at the individual resource and firm levels to determine their respective capacity investments. The resource or firm level optimization problem is conditional on the assignment offered by the program planner, the value (or cost) sharing mechanism offered, and the capacity levels of the other partners and resources. In Section 8.5, we analyze the equilibrium behavior of firms, via their resources, as they determine their optimal capacity investment levels, and demonstrate that the IRS, WIRS, and CRS value sharing mechanisms defined in Equations 6.18-6.20 can be restructured to coordinate the supply chain capacity investments.

Figure 6.3 explains the overall decentralized decision model and hierarchy that will be analyzed using the definitions of the decision variables, and performance measures defined in this section.

Before we analyze the network model, however, we analyze the assignment and capacity decisions for a single task considered in isolation. As we will see in the sections to follow, the insights derived here for a single task will translate, under certain subsets of program conditions, to the planning and coordination of broader network models.
6.5 The Fundamental Trade-offs in Programs for Innovation

In this section, we capture the fundamental trade-offs that determine success or failure in a project or program that has an innovation objective. On the one hand, innovation has some inherent market value, but typically this value is time-sensitive. In some situations, the value is sensitive to the processing time, but in others contexts, there are pre-determined schedules and the value from innovation is sensitive to the overall delay relative to the schedule targets. From the program manager’s perspective, it is the resources allocated to a task that are the primary determinant of program delays; the greater the capacity of resources that are assigned to a task, the faster the task is completed, and the greater the value derived from innovation. Still
there is also the matter of the cost of capacity; thus there is also trade-off between the cost of resources and their production or work rate. Finally, there is the issue of the collaborative inefficiencies (or synergies, as the case may be) within the resource pool.

Hence, in our model as we have defined it, the trade-off and interplay is between four variables: the rate of performance of work ($\mu_m$), the cost of capacity ($c^K_m$), the collaborative parameter that defines synergies or scale inefficiencies ($\theta_m$), and finally the value of time (or the time-sensitivity of value derived from innovation at a particular task($\beta_n$)). The optimal assignment of tasks to one or more resource pools is also explained through these fundamental trade-offs, and hence we devote the first few sections of this chapter to analyzing these trade-offs and deriving some basic insights into the interplay and impact of these different variables. Understanding these inherent trade-offs helps us identify some heuristics for optimal task assignment in program environments, even though for general program environments one still may need a brute-force computational approach to determine the optimal assignment strategies.

6.5.1 Deterministic program environment with processing time-sensitive value.

Consider a single task $n$ in isolation. Suppose there are $M_n$ candidate resources from the set $R_n = \{1, \ldots, M_n\}$ devoted to this task. The notation for the resource parameters, are the same as those defined in section 6.2.1; in particular, $\theta_m$ is the resource collaboration or scaling parameter; $K_m$ the capacity of the resource pool; $\mu^c_m$ is the resource processing rate; $c^K_m$ the constant marginal capacity sizing cost for resource $m$; and $c^O_m$ the marginal cost of the task demand $\lambda_n$ for a resource pool $m$. We assume that the program planner has to select one resource group from the set $R_n$ to process the task.
Suppose then that resource group \( m \) is selected for processing; the processing
time is then given by:

\[
\tau_{mn} = \frac{\lambda_n}{K_m \theta_m \mu_m^e}. \tag{6.25}
\]

Suppose further that the operating cost differences are negligible across resource
groups, the value derived from performing task \( n \) is given by:

\[
V_{mn} = \Pi_0 - \frac{\beta_n \lambda_n}{K_m \theta_m \mu_m^e} - c_m K_m. \tag{6.26}
\]

It can then be verified that \( V_{mn} \) is concave in \( K_m \); the optimal capacity \( K_m^* \) can
be determined by the first order condition:

\[
\frac{dV_{mn}}{dK_m} = \frac{\theta_m \beta_n \lambda_n}{K_m^{\theta_m+1} \mu_m^e} - c_m = 0
\]

\[
\Rightarrow \frac{\theta_m \beta_n \lambda_n}{K_m^{\theta_m+1} \mu_m^e} = c_m
\]

\[
\Rightarrow K_m^{\theta_m+1} = \frac{\theta_m \beta_n \lambda_n}{c_m \mu_m^e}.
\]

The optimal capacity for resource \( m \) is:

\[
K_m^* = \left( \frac{\theta_m \beta_n \lambda_n}{c_m \mu_m^e} \right)^{\frac{1}{\theta_m+1}}. \tag{6.27}
\]

Substituting into the value function, the optimal value given this particular as-
signment is:

\[
V_{mn}^* = \Pi_0 - c_m^O \lambda_n - \left( 1 + \theta_m \right) \left( \frac{c_m^K \theta_m}{\mu_m^e} \right)^{\frac{1}{\theta_m+1}} \left( \frac{\theta_m \lambda_n}{\mu_m^e} \right)^{\frac{1}{\theta_m+1}} \tag{6.28}
\]

\[
= \Pi_0 - c_m^O \lambda_n - \left( 1 + \theta_m \right) \left( \frac{c_m^K \theta_m}{\mu_m^e} \right)^{\frac{1}{\theta_m+1}} \left( \beta_n \lambda_n \right)^{\frac{1}{\theta_m+1}} \tag{6.29}
\]
In particular, when $\theta_m = 1$, we have the simpler expression:

$$V^*_{mn} = \Pi_0 - c_m^O \lambda_n - 2 \sqrt{\frac{c_m^K \beta_n \lambda_n}{\mu_m^e}} = \Pi_0 - c_m^O \lambda_n - 2 \sqrt{\frac{c_m^K}{\mu_m^e} \sqrt{\beta_n \lambda_n}}. \tag{6.30}$$

With this characterization, we can now examine the issue of task assignment to the candidate resources. Since our objective is to maximize the value, we will define the optimal program value over all possible assignments as:

$$V^*_n = \max_m \{V^*_{mn}\}; \tag{6.31}$$

and the program optimal assignment as:

$$m^*_n = \arg \max_{m \in R_n} \{V^*_{mn}\}; \tag{6.32}$$

Note that the gain sensitivity parameter $\beta_n$ and the task demand $\lambda_n$ are common to all resource groups in $\mathbf{R}_n$. Hence the assignment decision is simplified to a comparison, over the elements of $\mathbf{R}_n$, of a well-defined function in the variables $\{\theta_m^e, \mu_m^e, c_m^K, c_m^O\}$ for given task parameters $\{\beta_n, \lambda_n\}$.

In the special case where $(\theta_m^e, c_m^O) = (\theta_n^e, c_n^O)$ for all $m \in \mathbf{R}_n$ (that is resources share the same common capacity scaling and operating cost parameters), we have:

$$m^*_n = \arg \min_{m \in \mathbf{R}_n} \left\{ \left(\frac{c_m^K}{\mu_m^e}\right)^{\beta_n} \right\}. \tag{6.33}$$

Further, when $(\theta_m^e, c_m^O) = (1, c_n^O)$ for all $m$, the resources can simply be ranked in the order of the parameter $\frac{c_m^K}{\mu_m^e}$, and the candidate resource with the lowest ratio of marginal capacity cost to the processing rate is the optimal choice for task $n$.

We formalize this discussion above through the following two propositions:
**Proposition 6.5.1.** Considering in isolation a single task \( n \) in a deterministic program environment with processing time sensitive value, it is optimal to assign the task to the candidate resource

\[
    m^*_n = \arg \min_{m \in \mathbb{R}_n} \left\{ \frac{(1 + \theta_m)}{\theta_m^{\gamma_m + 1}} \left( \frac{(cK_m)^{\theta_m}}{\mu_m^e} \right)^{\frac{1}{\gamma_m + 1}} (\beta_n\lambda_n)^{\frac{1}{\gamma_n + 1}} + c_m^O\lambda_n \right\}.
\]

Furthermore, the optimal capacity invested in that task is given by:

\[
    K^*_n = \left( \frac{\theta_m^* \beta_n \lambda_n}{c_m^K \mu_m^e} \right)^{\frac{1}{\gamma_m^* + 1}}.
\]

**Corollary 6.5.2.** Considering in isolation a single task \( n \) in a deterministic program environment with processing time sensitive value and such that \((\theta_m, c_m^O) = (\theta_n, c_n^O), \forall m \), it is optimal to assign the task to the candidate resource

\[
    m^*_n = \arg \min_{m \in \mathbb{R}_n} \left\{ \frac{(cK_m)^{\theta_n}}{\mu_m^e} \right\}.
\]

**Corollary 6.5.3.** Considering in isolation a single task \( n \) in a deterministic program environment with processing time sensitive value and such that \((\theta_m, c_m^O) = (1, c_m^O), \forall m \in \mathbb{R}_n \), it is optimal to assign the task to the candidate resource

\[
    m^*_n = \arg \min_{m \in \mathbb{R}_n} \left\{ \frac{c_m^K}{\mu_m^e} \right\}.
\]

Furthermore, the optimal capacity invested in that task is given by:

\[
    K^*_n = \sqrt{\frac{\beta_n \lambda_n}{c_m^K \mu_m^e}}.
\]
See Figure 6.4 for an illustration of how the optimal capacity varies with the collaboration parameter $\theta_m$ and the offered load $\frac{\lambda_n}{\mu_m}$ for a specific choice of resource group $m$. Note that with higher loads, we always wish to have greater capacity, but more specifically, our need for greater capacity increases with greater value of time ($\beta_n$), or with lower resource cost ($c^K_m$). More interesting is how the optimal capacity is dependent on the collaboration parameter $\theta_m$. Recalling that capacity scales linearly when $\theta_m = 1$, observe that we may require greater capacity (for the same offered load) for moderate values of $\theta_m$ than in the extremes, except when the capacity is really expensive relative to the time value. The argument here is that if capacity is really expensive, or time is not of much value (for this task), then we only wish to increase capacity when there are sufficient scale benefits. Conversely, when capacity costs and time costs are comparable we observe (again for a fixed load) that marginal capacity at the extremes of $\theta_m$ is not as valuable. Also, in general, the cheaper the capacity, the more willing we are to tolerate collaborative inefficiencies.

Figure 6.5 reveals the impact of the collaborative parameter $\theta_m$ on time measures. Again, note that processing time is monotonic in the load $\frac{\lambda_n}{\mu_m}$. However, the impact of the workload on the processing time is different at different values of $\theta_m$; when a resource group does not exhibit good collaborative capabilities, the processing time is impacted much more severely by the load. Unfortunately, this effect is much more pronounced when we value processing time the greatest; i.e. at higher values of $\beta_n$. The good news (or bad, depending on the context) is that marginal improvements to collaborative capabilities within a resource pool can have much impact on processing time, since we observe that the returns in the processing time are diminishing with $\theta_m > 1$.

Figure 6.6 shows the impact of collaboration parameter $\theta_m$ on the net value
derived from completing the task, again for different sensitivities of task value to time. We compute the ratio of value for varying $\theta_m$ relative to the fixed case of linear capacity scaling case, i.e. relative to $\theta_m = 1$. What we find again affirms the importance of collaborative capabilities: when the resource is moderate to heavily loaded, the value is generally increasing with better scaling of capacity. Furthermore, the differential impact of collaboration is greater with increasing value of time ($\beta_n$).

What is surprising is the fact that when value is less sensitive to processing time, and $\beta_n$ is lower, the value is decreasing in $\theta_m$; even though the differentials are not as pronounced. The same effect is also observed at smaller loads relative to the capacity. What this suggests is that with smaller loads and with smaller time sensitivity, improved collaboration may not have much impact on derived value from the task, and indeed can diminish the value to a certain extent; this fact is somewhat counterintuitive. Essentially, the task can be completed as easily by one resource as by many.

6.5.2 Deterministic program environment with delay sensitive value.

Next, we consider environments where the program value is sensitive to delays in the task processing, as opposed to being dependent directly on the processing time. Suppose again that we are concerned with a single task $n$, but one whose slotted completion time interval is defined to be $\hat{\tau}_n$.

Then, suppose resource $m$ is selected for processing task $n$; the delay in processing this task is then:

$$\hat{\tau}_{mn} = \left[ \frac{\lambda_n}{K_m^{\theta_m} \mu_m^e} - \hat{\tau}_n \right]^+. \quad (6.34)$$
Further that the value derived from performing task $n$ is defined as:

$$
\hat{V}_{mn} = \Pi_0 - c_m^O \lambda_n - (\beta_n + c_m^P) \hat{\tau}_{mn} - c_m^K K_m
$$

(6.35)

$$
= \Pi_0 - c_m^O \lambda_n - (\beta_n + c_m^P) \left[ \frac{\lambda_n}{K_m \mu_m \hat{\tau}_n} - \hat{\tau}_n \right]^+ - c_m^K K_m.
$$

(6.36)

Now, consider what happens to the value function at different ranges of $K_m$. Observe first that $\hat{\tau}_{mn}$ is non-increasing in $K_m$; in particular it is strictly decreasing up to a level:

$$
\hat{K}_m = \left( \frac{\lambda_n}{\mu_m \hat{\tau}_n} \right)^{\frac{1}{\hat{\tau}_n}}.
$$

(6.37)

For $K_m \geq \hat{K}_m$, $\hat{\tau}_{mn} = 0$, and hence the value function reduces to $\hat{V}_{mn} = \Pi_0 - \ldots$
(a) $\beta_n = 0.1; c^K_m = 1.$

(b) $\beta_n = 0.5; c^K_m = 1.$

(c) $\beta_n = 1; c^K_m = 1.$

(d) $\beta_n = 2; c^K_m = 1.$

Figure 6.5: The optimal processing time $\tau^*_n$ with varying $\theta_m$ and $\frac{\lambda_n}{\mu_m}$ in Proposition 6.5.1.

(a) $\beta_n = 0.1; c^K_m = \mu^e_m = 1.$

(b) $\beta_n = 0.5; c^K_m = \mu^e_m = 1.$

(c) $\beta_n = 1; c^K_m = \mu^e_m = 1.$

(d) $\beta_n = 2; c^K_m = \mu^e_m = 1.$

Figure 6.6: The ratio of the optimal value $V^*_n$, to the optimal value with $\theta_m = 1$; for varying $\theta_m$ and $\lambda_n$ in Proposition 6.5.1.
\(c_m^O \lambda_n - c_m^K K_m\), which is strictly decreasing in \(K_m\). In other words, there is no incentive for the firm owning resource \(m\) to invest beyond \(\hat{K}_m\), and therefore we have a lower bound on the optimal value function as follows:

\[
\hat{V}^{*}_{mn} \geq \Pi_0 - c_m^O \lambda_n - c_m^K \left( \frac{\lambda_n}{\mu_n \hat{\tau}_n} \right)^{1/\theta_m}.
\] (6.38)

Furthermore, we can verify that the optimal capacity investment for resource \(m\) (conditional on its selection), is given by:

\[
K^*_m = \min \left[ \hat{K}_m, \left( \frac{\theta_m \lambda_n (\beta_n + c_m^P)}{c_m^K \mu_m^e \bar{\tau}_n} \right)^{1/\theta_m + 1} \right] \quad (6.39)
\]

\[
= \min \left[ \left( \frac{\lambda_n}{\mu_m^e \bar{\tau}_n} \right)^{1/\theta_m}, \left( \frac{\theta_m \lambda_n (\beta_n + c_m^P)}{c_m^K \mu_m^e \bar{\tau}_n} \right)^{1/\theta_m + 1} \right].
\] (6.40)

The optimal value when resource \(m\) is chosen is:

\[
\hat{V}^{*}_{mn} = \Pi_0 - c_m^O \lambda_n
\]

\[
- \min \left[ \left( \frac{\theta_m \lambda_n (\beta_n + c_m^P)}{\beta_n + c_m^P} \right)^{1/\theta_m + 1} \right] + \left( \frac{\theta_m \lambda_n (\beta_n + c_m^P)}{c_m^K \mu_m^e \bar{\tau}_n} \right)^{1/\theta_m + 1}.
\] (6.41)

In particular, when \((\theta_m, c_m^P) = (1, 0); \forall m\), the value function from assigning task \(n\) to resource \(m\) is given by the simpler expression:

\[
\hat{V}^{*}_{mn} = \Pi_0 - c_m^O \lambda_n - \min \left[ \left( \frac{\theta_m \lambda_n (\beta_n + c_m^P)}{\beta_n + c_m^P} \right)^{1/\theta_m + 1} \right] + \left( \frac{\theta_m \lambda_n (\beta_n + c_m^P)}{\beta_n + c_m^P} \right)^{1/\theta_m + 1}.
\] (6.42)
We can again formulate the value maximization problem as follows:

\[ \hat{V}^*_{n} = \max_{m \in R_n} \{ \hat{V}^*_{mn} \}; \]

with the optimal task assignment given by:

\[ m^*_n = \arg \max_{m \in R_n} \{ \hat{V}^*_{mn} \}. \]

Note again that the value sensitivity parameter \( \beta_n \), the task demand \( \lambda_n \), and the deadline \( \hat{\tau}_n \) are common to all resources in \( R_n \). Hence the assignment decision is simplified to a comparison, over the resources in \( R_n \), of a well-defined function in the resource variables \( \{ \theta_m, \mu^e_m, c^K_m, c^P_m, c^O_m \} \) for given task parameters \( \{ \beta_n, \lambda_n, \hat{\tau}_n \} \). Hence again, we observe a simple resource selection rule in the special case where \((\theta_m, c^K_m, c^O_m) = (\theta_n, c^K_n, c^O_n)\) for all \( m \) (that is the resources are not differentiated along these particular parameters; rather only through their marginal capacity cost and processing rates):

\[ m^*_n = \arg \min_{m \in R_n} \left\{ \frac{c^K_m \theta_n}{\mu^e_m} \right\}. \]

In the special case where \((\theta_m, c^K_m, c^O_m) = (1, c^K_n, c^O_n), \forall m; \forall n\); we can further reduce the assignment problem to:

\[ m^*_n = \arg \min_{m \in R_n} \left\{ \frac{c^K_m}{\mu^e_m} \right\}. \]

Thus, when the resource capacity scales linearly, the resources can again be ranked in the order of \( \frac{c^K_m}{\mu^e_m} \), and the candidate resource with the lowest ratio of marginal capacity cost to the processing rate is the optimal choice for task \( n \). The delay sensitivity (as opposed to processing time sensitivity) of the value function does not play any role whatsoever in this case.

We again summarize the discussion above through the following two propositions:
Proposition 6.5.4. Considering in isolation a single task \( n \) in a deterministic program environment with delay sensitive value, it is optimal to assign the task to the candidate resource:

\[
m_n^* = \arg \min_{m \in \mathbb{R}_n} \left\{ c_m^O \lambda_n + \min \left[ \left( \frac{(c_m^K)^{\theta_m} \lambda_n (\beta_n + c_m^P)}{\theta_m^m \mu_m^e} \right)^{\frac{1}{\nu_m + 1}} - (\beta_n + c_m^P) \hat{\tau}_n \right] + \right. \\
\left. \left( \frac{(c_m^K)^{\theta_m} \lambda_n (\beta_n + c_m^P)}{\theta_m^m \mu_m^e} \right)^{\frac{1}{\nu_m + 1}} \right] \}. 
\]

Furthermore, the optimal capacity invested in that task is given by:

\[
K_n^* = \min \left[ \left( \frac{\lambda_n}{\mu_n^{m_n^*} \hat{\tau}_n} \right)^{\frac{1}{\nu_n}} \left( \frac{\theta_n^{m_n^*} \lambda_n (\beta_n + c_n^P)}{c_{m_n^*}^K \mu_{m_n^*}^e} \right)^{\frac{1}{\nu_n + 1}} \right];
\]

Corollary 6.5.5. Considering in isolation a single task \( n \) in a deterministic program environment with delay sensitive value and such that \((\theta_m, c_m^P, c_m^O) = (\theta_n, c_n^P, c_n^O), \forall m\), it is optimal to assign the task to the candidate resource

\[
m_n^* = \arg \min_{m \in \mathbb{R}_n} \left\{ \frac{(c_m^K)^{\theta_n}}{\mu_m^e} \right\}.
\]

Corollary 6.5.6. Considering in isolation a single task \( n \) in a deterministic program environment with delay sensitive value and such that \((\theta_m, c_m^P, c_m^O) = (1, 0, c_n^O); \forall m \in \mathbb{R}_n\), it is optimal to assign the task to the candidate resource

\[
m_n^* = \arg \min_{m \in \mathbb{R}_n} \left\{ \frac{c_m^K}{\mu_m^e} \right\}.
\]

Furthermore, the optimal capacity invested in that task is given by:

\[
K_n^* = \min \left[ \frac{\lambda_n}{\mu_n^{m_n^*} \hat{\tau}_n}, \sqrt{\frac{\beta_n \lambda_n}{c_{m_n^*}^K \mu_{m_n^*}^e}} \right];
\]
and the optimal value from task completion is:

\[
V_n^* = \Pi_0 - \min \left[ \left( \frac{c K_m^*}{\mu_m^*} \right) \beta_n \lambda_n - \beta_n \hat{\tau}_n \right] + \sqrt{\frac{c K_m^*}{\mu_m^*} \beta_n \lambda_n} \left( \frac{c K_m^*}{\mu_m^*} \right) \left( \frac{\lambda_n}{\hat{\tau}_n} \right).
\]

Figure 6.7 illustrates how the optimal capacity \( \hat{K}_n^* \) depends on the schedule or target completion time \( \hat{\tau}_n \) and on the offered workload \( \frac{\lambda_n}{\mu_m^*} \). Notice that \( \hat{K}_m^* \) is non-increasing with \( \hat{\tau}_n \); the optimal capacity for a given \( \hat{\tau}_n = 10 \) is a lower bound on the optimal capacity for progressively shorter deadlines. Also, for a given load \( \frac{\lambda_n}{\mu_m^*} \), this differential in increasing with increasing sensitivity to delay represented by \( \beta_n \), or with lower capacity costs. However when the load increases beyond a certain level, imposing deadlines is not meaningful, especially when the deadlines are not credible (\( \beta_n \) is small), or when capacity is expensive relative to the value of time. This gives
Figure 6.8: The optimal value $\hat{V}_n^*$ for varying $\hat{\tau}_n$ and $\lambda_n$ in Corollary 6.5.6. us an idea of when deadlines can have an impact, and when they are not as fruitful.

Figure 6.8 shows the relative impact of deadlines on the optimal value derived from task completion. Again, notice that the value is in general non-increasing with the strictness of the deadlines; this is not surprising because deadlines can be viewed as a constraint for the capacity optimization. Secondly, we again observe that the value differential between alternative deadlines is increasing with the delay sensitivity, and with greater loading. Furthermore, the optimal value is less sensitive to loading with more relaxed deadlines; which again agrees with common intuition.

6.5.3 Task assignment in program environments subject to uncertainty.

Here we show that the simple task assignment or resource selection rules derived in corollaries 6.5.2, 6.5.5, 6.5.3, and 6.5.6 are still applicable when demand $\lambda_n$ is ran-
dom, and when the program value $\Pi_0$ is subject to uncertainty. Note however that the rules do not apply directly when the resource processing rate $\mu_e^m$ is uncertain. However, even in this latter case, it is relatively easy to extend the selection rules to situations when we maximize expected value in environments where the value is sensitive to processing time. The caveat is that when we deal with uncertain environments, the simple resource selection rules may be optimal only in the local sense, i.e., when the task is considered in isolation from the rest of the program network, and not in a global sense.

When $\lambda_n$ and $\Pi_0$ are subject to uncertainty, and with common parameters $(\theta_n, c_n^p, c_n^o)$ across the resource groups, we can observe that for every joint realization of the random variables $\lambda_n$ and $\Pi_0$ respectively, the capacity optimization problem is identical to the deterministic case. Hence, for a given joint realization $\{\lambda_n, \Pi_0\}$, and under these specific conditions, we prefer the resource group $m$ with the lowest value of $\frac{(c_K^m)^{\theta_m}}{\mu_e^m}$, and furthermore we maintain this preference for every realization of these random variables. Hence, the resource pool with the lowest ratio of $\frac{(c_K^m)^{\theta_m}}{\mu_e^m}$ again provides the maximum expected value. While we do not provide a more detailed proof, the following proposition makes this claims more formally. The proof involves a simple comparison of the expectation of the optimal value function across the resource groups.

**Proposition 6.5.7.** Consider in isolation a single task $n$ in a program environment with either delay or processing time sensitive value; subject to uncertainty in $\lambda_n$ and $\Pi_0$; and where resources share common parameters $(\theta_m, c_m^p, c_m^o) = (\theta_n, c_n^p, c_n^o), \forall m \in \mathbb{R}_n$. When maximizing expected value from the assignment, it is optimal to assign
this task \( n \) to the candidate resource

\[
m^*_n = \arg \min_{m \in R_n} \{ (c^K_m)^{\theta_n} \frac{\mu^e_m}{\mu^r_m} \}.
\]

Proposition 6.5.7 is of some comfort because it means that the assignment decision is considerably simplified, and furthermore is independent of the specific distribution of \( \lambda_n \) and the gain \( \Pi_0 \), respectively. In particular, we can accommodate a wide variety of distributions for \( \lambda_n \) (for example discrete, or continuous), and as long as the candidate resource groups are identical in terms of \( (\theta_m, c^P_m, c^O_m) \), the simple resource selection rule is still applicable.

Finally, we examine the case where the \( \mu_m \) are also subject to uncertainty. Recall that we have assumed that the distribution of \( \mu_m \) is given by \( G_m(\mu_m) \), where the distribution has finite moments. Further, we have assumed that the inverse \( \frac{1}{\mu_m} \) also has finite moments (where we make use of the assumption that the support for \( \mu_m \) is strictly positive). We can then make the following claim. The proof is again simply a matter of comparing the expectations of the optimal value function across the resource groups.

**Proposition 6.5.8.** Consider in isolation a single task \( n \) in a program environment with processing time sensitive value; subject to possible uncertainty in \( \lambda_n, \Pi_0 \) and \( \mu_m \); and such that \( (\theta_m, c^O_m) = (\theta_n, c^O_n), \forall m \in R_n \). When maximizing expected value from the assignment, it is optimal to assign this task \( n \) to the candidate resource

\[
m^*_n = \arg \min_{m \in R_n} \{ (c^K_m)^{\theta_n} E[\frac{1}{\mu^e_m}] \};
\]

where the expectation can be computed over either \( G_m(\mu_m) \), or \( G_m^{\text{inv}}\left(\frac{1}{\mu_m}\right)\).
6.5.4 Some remarks on globally optimal resource selection strategies.

In fact, the applicability of the \( \frac{(c^K)^\theta_n}{\mu^e_m} \) rule to broader networks will become apparent in later sections, when we discuss resource selection strategies for more general tree networks, but for a deterministic environment. As it turns out, this locally optimal assignment strategy is in fact globally optimal for the entire program network of tasks, under certain conditions on the constraints for the optimization. In other words, the \( \frac{(c^K)^\theta_n}{\mu^e_m} \) rule could well be a globally optimal strategy for a program with deterministic parameters \((\Pi_0, \lambda_n, \mu_m)\). The possibility of such a globally optimal assignment strategy, depends crucially on the following proposition, which we state and prove below.

**Proposition 6.5.9.** Consider a single task \( n \) in a deterministic program environment with either delay or processing time sensitive value; such that \((\theta_m, c^P_m, c^O_m) = (\theta_n, c^P_n, c^O_n), \forall m \in \mathbb{R}_n\). Then consider two candidate resource groups \( m, m' \in \mathbb{R}_n \).

For equivalent processing time or delay, if \( \frac{(c^K)^\theta_n}{\mu^e_m} \leq \frac{(c^K)^\theta_n}{\mu^e_{m'}} \), then resource group \( m \) completes the task at lower cost (and therefore generates greater value).

**Proof.** For the equivalent processing time case, we have:

\[
\frac{\lambda_n}{K^\theta_n \mu^e_m} = \frac{\lambda_n}{K^\theta_n \mu^e_{m'}} \implies K^\theta_n \mu^e_m = K^\theta_n \mu^e_{m'} \implies K_m = K_{m'} = K_m \left( \frac{\mu^e_m}{\mu^e_{m'}} \right)^{\frac{1}{\theta_n}}.
\]
Hence the cost difference if resource \( m \) performs the task over resource \( m' \) is:

\[
c_m^K K_m - c_{m'}^K K_{m'} = K_m \left( c_m^K - c_{m'}^K \left( \frac{\mu_m^e}{\mu_{m'}^e} \right) \right)
\]

\[
= K_m (\mu_m^e) \frac{1}{\theta_m} \left( \left( \frac{c_m^K \theta_n}{\mu_m^e} \right) \frac{1}{\theta_m} - \left( \frac{c_{m'}^K \theta_n}{\mu_{m'}^e} \right) \frac{1}{\theta_m} \right)
\]

\[
\leq 0.
\]

For equivalent delay, we can examine two sub-cases: in the first case, when the delay is zero for both resources; in which case we again claim that for any processing time achieved by resource \( m' \) that less than or equal to \( \hat{\tau}_n \), resource \( m \) completes the task at lower cost, while maintaining the zero delay condition. For any strictly positive delay given \( \hat{\tau}_n \), equivalent delay implies equivalent processing time for the two resources, and again resource \( m \) is more cost-efficient in achieving that delay, from the analysis above.

\[\square\]
7

Centralized Capacity Planning For Collaborative
Innovation Programs

7.1 Overview and Organization

In this chapter, we further develop the modeling framework of Chapter 6 to formulate a capacity planning problem for collaborative programs that are tasked with an agenda for innovation. The organization of this chapter is as follows: first we prove some basic convexity results for the processing time and delay performance measures in the capacity vector $K$, while also proving some important monotonicity properties of these measures. Then, assuming a fixed task-to-resource assignment $X$, we show the existence of a global optimal capacity vector that maximizes the net value to the program, under very general conditions including uncertainty in the key parameters of the program network. We are thus able to formulate a global, or centralized capacity planning problem for the program, given a fixed assignment strategy.

We also formulate an integrated task assignment and capacity planning problem, if only for the sake of illustrating the complexity of the integrated problem, and the
computational hurdles in solving large scale problems efficiently. We survey the literature on similar mixed-integer, non-linear problems and discuss the best possible generic algorithms that can be used for large scale programs.

7.2 Some Convexity And Monotonicity Properties Of Program Performance Measures.

In this section we claim and prove some convexity and monotonicity properties for the processing time and delay functions defined over the network $\Psi$, given a fixed assignment $X$. We also prove by correspondence the convexity and monotonicity properties of the associated cost measures.

**Proposition 7.2.1.** Given a fixed assignment $X$ (with $x_{mn} = 1$), the processing time function $\tau_{mn}(X, K_m)$, as defined in Equation 6.1, is non-increasing convex in $K_m$.

**Proof.** For $\theta_m > 0$, note that while the first derivative is negative, the second derivative of $\tau_{mn}(X, K_m)$ w.r.t. $K_m$ is non-negative.

**Corollary 7.2.2.** Given a fixed assignment $X$ (with $x_{mn} = 1$), the processing time function $\tau_{mn}(X, K_m)$, as defined in Equation 6.1, is convex in $K$ (through $K_m$).

**Proposition 7.2.3.** The expected processing time $E[\tau_{mn}(X, K_m)]$ is non-negative, non-increasing, and convex in $K$ (through $K_m$).

**Proof.** We will give the proof for the convexity of expected measures only once; in essence the expectation of $\tau_{mn}(X, K_m)$ is over the random variables $\lambda_n$ and $\mu_m$.
which are parameters for the function. Since we assume that the base demands and processing rates are non-negative under every realization of the random variables, the non-negativity property of the functions extend to their respective expectations. Note that the non-negativity property is guaranteed for every realization of the random variables. The same argument applies for the non-decreasing property of the expected values; the first derivative does not change sign over the positive realizations. Finally, the convexity property is preserved in expectation; see Boyd and Vandenberghe [22].

Proposition 7.2.4. For any \( n \), the delay \( D_n(X, K) \), as defined in Equation 6.6, is also non-negative, non-increasing, and convex in \( K \).

Proof. Note again that \( \tau^e_n(X, K) \), of Equation 6.5 is convex in \( K \). The delay function \( D_n(X, K) \) is the positive part of \( (\tau^e_n(X, K) - \hat{\tau}_n) \) which quantity is again convex in \( K \).

Proposition 7.2.5. The program completion time \( \tau(X, K) \), as defined in Equation 6.9 is convex and non-increasing in \( K \).

Proof. The time to complete task \( k \) that is a leaf on the program network, as defined in Equation 6.7 is convex in \( K \) from the corollary to Proposition 7.2.2.

Next, given a set \( \{k\} \) of tasks that are at the leaves of the network, and given the recursion defined in Equation 6.8, consider the time to complete task \( n \) such that \( \psi_{kn} = +1 \), for some such \( k \). The recursion function for \( n \) is the sum of a convex function \( \tau^e_n(X, \{K_{m \in R_n}\}) \) and the maximum of convex functions \( \tau^c_k(X, \{K_{m \in R_k}\}) \) such that \( \psi_{kn} = +1 \). Hence \( \tau^c_n(X, \{K_{m \in R_n}\}) \) is convex in \( K \).
In this way, the convexity propagates up the program network from the leaves to \( \tau_N^c(X, \{K_m \in \mathbb{R}_N\}) = \tau(X, K) \), which is the cumulative time to complete task \( N \) which is at the root of the program tree.

Proposition 7.2.6. The program delay function \( D(X, K) \), of Equation 6.12 is convex in \( K \).

Proof. The proof is analogous to the proof of Proposition 7.2.5.

Proposition 7.2.7. For a fixed assignment \( X \), the resource level cost \( C_m^R(X, K) \), the firm level cost \( C_f^F(X, K) \), and the program level cost \( C(X, K) \) as defined in Equation 6.15 are each convex in \( K \), and so are their respective expected values \( E[C_m(X, K_m)] \), \( E[C_f(X, K)] \) and \( E[C(X, K)] \).

Proof. The resource level \( C_m^R(X, K) \) is the sum of the operating cost \( c_m^O \lambda_{mn}(X) \), the penalty cost \( c_m^P D_n(X, \{K_m \in \mathbb{R}_n\}) \) and the capacity sizing cost \( c_m^K K_m \). The latter two quantities are convex in \( K \). Hence \( C_m^R(X, K) \) is also convex in \( K \). The firm level, and program level costs are the sum of appropriate resource level costs, and are again convex in \( K \).

Proposition 7.2.8. For fixed \( K \), the value functions \( \hat{V}(X, K) \), and \( V(X, K) \) as defined in Equation 6.17, and their expected values, are each concave in \( K \).

Proof. The program gain functions, as defined in Equations 6.16a and 6.16b, are concave in \( X \). From Proposition 7.2.7, the program cost \( -C(X, K) \) is concave in \( K \),
and we have the result.

**Proposition 7.2.9.** Assume a fixed assignment \( X \). Let \( \Upsilon(X) \), alternatively \( \hat{\Upsilon}(X) \), be the set of maximizers for \( E[V(X,K)] \), respectively \( E[\hat{V}(X,K)] \), over all feasible \( K \). Then, \( \Upsilon(X) \) and \( \hat{\Upsilon}(X) \) are each non-empty sets; in other words, there is at least one global maximizer \( K^*(X) \) of the value function \( E[V(X,K)] \), and at least one global maximizer \( \hat{K}^*(X) \) of the value function \( E[\hat{V}(X,K)] \). Furthermore, \( \Upsilon(X) \) and \( \hat{\Upsilon}(X) \) are each convex over the feasible vectors \( K \).

**Proof.** Follows from the concavity of the value functions in \( K \). See Boyd and Vandenberghe [22].

The significance of Proposition 7.2.9 is that it allows the program value maximization problem to be decomposed in a special way. Given a fixed (or existing) assignment decision specified by the program planner can determine the optimal assignment \( K^* \) for the partner firms corresponding to that specific assignment. Then, the globally optimal joint assignment and capacity vector can be determined through a search procedure that evaluates the feasible assignments of tasks to resources.

In the following sections, we proceed to decompose the overall centralized decision problem in precisely such a fashion. Assuming a fixed (or pre-conceived) assignment strategy, we first model the program planner’s centralized problem of determining the capacity vector that maximizes overall program value. Then, since the resource pool selection is a binary choice, we formulate the assignment problem as a search over all feasible task-resource assignments. The assignment problem is a hard one, since the size of the problem grows very quickly in the number of candidate resources.
for a task, and with the number of tasks in the program network. However, we out-
line conditions under which simpler and locally optimal selection rules, such as those
derived in Chapter 6, might be applicable for network problems.

Looking further, in a decentralized decision environment, if for a given assignment
\( X \), the program optimal investment is \( K^* \); and if the program planner can somehow
induce the individual firms to invest \( K^*_m \) in each resource \( m \), then the resulting de-
centralized solution \( K^* \) is by definition a globally optimal capacity investment for the
program, for this particular assignment strategy. Furthermore, if \( (X^*, K^*) \) is indeed
the joint optimal assignment and investment, the concavity of the program value
function in \( K \) allows for the possibility that the decentralized capacity investments
by the firms can still lead to program optimal outcomes. Utilizing this key insight,
we examine the critical issue of coordinating the supply chain investment in capacity
in Sections 8.2 and 8.5. In other words, is it possible to replicate the program optimal
capacity vector derived in a centralized fashion, but via a decentralized mechanism?

7.3 An Integrated Formulation of the Centralized Assignment and
Capacity Problem.

In this section, we show the coupled nature of the task assignment and the capacity
investment problem, when viewed from the perspective of the central planner. We
show that the integrated problem is a mixed integer non-linear program, which is
in general a hard problem to solve, with the non-linearities imposed by the delay
costs, and the combinatorial nature of the assignment problem. We do however,
show that the problem fits into the mixed integer non-linear programming (MINLP)
classification of problems, and therefore perhaps yields to both classical and more
recent algorithms to reduce the computation effort in its solution.
For the centralized capacity assignment and investment problem, the input to the planner for its optimization problem is then as follows:

- The set of tasks, and their base demand distributions $\lambda$.
- The set of resources, their capabilities $\hat{X}$, their ownership structure $\alpha$; the distribution of the resource processing rates $\mu_m$, and their collaborative strength $\theta_m$.
- The feasible set of capacities for each resource pool; we denote this set for a resource group by $\kappa_m = [K_m, K_m]$; we assume that this set is closed to preserve convexity of the feasible set ($K_m$ and $K_m$ are the lower and upper bounds for the capacity of each resource pool). Critically, we do not allow for integer values for $K_m$, whereas this is the most natural model for the size of the resource pool. Rather we model different environments where the size of the resource pool can be variable on a continuous scale (for e.g. where resources can be hired in fractional increments).
- The collaborative requirements matrix $\Delta$ for the tasks in the program.
- The cost structure of the individual firms in terms of their marginal resource capacity costs $c^K$, marginal operating costs $c^O$, and the resource penalty costs $c^P$ (although in some cases, the penalty costs could be a requirement set by the central planner in the form of transfer payments).
- The information about the market or the customer that defines how the program gain depends on the performance of the program; in other words, which of Equations 6.16a and 6.16b determines the program gain, and hence the value function to be used as an objective in the program optimization.
Finally, we could impose participation constraints for the firms involved so that no firm incurs a net expected loss from participating in the program. In certain situations, firms may indeed require a bit more, and require minimum profit levels (or payoffs) from their participation. Let us assume the firms have lower bounds $V_l$ on their expected value from participation under either gain function. It is also possible in some situations to simplify the discussion and only impose non-negativity constraints on the net value derived by individual firms (i.e. define $V_l = 0$).

The program planner’s value maximization problem can then be written as follows.

(CD: Delay Sensitive Value Maximization) :

$$E\hat{V}^* = \max_{K,X} E[\hat{V}(X,K)]$$

subject to : $\left(\sum_{n \in T} x_{mn}\right) K_m \leq K_m \leq \left(\sum_{n \in T} x_{mn}\right) K_m$, $\forall m \in R$

subject to : $\sum_{m \in R} x_{mn} = 1; \forall n \in T$

subject to : $x_{mn} \leq \hat{x}_{mn}; \forall m \in R, n \in T$

subject to : $x_{mn} \in \{0,1\}; \forall m \in R, n \in T$

subject to : $E[\hat{V}_l(X,K)] \geq \min \left(1, \sum_{n \in T} \sum_{m \in R} \alpha_{lm} x_{mn}\right) V_l; \forall l \in A$;
or:

(CL: Processing Time Sensitive Value Maximization):

\[
EV^* = \max_{K,X} E[V(X,K)]
\]

subject to : \(\left(\sum_{n \in T} x_{mn}\right)K_m \leq K_m \leq \left(\sum_{n \in T} x_{mn}\right)\bar{K}_m, \forall m \in R\)

subject to : \(\sum_{m \in R} x_{mn} = 1; \forall n \in T\)

subject to : \(x_{mn} \leq \hat{x}_{mn}; \forall m \in R, n \in T\)

subject to : \(x_{mn} \in \{0,1\}; \forall m \in R, n \in T\)

subject to : \(E[V_l(X,K)] \geq \min \left(1, \sum_{n \in T} \sum_{m \in R} \alpha_{lm}x_{mn}\right) V_l; \forall l \in A.\)

7.3.1 Some remarks on solution algorithms for the central planner’s integrated problem.

As can be observed the assignment decision \(x_{mn}\) in problems (\(CD, CL\)), is a binary variable. This renders the central planner’s integrated problem as a mixed-integer non-linear program (MINLP) which is in general quite a hard problem to solve (in fact problems in this category are known to be NP-complete, and therefore worse than NP-hard optimization problems). However, over the past three or so decades, there have been a number of algorithmic approaches that have been applied to these problems that have improved the average-case performance for a wide range of applications. In all cases, the convexity of objective function, and optionally, the participation constraint is a requirement for guaranteed convergence to the true optimal solution.

Many of the solution algorithms developed generally for MINLP models follow ideas developed for mixed integer programs (MIP), and therefore rely on effective
searching through the feasible space of the integer variables. Examples include branch and bound techniques, where now a non-linear program (NLP) is solved at each node of the branch and bound tree, instead of a linear program (see Gupta and Ravindran, [93]). The complexity therefore depends on the efficiency of the algorithm used to solve the NLP. Since efficient interior point methods have been developed in commercial solvers for non-linear problems, the branch and bound method appears to be conducive for solving problems \((CD, CL)\). Still, the combinatorial nature of the assignment problem in problems \((CD, CL)\) may require at worst an enumeration of all permissible and feasible assignments of tasks to resource groups.

