Integrating Consumer Flexibility in Smart Grid and Mobility Systems

An Optimization and Online Mechanism Design Approach

Zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften

(Dr. rer. pol.)

von der Fakultät für
Wirtschaftswissenschaften
am Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl. Wi.-Ing. Philipp Ströhle

Tag der mündlichen Prüfung: 16.12.2014
Referent: Prof. Dr. Christof Weinhardt
Koreferent: Prof. Dr. Wolf Fichtner
Prüfer: Prof. Dr. Thomas Setzer
2014 Karlsruhe

Acknowledgements

Abstract

Flexibility is considered a multifaceted, domain-specific, and difficult to generally define concept. Systems can better adapt to a changing environment if flexibility potentials are available. In service systems consumer flexibility may readily be available, but requires careful dispatch to balance the interests of service providers and consumers. This thesis is concerned with efficient dispatch of consumer flexibility in smart-grid and car-sharing systems, instances of energy and mobility services.

The proliferation of renewable energy sources introduces both uncertainty and volatility to the supply side of electric power systems. Novel economic coordination approaches leveraging flexibility potentials based on smart grid technology require appropriate incentive design. In particular, incentives should dynamically reflect scarcity. We consider the case of demand-shifting and shedding, and design two dominant-strategy incentive-compatible online mechanisms. The first model-free mechanism coordinates between flexible demand and uncertain supply from renewables. The second model-based mechanism employs a more elaborate model of demand, and can additionally resort to conventional generation. Furthermore, it relies on algorithms from online stochastic combinatorial optimization, which are modified to achieve thruthfulness.

Future multi-modal mobility systems need to orchestrate heterogeneous mobility services to provide consumers with fast, reliable and sustainable mobility options. Car-sharing schemes may gain in importance and assume the role of system enablers by providing the building block of individual mobility. In order to efficiently coordinate capacity utilization, consumer flexibility with respect to location, time, or vehicle type may be leveraged. We develop an online optimization algorithm that leverages spatial flexibility and advance reservation information to improve the economics of car-sharing. Interestingly, diversity in usage requirements enables adoption of diverse technologies, each suited to a different part of the demand spectrum. Parts of this spectrum may be served by electric vehicles. To explore this potential, we examine the economics of electric vehicle adoption in car-sharing fleets. We find that consumer flexibility can greatly improve the prospects of electrification.

This work is anchored in the fields of online optimization and online mechanism design and presents applications thereof in energy and mobility service systems. It contributes towards a better understanding of the value and limitations associated with consumer flexibility in online settings in both fields.

Марине

Contents

List of Figures

List of Tables

List of Abbreviations

AMI   Advanced Metering Infrastructure
ARIMA   AutoRegressive Integrated Moving Average
BEV   Battery Electric Vehicle
CAISO   California ISO
CCGT   Combined Cycle Gas Turbine
CDA   Continuous Double Auction
CDF   Cumulative Distribution Function
CG   Conventional Generation
CHP   Combined Heat and Power
CPP   Critical Peak Pricing
DAM   Day-Ahead Market
DR   Demand Response
DSIC   Dominant Strategy Incentive Compatible
DSM   Demand Side Management
DSO   Distribution System Operator
DSS   Decision Support System
EEX   European Energy Exchange
EV   Electric Vehicle
FCFS   First Come First Served
FMS   Flexible Manufacturing System
GAP   Generalized Assignment Problem
GT   Gas Turbine
HAN   Home Area Network
HMM   Hidden Markov Model
HVAC   Heating, Ventilation and Air-Conditioning
ICE   Internal Combustion Engine
ICT   Information and Communication Technology
IC   Incentive Compatibility
IR   Individual Rationality
IS   Information System
JPMVEC JP Morgan Ventures Energy Corporation

LUSF	Least Utilized Station First
MAE	Mean Absolute Error
MIP	Mixed-Integer-Program
MKP	Multi-Knapsack Problem
OMD	Online Mechanism Design
OM	Operations Management
OTC	Over-the-counter
QoS	Quality-of-Service
RES	Renewable Energy Sources
RMSE	Root Mean Squared Error
RM	Revenue Management
RTP	Real-Time Pricing
SoC	State-of-Charge
ToU	Time-of-Use
TSP	Traveling Salesman Problem