Another approach is to formulate an associated relaxed problem that is a mixed-integer-linear program (MILP) and then use well-known cutting plane methods to shrink the feasible set of the relaxation, and converge to the true optimal solution (Westerlund and Pettersson, [197]). Cutting plane methods to solve NLPs were developed first by Kelley [117], while Westerlund and Pettersson extend Kelley’s algorithm to problems with integer variables. However, Westerlund and Pettersson’s algorithm is suited for problems that exhibit “mild” non-linearity, whereas problems \((CD, CL)\) exhibit a high degree of non-linearity due to the nature of the delay function. Moreover, constructing the linear relaxation of the convex constraint set requires gradient based methods, while the participation constraints in problems \((CD, CL)\) lead to non-smooth convex sets that potentially have many non-differentiable extreme points. Removing the participation constraint, and supposing further that \(K_m = 0\) and \(\overline{K}_m = \infty, \forall m\), simplifies problems \((CD, CL)\) to maximizing a convex non-linear objective over a polyhedral feasible space, and hence the MILP relaxation based methods become directly applicable.

A related approach called the Outer Approximation (OA) (due to Duran and
Grossman [61]) involves solving an alternating sequence of MILP relaxations, and the non-linear program. First a relaxed MILP problem is solved with a linear approximation of the convex objective function and constraints, and for the resulting values of the integer (assignment) variables, the NLP can solved to yield the best corresponding resource capacity values. If the NLP is feasible, it implies that the original MINLP is also feasible, and therefore a new linear relaxation of the original problem can be constructed at this point. Successive solution of the MILP and the NLP problems leads us to a solution that converges to the true optimal solution. The OA method is more likely to yield the true optimal solution since the algorithm terminates with the solution to the NLP corresponding to a given solution for the integer variables. The MILP (relaxation) could help identify the best assignment of tasks to resource groups. Again, removing the participation constraint, and defining $K_m = 0$ and $\bar{K}_m = \infty, \forall m$, simplifies problems $(CD, CL)$ and hence requires only a linear approximation of the convex objective function to construct the MILP relaxation.

Geoffrion [82] proposes the generalized Benders’ decomposition (GBD) method which is quite similar to the OA method outlined above (see Flippo and Kan [74] for a more general discussion of decomposition methods including the Dantzig-Wolfe algorithm). The key difference being how the relaxed MILP problem is constructed from the original MINLP. In the GBD algorithm, this MILP relaxation is constructed from the dual representation of the convex feasible space. The main idea again is to fix the integer (assignment) variables, and create dual subproblems to solve for the continuous (capacity) variables. However, as shown in Grothey et al. [91], the convergence of the decomposition algorithm is not guaranteed for general non-linear programs, especially for constraint and objective functions such as in $(CD, CL)$ that are not twice differentiable. See however, Federgruen and Zipkin [70] for an applica-
tion of the GBD method to a combined inventory allocation and routing problem in the logistics and distribution literature.

Finally, there have been some recent developments in developing branch and bound methods for handling the integer constraints in conjunction with efficient interior point methods for solving the NLPs generated at every node of the algorithm (see Benson [13]). Interior point methods are the algorithms of choice for many commercial nonlinear optimization solvers given their speed and accuracy for a vast range of non-linear problems. However, there are some issues posed by the inability of interior point methods to fit within “bi-level” approaches, where a sequence of alternating problems (such as master and subproblem) are solved and where information from the past solutions in the sequence can be used creatively to speed up the algorithm. Interior point methods further suffer from an inability to detect infeasible problems quickly – a feature required in combinatorial branch and bound searches. They also pose numerical difficulties if the optimal set of dual solutions in any of the subproblems is unbounded – which occurs when variables are fixed as outside parameters, for instance. However, as shown in Benson [13], there are modifications to the interior-point algorithm that can smooth over some of these drawbacks.

The purpose of the discussion above has been to outline the algorithmic alternatives that are feasible within the context of the central planner’s problem. More broadly, our focus in this dissertation is on the structure of the assignment and capacity planning problems, and on how it changes between the centralized and decentralized formulations. As such we avoid a detailed discussion and comparison of solution algorithms for the central planner’s core problem above. It is also our belief that given the fundamental combinatorial nature of the assignment problem, any solution algorithm will have to surmount the difficulties of searching through the
space of feasible task-to-resource assignments. Moreover, given the advances in computational techniques and microprocessor speeds, the comparison between different algorithms, based on the current and past approaches discussed above, comes down to applicability to the specific instances of \((CD, CL)\) generated by their parameters.

We therefore leave the development of more tailored solution algorithms for the central planner’s integrated assignment and capacity planning problem to future research. One promising avenue is to examine how the network (or tree) structure of tasks and resource groups can be exploited to solve the NLP generated by any given choice of assignment variables. For example, parallel tasks have potentially greater inter-dependency in their capacity requirements than tasks that are serially arranged, at least with our model of task processing time. One could then devise algorithms that exploit optimality conditions for specific tree structures in order to compute the optimal capacity allocations. There is also scope to exploit the structure of the objective function and of the constraint set to develop collaborative methods to solve the decentralized problem, but we do not address such a framework in this dissertation, and leave the exploration of such collaborative algorithms for future research.

7.4 Centralized, Program Optimal Capacity Investment Given A Fixed Assignment Strategy.

In this section, assuming a given or fixed assignment strategy, we formulate the problem of capacity determination from the program planner’s (or centralized) perspective. We assume the program planner is neutral w.r.t. the gains and losses of the participating firms; the only objective of the centralized planner is to maximize the net value derived for the program. Situations where the assignment is fixed
or pre-determined are more common than would first seem possible. For example, task responsibilities may be constrained by the technological capabilities of firms, or indeed by certain process, or even business relationship considerations. Here, we examine the problem of capacity determination given such a fixed assignment strategy.

However, our immediate motivation here is to decouple the centralized assignment and capacity investment problems, so that we can better highlight the problem of coordination of capacity investments. Hence, for a fixed assignment strategy, we define the centralized capacity investment problem. Then for a fixed assignment, we examine how firms would approach the capacity investment decision in a decentralized fashion. Then, conditional again on a fixed assignment, we develop coordination mechanisms that can achieve coordination of the capacity investment.

For the centralized capacity investment problem, we assume a fixed, pre-conceived assignment strategy $X$. As discussed earlier, this assignment strategy could be determined based on technology or process considerations, or based on other partnership, or even market related concerns. The program planner’s (conditional) value maximization problem can then be written as follows.

\[
\hat{EV}^*(X) = \max_K E[\hat{V}(X, K)]
\]

subject to : $K_m \in \kappa_m, \forall m : \sum_{n \in T} x_{mn} = 1$

subject to : $E[V_l(X, K)] \geq V_l, \forall l : \sum_{m \in R} \alpha_{lm} \sum_{n \in T} x_{mn} \geq 1$;
or:

$$EV^*(X) = \max_K E[V(X, K)]$$

subject to : $$K_m \in \kappa_m, \forall m : \sum_{n \in T} x_{mn} = 1$$

subject to : $$E[V_i(X, K)] \geq V_i, \forall l : \sum_{m \in R} \alpha_{lm} \sum_{n \in T} x_{mn} \geq 1.$$

**Proposition 7.4.1.** If problems (D) and (L) are feasible for the same data, $$EV^*(X) \leq \hat{EV}^*(X).$$

**Proof.** The proof follows from the fact that for any $$X$$, and for specific task deadlines $$\hat{\tau} \geq 0$$, $$\tau(X, K) \geq D(X, K);$$ hence $$\Pi(X, K) \leq \Pi\hat{\Pi}(X, K),$$ and therefore $$V(X, K) \leq \hat{V}(X, K)$$ for every $$K.$$

Proposition 7.4.1 suggests (or confirms) that programs that are sensitive to the processing time, rather than to the delay in program completion yield lower value. Note that the value is lower not only in expectation, but also for every realization of the random environment. The output of the central planner’s optimization problem is therefore defined by:

1. A set of optimal capacity investment vectors $$\hat{\Upsilon}(X)$$ such that any $$(K^*) \in \hat{\Upsilon}(X)$$ is optimal for problem $$D;$$ alternatively, $$\Upsilon(X)$$ such that any $$(K^*) \in \Upsilon(X)$$ is optimal for problem $$L.$$

2. The maximum expected value $$EV^*(X),$$ alternatively, $$\hat{EV}^*(X),$$ for the program.
In Section 8.2, we translate this capacity optimization problem, solved from the perspective of the centralized program planner, to decentralized resource or firm level optimization problems. The objective of these decentralized problems is to determine the optimal capacity investment from the perspective of an independent resource owner, or indeed from the perspective of the participating firms. The broader goal is to derive mechanisms that ensure that the decentralized solution(s) align with this centralized solution(s), while satisfying the requirements and constraints of individual firms. Doing so would imply that the supply chain decisions on capacity investment are coordinated. Before doing so however, we first explore the potential for developing algorithms for the centralized task assignment/ resource selection problem in deterministic environments.

7.5 Remarks on centralized program optimal assignment in general networks.

In the previous sub-sections, we have essentially proved the existence of an optimal capacity vector, given a network-wide assignment strategy $X$ (conditional of course, on well-defined problem parameters). Thus, given an particular assignment, we can solve the central planner’s program for the set of optimal capacity vectors $\hat{\Upsilon}(X)$ or $\Upsilon(X)$ (as the case may be), and also compute the associated optimal value $\hat{EV}^*(X)$ or $EV^*(X)$, respectively. Let us denote the set of feasible assignments by $Z$. Then, we can, in theory, conduct a search over the feasible assignments in $Z$ to determine the optimal assignment $X^*$ that maximizes the program value $\hat{EV}^*(X)$ (or $EV^*(X)$) over all $X \in Z$.

The problem with the above approach is that the number of possible elements in the set $Z$ could be exponential in the number of resources and in the tasks; the total
number of such assignments could be as large as $\prod_{n=1}^{N} M_n$. Hence, it is appropriate to consider more efficient search procedures for optimizing over the feasible assignments.

One such avenue is offered from the previous discussion in Chapter 6. For a given task $n$ considered in isolation from the rest of the network, we have shown how the optimal policy, under a deterministic environment is to select the resource pool with the lowest ratio $\frac{(c^K_m)^\theta}{\mu_m}$, when the resource groups all exhibit identical cost parameters (other than the marginal capacity cost). One possibility is to examine variants of this policy for more general networks, subject to uncertainty in the task requirements, resource processing rate, or program gain. This is a topic of ongoing research by the author.

Here, we analyze the special case when the program environment is deterministic, and when there are no constraints on the problems $D$ and $L$ from the previous section. In other words the feasible set of capacities $\kappa_m$ is the non-negative reals, and firms do not have any participation constraints. Also, the analysis we present here is applicable to environments where the operating costs and resource specific penalty costs are assumed insignificant relative to the resource capacity costs. Then, we show here that the locally optimal strategy of selecting the resource with the lowest ratio $\frac{(c^K_m)^\theta}{\mu_m}$ is also a globally optimal one.

**Proposition 7.5.1.** Consider a program network with task set $T$ specified by precedence relationship $\Psi$ in a deterministic program environment with either delay or processing time sensitive gain; and assume that there are no separate resource penalties imposed for task delays. Suppose $(X^*, K^*)$ that is the jointly optimal assignment, and capacity vector, resp., for the program. Then, consider any task $n$ that has been
assigned to resource group \( m \) in the optimal assignment, i.e. \( x_{nm}^* = 1 \) is specified by the optimal assignment \( X^* \); and further suppose that \((\theta_m, c^P_m, c^O_m) = (\theta_n, c^P_n, c^O_n), \forall m \in R_n\). Then,

\[
\frac{(c^K_m)^{\theta_n}}{\mu_m^e} \leq \frac{(c^K_{m'})^{\theta_n}}{\mu_{m'}^e}, \forall m' \in R_n.
\]

Proof. The proof is via contradiction. Suppose for task \( n \) under consideration, the optimal policy \((X^*, K^*)\) assigned the task to a resource pool \( m' \) such that for some other \( m \in R_n \), we have \((c^K_m)^{\theta_n} \mu_m^e < (c^K_{m'})^{\theta_n} \mu_{m'}^e\).

Let us first examine the processing time sensitive program gain. Suppose the optimal processing time for task \( n \) processed by resource \( m' \) is \( \tau_n^* \). Then, from Proposition 6.5.9, for the same processing time \( \tau_n^* \), resource pool \( m \) completes task \( n \) at strictly lower cost. Furthermore, if we maintain the same processing time for the task, the overall program gain is unperturbed, hence resource \( m' \) cannot be an optimal assignment for task \( n \).

When the program gain is sensitive to delays instead, then similar to the proof of Proposition 6.5.9, there are two further scenarios to consider. If the optimal delay \( \hat{\tau}_n^* \) is zero, but the optimal processing time is equal to \( \tau_n^* \), then the same argument above is repeated to show the contradiction. On the other hand, if the optimal delay for task \( n \) is \( \hat{\tau}_n^* \), then again for the corresponding optimal processing time \( \tau_n^* \), we can again substitute resource \( m' \) with \( m \), and reduce costs, without any perturbation to the program gain.

What happens when the parameters \((\theta_m, c^P_m, c^O_m)\) vary across the resources? In
this case, there is no simple or elegant ranking characterization of the optimal re-
source selection or task assignment policy. Hence, the main trade-offs in task as-
signment are between greater capacity costs and the opportunity costs of processing
time and/or delays.
8

Decentralized Planning & Coordination
Mechanisms For Innovation Programs

8.1 Overview and Organization

In this chapter, we will assume throughout a fixed and centralized task assignment, and analyze instead on the decentralized capacity investment problem. Since the typical program is composed from multiple firms, the decentralized version of the capacity planning problem assumes special significance. Here we show the negative result that decentralized decisions by firms do not lead to the centralized or globally optimal capacity solution (this is perhaps readily understood, but important nevertheless). This points to inherent inefficiencies in the decentralized decision-making. Most critically, we demonstrate how even commonly used profit or revenue sharing mechanisms such as gain or work share formulae actually contribute to these inefficiencies when used naively.

We also demonstrate the existence (or at least the possibility) of equilibrium capacity investments by firms competing for gain share in stylized program settings.
However, we also show how these equilibrium capacity investments can again be inefficient compared to the centrally optimal capacity investments. The rest of the chapter, therefore, is devoted to mechanisms that can mitigate or eliminate such inefficiencies.

We propose some simple modifications to different gain share mechanisms, when used in conjunction with decision-hierarchies and rules for information exchange, can help coordinate the decentralized decisions. These decision-hierarchies and information rules while arguably simplistic, are nevertheless not beyond comprehension as governance principles for diverse and complex collaborative program environments. Essentially, they call for bilateral, rather than multi-lateral negotiation and information exchange, and grant special decision rights to the central planner. In effect they force firms to make the assumption that all of their partners are already investing capacity in accordance with a centralized or globally optimal solution. Without these simplifying assumptions, the problem escalates into a game-theoretic interaction between multiple firms that while interesting to model and analyze, is seldom governable from the point of view of sustaining coalitions in real world collaborative programs. We leave such game-theoretic formulations for future work, and devote our attention to analyzing more basic bilateral formulations that can provide insight into the behavior of readily implementable coordination mechanisms.

Finally, we also provide two alternative views of the coordination problem: the first is a resource level view, that captures how coordination can only be achieved by aligning incentives at the individual resource (and therefore individual task) levels. The resource level coordination problem helps us understand the basic requirements for decentralized decisions to align with the globally optimal solutions. The second is a firm level view that illustrates how these resource or task level incentives are used
by firms owning a collection of resources and contributing capacity to various tasks in the overall program. Most critically, the insight from our analysis here is that it is very difficult to define and set incentives at the level of entire firms, that can achieve alignment with the central planner’s objectives and decisions. Incentives that are provided only at the firm level can lead to firms maximizing their net value or gain share by allocating their capacities in ways that are again not conducive to the overall program. However, by defining incentives and enforcing penalties at the level of individual resource groups, the central planners can motivate firms to make capacity decisions consistent with the global optimal. These alternative perspectives, while laborious and somewhat repetitive in their presentation, are nevertheless a crucial insight into the important problem of coordinating decisions for the collaborative program environment.

8.2 Decentralized Resource Level Capacity Decisions For a Fixed Assignment.

To set up the firms’ capacity decision problems, the program planner has to specify the revenue or gain share mechanism. We will assume that program planners have the option of choosing a single gain share mechanism from those defined in Equations 6.18-6.20, and then offering them to all of the partner firms. In this way firms are all subject to the same gain and cost share mechanism offered by the program planner. As we will see later, there could be situations where the program planner can achieve coordination with all three mechanisms defined above, but that offering a uniform gain share mechanism to all firms is a pre-condition for coordination.

Consider a fixed assignment strategy and its corresponding optimal investment $K^* \in \Upsilon(X)$ derived from the global program optimization problem $L$ defined previ-
ously. The question we examine here is whether firms can independently determine their capacity investments and arrive at the same solution. To answer this question, we first look at a single firm’s capacity decision w.r.t. a single resource $m$.

Assume, for now, that each firm that owns only one resource $m$ in the program; we relax this assumption later. Suppose the program planners offers the assignment $X$ along with a guarantee that $K_j = K_j^*; \forall j \neq m$; i.e. the planner guarantees that all of the other resources will have program optimal capacity $K_j = K_j^*$. Suppose further that the planner offers a share $\gamma^R_m \in \{\gamma^{IRS}_m, \gamma^{WIRS}_m, \gamma^{CRS}_m\}$. Finally, suppose that the environment is described by a gain function that could depend on either the program level delays or the program processing time, but not both (i.e. the gain is defined by Equations 6.16b and 6.16a). In the analysis to follow, we will use these gain functions interchangeably, but it is easily verified that the analysis holds for both types of program environments.

Then, the firm solves one of the following problems:

$$
(\gamma^R_m\text{-D: Delay Sensitive Value}):$

$$
E \hat{V}^R_m(\gamma^R_m, X, K^*_{m-}) = \max_{K_m} E \left[ \gamma^R_m \hat{\Pi}(X, (K^*_m, K_m)) - C^O_m(X) - C^P_m(X, K_m) \right] - c^K_m K_m
$$

subject to : $K_m \in \kappa_m$.
\((\gamma_m^R - L): \text{Processing Time Sensitive Value}\):

\[
EV_m^{R_s}(\gamma_m^R, X, K_{m^*}) = \max_{K_m} E[\gamma_m^R \Pi(X, (K_{m^*}, K_m)) - C_m^{QR}(X) - C_m^{P}(X, K_m)]
\]

\[-c_m^K K_m
\]

subject to : \(K_m \in \kappa_m\).

**Proposition 8.2.1.** \(\gamma_m^R - D\) and \(\gamma_m^R - L\) are not convex programs, for any \(\gamma_m^R \in \{\gamma_{IRS}^m, \gamma_{WIRS}^m, \gamma_{CRS}^m\}\). However, both problems possess well-defined maximum values over the feasible range of \(K_m\).

**Proof.** We prove this fact by construction. Essentially, neither \(EV_m^{R_s}(\gamma_m^R, X, K_{m^*})\) nor \(EV_m^{R_s}(\gamma_m^R, X, K_{m^*})\) is guaranteed to be concave in \(K_m\). The non-concavity results from the fact that \(\gamma_m^R \Pi(X, (K_{m^*}, K_m))\) is the product of a ratio \(\gamma_m^R\), and the gain function \(\Pi(X, (K_{m^*}, K_m))\). Now, \(\gamma_{IRS}^m\) and \(\gamma_{WIRS}^m\) are each concave in \(K_m\). However, \(\gamma_{CRS}^m\) cannot be guaranteed to be concave in \(K_m\) in general.

Independent of whether \(\gamma_m^R \in \{\gamma_{IRS}^m, \gamma_{WIRS}^m, \gamma_{CRS}^m\}\) is concave, the product

\[
\gamma_m^R \Pi(X, (K_{m^*}, K_m))
\]

is not guaranteed to be concave in \(K_m\).

Note that there is always, however, a well-defined global maximum to both problems, since in the objective function, all of the terms converge to some constants w.r.t. \(K_m\) save for the capacity cost \(c_m^K K_m\). In other words, the firm’s gain, its operating cost, and its delay costs become invariant w.r.t \(K_m\) as \(K_m \to \infty\); since the capacity cost is strictly increasing for positive \(c_m^K\), we are ensured a global maximum for both problems above.
The implication of the proposition above is that coordination is hard to achieve, even with a single resource, because of the non-convex nature of resource level value functions. Effectively, even when the program planner can offer an assignment $X$ combined with the optimal resource investments $K_j^*$ for all the other resource groups, the firm operating resource $m$ may not “see” the global optimum $K_m^*$ if going by a marginal analysis approach; the firm may reach and prefer to invest in a locally optimum fashion.

Next, we consider the same question, but when the firms can somehow distinguish between local and global optimum in $\gamma^R_m - D$ and $\gamma^R_m - L$. Could the global optima of these two resource level sub-problems coincide with the global optima of the program level problems $D$ or $L$, for any $\gamma^R_m \in \{\gamma^{IRS}_m, \gamma^{WIRS}_m, \gamma^{CRS}_m\}$? As the next proposition claims, this is not necessarily the case for general program environments.

**Proposition 8.2.2.** $K^*_m$, a globally optimal resource investment for $m$ under program optimization problem $D$ (or $L$) may not be the globally optimum solution to the resource level problem $\gamma^R_m - D$ (or $\gamma^R_m - L$) for any $\gamma^R_m \in \{\gamma^{IRS}_m, \gamma^{WIRS}_m, \gamma^{CRS}_m\}$.

**Proof.** We prove this fact again by construction. Suppose first $K^* \in \Upsilon(X)$ is such that $K^*_m$ specified by the global optimum is the maximum over any $K^o_m$ such that $K^o \in \Upsilon(X)$.

Note that $C^{OR}_m(X)$ is independent of $K_m$ so the capacity decision has no influence on the processing cost. Next, note that as $K_m \to \infty$, $E[D^R_m(X, K_m)] \to 0$, and therefore $E[C^P_m(X, K_m)] \to 0$, independent of the assignment $X$ and independent of the capacities $K_{m-}$. Similarly, $E[D(X, K^*_m, K_m)] \to s$ for some constant $s$ as
\(K_m \to \infty\), since with large \(K_m\) any delays can only be due to \(K_{m-}^*\). Also, observe that \(\gamma_m^R \to 1\) as \(K_m \to \infty\) for problem \(\gamma_m^R - D\).

However, the program optimal gain share \(\gamma_m^*\) could be less than 1 under problem \(P\) which implies that beyond that level the program sees no benefit from deploying more of resource \(m\). Suppose that indeed is the case, and \(\gamma_m^* < 1\). Since, the program gain \(E[\Pi(X, (K_{m-}^*, K_m^*))]\) is non-decreasing in \(K_m\), \(E[C^O(X)]\) is independent of \(K_m\), and \(E[C^P(X, K_{m-}^*, K_m^*)]\) is non-increasing in \(K_m\), it must be that the first difference

\[
E[\Pi(X, K_{m-}^*, K_m^* + \epsilon)] - E[\Pi(X, K_{m-}^*, K_m^*)]
\]

\[
- E[C_m^P(X, K_m^* + \epsilon)] + E[C_m^P(X, K_m^*)] = c_k K_m \epsilon
\]

\[
\implies
E[\Pi(X, K_{m-}^*, K_m^* + \epsilon)] - E[\Pi(X, K_{m-}^*, K_m^*)]
= E[C_m^P(X, K_m^* + \epsilon)] - E[C_m^P(X, K_m^*)] + c_k K_m \epsilon,
\]

for some sufficiently small \(\epsilon > 0\).

However if \(\gamma_m^* < 1\), then \(K_m = K_m^* + \epsilon\) would imply that \(\gamma_m^R[K_m^* + \epsilon] > \gamma_m^*\), so the share of resource group \(m\) in the program gain is still increasing at \(K_m^*\). Thus,

\[
E[\gamma_m^R[K_m^* + \epsilon]\Pi(X, K_{m-}^*, K_m^* + \epsilon)] > E[\gamma_m^*\Pi(X, K_{m-}^*, K_m^*)]
\]

\[
\implies
E[\gamma_m^R[K_m^* + \epsilon]\Pi(X, K_{m-}^*, K_m^* + \epsilon)] - E[C_m^P(X, K_m^* + \epsilon)] - C_m^{QR}(X)
\]

\[
> E[\gamma_m^*\Pi(X, K_{m-}^*, K_m^*)] - E[C_m^P(X, K_m^*)] - C_m^{QR}(X)
\]

\[
\implies
E[\gamma_m^R[K_m^* + \epsilon]\Pi(X, K_{m-}^*, K_m^* + \epsilon)] - E[C_m^P(X, K_m^* + \epsilon)] - C_m^{QR}(X) + c_k K_m \epsilon
\]

\[
> E[\gamma_m^*\Pi(X, K_{m-}^*, K_m^*)] - E[C_m^P(X, K_m^*)] - C_m^{QR}(X),
\]

for some sufficiently small \(\epsilon > 0\).
Hence, if $\gamma^R_m[K^*_m + \epsilon] > \gamma^*_m$, then it could be advantageous for the resource to increase its capacity beyond $K^*_m$. In which case, the global optimum for the problem $D$ is not honored by the solution to resource problem $\gamma^R_m - D$. The above conditions are not ruled out by any of our assumptions, and indeed can occur the revenue or gain $\Pi_0$ is sufficiently large so that it is possible for a firm to justify its marginal cost of capacity at $K_m > K^*_m$, in order to garner more of the revenue share. In general, we expect to see this behavior in profitable programs.

Conversely, suppose that that $\Pi_0$ is small, while $\beta \geq 0$, Then, the trade-off is between increasing capacity $K_m$ and incurring negative gain share (or costs) through program level delays, for that individual resource (note that this can happen even while we satisfy the participation constraints for the firms). Now, if the marginal share of the program delay penalties for $R_m$ is greater than the marginal cost of capacity $c^K_m$ for some $K^o_m < K^*_m$, then the same is true for all $K_m \geq K^o_m$. This implies that it would be not be optimal for the firm operating $R_m$ to invest beyond $K^o_m$, and therefore it is optimal to under-invest in capacity relative to $K^*_m$ - the program optimal capacity level.

A similar argument can be used in the context of problems $L$ and $\gamma^R_m - L$.

See Figures 8.1 and 8.2, and Figures 8.3 and 8.4 for an illustration of Proposition 8.2.2 to different program settings (series vs. parallel tasks). This proposition implies that program planners need to worry about situations where there are incentives for firms to overinvest as well as situations that can foster under-investment. The degree of over-investment or under-investment depends, of course, on the specific program environment; in particular it depends on how profitable the program is for the participating firms. However, there is good news (in bad) for program planners
Figure 8.1: Example program with two tasks run by two firms in series: $c_1^K = 20, c_2^K = 10; \Pi_0 = 1000; \frac{\beta_i \lambda_i}{\mu_i} = 100; i = \{1, 2\}$. The firms overinvest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
Figure 8.2: Example program with two tasks run by two firms in series: $\Pi_0 = 200; Pr(\frac{A_1}{\mu_1} = 100) = 0.25; Pr(\frac{A_1}{\mu_1} = 1000) = 0.75; Pr(\frac{A_2}{\mu_2} = 100) = 0.75; Pr(\frac{A_2}{\mu_2} = 1000) = 0.25$. The firms under-invest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
Figure 8.3: Example program with two tasks run by two firms in parallel: $c_i^K = 20, c_i^K = 10; \Pi_0 = 1000; \frac{\lambda_i}{\mu_i} = 100; i = \{1, 2\}$. The firms overinvest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
Figure 8.4: Example program with two tasks run by two firms in parallel: $\Pi_0 = 200; \Pr(\lambda_1 = 100) = 0.25; \Pr(\lambda_1 = 1000) = 0.75; \Pr(\lambda_2 = 100) = 0.75; \Pr(\lambda_2 = 1000) = 0.25$. The firms under-invest w.r.t. to the program optimal capacity investment in resource groups 1 & 2 resp.
as they look to induce investments from the resources that maximize program value. Since the culprit behind lack of coordination is the gain share mechanism, and how it shares the program gain with individual investors, it leads us to believe that by suitably modifying the gain share mechanism, it could be possible to achieve coordination of capacity investment.

Proposition 8.2.2 also suggests that it could be no more easier to coordinate blockbuster programs than the less profitable ones; since firms may have less of an incentive to invest in less profitable programs, while they may exhibit a tendency to overinvest in capacity to garner greater revenues in the blockbuster opportunities.

Further, the above discussion raises the issue of what happens when firms own a combination of resources, as opposed to the single resource ownership case discussed up until now. With some relatively expensive resources in the context of the program, and some others less expensive, it would seem that firms would prefer to optimize their share of the value jointly over their set of resources. As it turns out this scenario is also problematic w.r.t. co-ordination in general.

8.3 The Firm Level Capacity Coordination Problem.

Consider a firm \( l \) that owns all resources \( m \) such that \( \alpha_{lm} = 1 \). Suppose the program planner offers the program optimal assignment \( X \) along with a guarantee that \( K_t = K_t^*; \forall t \) s.t. \( \alpha_{lt} \neq 1 \); i.e. the planner guarantees that all of the resources \( t \) not owned by firm \( l \) will have program optimal capacity \( K_t = K_t^* \).

Suppose further that the planner offers a share \( \gamma^F_l \in \{\gamma^{IRS}_l, \gamma^{WIRS}_l, \gamma^{CRS}_l\} \) as
defined in Equation 6.18-6.20. Then, the firm \( l \) solves one of the following problems:

**(\( \gamma^F_D \): Delay Sensitive Firm Value):**

\[
EV^F_l(\gamma^F_l, X, \{K^*_j: \alpha_{lj}=0\}) = \max_{K_m: \alpha_{lm}=1} \{ E[\gamma^F_l \Pi(X, \{K_{j: \alpha_{lj}=0}\}, \{K_{j: \alpha_{lj}=1}\})] - E[C^O_R(X)] - E[C^P_l(X, \{K_{j: \alpha_{lj}=0}\}, \{K_{j: \alpha_{lj}=1}\})] - \sum_{m=1}^M \alpha_{lm} c^K_m K_m \}
\]

subject to : \( K_m \in \kappa_m, \forall m : \alpha_{lm} = 1. \)

**(\( \gamma^F_L \): Processing Time Sensitive Firm Value):**

\[
EV^F_l(\gamma^F_l, X, \{K^*_j: \alpha_{lj}=0\}) = \max_{K_m: \alpha_{lm}=1} \{ E[\gamma^F_l \Pi(X, \{K_{j: \alpha_{lj}=0}\}, \{K_{j: \alpha_{lj}=1}\})] - E[C^O_R(X)] - E[C^P_l(X, \{K_{j: \alpha_{lj}=0}\}, \{K_{j: \alpha_{lj}=1}\})] - \sum_{m=1}^M \alpha_{lm} c^K_m K_m \}
\]

subject to : \( K_m \in \kappa_m, \forall m : \alpha_{lm} = 1. \)

**Proposition 8.3.1.** \( \{K^*_m: \alpha_{lm} = 1\} \), the program optimal resource investment strategies for resources \( m \) owned by firm \( l \) under program optimization problem \( D \) (or \( L \)), may not be a globally optimal solution to the firm level problem \( \gamma^F_l - D \) (or \( \gamma^F_l - L \)) for any \( \gamma^F_l \in \{\gamma^{IRS}_l, \gamma^{WIRS}_l, \gamma^{CRS}_l\} \).

**Proof.** The proof is again by construction; moreover it is just an extension of the proof of Proposition 8.2.2. Suppose now that \( K_j = K^*_j; \forall j : \alpha_{lj} = 1; j \neq m \); in other words all resources owned by firm \( l \) are set at capacity \( K^*_j \) except for resource \( m \).
Suppose again that $K^* \in \Upsilon(X)$ is such that $K_m^*$ is the maximum over any $K_m^o$ such that $K^o$ is part of a program optimal solution in the set $\Upsilon(X)$.

Then, Proposition 8.2.2 essentially constructs a scenario for us again where increasing $K_m$ beyond $K_m^*$ could be profitable for resource $m$ for some program environment. Conversely, the same proposition shows us how some $K_m < K_m^*$ could be an optimal solution to $\gamma_l^F - D$.

In section 8.5, we present some simple mechanisms (or algorithms, based on the perspective) that are guaranteed to coordinate the capacity investments for the supply chain program. Before presenting such mechanisms, we raise the issue of whether it is at all possible for firms to come to any sort of commonly agreeable or equilibrium capacity investment solution, given some fixed assignment strategy. As it turns out, the existence of an equilibrium capacity solution across resources or firms depends critically on the structure of the program network and also on the problem parameters. For instance, the equilibrium investment levels depend on how the firms are located on the program (tree) network, and how their individual capacity investments contribute to the program outcomes, and therefore to their share of the program revenues. We illustrate this notion of equilibrium investments through this next section, conditional on a given fixed assignment strategy.

8.4 Equilibrium Behavior of Investors in the Decentralized Program Environment.

The objectives of this section are three-fold. Firstly we wish to demonstrate how the existence of equilibrium capacity investment solution depends critically on the structure of the program network. Secondly, we wish to highlight the properties of the best response functions of individual firms when faced with the capacity strategies of their partners. We provide such illustration through the simplest (and smallest) possible program network that can allow participation by multiple firms: namely a program with just two tasks. Finally, we show that equilibrium investment be-
behavior is rather fragile – if it even exists – for even the smallest networks operating under the proportional revenue sharing mechanisms discussed in the previous section.

We separate the discussion based on whether these tasks are arranged in series or in parallel, in order to compose the program. The analysis, and the equilibrium behavior of firms is remarkably different for these two fundamental types of project sub-structures. We also assume that the gain or revenue share mechanisms in use is the basic IRS mechanism, which splits the program gain in proportion to the share of each firm in the program capacity (investment). Since the operating costs are independent of the capacity investment (in our original model construction), we do not consider their impact. Finally, we assume away the budget or maximum capacity constraints defined for general program environments.

8.4.1 Tasks in series.

Consider two firms 1 and 2 each assigned a task (say 1 and 2, resp.) in a program under a deterministic environment, where tasks 1 and 2 are arranged in series within the program network. Suppose firm 1 operates resource group 1, while firm 2 operates resource group 2. The task, resource, and program parameters are defined using the notation defined in Chapter 6. Hence with capacity $K_1$ and $K_2$ invested by firms 1 and 2 respectively, the program gain is given by

$$V(K_1, K_2) = \Pi_0 - \beta_1 \frac{\lambda_1}{K_1^{\theta_1} \mu_1^e} - \beta_2 \frac{\lambda_2}{K_2^{\theta_2} \mu_2^e}. \quad (8.1)$$

We will assume for the sake of simplicity that $\theta_1 = \theta_2 = 1$. Firm $i \in \{1, 2\}$ derives a fraction $\frac{c_i K_i}{c_i K_i + c_{i-1} K_{i-1}}$ of the program gain, while investing $c_i K_i$ of the total program capacity, and therefore derives value:
\[ V_i(K_i) = \frac{c_i K_i}{c_i K_i + c_i - K_i} \left( \Pi_0 - \beta_i \frac{\lambda_i}{K_i \mu_i} - \beta_i \frac{\lambda_i}{K_i \mu_i} \right) - c_i K_i \]  
(8.2)

\[ = \frac{c_i K_i}{c_i K_i + c_i - K_i} \left( \Pi_0 - \beta_i \frac{\lambda_i}{K_i \mu_i} \right) - \frac{\beta_i c_i \lambda_i}{c_i K_i + c_i - K_i} - c_i K_i. \]  
(8.3)

For a given investment \( K_i \), the term \( \Pi_0 - \beta_i \frac{\lambda_i}{K_i \mu_i} \) represents a constant, and therefore will be denoted by \( \Pi_i \) (the maximum gain available to firm \( i \) given the partner firms’ standing investment). Hence, for an investment \( K_i \) by the other firm, the firm maximizes the value function:

\[ V_i(K_i) = \Pi_i \frac{c_i K_i}{c_i K_i + c_i - K_i} - \frac{\beta_i c_i \lambda_i}{c_i K_i + c_i - K_i} - c_i K_i. \]  
(8.4)

The value function \( V_i(K_i) \) can be shown to be concave in \( K_i \geq 0 \), and therefore the firm \( i \) derives the maximum value by investing:

\[ K^*_i = \sqrt{\frac{c_i K_i}{c_i K_i + c_i - K_i}} \left( \Pi_0 - \beta_i \frac{\lambda_i}{K_i \mu_i} \right) - \frac{\beta_i c_i \lambda_i}{c_i K_i + c_i - K_i} - c_i K_i. \]  
(8.5)

For some sensitivity analysis, note that \( K_i \) is concave function of \( K_i \) while in the limit as \( K_i \to \infty, K_i \to 0 \). Furthermore, differentiating \( K^*_i \) w.r.t. \( K_i \), we obtain:

\[ \frac{dK^*_i}{dK_i} = \frac{c_i - K_i}{2c_i^2} \Pi_0 \left( \frac{c_i K_i}{c_i K_i + c_i - K_i} \left( \Pi_0 - \beta_i \frac{\lambda_i}{K_i \mu_i} \right) - \frac{\beta_i c_i \lambda_i}{c_i K_i + c_i - K_i} - c_i K_i \right) \]  
(8.7)

Negative values of \( K_i : i \in \{1, 2\} \) are meaningless; hence we can observe that \( K_i^* \) is not necessarily monotonic in \( K_i \) in a way that is independent of the parameter set. It is however true that \( \frac{dK_i^*}{dK_i} \) is monotonically decreasing in \( K_i \). Thus, for some problem parameters, where the ratios \( \frac{c_i}{c_i^2} - \Pi_0 \), or \( \frac{\beta_i \lambda_i}{c_i \mu_i} \) are relatively small, or indeed
where the ratio $\frac{c_i - \lambda_i - \mu_i}{c_i}$ is large, there is little incentive for firm $i$ to invest any capacity whatsoever in the program as a response to the other firms’ capacity offer. In the contrasting situations, it is worthwhile for the firm to invest up-to a certain amount of capacity before the returns begin to diminish.

Next, we consider the possibility of equilibrium capacity investments. We define an equilibrium as a joint capacity solution $(K^o_i, K^o_{i-})$ where $K^o_{i-}$ maximizes $V_i(\cdot)$, while simultaneously $K^o_i$ maximizes $V_i(\cdot)$. Hence with some algebra, and according to this definition of an equilibrium capacity investment, we can compose the equilibrium capacity investment(s) from the independent solution(s) to the equations:

$$\sqrt{\frac{c_i - \lambda_i + \beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i}} = \frac{c_i - \lambda_i}{c_i} \left( \sqrt{\frac{c_i - \lambda_i + \beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i}} - \frac{c_i - \lambda_i}{c_i} K^o_i \right) + K^o_i$$

Simplifying, we obtain:

$$\frac{c_i - \lambda_i}{c_i} \left( \sqrt{\frac{c_i - \lambda_i + \beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i}} - \frac{c_i - \lambda_i}{c_i} K^o_i \right) + K^o_i$$

or the quadratic equation in $K^o_i$:

$$\sqrt{\frac{c_i - \lambda_i + \beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i} + \frac{\beta_i - \beta}{c_i \lambda_i \mu_i}} - \frac{2c_i \beta_i - \lambda_i}{c_i \lambda_i \mu_i} K^o_i = \frac{\beta_i - \lambda_i}{c_i \lambda_i \mu_i} \left( \frac{\beta_i - \lambda_i}{c_i \lambda_i \mu_i} - \frac{\beta_i c_i \lambda_i}{c_i \lambda_i \mu_i} \right) \frac{\beta_i - \lambda_i}{c_i \lambda_i \mu_i} \left( \frac{\beta_i - \lambda_i}{c_i \lambda_i \mu_i} - \frac{\beta_i c_i \lambda_i}{c_i \lambda_i \mu_i} \right)$$

We look for the non-negative roots to this above equation, and can then verify that the conditions for the equilibrium are satisfied. It is indeed possible, given the
problem parameters to have multiple equilibria, up to a maximum of four distinct equilibrium solutions. More importantly, it is possible, given the program parameters, that there is no equilibrium capacity investment solution across the two firms. This happens when the best response functions of the two firms do not intersect in the feasible non-negative orthant \( \mathbb{R}_+^2 \). Finally, note that with uncertainty in the parameters \( \Pi_0, \lambda_i, \) and \( \mu^e_i \), the analysis can be carried out in very similar fashion based on the firms maximizing expected value given well-behaved probability distribution functions governing these parameters.

### 8.4.2 Tasks in parallel.

Let us again consider two firms 1 and 2 each assigned a task (say 1 and 2, resp.) in a program under a deterministic environment, where tasks 1 and 2 are arranged in parallel to compose the program network, and where the program gain is processing time (as opposed to delay) sensitive. Suppose firm 1 operates resource group 1, while firm 2 operates resource group 2. The task, resource, and program parameters are again defined using the notation defined in Chapter 6. Hence with capacity \( K_1 \) and \( K_2 \) invested by firms 1 and 2 respectively, the program gain is given by:

\[
V(K_1, K_2) = \Pi_0 - \beta \max \left[ \frac{\lambda_1}{K_1 \mu^e_1}, \frac{\lambda_2}{K_2 \mu^e_2} \right].
\]  

(8.8)

We will again assume, for the sake of simplicity that \( \theta_1 = \theta_2 = 1 \). More importantly, unlike the previous sub-section we will consider – for clarity of analysis – only the case where \( c_i = c_{i-} = c \). The asymmetric case is analyzed in a similar fashion, and is not presented here. Firm \( i \in \{1, 2\} \) therefore derives a fraction \( \frac{K_i}{K_i + K_{i-}} \) of the program gain, while investing \( cK_i \) of the total program capacity, and therefore derives value:

\[
V_i(K_i) = \frac{K_i}{K_i + K_{i-}} \left( \Pi_0 - \beta \max \left[ \frac{\lambda_i}{K_i \mu^e_i}, \frac{\lambda_{i-}}{K_{i-} \mu^e_{i-}} \right] \right) - cK_i.
\]  

(8.9)

354
For a given investment \( K_i \), the term \( \frac{\lambda_i}{K_i - \mu_i^e} \) represents a constant, and therefore will be denoted by \( \hat{\tau}_i \) (the processing time for the other firms’ task). Hence, given an investment \( K_i \) by the other firm, the firm maximizes the value function:

\[
V_i(K_i) = \frac{K_i}{K_i + K_{i-}} \left( \Pi_0 - \beta \max \left[ \frac{\lambda_i}{K_i \mu_i^e} \hat{\tau}_i \right] \right) - cK_i. \tag{8.10}
\]

First, note that the value function \( V_i(K_i) \) can be observed to be concave in \( K_i \leq \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} \). Secondly, note that the value function for \( K_i > \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} \), is simply:

\[
V_i(K_i) = \frac{K_i}{K_i + K_{i-}} (\Pi_0 - \beta \hat{\tau}_i) - cK_i. \tag{8.11}
\]

It can therefore be verified that \( V_i(K_i) \) is again concave in \( K_i > \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} \). Furthermore, with some analysis, it can be proved that:

\[
\frac{dV_i}{dK_i |_{K_i = \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}}} \geq \frac{dV_i}{dK_i |_{K_i > \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}}} \tag{8.12}
\]

Hence, the value function \( V_i(K_i) \) is concave in \( K_i \geq 0 \). As such, we can find a maximizer for \( V_i(K_i) \). There are two cases: first, the maximizer could be in the range \( \left[ 0, \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} \right] \); alternatively, the maximizer is in the range \( \left( \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}, \infty \right) \).

In the former case – which corresponds to situations where the firm \( i \) does not see an incentive to invest at a level beyond what is necessary to match the processing time target set by the partner firm – it is fairly easy to see that the maximizer is given by:

\[
K_i^o = \min \left[ \sqrt{\frac{\Pi_0 K_{i-} + \beta \lambda_i}{c}} - K_{i-}, \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} \right]. \tag{8.13}
\]

In the latter case – which now corresponds to situations where the firm \( i \) actually benefits by investing at a level where its own processing time is now lower than the
target set by the partner firm – the maximizer is given by:

\[ K_i^o = \max \left[ \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}, \sqrt{\frac{\Pi_0 K_{i-} - \beta \frac{\lambda_i}{\mu_i^e} }{c}} - K_{i-} \right]. \quad (8.14) \]

In this latter case, firm \( i \) sees an incentive to invest beyond the target level \( \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} \), primarily due to its own share of the program gain increasing in its investment.

For some sensitivity analysis, we can again observe that \( K_i^o \) is not necessarily monotonic in \( K_{i-} \) in a way that is independent of the parameter set, while at the same time \( \frac{dK_i^o}{dK_{i-}} \) is monotonically decreasing in \( K_{i-} \) (implying concavity of \( K_i^o \) in each the two ranges of \( K_{i-} \)). We can also observe that as the load \( \frac{\lambda_i}{\mu_i^e} \) resulting from a task increases, and as the gain multiple \( \frac{\Pi_0}{c} \) increases, or indeed as the delay sensitivity \( \beta \) increases, there is greater incentive for firm \( i \) to increase the capacity investment towards its task. Conversely, an increasing load factor \( \frac{\lambda_i}{\mu_i^e} \) for the partner firm tends to depress the capacity investment by firm \( i \).

The existence of an equilibrium depends now on the continuity of the response functions defined by Equations 8.13 and 8.14 above (it also depends on the shape of the response functions for the two firms and whether they intersect in the feasible non-negative orthant). Here we argue, informally, that the response functions are indeed continuous functions. In the under-investment case, suppose for some \( K_{i-} \), the optimal response for firm \( i \) is such that:

\[ K_i^o = \sqrt{\frac{\Pi_0 K_{i-} + \beta \frac{\lambda_i}{\mu_i^e} }{c}} - K_{i-} < \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}. \quad (8.15) \]

Then in an infinitesimal interval \( (K_{i-} - \epsilon, K_{i-} + \epsilon) \) for small \( \epsilon > 0 \), the optimal response is still characterized by Equation 8.15 above.
Next, suppose for some $K_{i-}$, the optimal response for firm $i$ is now:

$$K^o_i = \sqrt{\frac{\Pi_0 K_{i-} - \beta \frac{\lambda_i}{\mu_i} K_{i-}}{c} - K_{i-}}. \quad (8.16)$$

In an infinitesimal interval $(K_{i-} - \epsilon, K_{i-} + \epsilon)$ for small $\epsilon > 0$, the optimal response is characterized by Equation 8.16.

We will show that the best response curve for firm $i$ crosses the line $K^o_i(K_{i-}) = \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}$ only once. To see this, note that if for some $K_{i-}$, if $K^o_i(K_{i-}) < \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}$, then for any greater value of $K_{i-}$, the optimal response will still remain below this line. If this assertion is true, then firm $i$ is under-investing in its capacity relative to the required target completion time, despite potential incentives to increase its own capacity contribution and thereby not only increase its gain share, but also the total program gain that is being shared by the two firms. Greater values of $K_{i-}$ now further lower the target completion time for firm $i$ (thus requiring greater investments), while simultaneously requiring correspondingly higher investments from firm $i$ to maintain its gain share ratio. Thus, there is now less incentive for firm $i$ to increase its investment, relative to its marginal capacity cost. This can also be verified by differentiating the following expression derived from Equation 8.13, and observing that the expression below is increasing in $K_{i-}$ in that range.

$$\frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-} - \left( \sqrt{\frac{\Pi_0 K_{i-} - \beta \frac{\lambda_i}{\mu_i} K_{i-}}{c} - K_{i-}} \right). \quad (8.17)$$

Conversely, if for some $K_{i-}$, if $K^o_i(K_{i-}) > \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}$, then for any lower value of $K_{i-}$, the optimal response will still remain above this line. A simple differentiation of the following expression points to this fact:

$$\sqrt{\frac{\Pi_0 K_{i-} - \beta \frac{\lambda_i}{\mu_i} K_{i-}}{c} - K_{i-}} - \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_{i-}. \quad (8.18)$$
Furthermore, this expression above is concave in $K_{i-}$, implying that the optimal response of firm $i$ is guaranteed to decrease to $\frac{\lambda_i \mu_i e_i}{\lambda_i \mu_i e_i} K_{i-}$, if indeed it is above this line for some value of $K_{i-}$, while in general as $K_{i-} \to \infty$, the best response capacity investment of the partner firm tends to zero.

The question of the continuity of $K_i^\rho(K_{i-})$ rest now only on how the optimal response by firm $i$ switches from Equation 8.14 to 8.13, with increasing $K_{i-}$. Suppose for some $K_{i-}$, it is optimal for firm $i$ to invest at a level exactly equal to $\frac{\lambda_i \mu_i e_i}{\lambda_i \mu_i e_i} K_{i-}$.

Then, this means that the value function is maximized when the completion times of both tasks are equalized. Suppose now that we increase $K_{i-}$ by some small $\epsilon > 0$; in this neighborhood, consider the difference between the value function with the completion time of $\frac{\lambda_i}{(K_{i-}+\epsilon) \mu_i e_i}$, and with a completion time of $\frac{\lambda_i}{K_i \mu_i e_i}$:

$$\frac{K_i}{K_i + K_{i-}} \beta \left( \frac{\lambda_i}{K_i \mu_i e_i} - \frac{\lambda_i}{(K_{i-}+\epsilon) \mu_i e_i} \right) = \frac{K_i}{K_i + K_{i-}} \beta \left( \frac{\lambda_i}{K_{i-} \mu_i e_i} - \frac{\lambda_i}{(K_{i-}+\epsilon) \mu_i e_i} \right)$$

Clearly, this difference above $\to 0$, as $\epsilon \to 0$. Since the difference in value function is infinitesimal with infinitesimal increments in $\epsilon$, the maximizer $K_i^\rho(K_{i-})$ is continuous in such a neighborhood. Hence, we have proved that $K_i^\rho(K_{i-})$ is continuous in $K_{i-} \geq 0$.

The importance of such behavior of the best response function is that we can now entertain the possibility of equilibrium capacity investment solutions. Such equilibrium solutions, similar to the case when tasks are arranged in series, will occur when the best response functions of the two firms intersect. Once again, equilibrium solutions are not guaranteed, since best response functions need not intersect anywhere in the feasible non-negative orthant $\mathbb{R}_+^2$. If the best response functions do intersect, there are two scenarios corresponding to which of the two best response functions – given by Equation 8.14 and Equation 8.13 – of either firm intersect.
Thus in the first scenario, one firm over-invests (relative to the target completion time, this time in equilibrium) in accordance with Equation 8.14, and the other firm in equilibrium under-invests in equilibrium – as described by Equation 8.13 – relative to the completion time target set by the first firm. The second scenario corresponds to the diametrically opposite case. (There is no equilibrium possible, where both firms simultaneously under-invest or over-invest relative to the other.) For each of the other two scenarios: we would have to solve a pair of quadratic equations to determine whether one or more equilibria exist corresponding that scenario.

The second scenario corresponds to the solution of the equations:

\[
K_i^o = \sqrt{\frac{\Pi_0 \left( \sqrt{\frac{\Pi_0 K_i^o - \beta \lambda_i}{\mu_i^o}} - K_i^o \right) + \beta \lambda_i}{c} - \sqrt{\frac{\Pi_0 K_i^o - \beta \lambda_i}{\mu_i^o}} + K_i^o},
\]

or simply:

\[
\Pi_0 \sqrt{\frac{\Pi_0 K_i^o - \beta \lambda_i}{c}} = 2 \Pi_0 K_i^o - 2 \beta \lambda_i \mu_i^o
\] (8.19)

and for the partner firm that is overinvesting, the quadratic equation:

\[
\Pi_0 \left( \sqrt{\frac{\Pi_0 K_{i-}^o + \beta \lambda_i}{\mu_{i-}^o}} - K_{i-}^o \right) + \beta \lambda_i \mu_{i-}^o = \sqrt{\frac{\Pi_0 K_{i-}^o + \beta \lambda_i}{\mu_{i-}^o}}.
\] (8.20)

We need both Equations 8.19 and 8.20 to have non-negative solutions in order for an equilibrium of this type to exist. Depending on how many non-negative roots there are to each of the two equations, we could have as many as four equilibrium solutions of this type where the first firm under-invests relative to the other firm. Of course, it is entirely possible that there is no equilibrium, if neither of these equations above yield non-negative roots.
The first scenario calls for the solution of the following pair of quadratic equations:

\[
\Pi_0 \left( \sqrt{\frac{\Pi_0 K_i^o + \beta \frac{\lambda_i}{\mu_i}}{c} - K_i^o} \right) - \frac{\beta \frac{\lambda_i}{\mu_i}}{c} = \frac{\Pi_0 K_i^o + \beta \frac{\lambda_i}{\mu_i}}{c}; \quad (8.21)
\]

\[
\Pi_0 \left( \frac{\sqrt{\Pi_0 K_i^- + \beta \frac{\lambda_i}{\mu_i}}}{c} - K_i^- \right) - \frac{\beta \frac{\lambda_i}{\mu_i}}{c} = \sqrt{\frac{\Pi_0 K_i^- + \beta \frac{\lambda_i}{\mu_i}}{c}}. \quad (8.22)
\]

The second scenario could also independently yield up to four distinct equilibrium capacity investment solutions. There is a third scenario that captures a situation where both firms invest at a level where their individual task completion times are identical. For the last scenario to be realized, an equilibrium solution obtained from either of the above pairs of quadratic equations would necessarily have to be on the line \( K_i^o = \frac{\lambda_i \mu_i^e}{\lambda_i - \mu_i^e} K_i^- \).

With uncertain parameters \( \lambda_i \) or \( \mu_i^e \), the analysis now becomes more complex, since the value function for the firms may not be as well-behaved or yield to smooth functions that can yield smooth or continuous best response functions. One would have to consider specific distributions to examine the existence of equilibria. Further, for more complex networks with more than two tasks, the approach to deriving the best response functions and solving for equilibria is similar - only with larger networks the best response function would have to be pieced together over many more intervals and special cases. In summary, proving the existence of equilibria for more complex networks, under our specific modeling assumptions, is quite a tedious exercise. Nevertheless, through the section below, we argue that such equilibria even when they exist can be inefficient relative to the centralized optimal solutions. As a consequence, relying on the existence of equilibria to manage decentralized program networks is not an attractive alternative, at least from a normative point of view. Rather, as we set out to do in the subsequent section, it is important to devise some mechanisms that can either eliminate or minimize the inefficiencies in decentralized equilibria, or alternatively coordinate the capacity investments in the program.

360
network so that the decentralized and centralized optimal solutions are identical in performance. Fortunately, we are able to devise mechanisms that achieve both of these objectives.

8.4.3 Illustration of best response functions and decentralized equilibria.

Consider a project of two tasks in series: each assigned to an independent firm, with the parameters: \( c_1^K = c_2^K = 10; \Pi_0 = 1000; \frac{\beta \lambda_i}{\mu_i} = 100; i = \{1, 2\} \). Figure 8.5 evaluates the value function \( V_1(K_1) \) for the first firm at various levels of capacity committed by the partner firm towards the other task. We can observe that value function for firm 1 is concave in \( K_1 \). Furthermore its optimal capacity response is also concave in \( K_2 \) in its positive range. Since the problem is constructed with parameters that are symmetric across the firms, the best response capacity curve is identical for the second firm. Hence, it is possible to observe regions of the feasible range \( \mathbb{R}^2_+ \) where the capacity equilibria can exist. In this case, \((0,0)\) is readily an equilibrium, but there is also exactly one equilibrium in the positive region where the best response curves intersect. This can be observed by considering the best response curves of firms 1 and 2 together through Figure 8.5(b). The \( x - \) axis measures the best response of firm 1 w.r.t. to the capacity commitment of firm 2 along the \( y - \) axis.

Each of the associated figures starting with 8.5 also shows the location of the centralized optimal solution within the \( \mathbb{R}^2_+ \) (through the brown filled circle). We can easily observe how distant the positive equilibria can be in relation to the centralized optimal, but it is also possible to measure the inefficiency of the equilibrium solution w.r.t. to the centrally optimal solution.

Figures 8.6 and 8.7 together show the impact of increasing the load on one arc alone, while preserving the load on the partner firm. For the firm experiencing the higher load Figure 8.6(a) shows that the value is first rapidly increasing but then decreasing in the capacity commitment of the partner firm. Figure 8.7(a) shows that the value to the other firm is generally decreasing as the load on the partner firm
is increased. Figures 8.6(b) and 8.7(b) show in this case that the positive equilibria are biased with the greater capacity for the firm with the higher load. In this sense, the positive equilibria are fair in the sense of allocating greater capacity to the firm experiencing greater loads.

Figures 8.8 and 8.9 in turn show how the cost of capacity dampens the capacity commitment of the firm experiencing higher costs, but surprisingly does not have any significant impact on the partner firm whose costs are unchanged. The intuition here is that in a series system, each firm has independent ability to impact the program delays, and therefore its time-cost trade-offs are to a certain extent immune to the costs of the partner firm. Reduced program revenues $\Pi_0$, on the other hand, impacts both firms equally, and reduces both the value derived by each firm, as well as dampens the overall capacity committed by each firm in equilibrium (and overall). This is seen through Figure 8.10.

Next consider a program with parallel arcs with parameters: $c^K_1 = c^K_2 = 10; \Pi_0 = 1000; \lambda_i / \mu_i = 100; i = \{1, 2\}; \beta = 1$. Figure 8.13 illustrates the value function for one (or both) firms along with its best response capacity curve. The value function is generally decreasing in the capacity commitment of the partner firm when the program is lucrative: this is the result of the proportional gain share mechanism. The best response capacity curve is concave in a piecewise manner (but still continuous) based on which firm is responsible for the overall program delay. However, as shown in the preceding analysis, the system (through the governing gain share mechanism) still allows for positive equilibria to develop. Greater penalties for delays simply highlight the different value functions (and resulting best response curves) in each piece wise region, as seen in Figure 8.14. Increasing the load several times for one firm makes that firm responsible for the delays, and hence the value function (and the resulting best response curve) is smooth and concave, as seen in Figure 8.15. Figure 8.16 shows the opposite impact on one firm when the other firm is responsible for carrying the greater load. In this case the best response curve for a firm changes its
trajectory only after the partner firm has committed sufficient capacity so as not to be responsible for the overall program delay.

When a resource becomes expensive, a firm has an incentive to contribute less of that resource to the program. This is true even when the positive equilibria exist, as can be seen through Figures 8.17 and 8.18. Finally, lower program gain $\Pi_0$ simply reduces the incentive for firms to increase their capacity contribution – as seen in Figure 8.19 – leading to correspondingly diminished equilibrium solutions.

8.4.4 Efficiency of equilibrium investments, and equilibria in general program networks.

With parallel tasks, the investments of the two firms are complementary, and there is greater interdependence in the investments. However, it is important to recognize
Figure 8.6: Increased load in series for resource 1: $\frac{\beta_1 \lambda_1}{\mu_1} = 1000$.

Figure 8.7: Increased load in series for resource 2: $\frac{\beta_2 \lambda_2}{\mu_2} = 1000$. 

364
(a) $V_1(K_1)$

Figure 8.8: Increased cost in series of resource 1: $c_1^K = 20$.

(b) $K_1^q(K_2)$.

(a) $V_1(K_1)$

Figure 8.9: Increased cost in series of resource 2: $c_2^K = 20$.

(b) $K_1^q(K_2)$.

(a) $V_1(K_1)$

Figure 8.10: Impact of lower program gain in series network: $\Pi_0 = 500$. 
Figure 8.11: Impact of uncertain loads in series network: $Pr\left(\frac{\lambda_1}{\mu_1} = 100\right) = 0.25; Pr\left(\frac{\lambda_1}{\mu_1} = 1000\right) = 0.75; Pr\left(\frac{\lambda_2}{\mu_2} = 100\right) = 0.75; Pr\left(\frac{\lambda_2}{\mu_2} = 1000\right) = 0.25.$

Figure 8.12: Impact of uncertain loads and lower revenues in series network: $\Pi_0 = 200; Pr\left(\frac{\lambda_1}{\mu_1} = 100\right) = 0.25; Pr\left(\frac{\lambda_1}{\mu_1} = 1000\right) = 0.75; Pr\left(\frac{\lambda_2}{\mu_2} = 100\right) = 0.75; Pr\left(\frac{\lambda_2}{\mu_2} = 1000\right) = 0.25.$
that any centralized solution to the program with two tasks in parallel as discussed above, necessarily equalizes the processing times of the two tasks. Therefore, when there are two tasks in parallel, the only candidates for an efficient equilibrium solution are those on the line \( K^c_i = \frac{\lambda_i \mu_i}{\lambda_i - \mu_i} K^o_i \). Any other equilibrium solution in the non-negative orthant is by construction, inefficient as compared to the centralized optimal solution. Furthermore, even if they are on this "capacity efficient" line, they may not be efficient overall in terms of generating the same program gain and/or value as the centralized optimal solution. The arguments for this statement are very similar to Proposition 8.2.2. The primary cause of inefficiency of equilibrium solutions is the competitive interaction between firms responsible for parallel tasks: each firm is interested in maximizing its own share of the program gain, even though its capacity choices could hurt the program overall.
Figure 8.14: Increased penalty for delays in a parallel network; $\beta = 5$.

Figure 8.15: Increased load for resource 1 in parallel; $\frac{\lambda_1}{\mu_1} = 1000$. 
Figure 8.16: Increased load for resource 2 in parallel: $\frac{\lambda_2}{\mu_2} = 1000$.

Figure 8.17: Increased cost of resource 1 in parallel: $c_1^K = 20$.

Figure 8.18: Increased cost of resource 2 in parallel: $c_2^K = 20$. 
Figure 8.19: Impact of lower program gain in a parallel network: $\Pi_0 = 500$.

Figure 8.20: Impact of uncertain loads in a parallel network: $Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25$.

When tasks are arranged in series, the capacity investments into these tasks in some sense have a substitutive effect, while still contributing to the overall program gain. An efficient equilibrium solution matches the centralized solution in terms of program value. However, as we have shown in Proposition 8.2.2, for general program structures there is no guarantee that a centrally optimal capacity investment by a firm will elicit a best response by the partner firm that is also centrally optimal. While such a scenario is not ruled out by our analysis (for all problem parameters), for many environments the equilibrium solutions will be inefficient purely based on the competitive behavior of firms as they vie for their share of the program gain.
As noted earlier, the analysis in this section has been limited to the simplest form of interaction between two firms in a two-task program network. The general approach to analyzing the equilibrium behavior would be the same for more complex networks with multiple tasks and multiple agents, operating under uncertain environments. Characterizing the best response functions becomes much harder for more than two firms, or for general tree networks; and therefore so does the search for feasible equilibrium investment solutions.

Finally, as seen in Figures 8.11, 8.12, 8.20, 8.21, for the simpler networks studied for the purpose of illustration, uncertainty in the program environment, and in particular in the task loads does not appear to fundamentally alter the insights regarding how equilibrium solutions can develop from the collaborative interactions between firms. Program risk also does not seem to have any particular bearing, at
least from the examples we study, on the degree of inefficiency of the equilibrium solutions relative to the centralized optimal. Firms seem to be as likely to over-invest or under-invest relative to the centralized optimal, independent of the uncertainty in task loads: the primary driver for such inefficient behavior being the proportional gain share mechanism. Given that equilibrium solutions have been shown to be inefficient in general, our interest – from a normative standpoint – is to construct mechanisms that will induce firms to invest in accordance with the centrally optimal solution(s). This is the objective of the next two sections.

8.5 Resource Level Capacity Investment Coordination Mechanisms.

8.5.1 Why does proportional gain share fail to coordinate?

The main issue, as we mention earlier, with gain share mechanisms is that it presents the supplier or partner firms with an incentive to either overinvest, or under-invest in capacity relative to the needs or what is optimal for the program as a whole. Firms do this because they maximize their own value function, and firms will continue to invest capacity, as long as it is marginally profitable in terms of the additional gain share accrued by the firm; conversely firms may choose to stop investing below the program optimal solution, when it is no longer profitable, in a marginal sense, to do so. Thus, if it is marginally profitable to increment capacity beyond the program optimal, the gain share regime breaks down as a coordination mechanism. Conversely, if the value function for the firm decreases at some level below the program optimal, the gain share mechanism again fails in its purpose.

Hence, one work-around, at least in concept, is for the program planner to “switch off the tap”, figuratively speaking; and alter the gain share mechanism so that there are no incremental gains to be achieved by firms beyond the program optimal capacity. We present exactly such a restructured gain share mechanism, and prove that each of the IRS, WIRS, and CRS mechanisms achieves coordination with this “clause” included. However, this solves only part of the problem impeding
coordination— that of over-investment. There is also the issue of non-concavity of the resource or firm level capacity problems as discussed in Proposition 8.2.1, coupled with the possibility of the firm value function decreasing before the program optimal capacity level is reached.

For resolving the problem of possible under-investment, or for that matter that of a decreasing value function, we need to contrive for the firm level value function to be at least non-decreasing in its capacity investment until we achieve the program optimal level. When combined with a constraint on the gain share, we would modify the value function to be at least quasi-concave, and we would be ensured that the decentralized capacity decision problems would not have local maxima.

Achieving this redefinition of the firm’s value function can be done in multiple ways; we discuss three of them below:

1. One way is to ensure minimum gains from capacity investment at levels below the program optimal level; so that below $K^o_m$, the firm never has an incentive to stop incrementing capacity. We call this the guaranteed revenue or capacity insurance model of coordination.

2. The second mechanism involves setting minimum capacity levels for the firms so that they never encounter a range where the value function is decreasing in resource capacity, before the program optimal level is reached; we call this the minimum capacity requirement model of coordination. The real question here is at what level should the program planner specify the minimum capacity; the issue with this approach is that if the firms were amenable to such a requirement, it would make actually make sense for the program planner to specify the program optimal capacity as the minimum requirement.

3. A third approach is to levy a charge with the firm for investing at levels below the program optimal; this penalty could reflect the expected loss to the overall program value, relative to the program optimal, in case a firm decides to under-invest at a local maximum. Of course the belief is that introducing
this additional penalty will induce the firm to invest at program optimal levels, so at the end, there is no transfer payment. We call this the noncompliance penalty model of coordination. From an implementation point of view, this would seem the most appealing mechanism to ensure global maxima for the firms’ capacity problem.

The guaranteed revenue model works if the program is “sufficiently” profitable; in particular if the value of the resource to the program is greater than its capacity cost, then the program could ensure marginal gains to the owning firm at least equal to the marginal cost of capacity. The program would do this at all investment levels below and approaching the program optimal; so at investments below the program optimal, the firm never experiences a marginal loss from incrementing capacity.

One problem with this guaranteed revenue approach is that while in expectation there is no difference in the gains derived by the firm, for any given scenario it is possible that the program is actually promising gains that are non-existent. For example, while the program planner has guaranteed revenues to the firm at least equal to its capacity cost, for a given scenario it could be that the net revenue for the program is be less than the capacity cost incurred by the firm. This would therefore make the mechanism infeasible for a given realization of the demand $\lambda$ or resource productivity $\mu$, or for a given realization of program revenue $\Pi_0$.

Another issue is that the firm could derive higher value from the implementation of the guaranteed revenue model, when compared to its own share under the original proportional gain share regime. Where will this difference in value be compensated from? A possible implementation scheme would be for another firm in the partnership (or perhaps a collective of firms) to deduct this difference from their own share of the program value. This may be feasible in some cases, but may not be so in other situations; so the guaranteed revenue model has to be implemented selectively. Thus, this issue of some firms gaining profits at other firms’ expense needs to be considered more carefully.
The minimum capacity requirement is certainly feasible for any scenario the program could encounter; however this mechanism does not agree very well with the decentralized approach to capacity determination.

The non-compliance penalty model works within the decentralized framework, and lets the individual firms make the tradeoff between non-compliance penalty costs and increased gain share from capacity increments beyond the program optimal level. The key is to set the penalty costs in a way that coordinates the decentralized capacity decisions. As we will prove, setting the penalty cost equal to the positive difference between the expected program gain at the program optimal capacity solution, and the expected program gain at any capacity level chosen by the firm leads to a firm value function which coordinates the supply chain. In other words, the firm compensates the program an amount equal to the loss in gain for the program from its decision to under-invest in relation to the program optimal capacity. There is no loss in program gain from over-investment by the firm, and the penalties are restricted to the under-investment region; the rationale being that the cap on the gain share separately addresses the over-investment issue.

In the next subsection, we define these coordination mechanisms (and their combinations) more formally, and prove that every such class of mechanisms when properly structured achieves the coordination of capacity investment.

8.5.2 Restructuring the resource level gain share mechanism.

Recall that the first adjustment to the firm’s resource level value function involves modifying the maximum gain share that can be achieved. The second adjustment involves modifying the gain function for the firm or resource using one of the approaches just discussed. We formalize these two adjustments through the following proposition.

To start with, however, we define truncated gain share mechanisms corresponding
to the IRS, WIRS, and CRS mechanisms defined in section 6.2.1. to reflect these adjustments. Suppose that the program planner offers assignment $X$, and that $K^*$ is an optimal solution to problem $P$ that the planner wants to achieve through co-ordination.

1. With the truncated investment risk sharing ($TIRS$) mechanism:

$$\gamma^TIRS_m = \min \left[ \frac{c^K_m K_m}{c^{K'} K^*}, \frac{c^K_m K^*}{c^{K'} K^*} \right]$$  \hspace{1cm} (8.23)

2. In the truncated work and investment risk sharing ($WIRS$) mechanism:

$$\gamma^{TWIRS}_m = \min \left[ \frac{C^{OR}(X) + c^K_m K_m}{C^{O}(X) + c^{K'} K^*}, \frac{C^{OR}(X) + c^K_m K^*}{C^{O}(X) + c^{K'} K^*} \right]$$  \hspace{1cm} (8.24)

3. Finally, with truncated comprehensive risk sharing:

$$\gamma^{TCRS}_m = \min \left[ \frac{C^{OR}(X) + C^{P}(X, K_m) + c^K_m K_m}{C^{O}(X) + C^{P}(X, K^*_m, K_m) + c^K_m K_m}, \frac{C^{OR}(X) + C^{P}(X, K^*_m) + c^K_m K^*}{C^{O}(X) + C^{P}(X, K^*) + c^K_m K^*} \right]$$  \hspace{1cm} (8.25)

In order to ensure that the firms’ capacity problems are amenable to solution (i.e. the value functions avoid local optimal), we also redefine the gain function (at the resource level) for any $\gamma^R_m \in \{\gamma^TIRS_m, \gamma^{TWIRS}_m, \gamma^{TCRS}_m\}$, using the approaches described in the following three subsections.

8.5.3 Guaranteed revenue or capacity insurance model at the resource level.

Here, the program planner insures the capacity investments of the firm up-to an amount $c^K_m K^*_m$. Beyond that amount, the firm has no guarantee that the capacity investments will yield any fixed gain. Hence, the gain function of the firm for resource
$m$ is redefined as:

$$\Pi_m^o(X, K_{m-}, K_m) = \max [\gamma_m^R \Pi(X, K_{m-}, K_m), c_m^K \min[K_m, K_m^*]] ; \quad (8.27)$$

$$\hat{\Pi}_m^o(X, K_{m-}, K_m) = \max [\gamma_m^R \hat{\Pi}(X, K_{m-}, K_m), c_m^K \min[K_m, K_m^*]] . \quad (8.28)$$

There is a caveat here: if $E[\gamma_m^R \Pi(X, K_m^*)] \leq c_m^K K_m^*$, that is when the expected gain for firm from investing at the program optimal $K_m^*$ is less than the invested capacity cost we do not offer the modified gain function above, but instead maximize the value function in problems $\gamma^R - D$ or $\gamma^R - L$ from Section 8.2. If however, $E[\gamma_m^R \Pi(X, K_m^*)] \geq c_m^K K_m^*$, we proceed to redefine the resource value function as below:

$$V_m^o(X, K_{m-}, K_m) = \Pi_m^o(X, K_{m-}, K_m) - C_m^O(X) - C_m^P(X, K_m) - c_m^K K_m; \quad (8.29)$$

$$\hat{V}_m^o(X, K_{m-}, K_m) = \hat{\Pi}_m^o(X, K_{m-}, K_m) - C_m^O(X) - C_m^P(X, K_m) - c_m^K K_m. \quad (8.30)$$

**Proposition 8.5.1.** Let $K^* \in \mathcal{Y}(X)$. Then,

1. $K_m^*$ maximizes $E[V_m^o(X, K_{m-}, K_m)]$ and $E[\hat{V}_m^o(X, K_{m-}, K_m)]$;

2. For any $\gamma_m^R \in \{\gamma_m^{TIRS}, \gamma_m^{TWIRS}, \gamma_m^{TCRS}\}$, $E[V_m^o(X, K_{m-}, K_m)]$ and $E[\hat{V}_m^o(X, K_{m-}, K_m)]$ are both quasi-concave in $K_m$.

3. For any $\gamma_m^R \in \{\gamma_m^{TIRS}, \gamma_m^{TWIRS}, \gamma_m^{TCRS}\}$, the centralized optimal solution $K^*$ also represents an equilibrium capacity investment solution.

**Proof.** $V_m^o(X, K_{m-}, K_m)$ and $\hat{V}_m^o(X, K_{m-}, K_m)$ are both non-decreasing in $K_m \leq K_m^*$. To see this, note that the penalty cost $C_m^P(X, K_m)$ is non-increasing in $K_m$ while we are ensured a gain of at least $C_m^K$ for $K_m \leq K_m^*$. Meanwhile the expected operating cost given $X$ is independent of $K_m$. Hence, $V_m^o(X, K_{m-}, K_m)$ and therefore $E[V_m^o(X, K_{m-}, K_m)]$ are non-decreasing in $K \leq K_m^*$.

Now, beyond $K_m^*$, $V_m^o(X, K_{m-}, K_m)$ is non-increasing in $K_m$, since for any $\gamma_m^R \in \{\gamma_m^{TIRS}, \gamma_m^{TWIRS}, \gamma_m^{TCRS}\}$, we have the firm gain function

$$\gamma_m^R \Pi(X, K_{m-}, K_m) = \gamma_m^{R*} \Pi(X, K_{m-}, K_m).$$

377
Note that this function is indeed non-decreasing in $K_m$, but the overall value function is non-increasing beyond $K^*_m$. This can be shown by contradiction. Suppose this were not true, and that the value function:

$$E[V^o_m(X, K^*_m, K_m)] > E[V^o_m(X, K^*_m, K_m)].$$

$$\implies$$

$$E[\Pi^o_m(X, K^*_m, K_m)] - E[C^O_m(X)] - E[C^P_m(X, K^*_m, K_m)] - c^K_m K_m$$

$$\implies$$

$$E[\Pi(X, K^*_m, K_m)] - E[C^O_m(X)] - E[C^P_m(X, K^*_m, K_m)] - c^K_m K_m$$

$$\implies$$

$$E[\Pi(X, K^*_m, K_m)] - E[C^O_m(X)] - E[C^P_m(X, K^*_m, K_m)] - c^K_m K^*_m;$$

which is a contradiction, because $K^*$ is the program optimal capacity vector corresponding to $X$.

Thus, $V^o_m(X, K^*_m, K_m)$ is a non-decreasing function up to $K^*_m$ and is non-increasing beyond that; implying that it is quasi-concave. Furthermore the function achieves its maximum at $K^*_m$, so we have proved that any mechanism $\gamma^R_m \in \{\gamma^{TIRS}_m, \gamma^{TWRS}_m, \gamma^{TCRS}_m\}$ co-ordinates the capacity investment for resource group $m$ from the program planner’s perspective.

$$\square$$

See Figures 8.22 and 8.23 for an illustration of how the adjustments made to the IRS regime can coordinate the capacity investment of a resource. Note that the guaranteed revenue (or profit) could actually result in a firm gaining more than its share of the overall profits than defined under the $\{TIRS, TWRS, TCRS\}$ regimes (this can be seen through under-investment scenario within Figure 8.23). Hence, this modified gain share mechanism may be perceived as being biased towards under-investing firms who can now garner a higher share of the program gains relative to other firms who may have genuine incentives to invest either at or above the program optimal levels. However, the guaranteed revenue (or capacity insurance) mechanism
Figure 8.22: Example program with two tasks run by two over-investing firms in series: $c^K_1 = 20, c^K_2 = 10; \Pi_0 = 1000; \frac{\beta_i \lambda_i}{\mu_i} = 100; i = \{1, 2\}$. The truncated gain sharing mechanism coupled with either the guaranteed capacity investments, or with non-compliance penalties, ensures that the centralized optimal solution is also a decentralized equilibrium.
Figure 8.23: Example program with two tasks run by two under-investing firms in series: $\Pi_0 = 200; \Pr(\frac{\lambda_1}{\mu_1} = 100) = 0.25; \Pr(\frac{\lambda_1}{\mu_1} = 1000) = 0.75; \Pr(\frac{\lambda_2}{\mu_2} = 100) = 0.75; \Pr(\frac{\lambda_2}{\mu_2} = 1000) = 0.25$. The truncated gain sharing mechanism coupled with the guaranteed capacity investments.
is still feasible when the capacity costs of the under-investing firm are small relative to firms who may be investing and earning a much higher share of the program profits. As such, in the interests of the program profits, these dominant firms may be quite willing to insure the capacity costs of less solvent firms in the partnership. Nevertheless, it is useful to know that one can deploy the guaranteed revenue or capacity insurance model in select situations.

8.5.4 Non-compliance penalty model at the resource level.

In this model, the value function for the firm from its investment in resource group \( m \) with \( \gamma^R_m \in \{\gamma^TIRS_m, \gamma^{TWIRS}_m, \gamma^{TCRS}_m\} \), is given by:

\[
V^\circ_m(X, K^*_m, K_m) = \gamma^R_m \Pi_m(X, K^*_m, K_m) - C^Q_m(X) - C^P_m(X, K_m) - \epsilon^K_m K_m - \\
(\Pi(X, K^*_m, K_m) - \Pi(X, K^*_m, K_m))^+;
\]

\[
\hat{V}^\circ_m(X, K^*_m, K_m) = \gamma^R_m \hat{\Pi}_m(X, K^*_m, K_m) - C^Q_m(X) - C^P_m(X, K_m) - \epsilon^K_m K_m - \\
(\hat{\Pi}(X, K^*_m, K_m) - \hat{\Pi}(X, K^*_m, K_m))^+;
\]

Thus, for \( K_m \leq K^*_m \), we have:

\[
E[V^\circ_m(X, K^*_m, K_m)] = \\
E[(1 + \gamma^R_m)\Pi(X, K^*_m, K_m)] - E[C^Q_m(X)] - E[C^P_m(X, K_m)] - \epsilon^K_m K_m \tag{8.31}
\]

and similarly,

\[
E[\hat{V}^\circ_m(X, K^*_m, K_m)] = \\
E[(1 + \gamma^R_m)\hat{\Pi}(X, K^*_m, K_m)] - E[C^Q_m(X)] - E[C^P_m(X, K_m)] - \epsilon^K_m K_m \tag{8.32}
\]

It can be verified that for \( K_m \leq K^*_m \) the expected value functions above are non-decreasing in \( K_m \). Furthermore, for \( K_m > K^*_m \), the function is non-increasing.
in $K_m$ which implies that the functions are each quasi-concave in $K_m$. Hence, $K^*_m$ is a maximizer for the value functions in Equations 8.32 and 8.31. The proposition below states and proves this result more formally.

**Proposition 8.5.2.** Let $(K^*, X) \in \Upsilon$. Then,

1. $K^*_m$ maximizes $E[V_m^o(X, K^*_{m-}, K_m)]$ and $E[\hat{V}_m^o(X, K^*_{m-}, K_m)]$ as defined in Equations 8.32 and 8.31, resp.;

2. For any $\gamma^R_m \in \{\gamma^{TIRS}_m, \gamma^{TWIRS}_m, \gamma^{TCRS}_m\}$, $E[V_m^o(X, K^*_{m-}, K_m)]$ and $E[\hat{V}_m^o(X, K^*_{m-}, K_m)]$ are both quasi-concave in $K_m$.

3. For any $\gamma^R_m \in \{\gamma^{TIRS}_m, \gamma^{TWIRS}_m, \gamma^{TCRS}_m\}$, the centralized optimal solution $K^*$ also represents an equilibrium capacity investment solution.

**Proof.** The proof is similar to that of Proposition 8.5.1. We only provide the proof for the processing time sensitive value function; the proof is identical for the delay sensitive model.

Note that $V_m^o(X, K^*_{m-}, K_m)$ and $\hat{V}_m^o(X, K^*_{m-}, K_m)$ are both non-decreasing in $K_m \leq K^*_m$. To see this, we can write the value function as:

$$E[V_m^o(X, K^*_{m-}, K_m)] = \left( E[\Pi(X, K^*_{m-}, K_m)] \right)^{-1} - E[C^{OR}_m(X)] - E[C^p_m(X, K_m)] - c_m K_m$$

$$- E[\Pi(X, K^*_{m-}, K_m)] + E[\gamma^R_m \Pi(X, K^*_{m-}, K_m)];$$

The terms in the bracket above constitute a concave, non-decreasing function in $K_m \leq K^*_m$ from Proposition 7.2.9. The remaining terms are non-decreasing in $K_m$. Hence, $V_m^o(X, K^*_{m-}, K_m)$ and therefore $E[V_m^o(X, K^*_{m-}, K_m)]$ are non-decreasing, and at least quasi-concave in $K \leq K^*_m$.

Now, beyond $K^*_m$, $V_m^o(X, K^*_{m-}, K_m)$ is non-increasing in $K_m$, since for any $\gamma^R_m \in \{\gamma^{TIRS}_m, \gamma^{TWIRS}_m, \gamma^{TCRS}_m\}$, we have the firm gain function

$$\gamma^R_m \Pi(X, K^*_{m-}, K_m) = \gamma^R_m \Pi(X, K^*_{m-}, K_m).$$
Note that this function is indeed non-decreasing in $K_m$, but the overall value function is non-increasing beyond $K^*_m$. This can be shown by contradiction. Suppose this were not true, and that the value function:

$$E[V_m(X, K^*_m - K_m)] > E[V_m(X, K^*_m - K_m)] \implies$$

$$E[\Pi_m(X, K^*_m - K_m)] - E[C^R_m(X)] - E[C^P_m(X, K^*_m)] - c^K_m K_m$$

$$> E[\Pi_m(X, K^*_m - K_m)] - E[C^R_m(X)] - E[C^P_m(X, K^*_m)] - c^K_m K^*_m$$

$$\implies$$

$$E[\Pi_m(X, K^*_m - K_m)] - E[C^P_m(X, K^*_m)] - c^K_m K^*_m$$

$$> E[\Pi(X, K^*_m - K_m)] - E[C^P_m(X, K^*_m)] - c^K_m K^*_m$$

$$\implies$$

$$E[\Pi(X, K^*_m - K_m)] - E[C^P_m(X, K^*_m)] - c^K_m K^*_m$$

$$> E[\Pi(X, K^*_m - K_m)] - E[C^P_m(X, K^*_m)] - c^K_m K^*_m$$

which is a contradiction, because $K^*_m$ is the program optimal capacity vector corresponding to $X$.

\[\square\]

See Figures 8.24 and 8.25 for an illustration of how the non-compliance penalty model combined with the TIRS gain share regime can yield a decentralized equilibrium solution that is identical to the centralized optimal capacity investment.

8.6 Firm Level Capacity investment coordination mechanisms.

8.6.1 Restructuring the firm level gain share mechanism.

For the firm level coordination problem, we need to ensure a firm $l$ derives revenue share $\gamma^R_m \in \{\gamma^TIRS_m, \gamma^{TWIRS}_m, \gamma^{TCRS}_m\}$, separately, and for each individual resource $m$ that it owns, and further modify the firm level value function in a similar fashion as with the one dimensional coordination problem. In this subsection, we provide
Figure 8.24: Example program with two tasks run by two over-investing firms in parallel: $c_1^K = 20, c_2^K = 10; \Pi_0 = 1000; \frac{\beta \Delta}{\mu_i} = 100; i = \{1, 2\}$. The truncated gain sharing mechanism coupled with non-compliance penalties, ensures the firms' decentralized investments are in equilibrium and identical to the centralized solution.
(a) $V(K_1, K_2) \text{ vs. } V_1^{TIRS-G}(K_1, K_2^*) \text{ vs. } V_1^{TIRS-N}(K_1, K_2)$.