Part I

Introduction and Foundations

Chapter 1

Introduction

This dissertation addresses service-oriented micro-transactions in two domains, electric power and mobility. Both domains exhibit asset perishability and uncertainty as central aspects, introducing interesting economic coordination challenges. We aim to meet these challenges by focusing on efficiently employing consumer flexibility, the overarching topic of this thesis. In the spirit of Jordan and Graves (1995), we argue that small amounts of flexibility may be sufficient to foster significantly improved economic outcomes. To this end, Information and Communication Technology (ICT) provides the technical infrastructure for semi-automated or assisted decision making. The addition of ICT in the power systems domain enables monitoring and control and turns power distribution networks into smart grids. Shared mobility systems rely on ICT to ensure smooth information exchange between users and providers, fostering efficient system control. Our application of interest in the mobility domain is station-based car-sharing. We formulate optimization problems and design online mechanisms that rely on ICT infrastructure with the objective of efficiently harnessing dispersed demand-side flexibility.

This chapter first presents an overview of the relationship between uncertainty and flexibility, with incentives providing the necessary link between the two. We then outline the research questions in the smart grid and mobility domains, and finally present the overall dissertation structure.

1.1 Uncertainty, Flexibility, and Incentives

Decision makers perpetually face uncertainty in almost all areas of business and economics. Often, forecasting methods (Hyndman and Athanasopoulos, 2013) are used to reduce uncertainty, enabling better decisions. Uncertainty in smart grids, or power systems, respectively, has mainly temporal character, i.e., quantity forecasts (of both, supply and demand) may err with respect to two dimensions: Phase and amplitude (Giebel et al., 2011). Besides research on quantity forecasts, a large body of literature is concerned with modeling and forecasting electricity prices (cf. Conejo et al., 2005; Banal-Estañol and Micola, 2009; Keles et al., 2012). In the mobility domain, both temporal and spatial demand uncertainty play an important role. To complicate matters,

we are concerned with decision making under incomplete information about the future, i.e., online decision making.

The negative aspects of uncertainty, however, may be alleviated by means of consumer flexibility. In both domains, we assume that consumers can, to a varying degree, be considered flexible with respect to non-functional service requirements. An important question therefore concerns the efficient utilization of available flexibility, i.e., "how much flexibility should be provisioned by each consumer?" A second question concerns incentive design: How can consumers' flexibility be elicited and "used"? Flexibility comes in a rich variety of flavors: In the smart grid setting, temporal flexibility is of paramount importance to efficiently integrate large shares of supply from volatile Renewable Energy Sources (RES). While shifting demand to a later time can alleviate the burden on conventional generation, demand-side flexibility is not a perfect substitute for supply capacity. Conventional generation is still required even if all customers are flexible, when total quantity demanded exceeds renewable supply throughout an extended period of time. In (station-based) car-sharing, we place special emphasis on consumers' spatial flexibility to compensate for fluctuating and uncertain demand patters over space and time. In other words, the corresponding non-functional service requirement concerns flexibility with respect to the stations at which the desired type of vehicle is stationed.

Flexibility often exhibits a convex cost structure: Provision of a limited amount of flexibility may be near-costless, while requesting very high flexibility may be, in contrast, extremely costly. In this vein, Kwon and Østergaard (2014) show that up to one quarter of total demand in power systems is flexible within a 2 hour time frame while only 7% of demand could be shifted by a whole day, indicating a convex structure regarding the cost of flexibility provision. Hence, it may be more economical to leverage micro-flexibilities from a large set of agents compared to large amounts of flexibilities from a small set of agents (cf. Tang and Tomlin, 2008). This may be in particular true if transaction costs (communication, computation, and automation) are negligible due to ICT adoption. For portfolios of small flexibilities to attain significant value, efficient composition, i.e., chaining a la Jordan and Graves (1995) is essential.