(b) $V(K_1^*, K_2) \text{ vs. } V_2^{TIRS-G}(K_1^*, K_2) \text{ vs. } V_2^{TIRS-N}(K_1^*, K_2)$.

Figure 8.25: Example program with two tasks run by two under-investing firms in series: \( T_0 = 200; \ Pr\left(\frac{\lambda_1}{\mu_1} = 100\right) = 0.25; \ Pr\left(\frac{\lambda_1}{\mu_1} = 1000\right) = 0.75; \ Pr\left(\frac{\lambda_2}{\mu_2} = 100\right) = 0.75; \ Pr\left(\frac{\lambda_2}{\mu_3} = 1000\right) = 0.25. \) The truncated gain sharing mechanism coupled with guaranteed capacity investments, or with non-compliance penalties.
some details on how the properties of the resource level capacity investment problem translate to the firm-level, multi-resource case. Furthermore, we show how the mechanisms and value function modifications we defined for the resource level can still coordinate investments at the firm level.

Consider a firm \( l \) that owns all resources \( m \) such that \( \alpha_{lm} = 1 \). Suppose the program planner offers a fixed assignment strategy \( X \) along with a guarantee that \( K_j = K_j^* \); \( \forall t \) s.t. \( \alpha_{lj} = 0 \); that is, the planner guarantees that all of the resources \( j \) not owned by firm \( l \) will have program optimal capacity \( K_j = K_j^* \).

Suppose further that the planner offers a share \( \gamma_l^F \in \{ \gamma_l^{IRS}, \gamma_l^{WIRS}, \gamma_l^{CRS} \} \) as defined in Equation 6.18-6.20. Then, we have seen in Proposition 8.3.1 that the program optimal capacity levels for resources owned by firm \( l \) may not be optimal for the firm level value maximization problems defined by \( \gamma_l^F-D \) and \( \gamma_l^F-L \) of Section 8.2.

Now, suppose we use gain share mechanisms \( \gamma_l^F \in \{ \gamma_l^{TIRS}, \gamma_l^{TWIRS}, \gamma_l^{TCRS} \} \), where these gain share fractions are defined below:

1. Truncated investment risk sharing (TIRS) mechanism:

\[
\gamma_l^{TIRS} = \sum_{m=1}^{M} \alpha_{lm} \left( \min \left[ \frac{c_m K_m K_m^*}{c^{K^* K}}, \frac{c_m K_m^*}{c^{K^* K}} \right] \right) \tag{8.33}
\]

2. In the truncated work and investment risk sharing (TWIRS) mechanism:

\[
\gamma_l^{TWIRS} = \sum_{m=1}^{M} \alpha_{lm} \left( \min \left[ \frac{C_m^O(X) + c_m K_m}{c^{K^* K}}, \frac{C_m^O(X) + c_m K_m^*}{c^{K^* K}} \right] \right) \tag{8.34}
\]
3. Finally, with truncated comprehensive risk sharing (TCRS):

\[
\gamma_{TCRS}^l = \sum_{m=1}^{M} \alpha_{lm} \left( \min \left[ \frac{C_{m}^{OR}(X) + C_{m}^{P}(X, K_m) + c_{m}K_{m}}{C^{O}(X) + C^{P}(X, K_{m-1}, K_m) + c_{m}K_{m}}, \frac{C_{m}^{OR}(X) + C_{m}^{P}(X, K_m) + c_{m}K_{m}^{*}}{C^{O}(X) + C^{P}(X, K^{*}) + c_{m}K_{m}^{*}} \right] \right),
\]

(8.35)

At the firm level, for the sake of brevity, we only discuss in detail the non-compliance penalty, and the guaranteed revenue models of coordination described earlier for the one-dimensional resource problem.

8.6.2 Guaranteed revenue or capacity insurance model for firms.

Here, the program planner insures the capacity investments of the firm up-to an amount \( \sum_{m=1}^{M} \alpha_{lm} c_{m}K_{m}^{*} \). Beyond that amount, the firm has no guarantee that the capacity investments will yield any deterministic gain. Hence, the gain function of the firm \( l \) is redefined as:

\[
\Pi_{l}^{o}(X, \{ K_{j: \alpha_{ij}=0} \}, \{ K_{j: \alpha_{ij}=1} \}) = \sum_{m=1}^{M} \alpha_{lm} \max \left[ \gamma_{m}^{R} \Pi(X, \{ K_{j: \alpha_{ij}=0} \}, \{ K_{j: \alpha_{ij}=1} \}), c_{m}K_{m}, \min[K_{m}, K_{m}^{*}] \right];
\]

(8.36)

\[
\hat{\Pi}_{m}^{o}(X, \{ K_{j: \alpha_{ij}=0} \}, \{ K_{j: \alpha_{ij}=1} \}) = \sum_{m=1}^{M} \alpha_{lm} \max \left[ \gamma_{m}^{R} \hat{\Pi}(X, \{ K_{j: \alpha_{ij}=0} \}, \{ K_{j: \alpha_{ij}=1} \}), c_{m}K_{m}, \min[K_{m}, K_{m}^{*}] \right].
\]

(8.37)

If \( \gamma_{m}^{R} E[\Pi(X, K^{*})] \leq c_{m}K_{m}^{*} \), for some resource group \( m \), that is when the expected gain for firm from investing at the program optimal \( K_{m}^{*} \) is less than the invested capacity cost we do not insure the capacity for that resource. Otherwise, we proceed to redefine the firm’s value function as below:
\[ V^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) = \]
\[ \Pi^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) - \]
\[ M \sum_{m=1}^{M} \alpha_{lm} \left( C^O_m(X) - C^P_m(X, K_m) - c^K_m K_m \right) ; \]

\[ \hat{V}^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) = \]
\[ \hat{\Pi}^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) - \]
\[ M \sum_{m=1}^{M} \alpha_{lm} \left( C^O_m(X) - C^P_m(X, K_m) - c^K_m K_m \right) . \]

**Proposition 8.6.1.** Let \( K^* \in \Upsilon(X) \). Then,

1. \( \{K^*_m : \alpha_{lm} = 1\} \) maximizes \( E[V^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\})] \)
   
   and \( E[\hat{V}^o_m(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\})] \);

2. For any \( \gamma^F_l \in \{\gamma^TIRS_l, \gamma^TWIRS_l, \gamma^TCRS_l\} \),
   
   \( E[V^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\})] \) and \( E[\hat{V}^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\})] \) are both
   
   quasi-concave in \( K_m : \alpha_{lm} = 1 \).

3. For any \( \gamma^F_l \in \{\gamma^TIRS_l, \gamma^TWIRS_l, \gamma^TCRS_l\} \), the centralized optimal solution \( K^* \) also
   
   represents an equilibrium capacity investment solution for the firms.

**Proof.** \( V^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) \) and \( \hat{V}^o_m(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) \) are both non-decreasing in \( K_m \leq K^*_m \), for \( m : \alpha_{lm} = 1 \). To see this, note that the penalty cost

\( C^P_m(X, K_m) \) is non-increasing in \( K_m \) while we are ensured a gain of at least \( C^K_m \) for

\( K_m \leq K^*_m \). Meanwhile the expected operating cost given \( X \) is independent of \( K_m \).

Hence, \( V^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\}) \) and therefore \( E[V^o_t(X, \{K_{j:a_{ij}=0}\}, \{K_{j:a_{ij}=1}\})] \)

are non-decreasing in \( K_m \leq K^*_m : \alpha_{lm} = 1 \).
Next, beyond \( K^*_m \), \( V^o_1(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j:a_{ij}=1} \}) \) is non-increasing in \( K_m : \alpha_{im}=1 \).

Now, for any \( \gamma_l^F \in \{ \gamma_l^{TIRS}, \gamma_l^{TWIRS}, \gamma_l^{TCRS} \} \), we have the firm gain function

\[
\gamma_l^F \Pi(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j:a_{ij}=1} \}),
\]

non-decreasing in \( K_m \), but the overall value function \( V^o_1(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j:a_{ij}=1} \}) \) is non-increasing beyond \( K^*_m \). This can be shown by contradiction. Suppose this were not true, and that the value function:

\[
E[V^o_1(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j\neq m:a_{ij}=1} \}, K_m)] > E[V^o_1(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j\neq m:a_{ij}=1} \}, K^*_m)],
\]

\[
\implies
\]

\[
E[\Pi^o_1(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j\neq m:a_{ij}=1} \}, K_m)] - E[C^O (X)] - E[C^P(X, \{ K_{j\neq m:a_{ij}=1} \}, K_m)]
\]

\[
- \sum_{m=1: \alpha_{im}=1}^M c^K_m K_m
\]

\[
> E[\Pi^o_1(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j\neq m:a_{ij}=1} \}, K^*_m)] - E[C^O (X)] - E[C^P(X, \{ K_{j\neq m:a_{ij}=1} \}, K^*_m)]
\]

\[
- \sum_{m=1: \alpha_{im}=1}^M c^K_m K^*_m
\]

\[
\implies
\]

\[
E[\Pi(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j\neq m:a_{ij}=1} \}, K_m)] - E[C^O (X)] - E[C^P(X, \{ K_{j\neq m:a_{ij}=1} \}, K_m)]
\]

\[
- \sum_{m=1: \alpha_{im}=1}^M c^K_m K_m
\]

\[
> E[\Pi(X, \{ K_{j:a_{ij}=0} \}, \{ K_{j\neq m:a_{ij}=1} \}, K^*_m)] - E[C^O (X)] - E[C^P(X, \{ K_{j\neq m:a_{ij}=1} \}, K^*_m)]
\]

\[
- \sum_{m=1: \alpha_{im}=1}^M c^K_m K^*_m
\]

which again is a contradiction, because \( K^* \) is the program optimal capacity vector corresponding to \( X \).

\[\square\]
Note again, that firms that are offered this capacity (investment) insurance or guaranteed revenue, may actually accrue greater value under this modified gain share regime. The issue of where this additional value can be apportioned from within the program level value is a matter to be decided for specific program environments. One avenue could be for firms contributing greater investments, and therefore having dominant share of the program value to transfer some of their profits to less solvent firms through the guaranteed revenue model.

8.6.3 Non-compliance penalty model for firms.

In this model, the value function for the firm \( l \) from its investment in \( m : \alpha_{lm} = 1 \) with \( \gamma^F_l \in \{ \gamma^TIRS_l, \gamma^TWIRS_l, \gamma^TCRS_l \} \), is given by:

\[
V^*_l(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) = 
\gamma^F_l \Pi_m(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) - C^OF_l(X, \{ K_{m : \alpha_{lm} = 1} \}) - \sum_{m=1; \alpha_{lm}=1}^{M} c^K_m K_m - 
\left( \Pi(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) - \Pi(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) \right)^+;
\] 

(8.40)

\[
\hat{V}^*_l(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) = 
\gamma^F_l \hat{\Pi}_m(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) - C^OF_l(X, \{ K_{m : \alpha_{lm} = 1} \}) - \sum_{m=1; \alpha_{lm}=1}^{M} c^K_m K_m - 
\left( \hat{\Pi}(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) - \hat{\Pi}(X, \{ K_j : \alpha_{lj} = 0 \}, \{ K_{m : \alpha_{lm} = 1} \}) \right)^+.
\] 

(8.41)

It can be verified that for \( K_m \leq K^*_m \) for all \( m : \alpha_{lm} = 1 \), the expectations of the value functions above are each non-decreasing in \( K_m \). Furthermore, for \( K_m > K^*_m \), the function is non-increasing in \( K_m \), for each \( m : \alpha_{lm} = 1 \) which implies that the functions are each quasi-concave in \( \{ K_m : \alpha_{lm} = 1 \} \). Hence, \( \{ K^*_m : \alpha_{lm} = 1 \} \) is a maximizer for the expected firm value functions, as defined above. The proposition below proves this result.
Proposition 8.6.2. Let $K^* \in \Upsilon(X)$. Then,

1. $\{K^*_m : \alpha_{lm} = 1\}$ maximizes

$$E[V^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})], E[\hat{V}^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})];$$

as defined above in Equations 8.40-8.41;

2. For any $\gamma^F_l \in \{\gamma^{TIRS}_l, \gamma^{TWIRS}_l, \gamma^{TCRS}_l\}$ $E[V^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})]$, and $E[\hat{V}^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})]$ are both quasi-concave in $\{K_m : \alpha_{lm} = 1\}$.

3. For any $\gamma^F_l \in \{\gamma^{TIRS}_l, \gamma^{TWIRS}_l, \gamma^{TCRS}_l\}$, the centralized optimal solution $K^*$ also represents an equilibrium capacity investment solution for the firms.

Proof. We only provide the proof for the processing time sensitive value function; the proof is identical for the delay sensitive model. Note that $V^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})$ and $\hat{V}^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})$ are both non-decreasing in $\{K_m \leq K^*_m : \alpha_{lm} = 1\}$. To see this, we can write the value function as:

$$E[V^o_l(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})] =$$

$$E[\Pi(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})] - E[C^F_l(X)] -$$

$$E[C^P_l(X, \{K_{m,\alpha_{lm}=1}\})] - \sum_{m=1; \alpha_{lm}=1}^{M} c^K_m K_m$$

$$- E[\Pi(X, \{K_{j,\alpha_{lj}=0}\}, \{K^*_m, \alpha_{lm} = 1\})] + E[\gamma^F_l \Pi(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})];$$

The first four terms above constitute a concave, non-decreasing function in $K_m \leq K^*_m$ from Proposition 7.2.9. The remaining terms are non-decreasing in $K_m$. Hence, $E[V^o_m(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})]$ is non-decreasing, and at least quasi-concave in $\{K_m \leq K^*_m : \alpha_{lm} = 1\}$.

Now, beyond $K^*_m$, $V^o_m(X, \{K_{j,\alpha_{lj}=0}\}, \{K_{m,\alpha_{lm}=1}\})$ is non-increasing in such $K_m$, since for any $\gamma^F_l \in \{\gamma^{TIRS}_l, \gamma^{TWIRS}_l, \gamma^{TCRS}_l\}$, we have the firm gain or revenue function.
as equal to $\gamma^F \Pi(X, \{K_j: \alpha_{lj} = 0\}, \{K_m: \alpha_{lm} = 1\})$. Note that this function is in fact non-decreasing in $K_m$, but the overall value function is non-increasing beyond $K^*_m: \alpha_{lm} = 1$. This can again be shown by contradiction. Suppose this were not true, and that the value function in Equation 8.40 satisfies for such $\{K_m > K^*_m\}$:

$$
E[V^o_l(X, \{K_j: \alpha_{lj} = 0\}, \{K_j: \alpha_{lj} = 1\}, K_m)] > E[V^o_l(X, \{K_j: \alpha_{lj} = 0\}, \{K_j: \alpha_{lj} = 1\}, K^*_m)].
$$

$$
\Rightarrow
E[\Pi_l^o(X, \{K_j: \alpha_{lj} = 0\}, \{K_j: \alpha_{lj} = 1\}, K_m)] - E[C^QF_l(X)] - E[C^P_l(X, \{K_j: \alpha_{lj} = 1\}, K_m)]
- \sum_{m=1: \alpha_{lm} = 1}^{M} c^K_m K_m
> E[\Pi_l^o(X, \{K_j: \alpha_{lj} = 0\}, \{K_j: \alpha_{lj} = 1\}, K^*_m)] - E[C^QF_l(X)] - E[C^P_l(X, \{K_j: \alpha_{lj} = 1\}, K^*_m)]
- \sum_{m=1: \alpha_{lm} = 1}^{M} c^K_m K^*_m.
$$

$$
\Rightarrow
E[\Pi(X, \{K_j: \alpha_{lj} = 0\}, \{K_j: \alpha_{lj} = 1\}, K_m)] - E[C^QF_l(X)] - E[C^P_l(X, \{K_j: \alpha_{lj} = 1\}, K_m)]
- \sum_{m=1: \alpha_{lm} = 1}^{M} c^K_m K_m
> E[\Pi(X, \{K_j: \alpha_{lj} = 0\}, \{K_j: \alpha_{lj} = 1\}, K^*_m)] - E[C^QF_l(X)] - E[C^P_l(X, \{K_j: \alpha_{lj} = 1\}, K^*_m)]
- \sum_{m=1: \alpha_{lm} = 1}^{M} c^K_m K^*_m.
$$

which is a contradiction, because $K^*$ is the program optimal capacity vector corresponding to $X$.  

\[\square\]
8.7 Conclusions and Summary for Capacity Planning Mechanisms for Collaborative Programs

The objectives of this chapter were to provide a framework and formulation for the capacity planning problems for supply chain program environments that have an innovation agenda. The aim was also to highlight the critical trade-offs in the planning of network capacity: the trade-off is between increased capacity investment costs for the central planner (or for individual firms), while greater resource capacity reduces the processing time and/or the delay measures for such programs. Thus, this chapter can be viewed as a time-cost trade-off analysis at the level of an entire network or for individual firms.

One critical assumption we have made throughout this chapter is that the operating costs are independent of the resource capacity investment, which while true in some cases would not be accurate for many other environments. In other words, we have assumed that operating costs are proportional to the work content, and not dependent on the resources being used to perform the work. So, firms still have to pay for the work performed, but our modeling assumption does not permit us to see how resources can impact the operating costs.

The aim was not only to condense the decision-making in this environment to two of the more critical decisions that determine the program structure: that of task assignment and capacity sizing of the resource groups, but also to decouple the two decisions from a planner’s perspective. One could thus construct a decision-hierarchy where it is the central planner who makes the task-to-resource assignment decisions, perhaps in a stage preceding the time when capacity commitments have to be made. Alternatively, the central planner could determine the assignment decisions once the capacity investment levels are revealed and program performance can be estimated for every such assignment decisions. Both structures of decision-making are made possible by the analysis and the results presented in this chapter.

The overall theme guiding the modeling efforts in this chapter has been our need
to show or demonstrate the existence of simple and implementable mechanisms that can coordinate the capacity investment in the supply chain program, given a fixed assignment strategy. Now that the existence question for these coordination mechanisms has been resolved, we can delve, perhaps in future work, into the sensitivity analysis of the capacity determination and the task assignment problems described here and previously in Chapter 6. For example, one could conduct a design of experiments based analysis of how parameters such as variability in work content of tasks, or variability in resource speeds, or indeed the variability in the gain parameter can influence capacity and the assignment decisions. Chapter 6 performs some basic sensitivity analysis, but given our elaborate modeling infrastructure, it is certainly possible to develop a wide range of insights into how the assignment and capacity decisions are influenced by the characteristics of tasks, resources, and the cost structure of firms. In fact, given the scope of our models, this kind of analysis could form the basis for several streams of future research. For example one could possibly raise questions regarding the extent to which the collaborative program willing to tolerate relatively inexpensive but risky and unreliable resources? Conversely, we could look for special strategies in terms of either the assignment or capacity sizing that can mitigate the impact of task variability, unreliable resources, and uncertainty in program revenues?

However, the motive for this dissertation is to present a framework for such analysis and associated models that can allow decision-makers to develop such insights for their own complex collaborative environments. We do not claim a motive to present more detailed insight into the behavior of specific collaborative program environments: the sheer diversity of real world program environments, while tremendously exciting, also prevents us from doing full justice to providing such managerial or operational insights for specific program settings. We save the analysis and application to different program environments for future research.

In the remainder of this dissertation, therefore, we will preserve our attention for developing another such framework for the "process" domain, i.e. program environ-
ments that consist of repetitive but interconnected tasks that again require resource capacity contributions from one or more independent firms or organizations. In the following two chapters, we develop such a modeling framework for collaborative logistics programs, and provide mathematical programming formulations for the critical capacity investment decisions by firms in such environments. We again provide formulations for a central planner that wishes to determine the optimal capacity investment towards each task, and develop results for the decentralized firm level capacity problem that parallel those in this chapter. Again, similar to this chapter, we propose readily implementable coordination mechanisms that can mitigate the inherent inefficiencies in decentralized decision-making, while illustrating the equilibrium capacity investment behavior of firms partnering in a logistics network.
Planning Capacity within Collaborative Programs for Logistics and Fulfillment

9.1 Overview and Organization

The purpose of this chapter is to illustrate another application of the collaborative planning concepts we have developed through Chapter 3, to the logistics and transportation environment. In particular, we illustrate an important application of the five expanded dimensions of supply chain planning discussed in that chapter: namely distributed agents, dynamic redefinitions, information sets and access, decision-hierarchies, and finally incentive frameworks to manage distributed agents who invest in and share resources that are deployed within the logistics activity network. While there are many different aspects to logistics and distribution planning, we focus on the planning of logistics capacity, in keeping with the theme in Chapter 5. Further, we aim to show results for collaborative logistics and process environments similar to those derived in Chapter 7, where in turn we modeled one-off projects for innovation that require collaborative partnerships. The modeling in this chapter also attempts to create a credible instance of collaborative planning in a supply chain
setting, as in those chapters.

We again define collaboration at two levels in the logistics environment: the planning phase where the network capacity is determined, and in the execution phase, where the capacity is utilized to meet demand originating and terminating at different locations in the network. Similar to supply chain program environments, inefficiencies from collaboration can arise either at the execution phase, or in the planning phase. As for those environments, it is possible to observe how collaborative inefficiencies in execution can exacerbate inefficiencies in the collaborative capacity arrangements, while in turn inefficient capacity investments can force utilization of routes that increase the net cost of collaborative inefficiencies in execution. We also show here the ways in which decentralized capacity decisions can be inefficient relative to the centralized system where one decision maker determines capacity of the logistics network subject to demand requirements and the cost structure. We show this to be true even when firms have some common agreement to share the costs and rewards from their participation in some “equitable” fashion. However, again, we are able to demonstrate the existence of a combination of incentive and disincentive mechanisms that can help coordinate the partnership.

9.2 Model

9.2.1 A Dynamic intermodal logistics or processing network.

A set \( V = \{1, ..., V\} \) of geographical locations are modeled at \( \tau_v = \{\tau^1_v, ..., \tau^T_v\}; v \in V \) distinct points in time. The set of nodes is given by the union \( N = \bigcup_{v \in V} \tau_v \); \( \|N\| = N \). Note that \( \tau^i_v \) is a time epoch at which we observe location \( v \in V \). We let \( E = \{1, ..., M\} \) represent the set of directed transportation links or arcs. The arcs and nodes are related by an \( N \times M \) node-arc incidence matrix \( \Phi \); and an \( M \times M \)
arc-arc adjacency matrix $\Psi$. An element $\phi_{nm} = 1 \implies m \in E$ is an outgoing arc at node $n \in N$, and $\phi_{nm} = -1 \implies m$ is an incoming arc at node $n$. Element $\psi_{mk} = 1 \implies \exists n \in N$ such that $\phi_{nm} = -1; \phi_{nk} = 1$, for a pair $(m,k) \in E; \psi_{mk} = 0$, otherwise. Arcs that can only be directed forward in time, and further each location can be modeled for different, or even distinct, mutually exclusive sets of time-epochs.

Each arc $e \in E$ is endowed with some capacity $K_e$ for flow, generating an $M \times 1$ vector $K$ of arc capacities. The constant marginal cost of capacity for each arc is denoted by $c^K_e$. There is a set of firms $A = 1,...,L$ who own one or more arcs in the network, and every arc in the network belongs to a non-empty subset of firms operating in the network. We use an $L \times M$ ownership matrix $\alpha$ with elements $\alpha \in [0, 1]$ such that $\sum_{l=1}^{L} \alpha_{le} = 1, \forall e \in E$, while $\alpha_{le} > 0$ for at least one firm $l \in A$. Thus ownership of an arc or an arc can be shared among more than one firm, but each arc is owned by at least one firm. Let $K^{l}_e = \alpha_{le} K_e$ be the capacity contribution by any one firm $l$ to arc $e$, and let $K^{l}$ be the $M \times 1$ capacity vector contributed by firm $l$. In addition, firms have specific capabilities w.r.t. to the arcs: for example, some firms may not be able to operate under a certain time-window, or may not be capable of providing services between pairs of locations. We use an $L \times M$ matrix $Z$ to denote these firm capabilities: an element $z_{le} = 1; \iff$ firm $l \in A$ can contribute capacity to arc $e \in E$, and therefore process work flows on that arc. Hence $K^{l}_e > 0 \implies z_{le} = 1$.

We only consider arcs that go forward in time; back-haul transport could be considered in this model (see the discussion below). Note that each node in the network represents a geographic location at a certain time during the planning horizon. Arcs within a location represent inventory and warehousing over a time interval. See Figure 9.1 for an illustration of a network modeled within this dynamic framework.
A network modeled in this framework has several special characteristics:

1. Firstly, these networks model a fixed span of time defined by the difference between the smallest and the farthest time-epoch. As such, the logistics processes are modeled for a fixed time-interval that really defines a finite planning horizon. However, this could certainly be a rolling horizon from the perspective of planning, and therefore these networks really can represent repetitive processes that occur over time in a production-distribution, or service system.

2. Secondly, the network is acyclic, and in particular because flows are allowed only in a forward time direction. But it is certainly possible for flows to return...
to a given location given the logistics process requirements or outcomes.

3. Different arcs between two location-time pairs allows for the possibility of different capacities and cost-structures for the flows between any two nodes. Depending on the demand for flows between any two location-time pairs, the network may utilize multiple arcs for flow between two nodes.

4. This dynamic representation of the logistics network is a powerful and flexible modeling language. However, it does come at the expense of a state space that grows very quickly with the number of locations, the number of discrete time buckets. The size of the network representation also depends on the number of arcs that connect different location-time pairs.

5. While modeling a fixed time span, the network can still admit stochasticity in the network parameters (to be defined below).

6. We assume conservation of flow at every stage. This implies that flows entering a node (network), must leave the node (network) in their entirety.

9.2.2 Orders and their fulfillment strategies.

An $N \times N$ matrix $D$ represents the demand for transport or general logistics services (including perhaps manufacturing tasks at an aggregate level) in the network: $D_{ij} \in D$ represents customer orders that need to be transported or shipped between two nodes $i, j \in N; i \neq j$. (Clearly, we need $j$ to represent a location-time pair that is farther along the planning horizon than $i$.) In typical applications, this matrix $D$ would be sparse, as relatively few nodes would be the origin or the final destination of orders, at least relative to the intermediate locations and times. Orders preserve their integrity along all paths from node $i$ to $j$. That is, flow is conserved at the order level along all possible paths from $i$ to $j$. There could be alternative paths between nodes
$i$ and $j$, which could exhibit different characteristics of the transportation: for example, cost versus time. Order quantities are subject to uncertainty. We assume then when demand is stochastic, it is modeled by a suitable distribution that captures the variation in demand across each location-time pairs. We let $F_{ij}(.)$ denote the distribution of random variables $D_{ij}$ for a location-time pair $i, j \in \mathbb{N}; i \neq j$. These demands could further be correlated depending on the environment being modeled. Finally, we let $\delta_{ij}$ denote the portion of demand $D_{ij}$ between that specific $O-D$ pair that goes unfulfilled. This shortfall can happen either due to inadequate capacity or in other cases due to the flows in the network not being adequately profitable for that demand type.

Thus, the system can be viewed as a multi-commodity flow network with the each location-time pair $(i, j)$ representing a possible origin-destination pair. Here, an assumption is that an $O-D$ pair can accommodate only one type of product, but the case where differentiated product shipments or processing is combined can be dealt with by creating different origin locations (for the same time-epoch) for different products, or different destination locations for each type of customer. In other words, we assume every “commodity” in the network is uniquely represented by a location-time (or $O-D$) pair. Nevertheless, the capacity of an arc $e$, $K_e$ is potentially shared by multiple “commodities”. The marginal cost of sending a unit of flow originating from node $i$ and intended for node $j$ is assumed to be constant and is denoted by $c_{ij}^e$, and thus discounts the possibility of both congestion effects, and different marginal cost of flow across order types.

In some situations, one can assume that the flow generated by the $O-D$ pairs is homogenous: the capacity usage along an arc $e$ for a unit of flow is identical across each potential $O-D$ pair. However, in some others, the utilization of capacity per
unit of flow varies between order types. In those cases, we denote $\omega_{ij}^e$ to be the capacity consumed per unit of flow of type $(i, j)$ along arc $e$. Finally, we must allow for fractional flows if one requires the problem to have any measure of tractability.

In the next sub-section, we first provide some illustration of the range of environments that can be modeled by our network representation, and both the utility, as well as the justification of the specific assumptions of the modeling framework. Subsequently, we return to develop the model further to establish decision-making and planning mechanisms that can incorporate the motives of multiple participating firms.

9.2.3 Illustration of fulfillment strategies.

This simple dynamic representation is remarkably flexible in its ability to capture various combinations of fulfillment strategies. For example, Figure 9.2.3 illustrates a network with four locations each represented at three points in time, that follows several variations of the Just-in-Time (JIT) model of logistics and processing. Figure 9.2(a) depicts a network that does not carry any inventory (i.e., materials do not remain in any one location for extended periods of time). Figure 9.2(b) represents another variation on the JIT principle where inventory is only carried at the final location, while Figure 9.2(c) shows a similar system but where inventory is carried in two locations: an intermediate location such as a warehouse, and the final location. Finally, Figure 9.2(d) depicts a network that carries inventory only in an intermediate location. Note that inventory is a term here used synonymously with Work-In-Process, so any of these networks could also describe, albeit in abstraction, a production-distribution network.
Similarly, Figure 9.2.3 illustrates a similar network that follows several variations of the Pull model of logistics and processing. We differentiate Pull from JIT strategies loosely, by calling it a Pull system if the flow for a forward section of the network depends on the existence of an inventory buffer in the previous stages. Figure 9.3(a) shows Location 6 “pulling” in a JIT fashion, material from Location 1 that carries some inventory buffers. Figure 9.3(b) depicts again a Push-Pull system but where instead of a JIT system, the network uses expediting options to fulfill demand that occurs at a single point in time (or perhaps a relatively small time window). Figure 9.3(c) shows a system that could rely solely on JIT fulfillment but protects itself with some inventory at Location 1, and finally Figure 9.3(d) illustrates a network that has flexibility in satisfying demand both using JIT as well as with expedited deliveries, and on the supply side, can bring in JIT flows, as well as carry some inventory between time periods.
9.2.4 Revenues and costs in the network.

Returning to the development of our modeling framework, the constant marginal revenue earned by a unit of flow between $O - D$ pair $(i, j)$ is denoted as $r_{ij}$. This revenue per unit is split between the collection of firms $A$; a firm $l$ receives share $\gamma_l \geq 0$ of the revenue accrued from the transportation or processing of orders between nodes $i$ and $j$, where $\sum_{l=1}^{L} \gamma_l = 1$ for all allowable $O - D$ pairs $(i, j)$. This revenue sharing is consistent across order types: the revenue earned from every order type is shared by the firms in accordance with the same sharing formula (this is a rather strong assumption, but lends tractability to the resulting decision problems). Of course, the revenue split or gain share can be determined based on the parameters of
the network such as cost structure, realized demand, and based on the actions of the participating firms such as their decision to invest in network capacity. Conversely, the investment decisions of the firms would no doubt depend on the gain sharing mechanism that is in use.

Suppose the flow in the network is \( x_{ij}^e \) of product type \((i, j)\) along arc \(e\). Let \(I_j\) denote the set of incoming arcs, and \(O_j\) denote the outgoing arcs at node \(j \in N\). Then the total demand of type \((i, j)\) satisfied is equal to \(x_{ij}^e = D_{ij}^e - \delta_{ij}^e = \sum_{e \in I_j} x_{ij}^e\), and the revenue earned for that commodity is defined as equal to \(R_{ij}(D, K, X) = r_{ij} x_{ij}^e\). Let us assume that the shortage cost per unit of commodity \((i, j)\) is \(\beta_{ij}\). Commodity \((i, j)\) therefore incurs flow (or operational) costs \(C_{ij}^O(D, K, X) = \sum_{e \in E} c_{ij}^e x_{ij}^e\), and experiences penalty or shortage costs of \(C_{ij}^P(D, K, X) = \beta_{ij} \delta_{ij}\). Hence the profit earned (or loss incurred) per commodity is:

\[
\Pi_{ij}(D, K, X) = R_{ij}(D, K) - C_{ij}^O(D, K, X) - C_{ij}^P(D, K, X) \quad (9.1)
\]

\[
= r_{ij} \sum_{e \in I_j} x_{ij}^e - \sum_{e \in E} c_{ij}^e x_{ij}^e - \beta_{ij} \delta_{ij}. \quad (9.2)
\]

The total revenues can be written as:

\[
R(D, K, X) = \sum_{i \in N} \sum_{j \in N} R_{ij}(D, K, X) = \sum_{i \in N} \sum_{j \in N} r_{ij} \sum_{e \in I_j} x_{ij}^e,
\]

while the total flow costs in the network can be expressed as:

\[
C^O(D, K, X) = \sum_{i \in N} \sum_{j \in N} C_{ij}^O(D, K, X) = \sum_{i \in N} \sum_{j \in N} \sum_{e \in E} c_{ij}^e x_{ij}^e.
\]

The total network penalty costs, similarly, are

\[
C^P(D, K, X) = \sum_{i \in N} \sum_{j \in N} C_{ij}^P(D, K, X) = \sum_{i \in N} \sum_{j \in N} \beta_{ij} \delta_{ij}.
\]
The total profit earned (or loss incurred) in the network is simply

$$
\Pi(D, K, X) = R(D, K, X) - CO(D, K, X) - CP(D, K, X).
$$

Of the total costs, a firm $l$ incurs operating costs as follows:

$$
CO_l(D, K, X) = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \sum_{e \in \mathcal{E}} \alpha_{le} c_{ij} x_{ij}.
$$

It is now a matter of governance to determine how these network revenues and penalties are shared among the participating firms. For this, we once again define the following three types of profit, cost or risk sharing mechanisms that determine the incentives for the various players in the network. In all cases, we assume that the (operating) cost of flow along an arc is split among the firms in exactly the same proportion as their ownership stakes in that arc’s capacity. This reflects a consistency in the accounting mechanism which distributes the operating costs in the same proportion to the ownership stakes. However this approach may not always reflect fairness in that the share of the revenues from fulfilling orders may be different from the share of costs that are incurred in fulfilling those orders. On the other hand, the way the network revenues from fulfilling orders and the penalty costs resulting from unfulfilled demands are shared, is exactly the purview of these gain, cost, and risk sharing mechanisms we outline below. Similar to the modeling framework in Chapter 6, one can define gain, cost, or risk sharing mechanisms as follows.

1. In the first type called investment risk sharing (IRS) mechanism the firms just share the operating revenues, and further the penalty costs in proportion to their share of the network capacity or investment costs, i.e:

$$
\gamma_l^{IRS}(K) = \sum_{e \in \mathcal{E}} \left( \frac{c_{e}^K K_{e}^l}{c^K K} \right) \quad (9.3)
$$
Here \( c^K \) is the \( M \times 1 \) capacity cost vector for the \( M \) arcs in the network. Thus, with this risk sharing mechanism, the operating profit for firm \( l \) is

\[
\Pi_l^{IRS}(D, K, X) = \gamma_l^{IRS}(R(D, K, X) - C^P(D, K, X)) - C^O_l(D, K, X);
\]

note again that \( \sum_{l=1}^{L} \gamma_l^{IRS} = 1 \).

2. In the second type we call the comprehensive risk sharing (CRS) mechanism, the network operating revenues are shared between the firms participating in the program in proportion to their share in flow (or operating) costs plus the capacity costs, i.e. \( C^O(D, K, X) + (c^K)'K \). Firm \( l \) obtains a proportion \( \gamma_l^{CRS} \) of the program revenues, defined by:

\[
\gamma_l^{IRS}(D, K, X) = \sum_{e \in E} \left( \frac{c^K_x e \gamma_l^l + \sum_{i \in N} \sum_{j \in N} c_i^j x_i^j}{c^K_x + C^O(D, K, X)} \right) \tag{9.4}
\]

Thus, with this CRS sharing mechanism, the operating profit for firm \( l \) is

\[
\Pi_l^{CRS}(D, K, X) = \gamma_l^{CRS}(R(D, K, X) - C^P(D, K, X)) - C^O_l(D, K, X),
\]

with \( \sum_{l=1}^{L} \gamma_l^{IRS} = 1 \).

Thus, the net value for firm \( l \) under the two risk sharing regimes can be written as:

\[
V_l^g(D, K, X) = \gamma_l^g(R(D, K, X) - C^P(D, K, X)) - C^O_l(D, K, X) - \sum_{e \in E} \alpha_l e \left( c^K_x e \right), \tag{9.5}
\]

for \( g \in \{IRS, CRS\} \).

**9.2.5 Shared capacity and collaborative planning framework.**

In our model, resources and therefore transportation or processing capacity in the network are shared between different firms (or even organizational units). The critical assumption, however, is that the resources used towards defining capacity of
each arc operate independently, and therefore the capacity of each arc can be defined independent of the others. This is sometimes an assumption that is infeasible, as the same transportation fleet is used between alternate or parallel routes, and allocation decisions have to be made in response to realized demand. Nevertheless, this will be our operating assumption that the capacity $K_e$ of an arc $e$ has to be planned independently of any other arc that is a potential substitute. Certainly, the net outcome of the planning process could allocate more capacity to one arc versus another, but arc capacities are fixed in the planning phase, and are not subject to change in the execution phase. Rather in the execution phase, the firms are able to observe or realize the demand for transportation or processing across the network, and can make choices with respect to the utilization of the planned capacity of each arc.

This discussion leads to the development of a decision-hierarchy and corresponding decision-framework for planning and managing the logistics network. We propose a two-phase or split hierarchy of agents and decisions, as follows.

In the first phase, before the demands are observed, the firms plan or invest in network capacity in a decentralized fashion. The capacity of each arc $K_e$ is determined along with the proportion $\alpha_{le}$ for each firm $l \in A$. A simpler way to formulate the decision-problem is to consider the capacity $K^l_e$ contributed by firm $l$ to arc $e \in E$. Hence $K_e = \sum_{i=1}^{L} K^l_e$, and $\alpha_{le} = \frac{K^l_e}{K_e}$. There are four broad ways to determine the best capacity investment decisions.

1. **Centralized capacity planning and centralized flow determination:**
   
   The first is a centralized approach, where a core group of firms determines
the optimal capacity values considering the total network capacity investment costs (incurred in the first planning phase), and trading them off against the (expected) profit or value achieved in the second execution phase. The second stage flow determination problem is also approached in a centralized fashion: the capacity requirements are revealed in this second execution phase within each demand scenario, and a central agent determines the arc flows for each demand scenario, subject to the available capacity proposed in the first stage. Therefore one can construct a scenario-based capacity planning framework, that will lead to the optimal capacity decisions in the first stage. In particular, for every feasible capacity vector $K = \{K_1, ... K_M\}$, we can construct the (expected) marginal price (or value) of capacity along each arc, and then attempt to match (in the first stage) the fixed cost of capacity in each arc with its expected marginal price.

2. **Centralized capacity planning and decentralized execution:** In this decision-hierarchy and framework, the capacity investments are determined based on decentralized flow decisions to be effected in the execution stage. In the second stage, an agent who owns an arc $e$ independently determines the allocation of their capacity along that route towards fulfilling an order $D_{ij} \in D$, based on the revenue sharing mechanism. Any firm is allowed to use only its own capacity $K^i_{e}$ towards any flow it directs along that arc $e$. Thus, one would perhaps solve $L$ different multi-commodity flow networks for every possible demand scenario: one for each firm’s network. (For each of the $L$ decentralized problems, one would have to make suitable assumptions regarding the flow decisions of the other ($L - 1$) firms in the network.) Working backwards, this creates a pricing structure for the arcs (perhaps, in expectation) which helps the centralized or core group of agents to determine the (expected) marginal value
of a fixed network capacity contribution $\mathbf{K} = \{K_1, ..., K_M\}$, via the marginal value of the fixed contributions $\mathbf{K} = \{K^l_1, ..., K^l_M\}$ of each firm $l \in \mathbf{A}$. We tend not to favor a decentralized approach to the execution problem, as it has the potential to introduce uncertainty for any firm with regard to the flow decisions of their partner firms with imperfect information exchange between firms, and especially on a real-time basis.

3. Decentralized capacity planning and centralized flow determinations:
In this decision-framework, the flow determinations are made by solving exactly one multi-commodity flow problem for each demand scenario, but for a given choice of capacity vectors $\mathbf{K} = \{K^l_1, ..., K^l_M\}$ by each firm $l$ from the first phase of planning. The flow decisions are made in aggregate for an arc (by the central agent), and the resulting revenues are split between firms according to a pre-defined gain-share formula. This generates a price for each arc (perhaps, in expectation over multiple demand scenarios) for the first stage of the capacity planning. In the first stage, the individual firms have to make decentralized decisions with regard to their arc capacity contributions in such a way as to match the marginal costs of arc capacity with the marginal prices of the arcs. One again has to make suitable assumptions for a given firm $l$ regarding the capacity choices of the other firms in the network. Without such assumptions, the problem requires a game-theoretic formulation, which is by no means a tractable exercise.

4. Decentralized capacity planning and flow determinations: Finally, it is useful to note that at this stage of our research, we found it difficult to model, at least in any coherent fashion, the case where both the arc capacities and the arc flows are to be determined in a decentralized fashion. We do not attempt
to explain the complexities of this formulation. We leave an analysis of this framework for future research.

Based on the above discussion, we choose to model our network capacity planning problem under two different decision-frameworks: a centralized capacity planning and flow determination framework, and decentralized capacity planning with centralized flow determination. In both frameworks, we model a two-stage scenario-based planning problem: in the second stage, the flow in the network is determined by solving one multi-commodity flow problem for every possible demand scenario, for a fixed capacity vector. The second stage therefore provides the arc capacity prices for a choice of capacity vectors in the first stage, defined either in a centralized (or aggregate) manner \( K = \{ K_1, ... K_M \} \), or defined in a decentralized and individual manner for each firm as \( K^l = \{ K^l_1, ... K^l_M \} \). In the first stage of the problem, we solve for the arc capacities in either of two ways: a centralized approach that just computes the arc capacities in such a way as to equate the expected marginal capacity costs with the expected profits earned in the second stage with that marginal unit of arc capacity.