The value of optimization in many applications is indisputable. However, all-too-often engineers designing service systems neglect the question of incentive design in distributed systems involving self-interested agents. Rational agents *cannot* be expected to facilitate systems performance at their own expense. Instead, incentives for flexibility revelation and provision need to be established. Incentive design must ensure that revelation of agents' flexibility type renders them better off, i.e., provide benefits relative to non-disclosure. Then, individual agents can align behavior and incentives and thereby maximize individual utility. In short, Individual Rationality (IR) must be ensured for agents to participate in the mechanism. Furthermore, given the small value attached to individual transactions in the domains under consideration, special emphasis should be placed on ensuring Incentive Compatibility (IC) of the corresponding mechanisms. Mechanisms that are IC prevent agents from strategizing on

their type to increase payoffs from trade, and thus reduce complexity.[1] Briefly, "[..] in a [Bayesian-Nash] incentive-compatible mechanism, each individual can maximize his expected utility by reporting his true valuation, given that the other is expected to report honestly" (Myerson and Satterthwaite, 1983).

Achieving IC in the presence of competing design desiderata (Myerson and Satterthwaite, 1983) may be challenging. Accordingly, the design of sustainable markets requires economists to adopt engineering perspectives, understand issues of computational complexity, and have a solid foundation in the main results of mechanism design literature (cf. Roth, 2002) to avoid market failures. The market engineering framework (Weinhardt et al., 2003) structures the market design process into mutually exclusive subtasks that, as a whole, take a holistic approach towards the establishment of ICT-based markets. Hence it may serve as a structured blueprint for the design of markets and incentive schemes. We argue that incentives pose an essential building block towards effective flexibility provision by consumers in service systems.

1.2 Smart Grids

After the recent unprecedented growth in renewable generation capacity, RES are supplying up to 25.3% of total load in Germany (2013)[2]. The proliferation of RES with the goal of establishing more sustainable power systems, will have two major consequences: First, RES will inject additional uncertainty and volatility into power systems (Varaiya et al., 2011). Second, in contrast to traditional power systems architecture, low energy density of RES forces generation to be spread geographically (Smil, 2010), requiring upgrades of transmission and distribution networks that have historically been designed for centralized supply. Such decentralization renders maintaining the balance between supply and demand at all times increasingly difficult. In distribution networks, fluctuating power generation may additionally lead to serious power quality problems (voltage, jitter, etc.), and must be addressed appropriately.

Smart grid adoption, in contrast to traditional capacity over-provisioning, relies on (i) ICT upgrades to the existing power (distribution) networks (Farhangi, 2010), and, (ii) establishment of economic coordination schemes, potentially along the market engineering framework (Weinhardt et al., 2003; Weinhardt, 2012). On a related note, Bichler et al. (2010) coin the term "smart markets", which are supposed to facilitate the alignment of conflicting interests by means of electronic market-based coordination schemes. To foster acceptance of these coordination schemes in complex environments, such as the smart grid, the introduction of decision support systems guiding

---

[1]Nevertheless, compact type representation in combinatorial settings remains an inherently complex task.

[2]http://www.erneuerbare-energien.de/EE/Navigation/DE/Service/Erneuerbare_Energien_in_Zahlen/Zeitreihen/zeitreihen.html,   accessed October 2, 2014

human decision making may be required. Importantly, our interpretation of the smart grid as a techno-economic system (Nolden et al., 2013; Flath, 2013) reaches beyond a perspective that is solely concerned with the technical side (Amin and Wollenberg, 2005; Gellings, 2009), and, following Chassin and Kiesling (2008), focuses on the economics of demand-side activation.

Demand-side Flexibility and Energy Services   To compensate fluctuating renewable supply, demand-side flexibility may be an economic alternative to storage. At low penetration levels of RES, flexibility of conventional supply may be sufficient and can be re-dedicated to offset novel fluctuations due to RES (Strbac, 2008). However, guaranteeing reliability of supply without curtailing demand or shedding excess renewable generation becomes increasingly costly at higher RES penetration levels.