Our first objective is to show how the capacity choices under each of the two decision-frameworks (centralized vs. decentralized) can be very different, therefore leading to different outcomes in terms of payoffs to the firms. Our larger objective, or goal, is to demonstrate similar to the previous chapters on supply chain program planning, how revenue and cost sharing mechanisms can be used in conjunction with incentives (or penalties) to enable the coordination of capacity investment by the different firms under a decentralized decision-framework. We show how specific revenue sharing mechanisms can be used to provide incentives (or disincentives) to firms as they make their capacity decisions in a decentralized fashion. In the subsequent sec-
tions, we define and discuss the centralized version of the network capacity planning problem, followed by the decentralized version.

9.3 Model Literature Review

Before we delve deeper into the analysis of our collaborative logistics model, we discuss briefly the literature that is related to our own work, and discuss how our model represents a contribution to such literature, and also the points of departure from more conventional game-theoretic analysis of such collaborative problems. Recently, there has been a small but growing community of operations research scholars who have shown an interest in issues arising from multiple firms collaborating to deliver goods and services within a shared logistics network. Much of this work, however, has been in the flavor of cooperative multi-player games over flow networks. For example, Kalai and Zamel [112] considered a class of maximum flow games, and showed that it is always possible in multi-player cooperative game to find a core allocation of payoffs for the $n$ – player game and for any of its induced sub-games. This meant that for problems with this special structure, one could obtain a stable set of payoffs to firms from operating the max-flow network so that no coalition of players has an incentive to deviate from the centralized flow solution to the game. However, their class of maximum flow games is restricted to single-stage decision making where for example, a planner may be interested in coordinating the flow decisions in the model.

More recently, Agarwal and Ergun [2] build on such earlier results to show how to construct such core allocations (or payoffs) to individual firms in the context of a multi-commodity flow network. However, their multi-commodity network has a few key differences from our own set-up. For example, we assume that firms contribute capacity to one or more arcs in the network, but that the demands placed on the
network are not specific to any one firm. In their model, they assume that firms contribute such shared capacity, but then operate the network in a decentralized fashion to satisfy demands that are specific to each firm. Once again, their model represents single-stage decision-making, and does not include the first capacity problem, unlike ours. However, one idea from this paper that we explored is to construct payoffs for firms in the second stage that would be part of the core allocation of the multi-player game. Eventually, as discussed in the previous section, we have eschewed a decentralized second stage problem in its entirety, so the need for core allocations does not really arise in the second stage flow determination problem. Another interesting idea proposed in this paper is to solve an inverse multi-commodity flow problem that provides the centralized flow solution to the centralized problem. Using a primal-dual formulation, they compute the inverse optimal arc capacity costs that can cause the individual firms to agree with the centralized optimal multi-commodity flow values.

A related paper by Huang and Graves [101] models a two-stage decision problem, and provides a similar stochastic programming formulation as in this chapter. The objective of their model is to determine the first stage capacity values for various resources utilized by the demands, subject to an optimal demand fulfillment model in the second stage (that is solved as a linear program). However they do not consider a decentralized first stage problem (unlike our model), and their aim is to show how a central planner would choose between different forms of capacity including those that are purchased at fixed prices, and those that are more flexible, and purchased at option prices, but with separate and significant capacity utilization costs. Their paper also provides several algorithms for the efficient solution of the stochastic programming model formulation including sampling based methods that can be selective among the different demand scenarios and converge to the true optimal capacity solution. In this chapter, we largely ignore the computational issues arising from the
two-stage stochastic programming formulation. Our purpose is to illustrate insights related to the coordination of the capacity problem in the first stage, and hence our analysis is more structural and less dependent on extensive computation. It is certainly possible to develop future work that highlights the problem of selection from among different capacity pricing alternatives, and also looks at the efficient solution of the centralized capacity problem for real-world data sets.

We also note that the problem of partner selection, or the selection among multiple capacity contract types (as in Huang and Graves [101]) is addressed implicitly in our modeling framework. Given a specific network capacity configuration, the centralized solution to the flow determination problem in the second stage allocates demand in the form of arc flows that maximizes network profits. If a certain arc is not competitive in terms of flow costs it will not be utilized as much as competing arcs in the network in the second stage flow solution. Furthermore, the first stage capacity determination problem solves for the optimal capacity investment in the competing arcs based on the fixed investment costs, but also based, implicitly, on the second stage profits that take into account the arc flow costs. Hence, the trade-off between the fixed upfront cost of investment for a certain arc and greater utilization (or option) costs in the second stage is implicitly addressed in our modeling framework in cases where the central planner has to make such a choice between installing such types of capacity between pairs of nodes in the network.

Other than these two examples, we find that the literature is rather limited as it relates to our specific problem of determining arc capacities in a potentially decentralized fashion for a multi-commodity flow network. There are many examples of decentralized capacity planning in other domains such as queueing networks (see Masuda and Whang, [137]), in manufacturing capacity planning (see Karabuk and
Wu, [115], [114]), and in other domains such as Internet infrastructure management that depend on decentralized capacity commitments by firms. We find that our specific problem formulation, which is fairly straightforward, and does not depend on a game-theoretic setup is highly flexible and applicable to many different network environments where the capacity costs are linear and the second stage problem has a linear programming formulation. Moreover, it represents a significant addition to the sparse literature on this important problem. Further extensions of our modeling framework can also selectively address questions of importance to specific collaborative logistics environments, including issues of capacity pricing or partner selection strategies as discussed in the literature mentioned above.

Finally, we mention some related application areas for our two-stage formulation of the capacity and flow determination problems. Balakrishnan et al. [8] describe in detail the workings of a local access telecommunications network (based on technology from the latter part of the 20th century), and discuss modeling approaches that can capture the complexities of designing capacity, and further routing calls in such systems. They consider the problem of determining the optimal capacity (expansion) for the network in terms of the number of telephone lines between switches (or nodes), subject to flow conservation constraints on the nodes. They frame the network design problem as a “fixed charge” problem. Demands between different origin-destination pairs are assumed known, and the network is modeled with various arc capacity alternatives with a fixed cost of activating and routing calls along certain arcs. In this way, there is a fixed as well as a variable flow component to the network costs, not unlike our model two-stage formulation. As part of the optimization, a subset of arcs is chosen to satisfy the demands between pairs of nodes, and in this way the capacity and flow decisions are effected simultaneously to minimize the sum of fixed capacity charges and the variable demand fulfillment costs. In this
sense, their problem admits binary or integer variables, and is therefore similar to
the plant location problem in manufacturing and distribution applications. Balakrishnan et al. [9] describe decomposition methods utilizing Lagrangian relaxation to
solve the resulting large scale mixed integer programming models. For environments
where a continuous representation of arc capacities is not entirely valid, the fixed
charge problem becomes more relevant.

While our immediate logistics network design problem does not admit integer
variables, it fits into more traditional two-stage stochastic linear programming frame-
works that have been developed and implemented for a number of telecommunications applications (see Sen et al., [167]). The contribution of Sen et al. is to
demonstrate an application of sampling based methods to solve a two-stage linear
stochastic programming model under uncertainty to network capacity and planning.
This method reduces the computational effort in the second stage where the demand
uncertainty is resolved, and reduces the evaluation of the second stage decisions and
performance to a smaller subset of demand scenarios. Using well-known error bounds
that govern the rate of convergence of the sample statistics to the true expected sec-
ond stage performance, it is then possible to derive a sample size that could be much
more manageable from a computational standpoint. In contrast, we forgo – as in the
previous chapters on capacity planning for innovation programs – a detailed analysis
of computational methods for the models we illustrate. As in those chapters, the
objective here is to show some structural insights regarding the feasibility of collabor-
oration in the context of shared network capacity, and mechanisms to reduce such
incentive conflicts.
9.4 Centralized Planning of a Shared Multi-Commodity Network

Let us start with the second stage first: in the first stage, the central planners propose a certain capacity vector \( \mathbf{K} = \{K_1, \ldots, K_M\} \) for the arcs in the network. In the second stage, the demand vector \( \mathbf{D} = \{D_{ij} : i, j \in N\} \) is realized, and the central planners determine the flows along the arcs to fulfill the network demands in such a way as to maximize the network profits. Let \( x_{ij}^e \) denote the quantity of flow of commodity \((i, j)\) along arc \(e\); let \( \mathbf{X}_{ij} \) denote the \( M \times 1 \) vector of arc flows for commodity \((i, j)\).

When the demand vector \( \mathbf{D} \) is subject to uncertainty, we assume a finite number of scenarios \( S = \{\mathbf{D}^1, \ldots, \mathbf{D}^S\} \) with an associated probability vector \( \mathbf{P} = \{p^1, \ldots, p^S\} \).

Let \( \mathbf{OD}^s \) represent the vector of \( O-D \) pairs with positive demands under scenario \(s \in S\), and let \( \mathbf{X}_{s,ij} \) denote the flow vector under scenario \(s\). Hence, in the second stage of the planning problem, the central planners have to solve \( S \) different multi-commodity flow problems as follows:

\[
(CSP(s, \mathbf{K}) : \Pi^s(\mathbf{K}, \mathbf{D}^s) = \max_{\mathbf{X}_{ij} : (i, j) \in \mathbf{OD}^s} \sum_{(i, j) \in \mathbf{OD}^s} \left( r_{ij} \sum_{e \in I_{ij}} x_{ij}^e - \sum_{e \in E} c_{ij}^e x_{ij}^e - \beta_{ij}^s \delta_{ij}^s \right) \right)
\]

subject to (flow balance):
\[
\sum_{e \in I_{ij}} x_{ij}^e + \delta_{ij}^s = D_{ij}, \forall (i, j) \in \mathbf{OD}^s; \quad (9.7)
\]

subject to (flow balance):
\[
\sum_{e \in O_{ij}} x_{ij}^e + \delta_{ij}^s = D_{ij}, \forall (i, j) \in \mathbf{OD}^s; \quad (9.8)
\]

subject to (flow balance):
\[
\sum_{e \in I_{iv}} x_{ij}^e = \sum_{e \in O_v} x_{ij}^e, \forall (i, j) \in \mathbf{OD}^s, \forall v \in N, \neq i, j; \quad (9.9)
\]

subject to (capacity constraint):
\[
\sum_{(i, j) \in \mathbf{OD}^s} \omega_{ij}^s x_{ij}^s \leq K_e, \forall e \in \mathbf{E}; \quad (9.10)
\]

subject to (non-negativity):
\[
x_{ij}^e \geq 0, \forall e \in \mathbf{E}; \delta_{ij}^s \geq 0, \forall (i, j) \in \mathbf{OD}^s. \quad (9.11)
\]

The second stage problem \((CSP(s, \mathbf{K})\) is a linear program, and can therefore
be solved in polynomial time for each scenario. Of course, the number of scenarios in $S$ could be large, so computationally, the second stage could be expensive. The expected profit from the second stage, given a proposed capacity vector $K$ is then

$$\Pi(K) = \sum_{s \in S} p^s \Pi^s(K, D^s).$$

The optimal value $\Pi^s(K, D^s)$ is concave in $K$ for any scenario $s$, and hence expected profit $\Pi(K)$ is also concave in $K$. Thus, in the first stage, the central planner has to determine the capacity vector $K = \{K_1, ..., K_M\}$ that maximizes the net value:

$$V(K) = \Pi(K) - c^K.$$

The central planner solves the budget constrained problem as follows: Suppose the amount of capital available to the collection of firms is $B = \sum_{l \in A} B_l$, which is obtained by incorporating the capital budget $B_l$ available to any one firm $l$. Then the central planner solves:

$$(CFP): V^* = \max_{\{K_1, ..., K_M\}} \Pi(K) - c^K$$

subject to (budget constraints):

$$\sum_{e \in E} c^K e K_l e \leq B_l, \forall l \in A;$$

subject to (firm capabilities):

$$K_{le}^l = 0, \forall e \in E : z_{le} = 1; \forall l \in A.$$

The value function $V(K)$ is concave in the capacity vector, and hence the central planner can solve for the optimal capacity vector, at least in theory. The computational complexity of the capacity planning problem depends critically on the number of demand scenarios modeled in the second stage. We refer to Huang and Graves [101] for an excellent survey of gradient-based approaches to solving for the optimal capacity vector in an efficient manner. Our objective in this paper, on the other hand, is to contrast the centralized solution, as outlined above, with the decentralized version of the problem. Nevertheless, we provide below an alternative linear
programming formulation of the problem (CFP), combining the first and second stage problems, to demonstrate that the problem can indeed be solved in polynomial time for a finite number of demand scenarios.

\[
(CLP) : V^* = 
\max_{(K_1,\ldots,K_M), \mathbf{X^{s,ij};(i,j)\in OD^s}} \sum_{s\in S} p^s \left( \sum_{(i,j)\in OD^s} (r^{ij} \sum_{e\in I_j} x_e^{s,ij} - \sum_{e\in E} c_e^j x_e^{s,ij} - \beta^{ij} \delta^{s,ij}) - c^K K \right) 
\]

subject to (budget constraints): \[ \sum_{e\in E} c_e^l K^l_e \leq B_l, \forall l \in A; \] (9.16)

subject to (firm capabilities): \[ K^l_e = 0, \forall e \in E : z_{le} = 1; \forall l \in A; \] (9.17)

subject to (flow balance): \[ \sum_{e\in I_j} x_e^{s,ij} + \delta^{s,ij} = D^{s,ij}, \forall (i,j) \in OD^s; \forall s \in S; \] (9.18)

subject to (flow balance): \[ \sum_{e\in O_i} x_e^{s,ij} = D^{s,ij}, \forall (i,j) \in OD^s; \forall s \in S; \] (9.19)

subject to (flow balance): \[ \sum_{e\in I_v} x_e^{s,ij} = \sum_{e\in O_v} x_e^{s,ij}, \forall (i,j) \in OD^s, \forall v \in N, \neq i,j; \forall s \in S; \] (9.20)

subject to (capacity constraint): \[ \sum_{(i,j)\in OD^s} \omega^{ij}_e x_e^{s,ij} \leq K_e \forall e \in E; \forall s \in S; \] (9.21)

subject to (non-negativity): \[ x_e^{s,ij} \geq 0, \forall e \in E; \delta^{s,ij} \geq 0, \forall (i,j) \in OD^s; \forall s \in S. \] (9.22)

As can be seen, the centralized version of the problem assumes perfect knowledge of each firm’s budget. The cost structure is not really different for any firm. Let \( \mathcal{Y} \) be the set of all \( K \) that solve the problem (CFP), and therefore yield the maximum value for the network. Let \( K^* \) denote any particular solution in \( \mathcal{Y} \), \( K^*_e \) denote the
corresponding optimal capacity investment in arc $e$ for this particular optimal solution, and $K^*_l$ denote the corresponding optimal capacity vector contributed by firm $l$.

The firms receive a payoff based on the gain or risk-share model in effect. With the IRS regime, the firms receive revenues in the second stage in proportion to their capacity investment in the first stage: For a demand scenario $s$, and with a risk share mechanism $g \in \{IRS, CRS\}$ firm $l$ accrues profits in the second stage equal to:

$$
\Pi^s,g_l(K, D^s, X^{s,*}) = \gamma^g_l(K, D^s, X^{s,*})(R(D^s, K, X^{s,*}) - C^P(D^s, K, X^{s,*})) - C^O_l(D^s, K, X^{s,*}).
$$  (9.23)

Here $X^{s,*}$ denotes an *optimal* multi-commodity flow in the second stage conditional on the proposed capacity and the scenario demand. The expected profit for firm $l$ is then:

$$
\Pi^g_l(K) = \sum_{s \in S} p^s \Pi^s,g_l(K, D^s).
$$

The firm $l$ derives value, for any mechanism $g \in \{IRS, CRS\}$, equal to:

$$
V^{g,*}_l = \Pi^g_l(K) - \sum_{e \in E} c^K_e K^*_e.
$$

Note that there could be multiple optimal flow vectors under every demand scenario, for a proposed capacity vector. It is a matter of governance therefore for the central planner to choose one flow vector for every demand scenario, perhaps using some consistent selection rule. The implication being that different firms could receive different payoffs (profits) in the second stage, based on which of the optimal flow vectors is chosen by the central planner as the basis to divide the network profits.
Let us again start with the second stage first: Suppose the individual firms, taken together, propose a certain capacity vector $K = \{\sum_{l \in A} K^l_1, ..., \sum_{l \in A} K^l_M\}$ for the arcs in the network. In the second stage, the demand vector $D = \{D_{ij} : i, j \in N\}$ is realized, and the central planners determine the flows along the arcs in order to fulfill the network demands with the profit maximization objective. Let $x_{ij}$ denote the quantity of flow of commodity $(i, j)$ along arc $e$; let $X_{ij}$ denote the $M \times 1$ vector of arc flows for commodity $(i, j)$. When the demand vector $D$ is subject to uncertainty, we again assume a finite number of scenarios $S = \{D^1, ..., D^S\}$. Hence, in the second stage of the planning problem, the central planners have to solve $S$ different multi-commodity flow problems as follows:

$$\Pi^s(K, D^s) = \max_{X_{ij} : (i,j) \in OD^s} \sum_{(i,j) \in OD^s} r_{ij} \sum_{e \in I_j} x_{ij}^e - \sum_{e \in E} c_{ij}^e x_{ij}^e - \beta_{ij} \delta_{ij}$$

subject to (flow balance):
$$\sum_{e \in I_j} x_{ij}^e + \delta_{ij} = D_{ij}, \forall (i, j) \in OD^s; \quad (9.25)$$

subject to (flow balance):
$$\sum_{e \in O_i} x_{ij}^e + \delta_{ij} = D_{ij}, \forall (i, j) \in OD^s; \quad (9.26)$$

subject to (flow balance):
$$\sum_{e \in I_v} x_{ij}^e = \sum_{e \in O_v} x_{ij}^e, \forall (i, j) \in OD^s, \forall v \in N, \neq i, j; \quad (9.27)$$

subject to (capacity constraint):
$$\sum_{(i,j) \in OD^s} \omega_{ij} x_{ij}^e \leq \sum_{l \in A} K^l_{n \in E} \forall e \in E; \quad (9.28)$$

subject to (non-negativity):
$$x_{ij}^e \geq 0 \forall e \in E; \delta_{ij} \geq 0, \forall (i, j) \in OD^s. \quad (9.29)$$

The second stage centralized problem $(DSP(s, K))$ is identical to the centralized case $CSP(s, K)$ in all respects, except for the fact that the individual firms pro-
pose their capacity contribution towards each arc, and the proposed capacity vector (relevant to the second stage) is therefore the sum of the capacity contributions by individual firms. The expected profit from the second stage, given a proposed capacity vector $K = \{ \sum_{l \in A} K_l^1, ..., \sum_{l \in A} K_M^l \}$ is then $\Pi(K) = \sum_{s \in S} p^s \Pi^s(K, D^s)$. The optimal value $\Pi^s(K, D^s)$ is concave in $K$ for any scenario $s$, and hence the expected profit $\Pi(K)$ is also concave in $K$.

Stepping back into the first stage, an individual firm $l$ has to determine the capacity vector $K_l^l = \{ K_l^1, ..., K_M^l \}$ that maximizes the net value to the firm based on the particular revenue-sharing mechanism in place. Thus, with a regime $g \in \{ IRS, CRS \}$, a firm $l$ solves the following first stage problem:

$$\left( DP \left( \sum_{q \in A: \neq l} K^q \right) \right) : V_{l}^{g, \diamond} \left( \sum_{q \in A: \neq l} K^q \right) = \max_{K_l^l} \Pi_l^g(K^l + \sum_{q \in A: \neq l} K^q) - \sum_{e \in E} c^K_l K_e^l$$

subject to (budget constraint): $\sum_{e \in E} c^K_e K_e^l \leq B_l$; \hspace{1cm} (9.31)

subject to (firm capabilities): $K_e^l = 0, \forall e \in E : z_{te} = 1; \forall l \in A$. \hspace{1cm} (9.32)

The value function $V_{l}^{g, \diamond} \left( K^l + \sum_{q \in A: \neq l} K^q \right)$ is however no longer concave in the capacity vector $K^l$, for any $g \in \{ IRS, CRS \}$. Hence in contrast to the situation for the central planner, individual firms now face a capacity planning problem, where the local optimal are not the global optimal solutions for the first stage problem. However, as we show below, the problem $(DP)$ has finite maxima on the feasible range of $K^l$.

**Proposition 9.5.1.** Problem $\left( DP \left( \sum_{q \in A: \neq l} K^q \right) \right)$ is not a convex program, for any
However, it possesses well-defined maximum values over the feasible range of \( K_l \), as long as the second stage problem \((DSP(s, K))\) corresponding to any capacity vector proposed in the first stage is feasible and bounded, for every demand scenario \( s \in S \).

**Proof.** We prove this fact by construction. The non-concavity of results from the fact that:

\[
\Pi_{g,s}^l(K_l + \sum_{q \in A: \neq l} K^q, D^s) = \gamma_{g}^l(K, D^s) \left( R(D^s, K, X^s) - C^P(D^s, K, X^s) \right) - C^Q_l(D^s, K, X^s).
\]

The product term in the above expression is not necessarily concave in \( K_l \). Hence the expected profits \( \Pi_{g}^l(K_l + \sum_{q \in A: \neq l} K^q) \) over all the demand scenarios is also therefore not guaranteed to be concave in \( K_l \). Hence, the value function for firm \( l \), in problem \((DP(\sum_{q \in A: \neq l} K^q))\) is not concave in \( K_l \), leading to a non-convex optimization problem.

Note that there is always, however, a well-defined global maximum to both problems, since in problem \((DP(\sum_{q \in A: \neq l} K^q))\), the objective function of firm \( l \) is continuous over a positive domain for either \( g \in \{IRS, CRS\} \). The continuity of the objective function depends critically on the feasibility and bounded-ness of the second stage centralized problem \((DSP(s, K))\) for every demand scenario \( s \in S \). If \((DSP(s, K))\) can be solved and its optimal value is finite for every demand scenario for every proposed capacity vector \( K_l \) that lies within the budget set for firm \( l \), then the value function for firm \( l \) given by \( \Pi_{g}^l(K_l + \sum_{q \in A: \neq l} K^q) - \sum_{e \in E} c^K_e K^l_e \), is also continuous and bounded in the feasible range for \( K_l \) (this is a fundamental property of linear programs governing the behavior of the optimal value as a function of the RHS vectors that define the constraints). Thus, the conditions for Weierstrass’s Theorem [179] are satisfied, and we have well-defined minima and maxima over the
budget set for firm $l$.

Suppose now that any firm $l$ assumes that every other firm $q \in A; q \neq l$ invests in accordance with some optimal solution prescribed by the centralized capacity planning problem, and all the firms in the network operates under a common regime $g \in \{IRS, CRS\}$. That is for a regime $g$, suppose firms $q \neq l$ invest $K^*_q$ in accordance with some solution $K \in \Upsilon$. In other words, firm $l$ assumes that firm $q \neq l$ contributes capacity $K^{q,*}$ derived from one particular solution of the centralized planning problem $(CFP)$ \(^1\).

Under these specific conditions, let us denote the resulting optimal value derived by firm $l$ as $\hat{V}_l^g \left( \sum_{q \in A; q \neq l} K^{q,*} \right)$, and let $\Upsilon_l$ denote the set of capacity vectors $\hat{K}^{g,l}$ that maximize $\left( DP \left( \sum_{q \in A; q \neq l} K^{q,*} \right) \right)$. Then we have the following result that dictates that the firms solving for their capacity contributions in a decentralized fashion could arrive at different capacity decisions than those obtained through solving the centralized capacity problem.

**Proposition 9.5.2.** Suppose $K^{l,*}$, denoting the capacity contribution of firm $l$, is derived from $K^* = \sum_{q \in A} K^{q,*} \in \Upsilon$, in turn a globally optimal solution under the centralized program $(CFP)$. Then $K^{l,*}$ may not be a globally optimum solution to the decentralized problem $\left( DP \left( \sum_{q \in A; q \neq l} K^{q,*} \right) \right)$ for any $g \in \{IRS, CRS\}$, and therefore it is possible that $K^{l,*} \cap \Upsilon_l = \emptyset$.

\(^1\) Recall that there could be multiple optimal solutions to the centralized capacity planning problem.
Proof. We prove this fact again by construction, using a very simple network. Consider a network of four nodes \(\{v_1, v_2, v_3, v_4\}\), and two arcs \(E = \{1, 2\}\). Arc 1 connects nodes \(v_1\) and \(v_2\) and arc 2 connects \(v_3\) and \(v_4\). There are two firms \(A = \{1, 2\}\), such that \(z_{11} = z_{12} = 1\), while \(z_{21} = 0; z_{22} = 1\). Thus, firm 2 is not capable of contributing capacity to arc 2, while firm 1 can contribute capacity to both arcs. There are two commodities with deterministic demand and is equal to \(d\) along both arcs. In other words \(D_{v_1,v_2} = D_{v_3,v_4} = d\). The revenues for demand along the arcs are also equal with \(r_{v_1,v_2} = r_{v_3,v_4} = r > 0\), while being sufficiently large as to compensate for the capacity and flow costs. The firms have budgets of \(B_1 = d + \epsilon\) and \(B_2 = d\) respectively, where \(\epsilon > 0\) is small, while the penalty costs can be assumed to large for either commodity. The cost of capacity \(c^K_1 = c^K_2 = c^K = 1\) is constant, while the cost of flow is the same along both arcs, i.e.: \(c^O_1 = c^O_2 = c^O\). Thus, the revenue assumption here implies \(r > (c^K + c^O)\).

Then, the central planner would optimize the network by allocating the full capital budget of firm 2 to its only valid arc 2 and by allocating the budget of firm 1 to arc 1. This is to set capacities \(K_1 = K_2 = d\), and yield total value to the network of:

\[
V^* = 2rd - 2c^Od - 2d = 2d(r - c^O - 1).
\]

Each firm obtains revenue equal to:

\[
V^g_1 = V^g_2 = rd - c^Od - d = d(r - c^O - 1),
\]

under either gain share regime \(g \in \{IRS, CRS\}\); since the capacity and flow costs are the same for either firm, with no resulting penalty costs for either commodity. Further, this is the only optimal solution to the centralized capacity planning problem.

Next, consider the decentralized capacity planning problem. Now, for sufficiently large \(r\), firm 1 will actually invest capacity \(K_1 = d + \epsilon = B_1\), since the firm now
receives a share $\gamma_1^{CRS} = \frac{d+\epsilon+c^O}{2d+\epsilon}$, and $\gamma_1^{IRS} = \frac{d+\epsilon}{2d+\epsilon}$, under each of the revenue sharing regimes, respectively. Considering the CRS regime (the reasoning for the IRS regime is similar), the net value to firm 1 is now

$$V_{1}^{CRS} = \left( \frac{d + \epsilon + c^O}{2d + \epsilon + 2c^O} \right) (2rd - 2c^Od) - (d + \epsilon) > d(r - c^O - 1),$$

for sufficiently large $r$. This is the only decentralized optimal capacity solution for firm 1.

Conversely, consider another example, where there is actually a disincentive for the firms to invest in capacity in the decentralized problem. Suppose for the same network that the revenues $r^{v_1, v_2} = r^{v_3, v_4} = 0$, but now the penalties $\beta^{v_1, v_2} = \beta^{v_3, v_4} = \beta > 0$ are sufficiently severe so that $\beta > (c^K + c^O) = (1 + c^O)$. The other costs and demand parameters in the network are the same as above, except that now each firm has the same budget $B_1 = B_2 = d$.

In this case, the centralized solution is to set $K_1^1 = K_2^2 = d$, to receive optimal value $V^* = -2d(c^O + 1)$. Each firm’s share of the optimal value is simply $V_1^g = V_2^g = -d(c^O + 1)$, for any gain share regime.

However, in the decentralized problem, either firm now has an incentive to reduce capacity if the penalties are not sufficiently severe. Suppose firm $l \in \{1, 2\}$ decides to invest $(d - \epsilon)$, where $\epsilon > 0$ is a small number, while assuming the other firm invests in accordance with the centralized solution. Then it’s share of the network value (which is negative) is now:

$$V_l^{CRS} = - \left( \frac{(d - \epsilon)}{(2d - \epsilon)} \right) ((d - \epsilon)(1 + c^O) + (1 + c^O)d + \beta \epsilon).$$

When the shortfall penalty is such that $(1 + c^O) < \beta < \frac{2d - \epsilon}{d - \epsilon}(1 + c^O)$, either firm has the incentive to deviate from the centralized solution by $\epsilon$ since in such case $V_l^{CRS} > -d(c^O + 1)$. This is of course when the assumption is that the other firm
is investing capacity in accordance to the centralized solution. In fact, identical arguments and conditions on the penalty parameter can be derived for the IRS regime.

See Figures 9.4 and 9.5 for examples of over-investment in arc capacities in two different decentralized network environments.

9.6 Capacity Investment Coordination Mechanisms For A Logistics Network.

9.6.1 Why does proportional revenue sharing fail to coordinate the network?

The main issue again with revenue share mechanisms is that it presents the firms with an incentive to either overinvest or under-invest in capacity relative to the needs or what is optimal for the program as a whole. Under the seemingly restrictive assumption that other firms in the network invest in capacity according to the centralized solution, a firm will continue to invest capacity, as long as it is marginally profitable to do so relative to the additional capacity and flow costs that are incurred. Conversely firms may choose to stop investing below the program optimal solution, when the marginal penalties are not severe enough to deter under-investment (but severe enough for the central planner to commit greater capacity than a given firm’s preference).

Similar to the schemes we developed for development programs, one could then alter the revenue sharing mechanism so that there are no incremental gains to be achieved by firms beyond the program optimal capacity in the case where revenues dominate the costs providing perverse incentives to firms. There is also the issue of non-concavity of the firm’s value function in the decentralized prob-
Figure 9.4: Example network with two arcs in series operated by independent firms: $c^K_1 = c^K_2 = 1 = c^O_1 = c^O_2 = 1; r = 10; d = 100; \beta = 1$. The firms overinvest w.r.t. to the network optimal capacity investment in arcs 1 & 2 resp.
Figure 9.5: Example network with two arcs in parallel operated by two independent firms: \( c_1^K = c_2^K = 1 = c_1^O = c_2^O = 1; r_1 = r_2 = 10; d_1 = 200, d_2 = 100; \beta_1 = \beta_2 = 1. \) The firms overinvest w.r.t. to the program optimal capacity investment in arcs 1 & 2 resp.
lem \( DP \left( \sum_{q \in A, q \neq l} K^q \right) \). The revenue (and penalty) sharing formula needs to be suitably altered again to achieve a value function that lends itself to maximization. We present such restructured revenue sharing mechanisms for the logistics network, and prove that both the IRS and CRS mechanisms achieve coordination with these alterations. Of course, coordination here assumes a belief system among the firms regarding the choices made by other firms in the network: namely that those firms will invest capacity according to the centralized solution. For resolving the problem of possible under-investment, or for that matter that of a non-concave decentralized value function, we need to contrive for the firm’s value function to be at least non-decreasing in its capacity investment until we achieve the program optimal level. Hence, when combined with a constraint on the revenue sharing formula, we would modify the value function to be at least quasi-concave, and then the resulting decentralized firm level capacity problems would not have local maxima.

Achieving this redefinition of the firm’s value function can be done in exactly the same ways as was discussed in the context of supply chain development programs:

1. One way is to ensure minimum gains from capacity investment at levels below the program optimal, so that for \( K^l < K^l^* \), the firm always receives guaranteed revenues that will insure against negative return on investment. We call this the guaranteed revenue model of coordination. Unfortunately, this mechanism works only when there are sufficient revenues to be shared under every demand scenario.

2. A second approach is to levy an additional penalty with a firm for investing at levels below the program optimal; this penalty could reflect the expected loss to the overall program value, relative to the program optimal, in case a firm decides to under-invest at a local maximum. Of course the belief is that
introducing this additional penalty will induce the firm to invest at program
optimal levels, so at the end, there is no transfer payment. We call this the
noncompliance penalty model of coordination. From an implementation point
of view, this would seem the most appealing mechanism to ensure global max-
ima for the firms’ capacity problem.

The guaranteed revenue model works if the network is “sufficiently” profitable;
in particular if the value of the resource to the program is greater than its capacity
cost, then the program could ensure marginal gains to any firm at least equal to the
marginal cost of capacity. The program would do this at all investment levels below
and approaching the program optimal; so at investments below the program optimal,
the firm never experiences a marginal loss from incrementing capacity. However, the
problem with this approach is that while in expectation there is no difference in the
gains derived by the firm, for any given scenario, it is possible that the program is
actually promising gains that are non-existent. For example, while the program plan-
ner has guaranteed revenues to the firm at least equal to its capacity cost, for a given
scenario it could be that the net revenue for the program is be less than the capacity
cost incurred by the firm. This would therefore make the mechanism infeasible for
a given realization, D*, of the demand vector, or for a given revenue profile for orders.

The non-compliance penalty model works within the decentralized framework,
and lets the individual firms make the tradeoff between non-compliance penalty costs
and increased gain share from capacity increments that are still the program optimal
level. The key is to set the penalty costs in a way that coordinates the decentralized
capacity decisions. As we will prove, setting the penalty cost equal to the positive
difference between the program optimal capacity solution, and the expected program
gain at any capacity level chosen by the firm leads to a firm value function which coordinates the supply chain. In other words, the firm compensates the program an amount equal to the loss in gain for the program from its decision to under-invest in relation to the program optimal capacity. There is no loss in program gain from over-investment by the firm, and the penalties are restricted to the under-investment region; the rationale being that the cap on the gain share separately addresses the over-investment issue.

In the next section, we define these coordination mechanisms (and their combinations) more formally, and prove that every such class of mechanisms when properly structured achieves the coordination of capacity investment.

### 9.6.2 Restructuring the revenue sharing mechanisms.

To start with, we define truncated gain share mechanisms corresponding to the IRS, and CRS mechanisms defined in section 9.2.4. Suppose that $K^*$ is an optimal solution to problem (CFP) that the planner wants to achieve through coordination.

1. With the truncated investment risk sharing (TIRS) mechanism, we first define an arc level risk sharing formula for firm $l$ as follows (this is important in the analysis to follow):

$$
\gamma^{TIRS}_{l,e}(K) = \min \left[ \frac{c^K K^l}{c^K K^*}, \frac{c^K K^{l, \ast}}{c^K K^*} \right]
$$

(9.33)

The arc level risk sharing formula essentially defines the share of firms in the network revenues and penalties that result from their individual investment in
each arc. Then firm level risk sharing formula is then defined as:

$$\gamma_{TIRS}^l(K) = \sum_{e \in E} \left( \gamma_{TIRS}^{l,e}(K) \right)$$  (9.34)

2. In the truncated comprehensive risk sharing (TCRS) mechanism, we similarly define an arc level risk share formula:

$$\gamma_{TCRS}^{l,e}(D, K, X) = \min \left[ \alpha_{le} c^K e + \sum_{i \in N} \sum_{j \in N} c_{ij} e_x^{ij}, \alpha_{le}^* c^K e + \sum_{i \in N} \sum_{j \in N} c_{ij} e_x^{ij} \right]$$  (9.35)

This leads to a risk sharing formula for an entire firm \(l\) as:

$$\gamma_{TCRS}^l(D, K, X) = \sum_{e \in E} \left( \gamma_{TCRS}^{l,e} \right)$$  (9.36)

In order to ensure that the firms’ capacity problems are amenable to solution (i.e. the value functions avoid local optimal), we also redefine the gain function (at the resource level) for any \(g \in \{TIRS, TCRS\}\), using the approaches described in the following subsections.

9.6.3 Guaranteed revenue model for firms as a coordination mechanism.

Here, the program planner insures the capacity investments of the firm up-to an amount \(c^K l^* = \sum_{e \in E} c^K e l^*_e\). Further, the network or collective also guarantees the expected flow (or operational) costs for the firm at any level \(K_e l^* \leq K_e^*\). Beyond that centralized optimal investment level, the firm has no guarantee that the capacity investments will yield any fixed gain. Hence, the profit function of the firm \(l\), for demand scenario \(s \in S\) and for any of the truncated revenue sharing mechanisms
\( g \in \{TIRS, TCRS\} \), is redefined only for the range \( K^l_e \leq K^l_e^* \) as:

\[
\Pi^s_g(K, D^s, X^s) = \sum_{e \in E} \left( \max \left[ \eta^g_{l,e}(K, D^s, X^s)(R(D^s, K, X^s) - C^P(D^s, K, X^s)) - \right.ight.
\]
\[
\left. \sum_{i \in N} \sum_{j \in N} (\alpha^g_{le} e_{x_{ij,s}}), \left( c^g_{K_e K^l_e} e_{x_{ij,s}} \right) \right] \right) -
\]
\[
\sum_{e \in E} \left( \max \left[ \gamma^g_{l,e}(K, D^s, X^s) \right] \right) - \sum_{i \in N} \sum_{j \in N} \alpha^g_{le} e_{x_{ij,s}} \right). \tag{9.37}
\]

The caveat here is that if the expected profits over all the demand scenarios:

\[
\Pi^g_l(X^s) \leq c^{K'} K^l_e^* + \sum_{s \in S} p^g \sum_{i \in N} \sum_{j \in N} \alpha^g_{le} e_{x_{ij,s}},
\]

that is when the expected profits for firm from investing at the program optimal \( K^l_e^* \) is less than the invested capacity cost plus the expected operational costs, we do not insure the capacity for that firm. Otherwise, we proceed to redefine the value function for the firm and solve the following problem for firm \( l \), for each \( g \in \{TIRS, TCRS\} \), with the above modification of the profit function for the firm under each scenario \( s \in S \).

\[
V^g_{l} \left( \sum_{q \in A; \neq l} K^q \right) = \max_{K^l} \Pi^g_l(K^l + \sum_{q \in A; \neq l} K^q) - \sum_{e \in E} c^g_{K_e K^l_e} \tag{9.38}
\]

\[
\left( DP \left( \sum_{q \in A; \neq l} K^q \right) \right): V^g_{l} \left( \sum_{q \in A; \neq l} K^q \right) = \max_{K^l} \Pi^g_l(K^l + \sum_{q \in A; \neq l} K^q) - \sum_{e \in E} c^g_{K_e K^l_e} \tag{9.39}
\]

subject to (budget constraint):

\[
\sum_{e \in E} c^g_{K_e K^l_e} \leq B_l; \tag{9.40}
\]

subject to (firm capabilities):

\[
K^l_e = 0, \forall e \in E: z_{le} = 1; \forall l \in A. \tag{9.41}
\]

The next proposition shows, not surprisingly perhaps, that this modified and restructured revenue sharing mechanism indeed coordinates the capacity decisions in the network.
Proposition 9.6.1. Let $K^* \in \Upsilon$. Then, $K^{l, *} \text{solves } \left( DP \left( \sum_{q \in A, \neq l} K^{q, *} \right) \right)$, for any coordination mechanism $g \in \{ TIRS, TCRS \}$, implemented in conjunction with the guaranteed revenue model as defined in Equation 9.37.

Proof. The proof rests on the fact that the modified revenue sharing mechanism provides strong and unfailing incentives for a firm to match the centralized capacity solution. First, we discount the possibility of firms investing more than the centralized capacity level $K^{l, *}_e$ in any arc. Note that under any demand scenario, the truncated revenue sharing mechanism ensures that the firm does not obtain any increase in its total profit share $\gamma^{q, *}_{l, e}$ from investing at a level $K^{l, *}_e > K^{l, *}_e$. Let $\epsilon^l_e$ be an $M \times 1$ vector with zeros at all places except at the index for arc $e$ where it has value $\epsilon > 0$. For $K^l_e = K^{l, *}_e + \epsilon$, the share $\gamma^{q, *}_{l, e}(D^*, K^*, X^*) = \gamma^{q, *}_{l, e}(D^*, K^*, X^*)$, for any $g \in \{ TIRS, TCRS \}$, and for any scenario $s \in S$ from the definition of the truncated work share.

Next, note that the central planner has no additional incentive to invest at a level $K^* + \epsilon^l_e > K^*$, since $K^*$ is the optimal solution (i.e. the value function does not increase with $K_e > K^{*}_e$).

Suppose, without loss of generality, that $\gamma^{q, *}_{l, e}(D^*, K^*, X^*) < 1$. Further, for the sake of contradiction, suppose that it is possible for a firm to derive increased value from increasing its capacity contribution to $K^* + \epsilon^l_e$, from the optimal level $K^*$. If this last statement were true, then:

$$V^g_l(K^* + \epsilon^l_e) = \Pi^g_l(K^* + \epsilon^l_e) - c^{lK^l} - c^K \epsilon^l_e > \Pi^g_l(K^*) - c^K K^{l, *}_e = V^g_l(K^*).$$

Adding $\sum_{q \in A, \neq l} \Pi^g_q(K^* + \epsilon^l_e) - c^K K^{q, *}_q$ to both sides of the above inequality, this
The statement above implies:

\[
\sum_{q \in A} \Pi_q^g(K^* + \epsilon^l_e) - c^K e^l_e > \sum_{q \in A} \Pi_q^g(K^* - c^K e^l_e) = V^*(K^*).
\]

However, the last statement cannot be true since the risk share of every other firm \( q \neq l \) at \( K^* + \epsilon^l_e \) is identical to their risk share (for any demand scenario) at the optimal capacity \( K^* \), given our definition of \( g \in \{TIRS, TCRS\} \); which would in turn imply that

\[
V^*(K^* + \epsilon^l_e) > V^*(K^*),
\]

which contradicts the starting assumption that \( K^* \) is the optimal solution to the problem \( (CFP) \).

To summarize, we have proved that firms cannot gain any additional value in expectation by unilaterally increasing their capacity contribution to any arc \( e \in E \). Therefore, it is also true that firms cannot gain additional value, given our particular definition of the truncated risk share formulae, by increasing their capacity contribution anywhere, or in any aggregate manner across the arcs in the network. Next, we show that not only are there disincentives for the firms to invest less than \( K^l_{e^*} \) for any arc \( e \), but that the value function for firm \( l \) is non-decreasing in the range \( 0 \leq K^l_{e} \leq K^l_{e^*} \), so that overall, investing at level \( K^l_{e^*} \) is always optimal for firm \( l \) under regime \( g \in \{TIRS, TCRS\} \).

To see this, note the following behavior of the profit function \( \Pi_{l,s}^g(K, D^s, X^s) \) for the range \( K^l_{e} \leq K^l_{e^*} \): The penalty cost \( C^P(D^s, K, X^s) \) is non-increasing in \( K^l_{e} \) for the same range, while the revenues \( R(D^s, K, X^s) \) are non-decreasing in \( K^l_{e} \) under any scenario \( s \in S \). The same observation holds, therefore, for the expected profit function for the firm. Since the firm is assured of profits (per arc) under a scenario
of at least equal to
\[
\left( c_e^L K_e^L + \sum_{s \in S} p^s \sum_{i \in N} \sum_{j \in N} \alpha_{le} c_{ij} x_{ij,s} \right),
\]
the expected value function is also non-negative throughout the range \( K_e^L \leq K_e^{L,*} \).
Thus, the firm sees no incentives for investing at any level below \( K_e^{L,*} \).

9.6.4 The risk averse firm and the guaranteed profits model

Note also that the guaranteed revenue (or profit) model never actually results in a firm gaining more than its share of the overall profits than defined under \( g \in \{TIRS, TCRS\} \), since the firm does not see any incentives to invest at any level below the program optimal for every arc. So, while the incentive structure may appear to be biased in favor of under-investing firms, in equilibrium, the firms will never invest below the program optimal level with such strong incentives. On the other hand, it is possible to criticize this incentive mechanism as being overly simplistic, since a risk averse firm may actually prefer to invest slightly less than the program optimal level, and obtain guaranteed profits, rather than investing in the program optimal level, and lose that risk-free return! We do not delve into this anomaly, but rather assume that in many situations network participation could be profitable under every demand scenario, and therefore this issue of risk averse decision-makers does not pose a problem in such cases.

9.6.5 Non-compliance penalty model for firms.

The objective of this mechanism is to provide strong disincentives for firms to deviate from the centralized optimal, in conjunction with the truncated revenue sharing defined earlier. This involves redefining the profit function for the firm \( l \) from its invest-
ment in arc \( e \in E \) using the truncated revenue sharing regimes \( g \in \{TIRS, TCRS\} \) for any \( K^l_e \leq K^l_{e,*} \). Suppose that the second stage problem \((DSP(D^s, K, X))\) is feasible and has an optimal solution \( X^s \). Suppose further that \( \pi_e \) is the dual or shadow price associated with an arc \( e \) (corresponding to the capacity constraint \( \sum_{(i,j) \in OD^e} \omega^{ij}_e x^{ij}_e \leq \sum_{q \in A} K^l_q \)). The full vector of arc prices is denoted by \( \pi^s \). Complementary slackness conditions for linear programming suggest that the dual price \( \pi^s_e \) is positive if and only if for the proposed capacity vector \( K \), the capacity constraint is binding in the optimal solution \( X^s \) to the second stage problem \((DSP(D^s, K, X))\). If the capacity constraint is not binding, the dual price is zero for that particular arc. Finally, let the dual price for arc \( e \), under scenario \( s \in S \) for an optimal solution \( K^* \in \Upsilon \), be denoted by \( \pi^s_{e,*} \) while the vector of dual arc prices is denoted by \( \pi^{s,*} \).

The idea behind the non-compliance penalty model is to discourage investments towards an arc below the centralized optimal solution. The question therefore is what amount of penalty to levy with a firm at a given level of investment \( K^l \leq K^l_{e,*} \). Here we propose and prove that by charging a levy \( (\pi^{s,*} K^* - \pi^s K)\), the network can ensure that firm \( l \) invests according to the centralized optimal solution \( K^* \) towards every arc. Interestingly, however, we are only able to prove this for the \( TIRS \) gain share mechanism. The critical assumption to make this coordination mechanism work, again of course, is that firm \( l \) makes its investment decision assuming that all other firms \( q \in A; q \neq l \) invest at levels \( K^q_{e,*} \), i.e. according to the centralized optimal solution. Note that

\[
K^l \leq K^l_{e,*} \implies \pi^{s,*} K^* \geq \pi^s K,
\]

since relaxing the capacity constraint can only lead to greater second stage profits for any demand scenario, and since we only explicitly account for capacity costs in the first stage value function.
Note also that from the strong duality theorem of linear programming, assuming primal feasibility:

\[ \pi^{s,*} K^* = \Pi^*(D^s, K^*, X^{s,*}) \]

Thus, the penalty that is charged to firm \( l \) is simply the difference in the profits between any capacity level chosen by the firm, and the centralized optimal capacity solution \( K^* \), assuming that firm \( l \) is the only one to deviate from the centralized solution (by our rather stringent assumption). The second stage profit function for a firm, under any demand scenario, and for its own capacity choices (and conditional on other firms’ capacity investments) is then:

\[
\Pi^{s,TIRS}_l(K, D^s, X^s) = \sum_{e \in E} \gamma^{TIRS}_{l,e}(K, D^s, X^s) \left( R(D^s, K, X^s) - C^P(D^s, K, X^s) \right) - \sum_{(i,j) \in OD^s} \alpha_{ij} c^{ij,s}_{e}(K^* - \pi^{s,*} K^* - \pi^{s} K) . \]

(9.42)

In a manner identical to the guaranteed model, the firm’s value function can then be expressed as:

\[
V^{TIRS,\diamond}_l \left( \sum_{q \in A; \neq l} K^q \right) = \max_{K^l} \Pi^{TIRS}_l(K^l + \sum_{q \in A; \neq l} K^q) - \sum_{e \in E} c^K_e K^l_e . \]  

(9.43)

With this value function, the firm solves the first stage decentralized capacity problem again as under the guaranteed revenue model:

\[
\left( DP \left( \sum_{q \in A; \neq l} K^{q,*} \right) \right) : 
\]
\[
V_{TIRS}^l \left( \sum_{q \in A \neq l} K_{q,*}^l \right) = \max_{K^l} \Pi_{TIRS}^l (K^l + \sum_{q \in A \neq l} K_{q,*}^l) - \sum_{e \in E} c_e^K K_e^l, \quad (9.44)
\]

subject to (budget constraint): \[
\sum_{e \in E} c_e^K K_e^l \leq B_l; \quad (9.45)
\]

subject to (firm capabilities): \[
K_e^l = 0, \forall e \in E: z_{le} = 1; \forall l \in A. \quad (9.46)
\]

**Proposition 9.6.2.** Let \( K^* \in \Upsilon \). Then, \( K_{l,*}^l \) solves \( \left( DP \left( \sum_{q \in A \neq l} K_{q,*}^l \right) \right) \), for the TIRS coordination mechanism implemented in conjunction with the non-compliance penalty for firm \( l \) as defined in Equation 9.42.

**Proof.** The proof is very similar to the proof of Proposition 9.6.1. Identical to the guaranteed revenues model, we can discount the possibility of firms investing more than the centralized capacity level \( K_{e,*}^l \) in any arc. Recall that under any demand scenario, the truncated revenue sharing mechanism ensures that the firm does not obtain any increase in its total profit share \( \gamma_{TIRS,s}^l \) from investing at a level \( K_e^l > K_{e,*}^l \).

We do not repeat the proof of this assertion.

Next, we prove that that the profit function as defined in Equation 9.42 is non-decreasing in \( K_e^l \leq K_{e,*}^l \). To see this, we can rewrite the profit function in Equation 9.42 as:

\[
\Pi_{TIRS}^l (K, D^s, X^s) = \sum_{e \in E} \left( (1 + \gamma_{TIRS}^l)(K, D^s, X^s)(R(D^s, K, X^s) - C^P(D^s, K, X^s)) \right. \\
- \left. \sum_{(i,j) \in OD^s} (1 + \alpha_{le})c_{ij,e}^x x_{ij,s}^e \right) - \pi_{s,*}^l K^*. 
\]
Now, $\gamma_{TIRS}^{l,e} = \alpha_{l,e}$. To use this fact, we write the profit function as:

$$\Pi_{S,TIRS}^{l}(K, D^s, X^s) = \sum_{e \in E} \left( (1 + \gamma_{TIRS}^{l,e})(K, D^s, X^s)(R(D^s, K, X^s) - C^P(D^s, K, X^s)) \right)$$

$$- \sum_{(i,j) \in OD^s} (1 + \gamma_{TIRS}^{l,e})c_{ij}^{e}x_{ij,s}^{e} + \sum_{(i,j) \in OD^s} (\gamma_{TIRS}^{l,e} - \alpha_{l,e})c_{ij}^{e}x_{ij,s}^{e}$$

$$- \pi^{s,*}K^s$$

$$= (1 + \gamma_{TIRS}^{l,e})\Pi^*(K, D^s, X^s) - \pi^{s,*}K^s +$$

$$\sum_{e \in E} \sum_{(i,j) \in OD^s} (\gamma_{TIRS}^{l,e} - \alpha_{l,e})c_{ij}^{e}x_{ij,s}^{e}.$$

(9.47)

For the TIRS regime, the last term $(\gamma_{TIRS}^{l,e} - \alpha_{l,e})x_{ij,s}^{e} = 0$, and $(1 + \gamma_{TIRS}^{l,e})$ is increasing in $K_e^l \leq K_{l,*}^e$. Hence for any increase in $K^l_e (\leq K_{l,*}^e)$, the non-compliance penalty model essentially “rewards” the firm with an increasing multiple of the entire network’s profits $(1 + \gamma_{TIRS}^{l,e})\Pi^*(K, D^s, X^s)$ for just the firm’s increased capacity investment costs. Hence, if it is profitable for the central planner to invest at least $K_{l,*}^e$ (true by assumption), it is now profitable for the firm to invest at least at level $K_{l,*}^e$. In other words, $K_{l,*}^e$ solves the decentralized problem $\left(DP \left(\sum_{q \in A: \neq l} K_{q,*}^e\right)\right)$, for the range $K^l_e (\leq K_{l,*}^e)$, since the value function is non-decreasing in that range.

While for many problem settings, the TCRS risk share formula will also coordinate the capacity investment of firm $l$, it is difficult to prove that that it will in all situations. The reason being that in Equation 9.47, the term $\sum_{e \in E} \sum_{(i,j) \in OD^s} (\gamma_{TIRS}^{l,e} - \alpha_{l,e})c_{ij}^{e}x_{ij,s}^{e}$ is not in general non-decreasing in the capacity $K^l$. However, in problem settings where this is true, even the TCRS mechanism can be shown to coordinate the capacity investment when used in conjunction with the non-compliance penalty.
See Figures 9.6 and 9.7 for an illustration of how the truncated revenue sharing mechanism coupled with either the guaranteed revenue or the non-compliance penalty policies can yield a decentralized equilibrium that is identical to the centralized optimal capacity investment along each arc.

9.7 Chapter Summary.

In this chapter, our main contribution has been to develop a modeling framework for collaborative logistics programs. The dynamic multi-commodity flow representation of a logistics program is remarkably flexible in its ability to capture both the time sensitive nature of demands placed on the network resources, and furthermore the differentiation of logistics tasks based on their time schedule. The critical parameters of the network, as with many other multi-commodity flow formulations, are the demands between $O - D$ pairs in the network, the marginal cost of arc flows, and the revenues (penalty costs) associated with fulfilled (unmet) demand. The assumptions regarding the general model have been minimal: most notably, we assume that resources are dedicated to arcs and are not shared across arcs. The critical choices involve, not surprisingly, the arc flows; however, here we are also concerned with evaluating the optimal arc capacities, given the marginal capacity costs. This latter decision variable is especially important in a collaborative program environment, where firms have to independently decide their capacity contribution to the network. A broader decision or policy variable is the revenue and gain share rule that offers an avenue to individual firms to determine the capacity contribution that is optimal for their own objectives.

The results in this chapter are mainly structural: we again prove that sustaining
Figure 9.6: Example network with two arcs in series run by two over-investing firms: $c^K_1 = c^K_2 = c^O_2 = 1, c^K_2 = 4; r = 10; d = 100; \beta = 1$. The truncated gain sharing mechanism coupled with either the guaranteed capacity investments, or with non-compliance penalties, ensures that the centralized optimal solution is also a decentralized equilibrium.
Figure 9.7: Example network with two arcs in parallel run by two over-investing firms: $c_1^K = c_2^K = c_1^O = c_2^O = 1; r_1 = r_2 = 10; d_1 = 200, d_2 = 100; \beta_1 = \beta_2 = 1$. The truncated gain sharing mechanism can be coupled with either the guaranteed capacity investment, or with non-compliance penalty policies.
a coalition of firms in a purely decentralized setting is particularly challenging. What
is more, this inefficiency of the decentralized decision-hierarchy is caused by mech-
anism such as the proportional gain share regimes that have popular appeal given
their allusions to fairness and related notions of equity. However, there is good news
in bad here, as once again, as with the results shown previously for collaborative
innovation programs, these proportional gain share mechanisms can be refitted and
modified to achieve alignment of the firm level decisions with the centralized optimal
capacity investments.

The simplifying assumption again is that the information exchange and decision-
hierarchies are bilateral: in other words we forgo a game-theoretic formulation where
multiple firms interact with each other and formulate best responses or strategies
in response to every other firms’ actions. It is our belief that the simpler bilateral
model of interaction between a central agent (perhaps a lead firm) and individual
firms still yields several critical insights that can guide planners towards coordination
strategies. More importantly this bilateral model provides a template for real-world
program decision-making, especially since a great majority of collaborative logistics
programs have clearly defined leadership roles (that resemble those of a central plan-
ger) and a small set of key decision-makers who define the rules of engagement.

Another major structural insight is that coordination is only possible when incen-
tives (or disincentives) are presented at the level of individual tasks (or arcs). Since
firms typically carry out multiple tasks in a network, devising incentives at the small-
est (atomic) level of the work schedule is really critical to ensuring that firms make
decisions with respect to individual tasks that are in alignment with the central plan-
er’s decisions. In other words, the powerful insight here is that inefficiencies from
decentralization can only really be eliminated through the design and implementa-
tion of incentives at the task level (and therefore the individual resource level, since we assume that resources are not shared across tasks). Providing incentives at the firm level is not feasible since again, similar to collaborative programs for innovation, this creates perverse incentives for firms to optimize over their own task (or resource set) and hence make decisions that prove to be inefficient for the program as a whole.
In this chapter, we discuss the existence and conditions for equilibrium capacity investments by firms in the first stage of the two-stage stochastic programming model presented in the previous chapter. We continue to assume, as before, that the second stage multi-commodity flow problems are solved in a centralized fashion by a neutral agent in order to maximize the network profits. In the first stage, however, firms decide their capacity investments in a decentralized manner with the objective of maximizing their own value from investing capacity for demand fulfillment.

For the purpose of clear analysis, we consider the smallest possible network and collaborative logistics environment. We then show how the existence of equilibrium depends crucially on the structure of the network, and on the governing profit and revenue sharing mechanism. In particular we investigate two firms operating independently in a serial versus a parallel network to fulfill network demands. We assume to begin with that demand is deterministic, and then extend the analysis
to the case of random demand. We focus on the question of whether such collaborative environments can yield any sort of capacity investment equilibrium, and if so, under what conditions these are likely to develop. The overall objective is to show that equilibrium investment behavior is rather fragile and ephemeral – if they at all exist – for even the smallest networks operating under the proportional revenue sharing mechanisms discussed in the previous section. Another objective is to highlight the response of firms to their partners’ investment behavior. Finally, we develop an application of our collaborative logistics model to compare and analyze firm behavior under prevalent logistics strategies including Make-to-Stock systems and Just-in-Time systems; our aim is to demonstrate the versatility of our modeling and analytical framework for varied logistics and distribution networks.

10.1 Specific Network Structures: Arcs in series.

Consider a simple network of three nodes \( \{x, y, z\} \) with two arcs in series: arc 1 between nodes \( x \) and \( y \) and arc 2 between nodes \( y \) and \( z \). The arcs are owned and operated by independent firms 1 and 2, respectively. There is demand \( d_{xz} = d \) for a single commodity that flows between nodes \( x \) and \( z \). The profit earned by fulfilling unit demand for this commodity is \( r \); the shortfall penalty is \( \beta \); the cost of flow along an arc is \( c^O_i \); and finally the cost of unit capacity is \( c^K_i \) for arc \( i \in \{1, 2\} \).

Supposing that \( r + \beta > c^O_1 + c^O_2 \) all of the available capacity will be utilized to satisfy any demand \( d \leq \min[K_1, K_2] \); if \( r + \beta < c^O_1 + c^O_2 \), then the flow in the network is set to 0; this is the second stage centralized “optimal” flow solution. Assuming the former cost structure, the optimal centralized first stage solution is to invest \( K_1 = K_2 = d_{xz} \), and then fulfill all of the realized demand in the second stage. In the second stage, the first stage costs are assumed sunk, so we only measure the prof-
itability of satisfying demand based on the second stage flow costs. In the first stage, the firm also incorporates the costs of capacity in determining how much capacity to reserve for the demand (which we assume is deterministic in this example).

Now, consider the behavior of any one firm $i$ in response to the capacity investment $K_{i-}$ by the partner firm. Firm $i$ receives a share $\frac{c_i^K K_i}{c_i^K K_i + c_i^{K-} K_{i-}}$ of the total second stage revenues, and therefore maximizes the value function:

$$V_i(K_i) = \frac{c_i^K K_i}{c_i^K K_i + c_i^{K-} K_{i-}} \left( r \min[d, K_{i-}, K_i] - \beta (d - \min[K_i, K_{i-}])^+ \right)$$

$$- c_i^O \min[d, K_{i-}, K_i] - c_i^K K_i$$

We can observe that the value function is not concave in $K_i$. However, the value function is continuous in $K_i \geq 0$, and is bounded above by $rd$. So a well-defined maximum exists over the feasible range of $K_i$. There are two main scenarios to consider in maximizing this value function.

10.1.1 $K_{i-} < d$

First assume $K_{i-} < d$; in this case, for $K_i \leq K_{i-}$, the value function is:

$$V_i(K_i) = \frac{(\beta + r)c_i^K K_i^2}{c_i^K K_i + c_i^{K-} K_{i-}} - \frac{d\beta c_i^K K_i}{c_i^K K_i + c_i^{K-} K_{i-}} - K_i(c_i^O + c_i^K)$$

The first term in the value function is actually convex in $K_i$ and the last term is concave (without the negative sign), and therefore the value function is convex over $[0, K_{i-}]$. For $K_i \leq K_{i-}$, independent of whether $K_i < d$ or $K_i > d$, the value function can be observed to be concave in $K_i$: 449
\[ V_i(K_i) = \frac{(r + \beta)K_i - \beta d) c_i^K K_i}{c_i^K K_i + c_i^{KK} K_i} - c_i^O K_i - c_i^K K_i \] (10.2)

The two expressions above are identical for \( K_i = K_{i-} \), so it can be verified that the value function is continuous over \( K_i > 0 \). Hence, the maximum value of \( K_i \), given \( K_{i-} < d \) is never in the interior region \((0, K_{i-})\), and is either equal to 0 or lies in the region \([K_{i-}, \infty)\).

The optimal value \( K_i^* = 0 \), when either the cost of flow or the cost of capacity is very high relative to either \((\beta + r)\) (the opportunity costs). Otherwise, firm \( i \) chooses to commit at least as much capacity as the partner firm. Interestingly, the value function in Equation 10.1 is decreasing at \( K_i = 0 \). Hence, firm \( i \) chooses to commit \( K_i \geq K_{i-} \) if both of the following conditions hold:

1. The value function at \( K_i = K_{i-} \) is greater than at \( K_i = 0 \); which condition is expressed as:

\[
\frac{(\beta + r)c_i^{K} K_{i-} - \beta dc_i^{K}}{c_i^{K} + c_i^{KK} K_{i-}} - K_{i-}(c_i^O + c_i^K) > 0.
\] (10.3)

2. The value function is increasing at \( K_i = K_{i-} \), which first order condition is expressed (after some algebra) as:

\[
\frac{c_i^{K}(rK_{i-} - \beta (d - K_{i-}))}{K_{i-}(c_i^{K} + c_i^{KK})^2} > 1.
\] (10.4)

In this case, the optimal solution \( K_i^* \) (satisfying \( K_i > K_{i-} \)) can be obtained from the first order condition:

\[
\frac{c_i^{K} K_{i-}(rK_{i-} - \beta (d - K_{i-}))}{(c_i^{K} K_i + c_i^{KK} K_{i-})^2} = 1.
\] (10.5)
10.1.2 \( K_i > d \)

In this case, the value function is expressed generally as follows:

\[
V_i(K_i) = \frac{(r \min[K_i, d] - \beta(d - \min[K_i, d])c_i^K K_i)}{c_i^K K_i + c_i^{K-} K_i^-} - c_i^O \min[K_i, d] - c_i^K K_i \quad (10.6)
\]

When \( K_i \leq d \), the value function again simplifies to the expression (same as 10.1):

\[
V_i(K_i) = \frac{(\beta + r)c_i^K K_i}{c_i^K K_i + c_i^{K-} K_i^-} - \frac{d \beta c_i^K K_i}{c_i^K K_i + c_i^{K-} K_i^-} - K_i(c_i^O + c_i^K) \quad (10.7)
\]

Hence again in this case for \( K_i \leq d \), the value function is convex in its argument.

When \( K_i \geq d \) or indeed when \( K_i \geq K_i \), the value function is given by:

\[
V_i(K_i) = \frac{c_i^K K_i (rd)}{c_i^K K_i + c_i^{K-} K_i^-} - c_i^O d - c_i^K K_i \quad (10.8)
\]

Here, the value function is concave in \( K_i \geq d \). Hence, the optimal capacity commitment behavior is similar to that of the previous scenario where \( K_i < d \): The optimal value \( K_i^* = 0 \), when either the cost of flow or the cost of capacity is very high relative to either \((\beta + r)\) (the opportunity costs). Otherwise, firm \( i \) chooses to commit capacity at least equal to the demand \( d \). Again, the value function in Equation 10.7 is decreasing at \( K_i = 0 \). Hence, firm \( i \) chooses to commit \( K_i \geq d \) if both of the following conditions hold:

1. The value function at \( K_i = d \) is greater than at \( K_i = 0 \); which condition is expressed as:

\[
\frac{(\beta + r)(1 - \beta)c_i^K d^2}{c_i^K d + c_i^{K-} K_i^-} - d(c_i^O + c_i^K) > 0. \quad (10.9)
\]

2. The value function is increasing at \( K_i = d \), which first order condition is expressed (after some algebra) as:

\[
\frac{r d c_i^{K-} K_i^-}{(c_i^K d + c_i^{K-} K_i^-)^2} > 1. \quad (10.10)
\]
In this case, the optimal solution $K_i^*$ (satisfying $K_i > K_{i-}$) can be obtained from the first order condition:

$$\frac{rdc_i K_i - K_{i-}}{(c_i K_i + c_{i-} K_{i-})^2} = 1.$$  
(10.11)

10.1.3 Existence and characterization of equilibrium solutions.

In either case above, equilibrium solutions are not guaranteed to exist in general. Equilibrium solutions occur when the best response (functions) of either firm occur are coincidental. However, in each case, the best response function can exhibit discontinuities since the optimal solution for a firm $i$ could be to either invest $K_i = 0$, or to invest at a level greater than $K_{i-}$ or $d$ (depending on whether $K_{i-} < d$ or $K_{i-} > d$, respectively).

1. The only equilibrium possible in the convex regions of the firms' value functions is at $(0,0)$. In this case, each firm sees no incentive to invest in any capacity given that the arcs are arranged in series, and fulfilling any demand requires some minimal capacity commitment by the partner firm.

2. If one firm $i$ invests $K_i < d$, there could only be positive equilibria if the other firm invests at least $K_{i-} \geq K_i$. This follows from Equations 10.1 and 10.2. Even this equilibrium is not guaranteed, and is determined by the network parameters.

3. If one firm $i$ invests $K_i > d$, there could again be positive equilibria only if the other firm finds it beneficial to invest at least equal to $d$. This fact follows from Equations 10.7 and 10.8. Again, these equilibria are governed by network parameters.

In summary, equilibrium solutions are not guaranteed given the serial network structure and because of the non-concave objective functions of the firms (in turn,
these lead to discontinuities in the best response functions of the firms). When they exist, they depend on the problem parameters. In general, however, positive equilibria tend to be in regions where one or both firms overinvest in capacity relative to either the other firm, or relative to the demand. This behavior is driven by the proportional revenue sharing mechanism. In this sense, there are both drawbacks as well as benefits to deploying proportional revenue sharing: a convex objective function is the direct result of this mechanism, but given sufficiently profitable parameters, in turn it typically forces firms to invest at levels that are conducive to the program. Conditions 10.4 and 10.10 specify conditions that lead to inefficient equilibria, i.e. where the decentralized capacity investments do not lead to the centralized solution where both arcs have the same capacity.

10.2 Specific Network Structures: Arcs in Parallel.

Next, we consider the existence and characterization of equilibria with parallel arcs in a network. Consider a simple network of three nodes \{x, y, z\} with two arcs in parallel: arc 1 between nodes x and z and arc 2 between nodes y and z. The arcs are owned and operated by independent firms 1 and 2, respectively. There are two commodities 1 and 2 that flow across each arc, with demand \(d^{xz} = d_1; d^{yz} = d_2\). The profit earned by fulfilling unit demand for either commodity \(i \in \{1, 2\}\) is \(r_i\); the shortfall penalty is \(\beta_i\); the cost of flow along an arc is \(c_i^O\); and finally the cost of unit capacity is \(c_i^K\) for arc \(i \in \{1, 2\}\).

Supposing that \(r_i + \beta_i > c_i^O\) all of the available capacity will be utilized to satisfy any demand \(d_i \leq K_i\). If \(r_i + \beta_i < c_i^O\), then the flow in the network is set to 0; this is the second stage centralized “optimal” flow solution. Assuming the former cost structure, the optimal centralized first stage solution is to invest \(K_i = d_i\), and then
fulfill all of the realized demand in the second stage. As with arcs in series: in the second stage, the first stage costs are assumed sunk, so we only measure the profitability of satisfying demand based on the second stage flow costs. In the first stage, the firm also incorporates the costs of capacity in determining how much capacity to reserve for the demand – which we assume is deterministic in this example.

Now, consider the behavior of any one firm \( i \) in response to the capacity investment \( K_{i-} \) by the partner firm in the parallel arc. Firm \( i \) receives a share \( \frac{c_i^K K_i}{c_i^K K_i + c_i^K K_{i-}} \) of the total second stage revenues (and penalty costs), and therefore maximizes the value function:

\[
V_i(K_i) = \frac{c_i^K K_i}{c_i^K K_i + c_i^K K_{i-}} (r_i \min[K_i, d_i] + r_{i-} \min[K_{i-}, d_{i-}]
- \beta_i (d_i - K_i)^+ - \beta_{i-} (d_{i-} - K_{i-})^+) - c_i^O \min[d_i, K_i] - c_i^K K_i.
\]

We can again observe that the value function is not concave in \( K_i \). However, the value function is continuous in \( K_i \geq 0 \), so a well-defined maximum exists over the feasible range of \( K_i \). There are four scenarios to consider in maximizing this value function for any firm \( i \): (i) \( K_i \leq d_i, K_{i-} \leq d_{i-} \), (ii) \( K_i \geq d_i, K_{i-} \leq d_{i-} \), (iii) \( K_i \leq d_i, K_{i-} \geq d_{i-} \), and (iv) \( K_i \geq d_i, K_{i-} \geq d_{i-} \). We consider each of these scenarios below.

When \( K_i < d_i, K_{i-} < d_{i-} \), the value function for firm \( i \) is:

\[
V_i(K_i) = \frac{c_i^K K_i}{c_i^K K_i + c_i^K K_{i-}} ((r_i + \beta_i)K_i + (r_{i-} + \beta_{i-})K_{i-} - \beta_i d_i - \beta_{i-} d_{i-})
- (c_i^O + c_i^K) K_i
\]

Similar to the function in Equation 10.1, the value function above is convex in \( K_i \).
over $[0, d_i]$, and therefore the maximum is either at 0, or alternatively at least $d_i$. It can again be shown that this function is decreasing in the vicinity of $K_i = 0$.

For $K_i < d_i$, $K_i - < d_i$, the value function becomes

$$V_i(K_i) = \frac{c_i^K K_i}{c_i^K K_i + c_i^K K_i} (r_i d_i + (r_i + \beta_i)K_i - \beta_i d_i)$$

$$- c_i^O d_i - c_i^K K_i$$

The function is now concave in $K_i > d_i$, and is increasing in $K_i$ around $d_i$ if the following first order condition is satisfied:

$$\frac{c_i^K K_i}{(c_i^K d + c_i^K K_i -)^2} (r_i d_i + (r_i + \beta_i)K_i - \beta_i d_i) > 1. \quad (10.12)$$

For the diametrically opposite case, i.e. when $K_i < d_i$, $K_i - > d_i$, the value function starts out as convex in $K_i \leq d_i$:

$$V_i(K_i) = \frac{c_i^K K_i}{c_i^K K_i + c_i^K K_i} ( (r_i + \beta_i)K_i + r_i d_i - \beta_i d_i) - (c_i^O + c_i^K)K_i$$

For the last case, when $K_i > d_i$, $K_i - > d_i$, the value function is:

$$V_i(K_i) = \frac{c_i^K K_i}{c_i^K K_i + c_i^K K_i} (r_i d_i + r_i d_i) - c_i^O d_i - c_i^K K_i$$

10.2.1 Characterization of equilibrium solutions.

In either of the case ($K_i - \leq d_i$, or $K_i - > d_i$ above, equilibrium solutions are not guaranteed to exist in general. This is mainly since in each case, the best response function can exhibit discontinuities as the optimal solution for a firm $i$ could be to either invest $K_i = 0$, or to invest at a level greater than $d_i$.

1. The only equilibrium possible in the convex regions of the firms' value functions is at $(0, 0)$. Interestingly this equilibrium is likely to occur if the burden of carrying an incremental share of the partner firms' shortfall penalties is greater relative to the revenues to be made by a firm by increasing capacity from 0.
2. If one firm $i$ invests $K_i \geq d_i$, there could only be strictly positive equilibria if the other firm invests at least $K_{i-} \geq d_{i-}$. Even this equilibrium is not guaranteed, and is determined by the network parameters, as the other firm could find it beneficial to invest $K_{i-} = 0$.

Hence in summary, equilibrium solutions are not guaranteed for parallel network structures; mainly because of the non-concave objective functions of the firms (in turn, these lead to discontinuities in the best response functions of the firms). When they exist, they depend on the problem parameters. In general, however, positive equilibria tend to be in regions where one or both firms overinvest in capacity relative to the deterministic demand. This behavior is driven by the proportional revenue sharing mechanism. Thus, similar to the series network, there are both drawbacks as well as benefits to deploying proportional revenue sharing: a convex objective function is the direct result of this mechanism, but given sufficiently conducive parameters, it typically forces firms to invest at levels that are overall of benefit to a neutral central planner who seeks to maximize network value.

10.3 Illustration Of Best Response Functions And Equilibria for Just-in-Time Networks.

Consider first a Just-in-Time network (JIT) of two arcs in series delivering a single commodity (i.e. the two arcs form a single route and schedule for logistics) according to the following parameters: $c^K_1 = c^K_2 = c^O_1 = c^O_2 = 1; r = 10; d = 100; \beta = 1$. As discussed in the preceding analysis, the value function for a firm is not concave throughout in its capacity contribution. It is convex up-to a certain level, beyond which it becomes a concave function. This is observed in Figure 10.1, and in particular via Figure 10.1(a).
The value derived by either firm (this is a symmetric example) is not monotonic in the capacity contribution of the other (for any level of its own capacity). Furthermore, as seen in Figure 10.1(b) the best response function for one firm shows discontinuities because of the non-concave behavior of the value function (please refer to the analysis preceding this subsection for the intuition behind these discontinuities). Still, despite the discontinuity in the best response functions, it is possible to observe the existence of multiple equilibria in this simple case. It is also possible to observe the kink in the best response curve when the partner firm is able to match the demand for flow within its own arc, and the linear response curve preceding that point. In the base case, we also observe how the proportional gain share mechanism creates incentives for a firm to invest at levels much greater than the partner firm even though this provides no revenue increments overall to the network. Increased demand, as in Figure 10.2 simply extends the linear portion of the best response curve.

Each of the associated figures starting with 10.1 also shows the location of the centralized optimal solution within $\mathbb{R}^2_+$ (through the brown filled circle). We can easily observe that any existing positive equilibria can be quite distant (or different) in relation to the centralized optimal, while at the same time is also possible to measure the inefficiency of the equilibrium solution w.r.t. to the centrally optimal solution. In particular, it is possible to observe how firms will systematically prefer to over-commit capacity (or participate in "capacity gaming") in order to garner a greater share of the network revenues, at least in the profitable scenarios. With less profitable scenarios, firms will under-invest in resources as compared to the centrally optimal solution, leading to inefficient outcomes in equilibrium.

Figure 10.3 shows the impact of reduced revenues associated with the flow. In this
case, we observe the discontinuity of the best response curve on either end. When the other firm invests high, initially there is no incentive for a firm to invest any capacity whatsoever: the firm’s marginal share of the revenues (relative to the other firm) is not sufficient to warrant any capacity addition. At lower levels of partner capacity, there is again a disincentive to invest any capacity since the firm is now subject to an undue share of the penalties imposed by unsatisfied demand (because the other firm is the bottleneck for the flow and hence for revenues). Increased shortfall penalty – as in Figure 10.4 – lowers the tolerance of a firm to the reduced capacity commitment by the partner firm. This increased shortfall penalty also increases the slope of the linear portions of the best response curve (although it also serves to truncate it). Intuitively, the shortfall penalty amplifies the opportunity costs for the firm resulting from lack of capacity. As can be expected the shortfall penalty has no impact on situations where the service level is greater than 100%.

The impact of shortfall penalties provides a clue regarding how to effect and control the location of positive equilibria. If the shortfall penalties are too high, the equilibria may occur at very high levels of capacity which may not be very efficient from a central planner’s perspective. In fact the equilibria may not exist at all given discontinuities in the best response curve at the far end of the partners’ capacity. If the shortfall penalties are too low, the best response curves may not intersect if the best response curves are divergent.

When the capacity costs are high, the incremental gains from pure capacity gaming behavior at regions with greater than 100% service level are not sufficient to counterbalance capacity addition costs, and therefore neither firm invests beyond what is mandated by the demand, and therefore there is no excess capacity in the network. This is seen clearly in Figure 10.5.
What is interesting, however, is the impact of capacity cost increases for one firm on the best response of the partner firm, as observed in Figure 10.6. In this case, the partner experiences greater inertia at low levels of capacity, since its gain share ratio is now less elastic w.r.t. its own capacity contribution (given the relatively high cost of capacity of its partner). The same phenomenon leads to capacity withdrawal beyond the 100% service level region. Increased operational or flow costs similarly leads to greater inertia in the best response curve, as it tends to dampen the value curve for the impacted firm, as depicted in Figure 10.7, but as can be expected it has no bearing whatsoever on the behavior of the partner firm (Figure 10.8).

Figure 10.1: The base series JIT case: $c_1^K = c_2^K = c_1^O = c_2^O = 1; r = 10; d = 100; \beta = 1.$
Figure 10.2: Increased demand in a Series JIT network: $d = 200$.

Figure 10.3: Reduced margins in a Series JIT network: $r = 5$. 
Figure 10.4: Increased shortfall penalty in a Series JIT network: $\beta = 3$.

Figure 10.5: Increased cost of resource 1 in the Series JIT network: $c_1^K = 4$.

Figure 10.6: Increased cost of resource 2 in the Series JIT network: $c_2^K = 4$. 
Figure 10.7: Increased flow cost on arc 1 in the Series JIT network: $c_1^Q = 4$.

Figure 10.8: Increased flow cost of arc 2 in the Series JIT network: $c_2^Q = 4$.

Figure 10.9: Impact of uncertain loads in a Series JIT network: $Pr(d = 100) = 0.25; Pr(d = 200) = 0.75$. 

462
Next, consider a Just-in-Time (JIT) network with two parallel arcs serving two commodities each (a commodity in this case could be differentiated by transportation to different time-epochs, and therefore by the schedule or route of deliveries); and with the following parameters: \(c^K_i = c^O_i = 1; r_i = 10; d_i = 100; \beta_i = 1; i = \{1, 2\}\).

In the case of parallel JIT system, and given the particular ownership structure we investigate, each firm now has greater control of its revenue through its capacity contribution. Hence the value function is increasing in capacity even at less than 100% service levels as in Figure 10.11, and therefore we do not observe the discontinuities in the best response function at lower levels of partner contributions. The proportional revenue sharing still drives capacity gaming, and each firm sees an interest to invest more than its partner (at least in the symmetric case). The system and the
revenue sharing mechanism also allow for multiple positive equilibria, and in the base case which is symmetric, it is possible to observe two equilibria: one at low capacity levels, and the other at relatively higher capacity levels. It is also interesting to see that neither equilibrium is efficient and equal to the centralized optimal of a 100 units of capacity for each arc.

Increasing demand for either commodity (or route) leads to more pronounced capacity gaming, which is somewhat surprising. However, Figures 10.12 and 10.13 can be explained by the fact that increased demand for any commodity that can be fulfilled profitably amplifies the influence of the proportional gain share regime, and for both firms. Furthermore, it is possible to see in Figure 10.13(a) how the value function for a firm is monotonic in the capacity levels of the partner firm with the partner’s demand having doubled.

Reducing the marginal revenue for either route, as in Figures 10.14 and 10.14 increases the investment inertia of the firm directly impacted, but also makes the other firm less inclined to participate in capacity gaming. Surprisingly, increased shortfall penalties for either firm does not seem to have any pronounced impact other than to dampen the value function for both firms (Figures 10.16 and 10.16).

In contrast, Figure 10.18 illustrates how resource cost has a significant impact on investment inertia for the impacted firm, and strongly prohibits tendencies to invest excess capacity to garner greater revenue share by that firm. However, the partner firm now has a much more pronounced appetite for capacity gaming as seen in Figure 10.19; however this is only due to the example being that of a highly profitable network. With less profits to share, we are more likely to see under-investment behavior by the partner firm. In fact, this phenomenon is already observed in Figure
10.19 where firm 1 sees no incentive to invest in any capacity beyond a certain level of investment by firm 2 since at those levels, it is hard for firm 1 to realize any meaningful revenue share. Finally, through Figures 10.20 and 10.21, we see – similar to series JIT networks – that higher operational or flow costs for do not have any impact on the partner firms, but can dampen the value function considerably.
Figure 10.12: Increased demand for route 1 in the parallel JIT network: $d_1 = 200$.

Figure 10.13: Increased demand for commodity 2 in the Parallel JIT network: $d_2 = 200$. 
10.3.1 Equilibria in more general networks.

In the previous two subsections, we have only discussed the existence of equilibria in the smallest possible systems where the demand is deterministic. With more complex network structures, the analysis becomes more cumbersome, as there are additional dependencies that are introduced by the revenue (and penalty) sharing mechanism. One firm’s value function is now influenced by the capacity decisions of every other firm operating in the network. This leads to a combinatorial number of scenarios (similar to the series and parallel structures) based on the capacity choices made within every other arc in the network. Recall that in our general model, we also
Figure 10.16: Increased shortfall penalty for commodity 1 in the Parallel JIT network: $\beta_1 = 3$.

Figure 10.17: Increased shortfall penalty for commodity 2 in the Parallel JIT network: $\beta_2 = 3$.

Figure 10.18: Increased cost of resource 1 in the Parallel JIT network: $c^K_1 = 4$. 
Figure 10.19: Increased cost of resource 2 in the Parallel JIT network: \( c_2^K = 4 \).

Figure 10.20: Increased flow cost on arc 1 in the Parallel JIT network: \( c_1^O = 4 \).

Figure 10.21: Increased flow cost of arc 2 in the Parallel JIT network: \( c_2^O = 4 \).
Figure 10.22: Impact of uncertain loads in a Parallel JIT network: $Pr(d_1 = 100) = 0.25; Pr(d_1 = 200) = 0.75; Pr(d_2 = 100) = 0.75; Pr(d_2 = 200) = 0.25$.

Figure 10.23: Impact of uncertain loads and reduced revenues in a Parallel JIT network: $Pr(d_1 = 100) = 0.25; Pr(d_1 = 200) = 0.75; Pr(d_2 = 100) = 0.75; Pr(d_2 = 200) = 0.25; r_1 = 5$. 

470
allow for shared ownership of arcs.

However, with deterministic demand, the general principles hold: the firms do not see an incentive to invest in moderate levels of capacity. Rather they either reduce their investments and exit the network, or invest in a way that ensures the network at least has sufficient capacity; both phenomena that are attributes of the proportional revenue (and penalty) sharing class of mechanisms. The drawback is that firms now see an incentive to invest at levels that are inefficient when compared to the centralized optimal solution, leading to redundant capacity. Such propensity to over-invest or under-invest in capacity are the general outcomes from deploying the proportional gain sharing regimes.

Note that we have assumed throughout the previous section that demand is deterministic. With random demand, the value function is now the weighted average of (say) the value functions for discrete demand scenarios, and therefore the analysis is no longer as clear. In fact even the piece-wise convex or concave behavior of the value function that allowed us to predict equilibria in pair-wise interactions is no longer guaranteed. Hence, the general networks with stochastic demand scenarios are extremely hard to analyze in detail. However, for the simpler examples consisting of two arcs owned by two firms, uncertain demands for commodities do not alter how the network parameters impact the development of equilibria, and the efficiency of the equilibria relative to the centralized optimal solution (see Figures 10.9, 10.10, 10.22, and 10.23).

Moreover, as shown in the previous chapter, it is still possible to alter the value functions of firms to yield more predictable and efficient capacity investment behavior even in the presence of risk in the network environment. There we have shown
that by suitably modifying and restructuring the proportional revenue sharing mechanisms, we can achieve equilibria that are in fact also efficient in the sense that they replicate the centrally optimal network capacity solution.