The expected rise of Electric Vehicles (EVs) has, hence, provided an interesting research avenue into novel charging coordination approaches. The literature on price-based EV charging coordination has, to a large extent, been shaped by approaches to circumvent over-coordination and herding effects. Flath et al. (2013), for instance, propose spatial price differentiation in order to efficiently integrate distribution network constraints into charging regimes. Schuller et al. (2014) highlight the empirical savings potential resulting from smart EV charging and feeding energy from the vehicle back into the power grid (V2G).

In order to face the challenges associated with the current energy transition, economic coordination potentials associated with EV charging alone may be insufficient. Consequently, leveraging demand-side flexibility through so-called Demand Side Management (DSM) has received attention from academia and industry alike. This approach entails a paradigm shift in power system operations: DSM aims to shape demand to follow supply through shedding, shifting, curtailment, etc. Clearly, including demand-side flexibility opens up the search space and enables more favorable economic outcomes compared to an isolated reliance on supply side flexibility only. Implementing DSM requires a careful understanding of the associated costs and benefits on the individual customer level in order to efficiently reap DSM benefits. For instance, customers may be required to specify the extent to which their demand is temporally flexible and whether it can be shedded. The notion of *energy services* may assist in encapsulating the temporal aspects of energy consumption and therewith reduce cognitive overhead and transaction costs.

Traditionally, the term energy service is understood as "serving the need for electrical energy" (Sioshansi, 1995). In contrast, the energy services we refer to in this work are inspired by the seminal work of Schweppe et al. (1988):

> "Customers have no desire for electric energy per se (although people can get a charge out of it). Customers do desire the services provided by electric energy."

More specifically, customers do not value a certain amount of energy, but attach value to clean dishes, clean clothing, a well tempered home, or a sufficiently charged EV to provide mobility. The utility associated with functional service delivery might be influenced by non-functional characteristics, such as privacy, response time and reliability, thus introducing a notion of QoS. If service delivery is not required immediately, or the appliance providing the service can be modulated, the associated flexibility may be operationalized as a restricted form of storage.[3] Relative to physical storage devices (e.g., batteries) flexibility may be more cost-efficient.

In novel retail market designs, a QoS approach including reliability and timeliness of power service may play an important role in determining both allocation and payments. Due to the heterogeneity of QoS requirements for different demand classes, a granular approach is required. Then, customers are empowered to decide on the level of individual energy services regarding immediate or postponed service.

The Role of Information and Incentives   Efficient integration of RES via DSM requires economic coordination between a large number of stakeholders, i.e., network operators, generators, and consumers in the smart grid. The basic infrastructure building block for such economic coordination is the smart meter, which currently is introduced to the market. By means of complex pricing schemes it may convey accurate (in time and space) scarcity signals to consumers, and accordingly give rise to improved allocative efficiency. Nevertheless, in order to avoid over-coordination and herding effects from price signals, bidirectional communication and automation are required. The latter in particular requires a good understanding of individual preferences. Thus, information flow can be separated into two stages: From the consumer to an agent acting on his behalf, closely aligned with the individual consumer's preferences, and from the agent to a coordinating instance, e.g., a market.[4]

Given sufficiently developed automation technology, providing flexibility may be associated with low transaction costs, fostering the integration of even marginal amounts of flexibility. Similar to the manufacturing domain where small amounts of process flexibility can yield high value (Jordan and Graves, 1995), even small amounts of flexibility in the power system may significantly improve economic outcomes. Using flexibility efficiently requires not only information on flexibility endowments, but also the appropriate incentive and coordination structures.

Appropriate incentive structures need to be developed for the smart grid (Dash et al., 2003; Ramchurn et al., 2012), as self-interested consumers require compensation for flexibility provision. Simple posted price schemes may be quite efficient in static environments. However, at higher RES penetration levels and constrained

---

[3]For a discussion of qualities of demand-side flexibilities, see Petersen et al. (2013).

[4]A notable example of novel approaches regarding the first stage can for instance be seen in NEST's smart thermostats to control Heating, Ventilation and Air-Conditioning (HVAC) more efficiently (https://nest.com/). The information when consumers are at home, may prove beneficial in efficiently deciding on energy usage.

networks, this might no longer hold true: Static pricing rules cannot cope with the dynamism of fluctuating generation, possibly wasting large parts of readily available welfare-enhancing flexibility potentials. Hence, this work proposes the application of online mechanism design (Parkes, 2007) to bridge the gap between the dynamism present in modern power systems and the appropriate incentives to include demand-side flexibility in power system operations.

Research questions   Our first research question is concerned with the value of flexibility in a smart grid setting, where stochastic RESs are the only supply source. We employ a simple representation of demand and allow for limited demand shedding and shifting.

Research Question 1: *What is the value of flexibility under single-unit sheddable and shiftable demand in a local smart grid setting?*

Besides the potential value of flexibility, we are particularly interested by how much economic performance deteriorates in the presence of self-interested agents, i.e., under an incentive-compatible online mechanism.

Research Question 2: *What is the cost of ensuring IC in terms of social welfare?*

Our contribution to the literature is the adaption of the canonical online mechanism, presented in (Parkes, 2007), to the smart grid domain. More specifically, we introduce a generic supply model that captures realistic uncertainty attributes of RES, e.g., wind generators.

Expanding upon the results of these questions, we examine to what extent both, the introduction of a model of the future, and a slightly more complex model of demand, affect economic performance in Chapter 5. To this end, we consider multi-unit and non-preemptive demand. Hence, jobs can be shifted in time, but may not be interrupted after they have been started. Clearly, in the absence of conventional, dispatchable generation, poor scheduling decisions would lead to infeasible schedules. Therefore, we augment uncertain supply from RES with costly conventional generation to answer the following research question:

Research Question 3: *What is the value of shiftable and sheddable non-preemptive demand under uncertain future supply from RES in the presence of costly conventional supply?*

However, as mentioned before, the answer to the previous question is only an indi-

cator of the value of flexibility, which presumably cannot be achieved, as the proper incentives to reveal flexibility are lacking. To ensure IC, further constraints are added that reduce the set of feasible solutions, hence reducing overall welfare. We are interested in the cost associated with these IC constraints.

Research Question 4: *To what extent does flexibility affect the welfare gap between an online planner and an online mechanism ensuring IC with multi-unit, non-preemptive demand in the model-based case?*

Additionally, we explore the value of information with respect to both, overall welfare and the cost of IC:

Research Question 5: *To what extent can more accurate information regarding future uncertainty, e.g., a larger number of scenarios and longer scenario horizon, contribute to higher efficiency and reduce the cost of IC?*

Our contribution in Chapter 5 is twofold: First, we adapt the Expectation (Chang et al., 2000a) and Consensus (Bent and Van Hentenryck, 2005) algorithms from online scheduling to the multi-machine setting, such that multiple jobs may now be allocated at the same (discrete) time. Second, based on these adapted algorithms, we design a novel incentive-compatible online mechanism for allocating flexible demand to fluctuating renewable supply under uncertainty at high efficiency. To ensure IC the mechanism is organized in two stages. An initial consensus stage, that ensures monotonicity in allocation decisions through "pre-commitment" (cf. Gerding et al., 2011) and thus renders revelation of flexibility a dominant strategy, and a second stage concerned with improving economic efficiency, subject to the allocation decision fixed before. Importantly, arrival and flexibility do *not* affect allocation decisions, but *do affect* payments. Ceteris paribus, payments from conusmers to the mechanism are monotonously decreasing in flexibility.

1.3 Mobility Systems

Multimodal approaches to passenger transportation have received a great amount of attention in the recent academic literature (cf. Nobis, 2006, 2007; Kuhnimhof et al., 2012). Situation-based usage of the most appropriate means of transportation can assist in overcoming congestion, poor air-quality, accidents and other externalities of individual transportation. Not surprisingly, in the presence of a well-developed public transportation scheme, the relevance of personal vehicles is often driven by the need for sporadic access to individual transportation (think of shopping trips or

weekend travel). Hence, an important building block of multi-modal transportation solutions is access to point-to-point connectivity in order to be able to complete such individual trips. For example, mobility consumers may be satisfied to board suburban trains, trams, or buses for their everyday commute, but may sporadically require a different means of transport. The proliferation of ICT may attach the convenience commonly associated with ownership, to service solutions (which separating usage and ownership), and hence foster adoption of multimodal mobility services.