10.4 Illustration Of Best Response Functions And Equilibria for Make-to-Stock Networks.

In this section, we turn our attention to a more detailed analysis of Make-to-Stock (MTS) logistics strategies, and characterize the capacity investment behavior of firms collaborating under such regimes. We introduce a few additional network cost parameters, and therefore the approach to analysis depends on a more elaborate design of experiments involving these factors. Consider again a simple network of three nodes \( \{x, y, z\} \) with two arcs in series: arc 1 between nodes \( x \) and \( y \) and arc 2 between nodes \( y \) and \( z \). The arcs are owned and operated by independent firms 1 and 2, respectively. For the MTS system, however, each arc has a different purpose in the network: the first arc is a transportation (and implicitly also a storage arc), whereas the second arc serves purely as an inventory storage function. Demands in the network are for two “commodities”, each commodity representing the time-epoch for consumption of the material or goods being handled through the network. Demand is either for delivery of goods at node \( y \) – denoted by \( d_1 \), or for delivery after a certain finite storage period \( d_2 \). Demands at both epochs have to be routed through arc 1, and therefore consume capacity along the first arc. Arc 2 does not contribute any capacity towards commodity 1, and is only responsible for storage of goods for the period of time between the two demand epochs.

The profit earned by fulfilling unit demand for either commodity \( i \in \{1, 2\} \) is a uniform \( r \), whereas the shortfall penalty is a uniform \( \beta \) for each arc. The variable
cost of material flow along arc 1 is \( c_1^O \); the storage or inventory cost is represented as a fraction \( h^O \) of the variable cost \( c_1^O \). Hence the variable cost per unit of inventory stored along arc 2 is equal to \( h^O c_1^O \). Similarly, the cost of unit storage capacity for arc 2 is a fraction \( h^K \) of the capacity cost \( c_1^K \). This models MTS systems where customers may not be particularly sensitive to the time-epoch they receive the product, but where competitive offerings impose lost sales penalties that are invariant across time. Modeling time-sensitive customers is a fairly straightforward extension.

We will assume throughout that \( (r + \beta) > c_1^O (1 + 2h^O) + c_1^K (1 + 2h^K) \), so that even a centralized planner has an incentive to commit some capacity along each arc. For this setup, we can write the central planner’s value from investing capacity \( K_i; i \in \{1, 2\} \) as follows:

\[
V^{MTS}(K_1, K_2) = r \min[d_1, K_1] + r \min[d_2, K_2, (K_1 - d_1)^+] \\
- \beta (d_1 - K_1)^+ - \beta (d_2 - \min[K_2, (K_1 - d_1)^+] \\
- (1 + h^O) c_1^O \min[d_1, K_1] - h^O c_1^O \min[d_2, K_2, (K_1 - d_1)^+] \\
- (1 + h^K) c_1^K K_1 - h^K c_2^K K_2.
\]

For the stated assumptions, the value function above is concave in \((K_1, K_2)\); the same observation holds when the demands are subject to uncertainty (at least for well-behaved discrete demand distributions as modeled in this chapter). The central planner, in theory, can solve for the optimal capacity investment along each arc for some allowable set of network cost and demand parameters: See Tables 10.4.3 and 10.4.3 for the centrally optimal capacity investment given some sample parameters. Next, consider the behavior of any the firm 1 (which represents the transporter or material processor) in response to the capacity investment \( K_2 \) by the partner firm which performs only the inventory storage function. Firm 1 receives a
share \( \frac{(1+h^K)c^K_{K_1}}{(1+h^K)c^K_{K_1}+h^Kc^K_{K_2}} = \frac{(1+h^K)K_1}{(1+h^K)K_1+h^KK_2} (= \gamma_1(K_1, K_2)) \) of the total second stage revenues, and therefore maximizes the value function:

\[
V_{MTS}^1(K_1) = \gamma_1(K_1, K_2) \left( r \min[d_1, K_1] + r \min[d_2, K_2, (K_1 - d_1)^+] \right)
- \beta(d_1 - K_1)^+ - \beta(d_2 - \min[K_2, (K_1 - d_1)^+])
- (1 + h^O)c^K_{K_1} \min[d_1, K_1] - c^K_{K_1}.
\]

The warehousing firm derives corresponding value:

\[
V_{MTS}^2(K_2) = (1 - \gamma_1(K_1, K_2)) \left( r \min[d_1, K_1] + r \min[d_2, K_2, (K_1 - d_1)^+] \right)
- \beta(d_1 - K_1)^+ - \beta(d_2 - \min[K_2, (K_1 - d_1)^+])
- h^Oc^K_{K_2} \min[d_2, K_2, (K_1 - d_1)^+] - h^Kc^K_{K_2}.
\]

Once again it is possible to show that the value function for either firm \( i \) is not concave in \( K_i \). However, the value function is continuous in \( K_i \geq 0 \), and is bounded above by \( r(d_1 + d_2) \) (in the deterministic case). So a well-defined maximum exists over the feasible range of \( K_i \), for finite demand distributions.

The next several sets of figures attempt to characterize the equilibrium investment behavior of the two firms operating in a MTS regime, via numerical design of experiments. At the outset, it is useful to note that the conclusions are specific to the parameters studied, but it is still possible to observe how the equilibrium (or best response) capacity investment behavior of the firms is influenced by the revenue, cost, and demand parameters.
The base model has parameters defined at $r = 15; \beta = 1; c^O = c^K = 1; h^O = h^K = 0.05$, at two levels of demand variability. The lower demand variability is modeled by a distribution $Pr(d_i = 100) = \frac{1}{5}; Pr(d_i = 200) = \frac{3}{5}; Pr(d_i = 300) = \frac{1}{5}; i = \{1, 2\}$, while the higher variance case is given by $Pr(d_i = 100) = \frac{1}{3}; Pr(d_i = 200) = \frac{1}{3}; Pr(d_i = 300) = \frac{1}{3}; i = \{1, 2\}$. For each demand scenario, we then consider a total of 64 observations by increasing in turn each of the parameter values (independently) to a different level: $r = 20; \beta = 5; c^O = 5; c^K = 5; h^O = 0.25; h^K = 0.25$. Finally, we arbitrarily set the maximum capacity contribution by either firm at 5000; this is to keep the resulting plots clear enough for analysis. We can then make the following observations based on this limited, but still fairly general computation. Each of the figures 10.34 - 10.45 shows the best response capacity investment of each firm as a function of the proposed investment by the partner firm. The green circle shows the location of the centralized optimal capacity solution relative to these best response functions.
10.4.1 Which firm is best suited for the central planner’s role?

Observing Figures 10.34 - 10.45, it is clear that the firm with the higher marginal capacity cost is the best suited for the role of the central planner. This observation is highly specific to the revenue (and risk) sharing mechanism in place: for example, if we considered the comprehensive risk sharing regime, then the firm with larger capacity plus variable costs would be the ideal candidate for the central planner’s role. In each of those figures, we observe that the capacity choices of Firm 1 closely mirrors the optimal planner’s behavior, since it carries the bulk of the capacity costs for the network. Greater variance in demand, skewing the distribution to higher demand values, nudges Firm 1’s capacity (best) response closer to the central planner’s. The only exception is Figure 10.38, where the combined effect of higher variable and inventory storage costs forces firm 1 to compensate by engaging in capacity gaming.
(i.e. investing more than a central planner would in response to a capacity offer by the partner firm), even with higher demand variance.

10.4.2 *Is it really wise to out-source the warehousing function?*

The answer, at least in the context of our limited design of experiments, is no. In particular, when the warehousing capacity costs are relatively small w.r.t. the transportation or processing capacity costs, the warehousing firm is forced to engage in capacity gaming behavior in order to increase its share of the network profits. This is apparent through Figures 10.34 - 10.45, where firm 2 very quickly has an incentive to maximize its capacity investment to the allowable limit in order to compete for profit share with the firm contributing more expensive capacity in the form of transportation or processing. On the flip side, we can observe that firm 2 is more nervous in response to increased shortfall penalties (see Figure 10.30, and does not invest in any capacity until at least a certain minimum capacity contribution is guaranteed by Firm 1. Once that minimum capacity is guaranteed, Firm 2 increases its capacity contribution to the maximum permissible level in response to any further investment by the dominant partner. Higher revenues, or a more profitable network, simply increases the incentive for firm 2 to engage in capacity gaming.

In short, outsourcing the warehouse function, at least with revenue sharing mechanisms in place, is not warranted, at least from a purely economic analysis. If the storage capacity costs are relatively low, the transportation or manufacturing firm (or lead firm, as the case may be), should look for opportunities to keep the warehousing function in-house wherever possible. Conversely, if outsourcing is technologically inevitable, then the lead firm should look at other payment mechanisms that do not split the profits based on capacity contribution; rather, a variable payment scheme
based on services rendered would seem more suitable.

10.4.3 How do we deter firms from capacity gaming?

With revenue sharing mechanisms, there appear two intuitive approaches to eliminating capacity gaming behavior. Capacity gaming involves deviation from the centrally optimal solution, either by over-investment in capacity to maximize profit share, or investment inertia based on significant penalty risks. One approach would be to levy a tax on capacity contribution, making it more expensive for firms to over-invest, and the other to reduce the penalties accrued to a firm reluctant to invest in any capacity at all because of disproportionate risk. As we can see Figures 10.26 and 10.27, neither strategy is successful in eliminating capacity gaming behavior on its own. With increased storage capacity costs, firm 2 is now tempered in its capacity gaming, but now firm 1 sees an incentive to over-invest to maintain its own profit share (Figure 10.26). Similarly, significantly lower shortfall penalties do not seem to deter firm 2 from capacity gaming at all.

Hence, the only avenue to curb capacity gaming behavior seems to be the incentive mechanisms outlined in Section 9.6, where the lead firm, or central planner must use a combination of limited profit shares, and discounts on capacity investments to induce the firms to develop an equilibrium solution that is identical to the centralized optimal. Either approach on its own does not benefit the network performance, overall.
Figure 10.26: MTS system with more expensive storage capacity.

Figure 10.27: MTS system with more expensive storage capacity, and higher revenues.
Figure 10.28: MTS system with higher holding costs.

Figure 10.29: MTS system with higher holding costs, and higher revenues.
(a) $K_1^o(K_2)$ (low variance).

(b) $K_2^o(K_1)$ (low variance).

(c) $K_1^o(K_2)$ (high variance).

(d) $K_2^o(K_1)$ (low variance)

**Figure 10.30:** MTS system with higher shortfall penalties.

(a) $K_1^o(K_2)$ (low variance).

(b) $K_2^o(K_1)$ (low variance).

(c) $K_1^o(K_2)$ (high variance).

(d) $K_2^o(K_1)$ (low variance)

**Figure 10.31:** MTS system with higher shortfall penalties, and higher revenues.
Figure 10.32: MTS system with higher transportation (or processing) capacity costs.

Figure 10.33: MTS system with higher transportation (or processing) capacity costs, and higher revenues.
10.5 Chapter Summary and Conclusions

There is much scope for future work that utilizes the very same modeling framework as presented here. Foremost among such possibilities is the problem of matching specific firm and resource characteristics to specific categories of arcs (or tasks) in the network. This really addresses in a more concrete way, the issue of partner selection and assignment of work among the capable resources (and therefore firms). Other avenues for future work – ones that could directly utilize our modeling framework – include an analysis of the choice of capacity contracts for different types of tasks. For example, would programs be better served by choosing more flexible, but perhaps more expensive forms of capacity for serving critical tasks that are time sensitive? If so, what are the conditions applicable to the task, but also to the network environment that make one type of capacity contract choice more efficient than others for that context? As with the modeling framework for innovation programs, here
Figure 10.35: MTS system with higher variable (or operational) costs, and higher revenues.

Figure 10.36: MTS system with higher variable costs, storage capacity costs.
Figure 10.37: MTS system with higher variable costs, storage capacity costs, and higher revenues.

Figure 10.38: MTS system with higher variable costs and holding costs.
Figure 10.39: MTS system with higher variable, holding costs, and higher revenues.

Figure 10.40: MTS system with higher shortfall penalties and capacity storage costs.
Figure 10.41: MTS system with higher shortfall penalties, capacity storage costs, and revenues.

Figure 10.42: MTS system with higher shortfall penalties and variable costs.
Figure 10.43: MTS system with higher shortfall penalties, variable costs, and revenues.

Figure 10.44: MTS system with higher shortfall penalties, variable, and holding costs.
In this chapter, we have taken the first steps towards generating such insights. We analyze the investment behavior of individual firms in specific decentralized but collaborative logistics networks, and attempt to characterize the equilibrium capacity investments when firms interact within different network structures. We also compare such equilibrium behavior to the centralized planners decision-making. Finally, we close the discussion on collaborative logistics networks by analyzing the equilibrium behavior of firms, and the centralized optimal decisions when the network is operated under different logistics strategies. In particular we study the differences in firm behavior under the popular Just-in-Time versus the Make-to-Stock strategies.
for managing a logistics network.
Table 10.1: Centralized Optimal Capacity Investment in a Make-to-Stock (MTS) System; \( r = 15 \). Low demand variance: \( Pr(d_i = 100) = 0.2; Pr(d_i = 200) = 0.6; Pr(d_i = 300) = 0.2; i = \{1, 2\} \); High demand variance: \( Pr(d_i = 100) = \frac{1}{3}; Pr(d_i = 200) = \frac{1}{3}; Pr(d_i = 300) = \frac{1}{3}; i = \{1, 2\} \).

<table>
<thead>
<tr>
<th>MTS Parameters</th>
<th>Low Variance</th>
<th>High Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>( c^O )</td>
<td>( h^O )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Table 10.2: Centralized Optimal Capacity Investment in an MTS System with Higher Revenues; $r = 20$. Low demand variance: $Pr(d_i = 100) = 0.2; Pr(d_i = 200) = 0.6; Pr(d_i = 300) = 0.2; i = \{1, 2\}$; High demand variance: $Pr(d_i = 100) = \frac{1}{3}; Pr(d_i = 200) = \frac{1}{3}; Pr(d_i = 300) = \frac{1}{3}; i = \{1, 2\}$.

<table>
<thead>
<tr>
<th>MTS Parameters</th>
<th>Low Variance</th>
<th>High Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$c^O$</td>
<td>$h^O$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.25</td>
</tr>
</tbody>
</table>
11 Conclusions and Avenues for Further Work

11.1 Models for Collaborative Logistics Alliances.

The models of Chapter 9 and 10 are a first attempt to understand the capacity investment behavior of firms in logistics networks. With a fairly general set of assumptions, we construct a multi-agent dynamic, multi-commodity flow network to describe the collaborative logistics environment. Using this model, we develop a two-stage decision framework for determining the arc capacities, and later in the second stage, to determine the arc flows within the network to maximize network profits. We formulate both a centralized as well as a decentralized capacity problem subject to the second stage centralized multi-commodity flow optimization.

Though those essays, we consider the impact of collaboration exclusively in the first stage where the capacity contribution of each firm to the network has to be determined. Collaboration in the first stage requires a revenue and profit sharing mechanism that splits the overall profits accrued in the second stage - for each demand scenario - between the firms contributing capacity towards the network in the
first stage. We model several revenue and profit sharing mechanisms based on commonly used contractual arrangements, including risk sharing contracts that allocate profits to firms in proportion to their share of the overall capacity investment in the network, and related mechanisms that also factor in the operational costs of flow incurred by each firm.

For such revenue and profit sharing mechanisms, we first show how a decentralized formulation of the first stage capacity investment problem can lead to solutions that are inefficient relative to a central planner’s capacity plan. We show that a risk sharing mechanism can lead to (misaligned) incentives for firms to sometimes overinvest relative the centralized optimal when it is profitable to do so. In other situations, these revenue sharing mechanisms can cause a firm to under-invest relative to the centralized solution when the second stage multi-commodity flow allocation leads to increased operational costs, or indeed when a firm is forced – with the incremental capacity contribution – to satisfy demands that are not as lucrative. In Chapter 10, we also investigate equilibrium capacity investment solutions in smaller scale instances of the network and show how they too can be inefficient relative to the centralized capacity plan derived from solving the first stage of the stochastic program. Such behavior is not inconsistent with the well-known result that (Nash) equilibrium solutions in decentralized systems can be sub-optimal, and far away from the Pareto optimal frontier. Equilibrium capacity solutions are generated when the best response capacity plans for any given firm - conditional on the other firms’ capacity investments - are mutually reinforcing.

Such comparisons lead to the need for revenue and profit sharing mechanisms that can lead to a sustainable collaboration between the firms participating in the network. We show how the risk sharing mechanisms can be restructured, in remark-
ably simple ways, in order to eliminate incentives for any firm to under-invest or overinvest relative to the centralized optimal capacity plan. Critically, we show how these risk sharing mechanisms have to be restructured and enforced at the task level (or the level of each arc), since incentives defined at the firm or networks level fail to coordinate the capacity investments. At the same time, the advantage of capacity investment risk sharing contract mechanisms is that they are easily implementable, in contrast to (demand) scenario specific profit or revenue sharing mechanisms, and therefore the mechanisms we investigate are applicable to a wider range of real-world logistics environments.

We illustrate the risk and profit sharing mechanisms we propose for different kinds of collaborative logistics environments, including those that deploy Just-in-Time and Make-to-Stock logistics strategies. Such examples also indicate where or in which parts of such networks a lead firm may have an incentive to develop in-house capacity versus favoring incorporating outsourced capacity. For example, for illustrative networks, while sharing capacity investment risk could prove to be fruitful in a make-to-stock system with uncertain demand, it may be counter-productive in say a Just-In-Time network with fairly stable demand: capacity gaming can lead to under-capacitated arcs in critical portions of the network. In this fashion, our more detailed network model allows us to provide insights on tailoring contract types - or risk sharing mechanisms - to specific choices made by the lead firms with regard to the logistics or demand fulfillment strategy.

Finally, we also show how and why revenue sharing mechanisms would be beneficial to a lead logistics firm responsible for generating and sustaining such alliances of logistics firms. In situations where the lead firm can garner a share of the network profits that is independent of the capacity proportional revenue sharing mechanisms,
we show that as long as the remaining margins are still substantial, the participating firms would still see incentives to over-invest relative to the centralized optimal. In other words, the lead logistics firm could still anticipate redundant capacity and excellent customer service even with reduced profits for the other firms to share. These and other insights are obtained by solving a number of smaller and specific instances of the more general network models introduced in Chapter 9.

For future work, it is possible to relax some of our governing assumptions as a starting point. For example, our modeling framework where discrete demand scenarios are independently specified across O-D pairs in the network, is just shy of a stochastic programming formulation. Within a full-scale stochastic programming formulation, we can allow the demand scenarios to develop in a tree-like fashion across the nodes in the network. This is a richer representation of how demand in generated within the logistics network that also closely models real-world manufacturing and distribution applications. Thus demand across certain O-D pairs is generated due to the demand generated earlier in the dynamic network involving adjacent nodes.

It is also possible to allow stochastic generation of orders and their routes within the logistics network based on exigencies or other external factors. Formulation of a stochastic program within our network appears to be a straightforward extension of the linear programming (or multi-commodity flow) formulation we have shown for solving the discrete and independent demand scenarios.

Secondly, it is also possible to relax – albeit at the expense of computational and conceptual complexity – the assumption that the multi-commodity flow in the second stage is a centralized problem to be solved in aggregate across all the firms. Relaxing this assumption would require us to solve decentralized versions of the multi-commodity flow problem, where each firm would have control over their own
capacity and would determine how to allocate that capacity across the orders presented in the dynamic network. Here we could apply the recent ideas developed in the literature regarding how to coordinate such decentralized flow decisions using co-operative game theory. These coordinated flows in the second stage would form the basis for deploying revenue sharing mechanisms that we have discussed at length in this thesis, and ultimately to co-ordinate the capacity contributions of the firms.

In terms of application, there are again several avenues to consider. Firstly, we can investigate, using our models, the conceptual framework presented in Figure 5.13; and either validate or disprove our hypotheses regarding which type of logistics capacity contracts would be viable or indeed beneficial under different logistics strategy regimes. Conversely, for certain types of logistics contracts that are implemented, would certain strategy regimes such as Just-in-Time or Make-to-Stock dominate the set of strategies under consideration? Here we can simplify our representation of logistics contracts based on the relative significance of the fixed versus variable capacity or flow costs in the network. Variable cost capacity contracts would typically be those offered by external or third party logistics providers, while internal assets would be modeled better through fixed price contracts. Another avenue for further research is to investigate one of the value propositions offered by the so-called lead logistics providers. These firms aggregate capacity within different parts of the logistics network potentially lowering the capacity costs, while requesting a share of the gains accrued in this aggregation process. Future research efforts can investigate the conditions under which a gain sharing approach would be feasible, and what conditions would either boost or diminish the gains achieved by such capacity aggregation.
11.2 Models for Collaborative Alliances for Innovation.

In Chapters 6 - 8, we provided models, analogous to those for collaborative logistics programs, for the purpose of planning and managing decentralized projects. The basic idea of these models is to present an aggregate time-cost trade-off for tree-like program networks, composed of tasks that need one or more resource groups provided by various firms. Firms are now allowed to invest in resources that can compress the task completion times, and thus lead to shorter completion times or delays at the program level. It is a central planner’s role to assign the tasks to the candidate resource pools. Under very general assumptions regarding the form of the task processing time functions, we develop an integrated mixed-integer non-linear programming model for the central planner’s task assignment and capacity investment problem. We further show some simple rules for task assignment based on the internal collaborative capabilities within firms, their efficiency, and their marginal capacity costs.

In the associated decentralized version of the capacity problem, we start with a pre-specified task-resource assignment strategy. The partner firms, as a response, independently determine their resource capacity investments, assuming prior knowledge of their partners’ cost structures and other network parameters. For commonly used gain sharing mechanisms, where the program profits are shared in proportion to the capacity (investment) contributed, the decentralized capacity investment problem can be non-convex depending on the network instance being analyzed. Furthermore, these proportional revenue sharing mechanisms cause firms to over-commit or under-commit capacity relative to the central planner’s (optimal) capacity solution. Equilibrium capacity investments may not exist in many program environments, and even when they do occur, they could be sub-optimal relative to the centralized
planner’s optimal capacity requirements.

Given that equilibrium solutions are not guaranteed, and are sub-optimal in most cases, we propose modifications to the gain share mechanisms that can lead to a guaranteed equilibrium solution that is also optimal for the program overall. Thus, these newly constructed equilibria coordinate the firms’ decentralized capacity investments. These equilibria can only be achieved under the specialized condition where we disallow lateral capacity gaming behavior across firms. Rather each firm responds to incentive mechanisms offered by a central planner, and this is admittedly a limitation of our equilibrium concept. Nevertheless, it is a first step towards remedying the misaligned incentives that could be inherent to decentralized project environments.

Through our observations of smaller scale instances of the model that the so called “lead firms” actually have some perverse incentives to foster revenue or gain sharing mechanisms among their suppliers, while simultaneously deriving a share of profits that is not proportional to their own capacity contributions. When the remaining profits to be shared are substantial enough, it causes enough competition among the supplier firms to still over-invest capacity towards their assigned tasks, relative to the centrally optimal capacity solution. This perhaps explains the recent trends towards outsourcing manufacturing and distribution capacity by major product firms in many industrial sectors including electronics, software development, automobiles, and aviation. These results also parallel the findings in the traditional quantitative supply chain models where competition among supply chain agents leads to greater production and output levels in the supply chain, in general.

Finally, our models allow for uncertainties in the maximum program revenues (achievable), the processing rate of resources, and in the work content within each
task. Throughout, we maximize the expected value (expected profits minus capacity costs) of the task assignment and the resource capacity decision variables. We present some initial ideas for the decomposition of the integrated task assignment and capacity investment problem. For certain conditions satisfied by the program environment, we show that simple local (or task level) resource selection rules are also globally optimal. Eventually, we hope to use these resource selection rules to develop tailored and efficient algorithms to solve the integrated problem.

There are several avenues of research in the area of innovation or project alliances, and we mention only a few of them below here:

1. We have assumed, even in the decentralized capacity problem, that the task assignment is performed in a central fashion by a single decision-maker. One extension is to investigate bidding mechanisms or some other decentralized framework to allocate task rights to firms (and their resources). The bidding mechanism can be constructed in terms of one or more of the following parameters: contributed capacity, the share of the capacity cost the firm is willing to undertake, and the share of the program revenues desired by the firm in return for contributed capacity; the winning bid would be one that generates the most profits for the central planner. We can perhaps again compare how such decentralized mechanisms compare with a central planner’s optimal task assignment in terms of program outcomes.

2. We can relax the assumption that there is no lateral capacity gaming among suppliers. This leads to less tractable multi-player capacity (and/or assignment) games in a co-operative setting. However, for smaller scale instances, these problems can be analyzed for insights into the general behavior by firms when cross-firm interactions are allowed.
3. In our models, we assume that capacity investments are fixed over the time duration of the program. Allowing for these capacity decisions to be altered over time is a closer representation of many projects. This allows for the re-allocation of resources after observing the work-content (or as it evolves) across the program. We can examine whether the same results hold regarding decentralized behavior in programs with proportional revenue sharing. It is not entirely clear if such a framework has been even explored in significant detail at a centralized level.

4. Also, from a task assignment point of view, we have shown resource selection rules that work at local and global levels, but with the expected value criterion. It would be useful to look at selection rules when the criterion for optimization is some combination of expected value and the variance of value. Such rules would provide insights on partner selection when the central planner (or the lead firm) is risk averse.

These are just a few of the research issues raised by the modeling efforts of Chapters 6 - 8. There several other issues of interest to program environments in general, as outlined in Figure 4.8, that we do not discuss in any detail in this dissertation, but that could motivate and contribute to various streams of literature. This dissertation, on the other hand, serves to highlight the basic constructs in decentralized programs and alliances, and selectively address only one or two of the myriad and inter-connected decision problems in such environments. While it provides a model to capture the workings of program environment, much work has to be done in applying this basic model to solve different problems of interest to program planners, either selectively, or in some integrated fashion.
11.3 Collaborative Planning Frameworks for the 21st Century.

Through Chapters 2-3, we have made the case, albeit in essay form and in perhaps less rigorous language, for a new framework for managing collaborative alliances. The planning frameworks and functional models shown in subsequent chapters have been developed independently, but nevertheless work as instances of the type of collaborative frameworks we require to plan and execute business ideas and ventures in this Information Age. In Chapter 3, we present an evolutionary perspective of how planning frameworks evolved from their roots in manufacturing management and industrial administration, to the present day when they are synonymous with Information Technology and Decision-support Software. We chronicle in these essays how planning frameworks started as indistinguishable from their underlying decision-models, whereas in today’s architectures, planning systems work in a much more modular fashion, hiding the scope and complexities of the underlying decision models from the software front ends that end-users have come to be familiar with.

The task before 21st century researchers and practitioners alike is to translate these firm-level planning tools to become more collaborative and inclusive of multiple decision-makers and multiple decision environments. To that end, we highlight in some detail the pitfalls of applying the immensely successful planning frameworks (including decision-models and associated software) to environments that require collaboration and co-operation. These pitfalls include ill-defined or ill-suited objectives and scope, lack of standards both in design and in use across different firms that are in collaboration, the inadequate transition from a predominantly industrial and manufacturing focus to more service oriented business models, rigid decision hierarchies, competing legacy alternatives, limited risk management capabilities, and ultimately limited collaborative planning (as opposed to communication) capabilities.
or technologies.

We then propose an expansion of the same planning concept to work in collaborative environments; in particular we propose that planning should now \( (i) \) be amenable to dynamic re-definition of tasks to account for supply chain uncertainties, \( (ii) \) allow for distributed agents and stake-holders to define its objectives, \( (iii) \) allow for broader and more permeable information sets so as to allow greater (but still selective) transparency and communication, \( (iv) \) be carried out with not just one decision-maker, but with a hierarchy of decision-makers with different degrees of access and control, and finally \( (v) \) have accompanying incentive mechanisms to co-ordinate the decisions of the multiple controlling agents.

This may seem quite complex, and also perhaps a redundant extension of planning: both as a concept and how it is applied in today’s global business environments which are grappling with unprecedented levels of information access and exchange. However, the fact (or hypothesis, depending on the burden of proof required) remains that firms have only now begun to awaken to the pitfalls posed by adapting traditional planning ideologies and frameworks to their present day realities. Furthermore, as communications technologies and basic computing infrastructure have advanced, it has been easy to ignore the core and underlying decision-making aspects of planning as a construct. The fundamental idea of planning as resource allocation over time to competing priorities is now just one small and often ignored “module” in many of today’s corporate planning frameworks and systems. These systems while becoming adept at capturing and exchanging tera-bytes of data across fibre-optic networks straddling global supply chains, have not really begun grasping the idea that all said and done, it is careful decision making that drives business and shareholder success. While information and communications technologies resolve much
uncertainty in the collaborative business environment, they are still limited to the role of enablers of effective decision-making.

While human beings as executives are still some of the most flexible and capable decision-makers, they still require, at least in some situations, the assistance of planning frameworks such as those developed in Chapters 4-9; in order to develop a set of decision alternatives supported by objective criteria; and ultimately act upon them for the benefit of their shareholders and employees. Developing such frameworks that can provide insight and data for collaborative decision-making; by agents representing their firms and shareholders; has therefore been the guiding purpose of this dissertation.

If not already evident from our selective and at times disconnected presentation, much work remains to be done to translate these broad and fuzzy proposals on how to redefine planning concepts for the age of information and collaboration, to real world planning systems. These systems have to work within the technological and managerial constraints imposed by the operation of a profit-seeking firm. As with the vast majority of planning approaches, individual ideas and proposals can rarely gain traction unless they are tested through cycles of adoption and failures. The success of planning constructs such as those outlined in this thesis also depend critically on whether collective standards can emerge on how best to allow for collaborative planning and decision frameworks to bridge firm boundaries, and achieve the business and profit objectives for the supply chain. One approach to examining the feasibility of these proposals is to develop a prototype implementation based on functional models similar to those presented in this dissertation, and then to deploy the prototype within a real-world collaborative or supply chain environment. Observing the successes and failures from such deployments, and working through the inevitable
implementation hurdles, could provide clues on whether our proposed frameworks are adequate, limited, or indeed excessive in relation to the needs of planning within a collaborative business environment.
A.1 Mathematica Code used in Chapter 6.

\[
\begin{align*}
  k(t, b, l) &= (t^2 + t)^{1/(t + 1)} \\
  L(t, b, l) &= 1 + b / (k(t, b, l)^t) \\
  V(t, b, l) &= 100 - (1 + t)^{1/(t + 1)} * (t/(t + 1)) * (1 + b)^{1/(t + 1)} \\
\end{align*}
\]

\textbf{Needs("PlotLegends")}

\textbf{Plot3D}(k(t, 0.5, l), \{t, 0, 2\}, \{l, 0, 50\},
\textbf{AxesLabel} -> \{
\text{\[Theta\]}, \text{\[Lambda]/\[Mu]\}, \text{K^*}\}, \textbf{ColorFunction} -> \text{Hue},
\textbf{PlotLabel} -> \text{\[Beta\]/c=0.5}
\textbf{Plot3D}(L(t, 0.1, l), \{t, 0, 2\}, \{l, 0, 50\},
\textbf{AxesLabel} -> \{
\text{\[Theta\]}, \text{\[Lambda]/\[Mu]\}, \text{\[Tau\]^*}\}, 
\textbf{ColorFunction} -> \text{Hue}, \textbf{PlotLabel} -> \text{\[Beta\]=0.1}
\textbf{Plot}(\{L(0.5, 2, l), L(1, 2, l), L(1.5, 2, l)\}, \{l, 0, 50\},
\textbf{AxesLabel} -> \{
\text{\[Lambda]/\[Mu]\}, \text{\[Tau\]^*}\}, \textbf{PlotLabel} -> \text{\[Beta\]=2},
\textbf{PlotStyle} -> \{\text{Thin}, \text{Thin, Dashed}, \text{Thick, Dashed}\},
\textbf{PlotLegend} -> \{\text{\{\[Theta\]=0.5\}, \text{\{\[Theta\]=1\}, \text{\{\[Theta\]=1.5\}}},
\textbf{LegendPosition} -> \{0.6, -0.2\}
\textbf{Plot}(\{V(0.5, 0.1, l)/V(1, 0.1, l), L(1, 2, l), L(1.5, 2, l)\}, \{l, 0, 50\},
\textbf{AxesLabel} -> \{
\text{\[Lambda]/\[Mu]\}, \text{\[Tau\]^*}\}, \textbf{PlotLabel} -> \text{\[Beta\]=2},
\textbf{PlotStyle} -> \{\text{Thin}, \text{Thin, Dashed}, \text{Thick, Dashed}\},
\textbf{PlotLegend} -> \{\text{\{\[Lambda\]=0.5\}, \text{\{\[Lambda\]=1\}, \text{\{\[Lambda\]=1.5\}}},
\textbf{LegendPosition} -> \{0.6, -0.2\}
\textbf{Plot3D}(V(t, 0.1, l)/V(1, 0.1, l), \{t, 0, 2\}, \{l, 0, 50\},}
AxesLabel -> {\[Theta], \[Lambda]}, ColorFunction -> Hue,
PlotLabel -> "V^*(\[Theta])/V^*(1); \[Beta]=0.1"

khat[\[Theta]_, b_, l_] = Min[1/\[Theta], Sqrt[\[Theta] + 1]]

Plot[khat[0.01, 0.1, l], {l, 0, 50},
AxesLabel -> {\[Theta]^hat, \[Lambda]/\[Mu], K^hat},
PlotLabel -> "\[Beta]=0.1"

Vhat[\[Theta]_, b_, l_] =
100 - Min[1/\[Theta], Max[0, Sqrt[\[Beta] - \[Theta]] + Sqrt[\[Beta] + \[Theta]]]]

Plot[Vhat[0.01, 1, l], {l, 0, 50},
AxesLabel -> {\[Theta], \[Lambda], V^hat},
ColorFunction -> "Rainbow", PlotLabel -> "\[Beta]=0.5"
<< PlotLegends';
Plot[{khat[0.01, 0.1, l], khat[5, 0.1, l], khat[10, 0.1, l]}, {l, 0, 50},
AxesLabel -> {\[Lambda], K^hat}, PlotLabel -> "\[Beta]=0.1",
PlotStyle -> {Thin, {Thin, Dashed}, {Thick, Dashed}},
PlotLegend -> {\[Theta]^hat=0.01, \[Theta]^hat=5, \[Theta]^hat=10},
LegendPosition -> {0.6, -0.2}
Plot[{Vhat[0.01, 2, l], Vhat[5, 2, l], Vhat[10, 2, l]}, {l, 0, 50},
AxesLabel -> {\[Lambda], V^hat}, PlotLabel -> "\[Beta]=2",
PlotStyle -> {Thin, {Thin, Dashed}, {Thick, Dashed}},
PlotLegend -> {\[Theta]^hat=0.01, \[Theta]^hat=5, \[Theta]^hat=10},
LegendPosition -> {0.6, -0.2}

A.2 Mathematica Code used in Chapter 8.

A.2.1 Analysis of series systems

c1 = 10; c2 = 10; P0 = 1000; l1 = {100, 1000}; l2 = {100, 1000}; p1 = 1; p2 = 1; integer i; integer j; pr1 = {p1, 1 - p1}; pr2 = {p2, 1 - p2}; b1 = 1; b2 = 1; mu1 = 1; mu2 = 1;

val[k1_, k2_] =
Sum[pr1[[i]]*pr2[[j]]*(P0 - b1*l1[[i]]/(k1*mu1) - b2*l2[[j]]/(k2*mu2)), {i, 1, 2}, {j, 1, 2}];

optval = Maximize[{val[k1, k2], k1 >= 0, k2 >= 0}, {k1, k2}]
data = List[(optval[[2, 1, 2]], optval[[2, 2, 2]]])

opt = ListPlot[data, PlotRange -> {{0, 100}, {0, 100}},
PlotStyle -> {Brown, PointSize[0.02]}]

irs1[k1_, k2_] = c1*k1/(c1*k1 + c2*k2);

irs2[k1_, k2_] = 1 - irs1[k1, k2];
\[ \text{tirs1}[k1, k2] = \text{Min}[\text{irs1}[k1, k2], \text{irs1}[\text{optval}[2, 1, 2], \text{optval}[2, 2, 2]]]; \]
\[ \text{tirs2}[k1, k2] = \text{Min}[\text{irs2}[k1, k2], \text{irs2}[\text{optval}[2, 1, 2], \text{optval}[2, 2, 2]]]; \]
\[ \text{val1}[k1, k2] = \sum \text{pr1}[i] \cdot \text{pr2}[j] \cdot (\text{tirs1}[k1, k2] \cdot (P0 - b1 \cdot l1[i]/(k1 \cdot \mu1) - b2 \cdot l2[j]/(k2 \cdot \mu2)), \{i, 1, 2\}, \{j, 1, 2\}) - c1 \cdot k1; \]
\[ \text{val1irsguar}[k1, k2] = \sum \text{pr1}[i] \cdot \text{pr2}[j] \cdot (\text{Max}[\text{tirs1}[k1, k2] \cdot (P0 - b1 \cdot l1[i]/(k1 \cdot \mu1) - b2 \cdot l2[j]/(k2 \cdot \mu2)), \{i, 1, 2\}, \{j, 1, 2\}) - c1 \cdot k1; \]
\[ \text{val1irsncomp}[k1, k2] = \sum \text{pr1}[i] \cdot \text{pr2}[j] \cdot (\text{tirs1}[k1, k2] \cdot (P0 - b1 \cdot l1[i]/(k1 \cdot \mu1) - b2 \cdot l2[j]/(k2 \cdot \mu2)) - (\text{optval}[1] - \text{val}[\text{optval}[2, 1, 2], k2])), \{i, 1, 2\}, \{j, 1, 2\}) - c1 \cdot k1; \]
\[ \text{val2}[k1, k2] = \sum \text{pr1}[i] \cdot \text{pr2}[j] \cdot (\text{tirs2}[k1, k2] \cdot (P0 - b1 \cdot l1[i]/(k1 \cdot \mu1) - b2 \cdot l2[j]/(k2 \cdot \mu2)), \{i, 1, 2\}, \{j, 1, 2\}) - c2 \cdot k2; \]
\[ \text{val2irsguar}[k1, k2] = \sum \text{pr1}[i] \cdot \text{pr2}[j] \cdot (\text{Max}[\text{tirs2}[k1, k2] \cdot (P0 - b1 \cdot l1[i]/(k1 \cdot \mu1) - b2 \cdot l2[j]/(k2 \cdot \mu2)), \{i, 1, 2\}, \{j, 1, 2\}) - c2 \cdot k2; \]
\[ \text{val2irsncomp}[k1, k2] = \sum \text{pr1}[i] \cdot \text{pr2}[j] \cdot (\text{tirs2}[k1, k2] \cdot (P0 - b1 \cdot l1[i]/(k1 \cdot \mu1) - b2 \cdot l2[j]/(k2 \cdot \mu2)) - (\text{optval}[1] - \text{val}[\text{optval}[2, 1, 2], k2])), \{i, 1, 2\}, \{j, 1, 2\}) - c2 \cdot k2; \]
\[ \text{maxval1cent}[k2] = \text{ArgMax}[\text{val}[k1, k2], 0.001 \leq k1 \leq 1000, \{k1\}]; \]
\[ \text{maxval1}[k2] = \text{ArgMax}[\text{val1}[k1, k2], 0.001 \leq k1 \leq 1000, \{k1\}]; \]
\[ \text{maxval2cent}[k1] = \text{ArgMax}[\text{val}[k1, k2], 0.001 \leq k2 \leq 1000, \{k2\}]; \]
\[ \text{maxval2}[k1] = \text{ArgMax}[\text{val2}[k1, k2], 0.001 \leq k2 \leq 1000, \{k2\}]; \]
\[ m1 = \]
maxval1[optval[[2, 2, 2]]]; m2 = maxval2[optval[[2, 1, 2]]]; m2cent =
maxval1cent[optval[[2, 2, 2]]]; m2cent =
maxval2cent[optval[[2, 1, 2]]];

exprvalue = ToExpression["Value", TeXForm]; exprval1 =
ToExpression["V_1(K_1,K^*_2)", TeXForm]; exprval1 =
ToExpression["V(K_1,K^*_2)", TeXForm]; Plot[
{val1[k1, optval[[2, 2, 2]]],
val1[k1, optval[[2, 2, 2]]]}, {k1, 0, 10},
PlotStyle -> {Thick, (Thick, Dashed)},
AxesLabel -> {exprk1, exprvalue},
PlotLegend -> {exprval1, exprval1}, LegendPosition -> (0.4, -0.9),
LegendShadow -> None, LegendTextSpace -> 5]

exprvalue = ToExpression["Value", TeXForm]; exprval2 =
ToExpression["V_2(K^*_1,K_2)", TeXForm]; exprval2 =
ToExpression["V(K^*_1,K_2)", TeXForm]; Plot[
{val2[optval[[2, 1, 2]], k2],
val2[optval[[2, 1, 2]], k2]}, {k2, 0, 10},
AxesLabel -> {exprk2, exprvalue},
PlotLegend -> {exprval2, exprval2}, LegendPosition -> (0.4, -0.9),
LegendShadow -> None, LegendTextSpace -> 5]

exprvalue = ToExpression["Value", TeXForm]; exprval1 =
ToExpression["V_1(K_1,K^*_2)", TeXForm]; exprval1 =
ToExpression["V(K_1,K^*_2)", TeXForm]; exprval1 =
ToExpression["V^{TIRS-G}_1(K_1,K^*_2)", TeXForm]; exprval1 =
ToExpression["V^{TIRS-N}_1(K_1,K^*_2)", TeXForm]; Plot[
{val1[k1, optval[[2, 2, 2]]],
val1irsguar[k1, optval[[2, 2, 2]]],
val1irsncomp[k1, optval[[2, 2, 2]]]}, {k1, 0, 10},
PlotStyle -> {Thick, (Thick, Dashed), (Thick, Dashing[Large]), (Thick, Dashing[Tiny])},
AxesLabel -> {exprk1, exprvalue},
PlotLegend -> {exprval1, exprval1, exprval1guar, exprval1ncomp},
LegendPosition -> (0.4, -0.9), LegendShadow -> None,
LegendTextSpace -> 5]
\begin{verbatim}
TeXForm; Plot[{val[optval[[2, 1, 2]], k2],
val2[optval[[2, 1, 2]], k2], val2irsgaur[optval[[2, 1, 2]], k2],
val2irsncomp[optval[[2, 1, 2]], k2]}, {k2, 0, 10},
PlotStyle -> {Thick, (Thick, Dashed), (Thick,
Dashing[Large]), (Thick, Dashing[Tiny])},
AxesLabel -> {exprk2, exprvalue},
PlotLegend -> {exprval2, expr2val2, expr1val2guar, expr1val2ncomp},
LegendPosition -> {0.4, -0.9}, LegendShadow -> None,
LegendTextSpace -> 5]
datac = List[optval[[2, 2]], optval[[2, 2]]]; data1 =
List[m1[[1]], optval[[2, 2]]];
data2 =
List[optval[[2, 1, 2]], m2[[1]]]; uoinvest =
ListPlot[{datac, data1, data2}, PlotRange -> All,
PlotStyle -> {{Brown, PointSize[0.03]}, {Red,
PointSize[0.03]}, {Blue, PointSize[0.03]}},
AxesLabel -> {exprk1, exprk2}]
exprk1 = ToExpression["K_1", TeXForm]; exprk2 =
ToExpression["K_2", TeXForm];
exprk1stark2star =
ToExpression["\{K^*_1,K^*_2\", TeXForm]; exprk1star2o =
ToExpression["\{K^1,K^o_2\", TeXForm]; exprk1o2star =
ToExpression["\{K^o_1,K^*_2\", TeXForm];
exprval1 = ToExpression["V_1(K_1)", TeXForm]; exprval2 =
ToExpression["K_1", TeXForm]
exprval =
ToExpression["(V(K^*_1,K^*_2)-V(K_1,K_2))/(Abs(V(K^*_1,K^*_2)))", TeXForm]
exprk20 = ToExpression["K_2=0", TeXForm]; exprk225 =
ToExpression["K_2=25", TeXForm]; exprk250 =
ToExpression["K_2=50", TeXForm]; exprk2100 =
ToExpression["K_2=100", TeXForm]; exprk2150 =
ToExpression["K_2=150", TeXForm]; exprk2200 =
ToExpression["K_2=200", TeXForm];
(optval[[1]] - val180, 50)/Abs[optval[[1]]]
valk1diff =
Plot[{{optval[[1]] - val[k1, 1],
Abs[optval[[1]]], (optval[[1]] - val[k1, 25])/ Abs[optval[[1]]], (optval[[1]] - val[k1, 50])/ Abs[optval[[1]]], (optval[[1]] - val[k1, 100])/ Abs[optval[[1]]], (optval[[1]] - val[k1, 200])/

510
\end{verbatim}
A.2.2 Analysis of parallel systems

c1 = 10; c2 = 10; P0 = 1000; p1 = 1; p2 = 1; integer i; integer j; 
pr1 = {p1, 1 - p1}; pr2 = {p2, 1 - p2}; l1 = {100, 1000}; l2 = {100, 1000}; b = 1; mu1 = 1; mu2 = 1;

```
Needs("PlotLegends")
```

val[k1_, k2_] = 
Sum[pr1[[i]]*pr2[[j]]*(P0 - b*Max[l1[[i]]/(k1*mu1), l2[[j]]/(k2*mu2)]), {i, 1, 2}, {j, 1, 2}] - c1*k1 - c2*k2

```
optval = Maximize[{val[k1, k2], k1 >= 0, k2 >= 0}, {k1, k2}]
```

data = List[{{optval[[1, 1]]}, {optval[[1, 2]]}}]

```
opt = ListPlot[data, PlotRange -> {{0, 100}, {0, 100}},
PlotStyle -> {Brown, PointSize[0.02]}]
```

irs1[k1_, k2_] = c1*k1/(c1*k1 + c2*k2);

irs2[k1_, k2_] = 1 - irs1[k1, k2];
\[
tirs1[k_1, k_2_] = \text{Min}[irs1[k_1, k_2], \text{irs}[\text{optval}[[2, 1, 2]], \text{optval}[[2, 2, 2]]]];
\]
\[
tirs2[k_1, k_2_] = \text{Min}[irs2[k_1, k_2], \text{irs2}[\text{optval}[[2, 1, 2]], \text{optval}[[2, 2, 2]]]];
\]
\[
val1[k_1, k_2_] = \sum \text{pr1}[i] \times \text{pr2}[j] \times (tirs1[k_1, k_2] \times (P0 - b \times \text{Max}[l1[i]/(k_1 \times \mu1), l2[j]/(k_2 \times \mu2)]), \{i, 1, 2\}, \{j, 1, 2\}) - c1 \times k1;
\]
\[
val1irsguar[k_1, k_2_] = \sum \text{pr1}[i] \times \text{pr2}[j] \times (\text{Max}[tirs1[k_1, k_2] \times (P0 - b \times \text{Max}[l1[i]/(k_1 \times \mu1), l2[j]/(k_2 \times \mu2)]), \text{ci} \times \text{Min}[k_1, \text{optval}[[2, 1, 2]]], \{i, 1, 2\}, \{j, 1, 2\}) - c1 \times k1;
\]
\[
val1irsncomp[k_1, k_2_] = \sum \text{pr1}[i] \times \text{pr2}[j] \times (tirs1[k_1, k_2] \times (P0 - b \times \text{Max}[l1[i]/(k_1 \times \mu1), l2[j]/(k_2 \times \mu2)]) - (\text{optval}[[1]] - val[k_1, \text{optval}[[2, 2, 2]]]), \{i, 1, 2\}, \{j, 1, 2\}) - c1 \times k1;
\]
\[
val2[k_1, k_2_] = \sum \text{pr1}[i] \times \text{pr2}[j] \times (irs2[k_1, k_2] \times (P0 - b \times \text{Max}[l1[i]/(k_1 \times \mu1), l2[j]/(k_2 \times \mu2)]), \{i, 1, 2\}, \{j, 1, 2\}) - c2 \times k2;
\]
\[
val2irsguar[k_1, k_2_] = \sum \text{pr1}[i] \times \text{pr2}[j] \times (\text{Max}[tirs2[k_1, k_2] \times (P0 - b \times \text{Max}[l1[i]/(k_1 \times \mu1), l2[j]/(k_2 \times \mu2)]), \text{ci} \times \text{Min}[k_2, \text{optval}[[2, 2, 2]]], \{i, 1, 2\}, \{j, 1, 2\}) - c2 \times k2;
\]
\[
val2irsncomp[k_1, k_2_] = \sum \text{pr1}[i] \times \text{pr2}[j] \times (tirs2[k_1, k_2] \times (P0 - b \times \text{Max}[l1[i]/(k_1 \times \mu1), l2[j]/(k_2 \times \mu2)]) - (\text{optval}[[1]] - val[\text{optval}[[2, 1, 2]], k_2]), \{i, 1, 2\}, \{j, 1, 2\}) - c2 \times k2;
\]
\[
\text{maxval1cent}[k_2_] = \text{ArgMax}[\text{val}[k_1, k_2] \times 0.001 \leq k_1 \leq 1000, \{k_1\}];
\]
\[
\text{maxval1}[k_2_] = \text{ArgMax}[\text{val1}[k_1, k_2] \times 0.001 \leq k_1 \leq 1000, \{k_1\}];
\]
\[
\text{maxval2cent}[k_1_] = \text{ArgMax}[\text{val}[k_1, k_2] \times 0.001 \leq k_2 \leq 1000, \{k_2\}];
\]
\[
\text{maxval2}[k_1_] = \text{ArgMax}[\text{val2}[k_1, k_2] \times 0.001 \leq k_2 \leq 1000, \{k_2\}];
\]
\[
m1 = 512
\]
val2irsncomp[optval[[2, 1, 2]], k2]}, {k2, 0, 10},
PlotStyle -> {Thick, Thick, Dashed, Thick,
  Dashing[Large], Thick, Dashing[Tiny])},
AxesLabel -> {exprk2, exprvalue},
PlotLegend -> {exprval2, expr2val2, exprval2guar, exprval2ncomp},
LegendPosition -> {0.4, -0.9}, LegendShadow -> None,
LegendTextSpace -> 5]
datac = List[optval[[2, 1, 2]], optval[[2, 2, 2]]];
data1 = List[m1[[1]], optval[[2, 2, 2]]];
data2 = List[optval[[2, 1, 2]], m2[[1]]];
unoinvest = ListPlot[{data, data1, data2}, PlotRange -> All,
  PlotStyle -> {{Brown, PointSize[0.03]}, {Red, PointSize[0.03]}, {Blue, PointSize[0.03]}},
  AxesLabel -> {exprk1, exprk2}]
expr1val1 = ToExpression["V_1(K_1)", TeXForm]; expr2val1 =
  ToExpression["K_1", TeXForm];
expr1k20 = ToExpression["K_2=0", TeXForm]; expr1k225 =
  ToExpression["K_2=25", TeXForm]; expr1k250 =
  ToExpression["K_2=50", TeXForm]; expr1k2100 =
  ToExpression["K_2=100", TeXForm]; expr1k2150 =
  ToExpression["K_2=150", TeXForm]; expr1k2200 =
  ToExpression["K_2=200", TeXForm];
exprval =
  ToExpression["(V(K^*_1,K^*_2)-V(K_1,K_2))/(Abs(V(K^*_1,K^*_2)))", TeXForm];
val1k1diff =
  Plot[{(optval[[1]] - val[k1, 1])/Abs[optval[[1]]], (optval[[1]] - val[k1, 25])/Abs[optval[[1]]],
  (optval[[1]] - val[k1, 50])/Abs[optval[[1]]], (optval[[1]] - val[k1, 100])/Abs[optval[[1]]],
  (optval[[1]] - val[k1, 200])/Abs[optval[[1]]], (optval[[1]] - val[k1, 1])}/Abs[optval[[1]]],
  {k1, 0, 100},
  PlotStyle -> {Thick, Blue, Dashed, Thick, Red, Dashed, Thick, Green, Dashed, Thick, Brown, Dashed, Thick, Orange, Dashed},
  AxesLabel -> {expr2val1, expr1val}]
val1kt =
  Plot[{val[k1, 1], val1[k1, 25], val1[k1, 50], val1[k1, 100],
  val1[k1, 200]}, {k1, 0, 100},
  PlotLegend -> {expr1k20, expr1k225, expr1k250, expr1k2100},
  AxesLabel -> {expr2val1, expr1val}]
514
A.3 Mathematica Code used in Chapter 10.

A.3.1 Analysis of series systems

c1 = 1; c2 = 1; r = 5; p1 = 0.25; p2 = 1; integer i; integer j; pr1 = 
{p1, 1 - p1}; pr2 = {p2, 1 - p2};
d = {100, 200}; b = 1; co1 = 1; co2 = 1;
Needs["PlotLegends'"]

val[k1_, k2_] =
Sum[pr1[[i]]*(r*Min[d[[i]], k1, k2] - b*Max[d[[i]] - Min[k1, k2], 0] - co1*Min[d[[i]], k1, k2] -
co2*Min[d[[i]], k1, k2]), {i, 1, 2}] - c1*k1 - c2*k2

optval = Maximize[{val[k1, k2], k1 >= 0, k2 >= 0}, {k1, k2}]
data = List[{optval[[2, 2, 2]], optval[[2, 1, 2]]}]

opt = ListPlot[data, PlotRange -> {{0, 500}, {0, 500}},
PlotStyle -> {Brown, PointSize[0.02]}]

irs1[k1_, k2_] = c1*k1/(c1*k1 + c2*k2);
irs2[k1_, k2_] = 1 - irs1[k1, k2];
tirs1[k1_, k2_] =
Min[irs1[k1, k2], irs1[optval[[2, 1, 2]], optval[[2, 2, 2]]]];
tirs2[k1_, k2_] =
Min[irs2[k1, k2], irs2[optval[[2, 1, 2]], optval[[2, 2, 2]]]];

val1[k1_, k2_] =
Sum[pr1[[i]]*(irs1[k1, k2]*(r*Min[d[[i]], k1, k2] - b*Max[d[[i]] - Min[k1, k2], 0]) -
\[\text{val1irsguar}[k_1, k_2] = \sum_{i=1}^{2} \text{pr}_{1_i} \left( r \cdot \min\{d[i], k_1, k_2\} - b \cdot \max\{d[i] - \min\{k_1, k_2\}, 0\} \right) - \text{co}_1 \cdot \min\{d[i], k_1, k_2\} - c_1 \cdot k_1;\]

\[\text{val1irsncomp}[k_1, k_2] = \sum_{i=1}^{2} \text{pr}_{1_i} \left( \max\{tirs_1[k_1, k_2] \cdot (r \cdot \min\{d[i], k_1, k_2\} - b \cdot \max\{d[i] - \min\{k_1, k_2\}, 0\}), c_1 \cdot \min\{k_1, \text{optval}[2, 1, 2]\} + \text{co}_1 \cdot \min\{d[i], k_1, k_2\}\right) - \text{co}_1 \cdot \min\{d[i], k_1, k_2\} - (\text{optval}[1] - \text{val}[k_1, \text{optval}[2, 2, 2]]), (i, 1, 2) - c_1 \cdot k_1;\]

\[\text{val2}[k_1, k_2] = \sum_{i=1}^{2} \text{pr}_{1_i} \left( \text{irs}_2[k_1, k_2] \cdot (r \cdot \min\{d[i], k_1, k_2\} - b \cdot \max\{d[i] - \min\{k_1, k_2\}, 0\}) - \text{co}_2 \cdot \min\{d[i], k_1, k_2\}\right) - c_2 \cdot k_2;\]

\[\text{val2irsguar}[k_1, k_2] = \sum_{i=1}^{2} \text{pr}_{1_i} \left( \max\{tirs_2[k_1, k_2] \cdot (r \cdot \min\{d[i], k_1, k_2\} - b \cdot \max\{d[i] - \min\{k_1, k_2\}, 0\}), c_2 \cdot \min\{k_2, \text{optval}[2, 2, 2]\} + \text{co}_2 \cdot \min\{d[i], k_1, k_2\}\right) - \text{co}_2 \cdot \min\{d[i], k_1, k_2\}, (i, 1, 2) - c_2 \cdot k_2;\]

\[\text{val2irsncomp}[k_1, k_2] = \sum_{i=1}^{2} \text{pr}_{1_i} \left( tirs_2[k_1, k_2] \cdot (r \cdot \min\{d[i], k_1, k_2\} - b \cdot \max\{d[i] - \min\{k_1, k_2\}, 0\}) - \text{co}_2 \cdot \min\{d[i], k_1, k_2\} - (\text{optval}[1] - \text{val}[\text{optval}[2, 2, 2], k_2]), (i, 1, 2) - c_2 \cdot k_2;\]

\[\text{maxval1cent}[k_2] = \text{ArgMax}\{\text{val}[k_1, k_2], 0 \leq k_1 \leq 1000, (k_1)\};\]

\[\text{maxval1}[k_2] = \text{ArgMax}\{\text{val}[k_1, k_2], 0 \leq k_1 \leq 1000, (k_1)\};\]

\[\text{maxval2cent}[k_1] = \text{ArgMax}\{\text{val}[k_1, k_2], 0 \leq k_2 \leq 1000, (k_2)\};\]

\[\text{maxval2}[k_1] = \text{ArgMax}\{\text{val}[k_1, k_2], 0 \leq k_2 \leq 1000, (k_2)\};\]

\[\text{m}_1 = \text{maxval}[\text{optval}[2, 2, 2]]; \text{m}_2 = \text{maxval}[\text{optval}[2, 1, 2]]; \text{nicent} = \text{maxvalcent}[\text{optval}[2, 2, 2]]; \text{n2cent} = \text{maxval2cent}[\text{optval}[2, 2, 2]];\]

\[\text{exprk1} = \text{ToExpression}["\text{K}_1", \text{TeXForm}]; \text{exprk2} = \text{ToExpression}["\text{K}_2", \text{TeXForm}];\]

\[\text{exprvalue} = \text{ToExpression}["\text{Value}", \text{TeXForm}]; \text{exprval1} = \text{ToExpression}["\text{V}_1(\text{K}_1, \text{K}_2)\text{\^*}_2", \text{TeXForm}]; \text{exprval2} = \text{ToExpression}["\text{V}_2(\text{K}_1, \text{K}_2)\text{\^*}_2", \text{TeXForm}].\]
List[{m1[[1]], optval[[2, 2, 2]]}]; data2 = List[optval[[2, 1, 2]], m2[[1]]]; woinvest = ListPlot[{data1, data2}, PlotRange -> All, PlotStyle -> {{Brown, PointSize[0.03]}, {Red, PointSize[0.03]}, {Blue, PointSize[0.03]}}, AxesLabel -> {exprk1, exprk2}]
exprval1 = ToExpression["V_1(K_1)", TeXForm]; expr2val1 = ToExpression["K_1”, TeXForm]; expr1k20 = ToExpression["K_2=0", TeXForm]; expr1k225 = ToExpression["K_2=25", TeXForm]; expr1k250 = ToExpression["K_2=50", TeXForm]; expr1k2100 = ToExpression["K_2=100", TeXForm]; expr1k2200 = ToExpression["K_2=200", TeXForm];
exprval = ToExpression["(V(K^*_1,K^*_2)-V(K_1,K_2))/(Abs(V(K^*_1,K^*_2)))", TeXForm]
valk1idiff = Plot[{(optval[[1]] - val[k1, 1])/Abs[optval[[1]]], (optval[[1]] - val[k1, 25])/Abs[optval[[1]]], (optval[[1]] - val[k1, 50])/Abs[optval[[1]]], (optval[[1]] - val[k1, 100])/Abs[optval[[1]]], (optval[[1]] - val[k1, 200])/Abs[optval[[1]]], (optval[[1]] - val[k1, 1])/Abs[optval[[1]]]}, {k1, 0, 100}, PlotStyle -> {{Thick, Blue, Dashed}, {Thick, Red, Dashed}, {Thick, Green, Dashed}, {Thick, Brown, Dashed}, {Thick, Orange, Dashed}}, AxesLabel -> {expr2val1, expr1val}]
val1k1 = Plot[{val1[k1, 0], val1[k1, 25], val1[k1, 50], val1[k1, 100], val1[k1, 200]}, {k1, 0, 300}, PlotLegend -> {exprk20, exprk225, exprk250, exprk2100, exprk2200}, AxesLabel -> {expr2val1, exprval1}]
valk1 = Plot[{val[k1, 0], val[k1, 25], val[k1, 50], val[k1, 100], val[k1, 200]}, {k1, 0, 300}, PlotLegend -> {exprk20, exprk225, exprk250, exprk2100, exprk2200}];
expr1 = ToExpression["K_1^o(K_2)", TeXForm]; expr2 = ToExpression["K_2", TeXForm];
expr3 = ToExpression["K_2^o(K_1)", TeXForm]; expr4 = ToExpression["K_1", TeXForm];

maxval1plot = Plot[maxval1[k2], {k2, 0, 500}, AxesLabel -> {expr2, expr1}, PlotStyle -> {Thick, Red}, PlotRange -> {{0, 500}, {0, 500}}];
Show[maxval1plot, opt];

maxval2plot = Plot[maxval2[k1], {k1, 0, 500}, AxesLabel -> {exprk1, expr3}, PlotStyle -> {Thick, Blue}, PlotRange -> {{0, 500}, {0, 500}}];

A.3.2 Analysis of Parallel systems

\[\text{c1} = 1; \text{c2} = 1; \text{r1} = 10; \text{r2} = 5; \text{d1} = \{100, 200\}; \text{d2} = \{100, 200\}; \text{p1} = 1; \text{p2} = 1; \text{integer} \ i; \text{integer} \ j;\]

\[\text{pr1} = \{\text{p1}, 1 - \text{p1}\}; \text{pr2} = \{\text{p2}, 1 - \text{p2}\};\]

\[\text{b1} = 1; \text{b2} = 1; \text{co1} = 1; \text{co2} = 1;\]

\[\text{val}[k1_, k2_] = \sum_{i=1}^{2} \sum_{j=1}^{2} \text{pr1}[i] \times \text{pr2}[j] \times (\text{r1} \times \text{Min}[d1[i], k1] + \text{r2} \times \text{Min}[d2[j], k2] - \text{b1} \times \text{Max}[d1[i] - k1, 0] - \text{b2} \times \text{Max}[d2[j] - k2, 0] - \text{co1} \times \text{Min}[d1[i], k1] - \text{co2} \times \text{Min}[d2[j], k2]), \{i, 1, 2\}, \{j, 1, 2\} - \text{c1} \times k1 - \text{c2} \times k2\]

Maximize[{val[k1, k2], k1 >= 0, k2 >= 0}, {k1, k2}];
optval = Maximize[{val[k1, k2], k1 >= 0, k2 >= 0}, {k1, k2}];
data = List[{optval[[2, 1, 2]], optval[[2, 2, 2]]}];

e = ListPlot[data, PlotRange -> {{0, 500}, {0, 500}}];

irs1[k1_, k2_] = \frac{\text{c1} \times k1}{\text{c1} \times k1 + \text{c2} \times k2};
irs2[k1_, k2_] = 1 - irs1[k1, k2];
tirs1[k1_, k2_] = Min[irs1[k1, k2], irs1[optval[[2, 1, 2]], optval[[2, 2, 2]]]];
tirs2[k1_, k2_] = Min[irs2[k1, k2], irs2[optval[[2, 1, 2]], optval[[2, 2, 2]]]];

val1[k1_, k2_] = \sum_{i=1}^{2} \sum_{j=1}^{2} \text{pr1}[i] \times \text{pr2}[j] \times (\text{irs1}[k1, k2] \times (\text{r1} \times \text{Min}[d1[i], k1] + \text{r2} \times \text{Min}[d2[j], k2] - 5) - \text{c1} \times k1 - \text{c2} \times k2\]
\begin{verbatim}

val1irsguar[k1_, k2_] = 
  Sum[pr1[[i]]*pr2[[j]]*(Max[
    tirs1[k1, k2]*r1*Min[d1[[i]], k1] + r2*Min[d2[[j]], k2] -
    b1*Max[d1[[i]] - k1, 0] - b2*Max[d2[[j]] - k2, 0],
    c1*Min[k1, optval[2, 1, 2]] + co1*Min[d1[[i]], k1]] -
    c1*k1, {i, 1, 2}, {j, 1, 2}] - c1*k1;

val1irsncomp[k1_, k2_] = 
  Sum[pr1[[i]]*pr2[[j]]*(tirs1[k1, k2]*r1*Min[d1[[i]], k1] + r2*Min[d2[[j]], k2] -
    b1*Max[d1[[i]] - k1, 0] - b2*Max[d2[[j]] - k2, 0]) -
    c1*Min[k1, optval[2, 1, 2]] - (optval[1] -
    val[k1, optval[2, 2, 2]]), {i, 1, 2}, {j, 1, 2}] - c1*k1;

val2[k1_, k2_] = 
  Sum[pr1[[i]]*pr2[[j]]*(irs2[k1, k2]*r1*Min[d1[[i]], k1] + r2*Min[d2[[j]], k2] -
    b1*Max[d1[[i]] - k1, 0] - b2*Max[d2[[j]] - k2, 0]) -
    c2*Min[k2, optval[2, 2, 2]] + co2*Min[d2[[j]], k2]], {i, 1, 2}, {j, 1, 2}] - c2*k2;

val2irsguar[k1_, k2_] = 
  Sum[pr1[[i]]*pr2[[j]]*(Max[
    tirs2[k1, k2]*r1*Min[d1[[i]], k1] + r2*Min[d2[[j]], k2] -
    b1*Max[d1[[i]] - k1, 0] - b2*Max[d2[[j]] - k2, 0],
    c2*Min[k2, optval[2, 2, 2]] + co2*Min[d2[[j]], k2]] -
    c2*k2, {i, 1, 2}, {j, 1, 2}] - c2*k2;

val2irsncomp[k1_, k2_] = 
  Sum[pr1[[i]]*pr2[[j]]*(tirs2[k1, k2]*r1*Min[d1[[i]], k1] + r2*Min[d2[[j]], k2] -
    b1*Max[d1[[i]] - k1, 0] - b2*Max[d2[[j]] - k2, 0]) -
    c2*Min[k2, optval[2, 2, 2]] - (optval[1] -
    val[optval[2, 2, 2], k2]), {i, 1, 2}, {j, 1, 2}] - c2*k2;

maxval1cent[k2_] = ArgMax[val[k1, k2], 0 <= k1 <= 1000, {k1}];

maxval1[k2_] = ArgMax[val1[k1, k2], 0 <= k1 <= 1000, {k1}];

maxval2cent[k2_] = ArgMax[val2[k1, k2], 0 <= k1 <= 1000, {k1}];

maxval2[k2_] = ArgMax[val2[k1, k2], 0 <= k1 <= 1000, {k1}];

\end{verbatim}
maxval2cent[k1_] = ArgMax[val[k1, k2], 0 <= k2 <= 1000, {k2}];
maxval12[k1_] = ArgMax[val2[k1, k2], 0 <= k2 <= 1000, {k2}]; m1 =
maxval1cent[optval[{2, 2, 2}]]; m2 = maxval2[optval[{2, 1, 2}]]; n1cent =
maxval1cent[optval[{2, 2, 2}]]; n2cent =
maxval2cent[optval[{2, 1, 2}]];
exprk1 = ToExpression["K_1", TeXForm]; exprk2 =
ToExpression["K_2", TeXForm]; exprvalue = ToExpression["Value", TeXForm]; exprval1 =
ToExpression["V_1(K_1,K^*_2)", TeXForm]; exprval2 =
ToExpression["V(K^*_1,K_2)", TeXForm]; Plot[{val[k1, optval[{2, 2, 2}]],
val1[k1, optval[{2, 2, 2}]], {k1, 0, 500},
AxesLabel -> {exprk1, exprvalue},
PlotLegend -> {exprval1, exprval1}, LegendPosition -> (0.4, -0.9),
LegendShadow -> None, LegendTextSpace -> 5]
exprvalue = ToExpression["Value", TeXForm]; exprval2 =
ToExpression["V_2(K^*_1,K_2)", TeXForm]; exprval2 =
ToExpression["V(K^*_1,K_2)", TeXForm]; Plot[{val[optval[{2, 1, 2}]],
val2[optval[{2, 1, 2}], k2], {k2, 0, 500},
AxesLabel -> {exprk2, exprvalue},
PlotLegend -> {exprval2, exprval2}, LegendPosition -> (0.4, -0.9),
LegendShadow -> None, LegendTextSpace -> 5]
exprval1guar =
ToExpression["V^{TIRS-G}_1(K_1,K^*_2)", TeXForm]; exprval1ncomp =
ToExpression["V^{TIRS-N}_1(K_1,K^*_2)", TeXForm]; Plot[{val[k1, optval[{2, 2, 2}]],
val1[k1, optval[{2, 2, 2}]], val1irsguar[k1, optval[{2, 2, 2}]],
val1irsncomp[k1, optval[{2, 2, 2}]], {k1, 0, 500},
AxesLabel -> {exprk1, exprvalue},
PlotLegend -> {exprval1, exprval1, exprval1guar, exprval1ncomp},
LegendPosition -> (0.4, -0.9), LegendShadow -> None,
LegendTextSpace -> 5]
exprval2guar =
ToExpression["V^{TIRS-G}_2(K^*_1,K_2)", TeXForm]; exprval2ncomp =
ToExpression["V^{TIRS-N}_2(K^*_1,K_2)", TeXForm];
\text{TeXForm}; \text{Plot}\{\{\text{val}[[2, 1, 2]], k2\},

\text{val2}(\text{optval}[[2, 1, 2]], k2), \text{val2irsguar}(\text{optval}[[2, 1, 2]], k2),

\text{val2irsncomp}(\text{optval}[[2, 1, 2]], k2)\}, \{k2, 0, 500\},

\text{PlotStyle} \to \{\text{Thick}, \{\text{Thick}, \text{Dashed}\}, \{\text{Thick},

\text{Dashing}[\text{Large}]\}, \{\text{Thick}, \text{Dashing}[\text{Tiny}]\}\},

\text{AxesLabel} \to \{\text{exprk2}, \text{exprvalue}\},

\text{PlotLegend} \to \{\text{exprval2}, \text{expr2val2}, \text{expr1val2guar}, \text{expr1val2ncomp}\},

\text{LegendPosition} \to \{0.4, -0.9\}, \text{LegendShadow} \to \text{None},

\text{LegendTextSpace} \to 5\}

\text{datac = List}\{\{\text{optval}[[2, 1, 2]], \text{optval}[[2, 2, 2]]\}\}; \text{data1 =}

\text{List}\{\{\text{m1}[[1]], \text{optval}[[2, 2, 2]]\}\}; \text{data2 =}

\text{List}\{\{\text{optval}[[2, 1, 2]], \text{m2}[[1]]\}\}; \text{uininvest =}

\text{ListPlot}\{\text{data, data1, data2}, \text{PlotRange} \to \text{All},

\text{PlotStyle} \to \{\{\text{Brown}, \text{PointSize}[0.03]\}, \{\text{Red},

\text{PointSize}[0.03]\}, \{\text{Blue}, \text{PointSize}[0.03]\}\},

\text{AxesLabel} \to \{\text{exprk1}, \text{exprk2}\}\}

\text{exprval1 = ToExpression["V_1(K_1)", \text{TeXForm}]}; \text{expr2val1 =}

\text{ToExpression["K_1", \text{TeXForm}]}; \text{expr1k20 =}

\text{ToExpression["K_2=0", \text{TeXForm}]}; \text{expr1k225 =}

\text{ToExpression["K_2=25", \text{TeXForm}]}; \text{expr1k250 =}

\text{ToExpression["K_2=50", \text{TeXForm}]}; \text{expr1k2100 =}

\text{ToExpression["K_2=100", \text{TeXForm}]}; \text{expr1k2150 =}

\text{ToExpression["K_2=150", \text{TeXForm}]}; \text{expr1k2200 =}

\text{ToExpression["K_2=200", \text{TeXForm}]};

\text{Needs["PlotLegends"]}

\text{exprval =}

\text{ToExpression["(V(K^*_1,K^*_2)-V(K_1,K_2))/(Abs(V(K^*_1,K^*_2)))", \text{TeXForm}]},

\text{val1k1diff =}

\text{Plot}\{\{\text{optval}[[1]] - \text{val}[k1, 1]/

\text{Abs}[\text{optval}[[1]]], \{\text{optval}[[1]] - \text{val}[k1, 25]/

\text{Abs}[\text{optval}[[1]]], \{\text{optval}[[1]] - \text{val}[k1, 50]/

\text{Abs}[\text{optval}[[1]]], \{\text{optval}[[1]] - \text{val}[k1, 100]/

\text{Abs}[\text{optval}[[1]]], \{\text{optval}[[1]] - \text{val}[k1, 200]/

\text{Abs}[\text{optval}[[1]]], \{\text{optval}[[1]] - \text{val}[k1, 1]/

\text{Abs}[\text{optval}[[1]]]\}, \{k1, 0, 100\},

\text{PlotStyle} \to \{\{\text{Thick}, \text{Blue}, \text{Dashed}\}, \{\text{Thick}, \text{Red},

\text{Dashed}\}, \{\text{Thick}, \text{Green}, \text{Dashed}\}, \{\text{Thick}, \text{Brown},

\text{Dashed}\}, \{\text{Thick}, \text{Orange}, \text{Dashed}\}\},

\text{AxesLabel} \to \{\text{expr2val1}, \text{exprval1}\}\}

\text{valkt =}
A.3.3 Analysis of Make-to-Stock systems

\[d_1 = \{100, 200, 300\}; d_2 = \{100, 200, 300\};\]
\[rvalsmts = \{15, 20\}; bvalsmts = \{1, 5\}; covalsmts = \{1, 5\}; hcostfactor = \{0.05, 0.25\}; cvalsmts = \{1, 5\}; icapfactor = \{0.05, 0.25\};\]
\[p11 = 0.2; p12 = 0.6; p13 = 0.2; p21 = 0.2; p22 = 0.6; p23 = 0.2; pr1 \]
\[= \{p11, p12, p13\}; pr2 = \{p21, p22, p23\};\]

integer i; integer j; integer k; integer l; integer m; integer n; integer o; integer p; mtsoptdata = List[]; Needs["PlotLegends"];
For\[k = 1, k <= 2, k = k + 1,\]
For\[l = 1, l <= 2, l = l + 1,\]
For\[m = 1, m <= 2, m = m + 1,\]
For\[n = 1, n <= 2, n = n + 1,\]
For\[o = 1, o <= 2, o = o + 1,\]
For\[p = 1, p <= 2, p = p + 1,\]
mtsval[k1_, k2_, k_, l_, m_, n_, o_, p_] :=
Sum[pr1[[i]]*pr2[[j]]*rvalsmts[[k]]*Min[4t[[i]], k1] -
bvalsmts[l]*Max[d1[i] - k1, 0] - (1 + hcostfactor[n])*covalsmts[m]*Min[d1[i] + d2[j], k1] - 
  rvalsmts[k]*Min[d2[j], k2, Max[k1 - d1[i], 0]] - 
  bvalsmts[l]*Max[d2[j] - Min[k2, Max[k1 - d1[i], 0]], 0] - 
  hcostfactor[n]*covalsmts[m]*Min[d2[j], k2, Max[k1 - d1[i], 0]], {i, 1, 3}, {j, 1, 3}) - (1 + icapfactor[p])*cvalsmts[o]*k1 - 
  icapfactor[p]*cvalsmts[o]*k2;

oldMaxIt = Options[NMaximize, MaxIterations];
SetOptions[NMaximize, MaxIterations -> 500];

mtsoptval = 
  Maximize[{mtsval[k1, k2, k, l, m, n, o, p], k1 >= 0, k2 >= 0, 
    k2 <= k1}, {k1, k2}];
SetOptions[NMaximize, oldMaxIt[[1]]];

Options[NMaximize, MaxIterations];

mtdata = List[mtsoptval[2, 2], mtsoptval[2, 2]];
mtsoptdata = Append[mtsoptdata, mtdata];

mtsopt = 
  ListPlot[mtsoptdata, PlotRange -> {{0, 500}, {0, 500}}, 
    PlotStyle -> {Green, PointSize[0.02]}];

mtsval1[k1_, k2_, l_, n_, n_, o_, p_] := 
  Sum[pr1[i] * pr2[j] * ((1 + icapfactor[p])*cvalsmts[o]*
    k1/((1 + icapfactor[p])*cvalsmts[o]*k1 + 
    icapfactor[p]*cvalsmts[o]*k2)*rvalsmts[k]*
    Min[d1[i], k1] - 
    bvalsmts[l]*Max[d1[i] - k1, 0] + 
    rvalsmts[k]*Min[d2[j], k2, Max[k1 - d1[i], 0]] - 
    bvalsmts[l]*Max[d2[j] - Min[k2, Max[k1 - d1[i], 0]], 0] - 
    (1 + hcostfactor[n])*covalsmts[m]*Min[d1[i] + d2[j], k1], {i, 1, 3}, {j, 1, 3}) - (1 + icapfactor[p])*cvalsmts[o]*k1;

mtsval2[k1_, k2_, l_, n_, n_, o_, p_] := 
  Sum[pr1[i] * 
    pr2[j] * (icapfactor[p])*cvalsmts[o]*
    k2/((1 + icapfactor[p])*cvalsmts[o]*k1 + 
    icapfactor[p]*cvalsmts[o]*k2)*rvalsmts[k]*

524
Min[d1[[i]], k1] - bvalsmts[[l]]*Max[d1[[i]] - k1, 0] + rvalsmts[[k]]*Min[d2[[j]], k2, Max[k1 - d1[[i]], 0]] - bvalsmts[[l]]* Max[d2[[j]] - Min[k2, Max[k1 - d1[[i]], 0]], 0] - hcostfactor[[n]]*cvalsmts[[m]]* 
Min[d2[[j]], k2, Max[k1 - d1[[i]], 0])]. {i, 1, 3}, {j, 1, 3}) - icapfactor[[p]]*cvalsmts[[m]]*k2;

mtsmaxval1[k2_, k_, l_, m_, n_, o_, p_] := 
ArgMax[mtsval1[k1, k2, k, l, m, n, o, p], 
0 <= k1 <= 5000, {k1}];

mtsmaxval2[k1_, k_, l_, m_, n_, o_, p_] := 
ArgMax[mtsval2[k1, k2, k, l, m, n, o, p], 
0 <= k2 <= 5000, {k2}];

exprval1 = ToExpression["V_1(K_1)", TeXForm];
exprval21 = ToExpression["K_1", TeXForm];
jitexprval1 = ToExpression["V^{JIT}_1(K_1)", TeXForm];
mtsexprval1 = ToExpression["V^{MTS}_1(K_1)", TeXForm];

expr1k20 = ToExpression["K_2=0", TeXForm];
expr1k225 = ToExpression["K_2=25", TeXForm];
expr1k250 = ToExpression["K_2=50", TeXForm];
expr1k2100 = ToExpression["K_2=100", TeXForm];
expr1k2150 = ToExpression["K_2=150", TeXForm];
expr1k2200 = ToExpression["K_2=200", TeXForm];

mtsexpr = 
ToExpression[ 
"(V^{MTS}(K^*_1,K^*_2)-V^{MTS}(K_1,K_2))/V^{MTS}(K^*_1,K^*_2)" , TeXForm];

diffname = "lmtsdiff_" <> "k" <> ToString[k] <> "l" <> ToString[l] <> 
"m" <> ToString[m] <> "n" <> ToString[n] <> "o" <> ToString[o] <> 
ToExpression[p] <> "p" <> ToExpression[p] <> ".eps";

valname = "lmtsval1_" <> "k" <> ToString[k] <> "l" <> ToString[l] <> 
"m" <> ToString[m] <> "n" <> ToString[n] <> "o" <> ToString[o] <> 
ToExpression[p] <> "p" <> ToExpression[p] <> ".eps";

max1name = "lmtsmax1_" <> "k" <> ToString[k] <> "l" <> ToString[l] <> 
"m" <> ToString[m] <> "n" <> ToString[n] <> "o" <> ToString[o] <> 
ToExpression[p] <> "p" <> ToExpression[p] <> ".eps";

max2name =
```
"lmtsmax2_" <> "a" <> ToString[k] <> "l" <> ToString[l] <>
"m" <> ToString[m] <> "n" <> ToString[n] <> "o" <>
ToString[o] <> "p" <> ToString[p] <> ".eps"

mtsdiff =
Plot[{(mtsoptval[[1]] - mtsval[k1, 1, k, l, m, n, o, p])/
Abs[mtsoptval[[1]]], (mtsoptval[[1]] -
mtsoval[k1, 25, k, l, m, n, o, p])/Abs[mtsoptval[[1]]], (mtsoptval[[1]] -
mtsoval[k1, 50, k, l, m, n, o, p])/Abs[mtsoptval[[1]]], (mtsoptval[[1]] -
mtsoval[k1, 100, k, l, m, n, o, p])/Abs[mtsoptval[[1]]], (mtsoptval[[1]] -
mtsoval[k1, 200, k, l, m, n, o, p])/Abs[mtsoptval[[1]]],
{mtsoptval[[1]]}[[k1]], (mtsoptval[[1]] -
mtsoval[k1, 1, k, l, m, n, o, p])/Abs[mtsoptval[[1]]],
Abs[mtsoptval[[1]]], {k1, 0, 100},
PlotStyle -> {{Thick, Blue, Dashed}, {Thick, Red,
Dashed}, {Thick, Green, Dashed}, {Thick, Brown,
Dashed}, {Thick, Orange, Dashed}},
AxesLabel -> {expr2val1, mtsexpr}
];

mtsoval1k1 =
Plot[{mtsoval1[k1, 0, k, l, m, n, o, p],
mtsoval1[k1, 25, k, l, m, n, o, p],
mtsoval1[k1, 50, k, l, m, n, o, p],
mtsoval1[k1, 100, k, l, m, n, o, p],
mtsoval1[k1, 200, k, l, m, n, o, p]}, {k1, 0, 400},
PlotLegend -> {expr1k20, expr1k225, expr1k250, expr1k2100,
expr1k2200},
PlotStyle -> {{Thick, Blue, Dashed}, {Thick, Red,
Dashed}, {Thick, Green, Dashed}, {Thick, Brown,
Dashed}, {Thick, Orange, Dashed}},
AxesLabel -> {expr2val1, mtsexprval1}
];

expr1 = ToExpression["K_1^o(K_2)", TeXForm];
expr2 = ToExpression["K_2", TeXForm];
expr3 = ToExpression["K_2^o(K_1)", TeXForm];
expr4 = ToExpression["K_1", TeXForm];
mtkl1 = Show[
Plot[mtskmaxval1[k2, k, l, m, n, o, p], {k2, 0, 500},
AxesLabel -> {expr2, expr1}, PlotStyle -> {Thick, Red} ],
mtsopt];
```
mtsk2 = Plot[mtsmval2[k1, k, l, m, o, p], {k1, 0, 500},
AxesLabel -> {expr4, expr3}, PlotStyle -> {Thick, Blue}];
Export [diffname, mtsdiff]; Export [val1name, mtsval1k1];
Export [max1name, mtsk1]; Export [max2name, mtsk2]; ]]]]]}
Bibliography


Biography

Suri Gurumurthi (born November 23rd 1975) obtained his Bachelor’s degree in Mechanical Engineering at the Birla Institute of Technology, India, with major coursework in design and manufacturing. The same interests led him to a year of training in production and quality planning at Tata Motors in Pune, India. Subsequently, he obtained a Masters degree in IEOR at the University of Minnesota, while successfully completing a semiconductor fabrication simulation and capacity planning project in Bloomington, MN. He then joined the Alcoa Technical Center as a manufacturing systems engineer, and developed OR solutions for large scale process management problems, and also helped architect the implementation of the ORACLE 11i APS module within Alcoa’s Milled Products Division. Pursuing further graduate work in operations at MIT Sloan, he broadened his interests with a project sponsored by UPS Supply Chain Solutions Division, and as a product development research associate role at McKinsey and Co. He also served as a Teaching Fellow for the MIT Sloan Fellows program. Upon entering the Fuqua School Ph.D. program, he helped develop course modules for the day-time MBA program, and later – over three years – taught a graduate course on supply chain management in the Engineering Management program. He also worked briefly as a consultant for Pratt & Whitney Engine Services, with the support of the faculty at Duke. He is currently serving an appointment as Lecturer of Process Management, in the Department of Business Administration at the University of Illinois, Urbana Champaign.