Car-sharing The neighborhood rental of vehicles with automated vendor interaction, i.e., car-sharing, is suited to economically provide such sporadic point-to-point connectivity and hence may serve as an ideal complement to public transportation. Alternatively, car-sharing may provide mobility to families instead of forcing ownership to cover sporadically arising mobility needs. The economics of car-sharing are in favor for low intensity users of individual transportation, as fixed and maintenance costs are distributed over a larger user base. Improved information access via connected mobile devices reduces transaction costs and allows for easier resource sharing. Hence, it may render novel, more dynamic forms of sharing possible that so far encountered prohibitive transaction costs.

Improving Fleet Utilization   Similar to the case of energy services, mobility services require multi-dimensional characterization. Customers may be flexible with respect to the exact type of vehicle to use for the trip under consideration, the time of service delivery, the station used for vehicle pick-up and drop-off, or a combination thereof. As we will see, customers' spatial flexibility turns out to be decisive with respect to achieving high-quality economic outcomes in the dimensions of fleet utilization and quality-of-service for customers, at least in the particular instance of car-sharing examined in Chapter 7.

An interesting intersection of the two domains of power and mobility systems can be found in electrified car-sharing. We will explore this topic towards the end of this thesis.

Based on a unique car-sharing data set, we examine the value of temporal and spatial flexibility and derive potential fleet size reductions due to customer flexibility in the offline case. We demonstrate that flexibility induces pooling effects which may be leveraged to improve operations efficiency. Accordingly, we pose the following question.

Research Question 6: *To what extent can consumer flexibility improve the economics of car-sharing?*

Our contribution is an adaptation of the widely known bin-packing problem (cf. Nemhauser and Wolsey, 1988) that we employ to determine the minimal number of

vehicles required to serve all customer requests. Consumer flexibility enlarges the set of vehicles a reservation can be assigned to and hence allows to reduce overall fleet size through denser scheduling of reservations on fewer vehicles.

In the online case, given an exogenously defined fleet of reduced size, we employ flexibility to improve QoS, i.e., the shares of reservations served. In this case, the challenge lays in assigning reservations to vehicles in such a way as not to block future reservations assignments[5], hence we ask:

Research Question 7: *What is the value of spatial consumer flexibility in terms of QoS?*

To answer this question, we present an Expectation-inspired (Hentenryck and Bent, 2009) online planner which trades-off QoS and consumer walkways and thereby outperforms greedy benchmarks selecting either the closest (on a First Come First Served (FCFS) basis) or the historically least-utilized station.

Fostering Sustainability   EVs are, due to the high cost of batteries, more expensive to purchase, but, in exchange, offer reduced operating expenses. This combination calls for intensive usage in order for EVs to be economically justified. Thus, it nicely maps into the scenario of car-sharing, where frequent vehicle utilization and mostly short distance trips renders EVs an interesting complement. As EVs are locally emission-free, they provide an additional advantage for passenger transportation in densely populated urban areas. In shared mobility scenarios, drivers can avoid the issue of range anxiety (Eberle and von Helmolt, 2010) by adapting the choice of vehicle technology to trip distance.

First, we ask to what extent reservations can be satisfied from range-restricted electric vehicles. By giving customers the choice for a vehicle for each trip, a more diverse range of technologies can be employed. The availability of multiple technologies may accredit sharing schemes with tangible advantages relative to more traditional individual ownership schemes. In the latter, the corresponding vehicle must be designed such that even the most extreme mobility demands (e.g., range, number of seats) can be met by the vehicle, often dominating other aspects of mobility choice. As reservation distance and duration are distributed heterogeneously, a significant amount of reservations may be served using EVs. Again, we are interested in the impact of consumers' spatial flexibility on fleet size, if EVs are to be included.

Research Question 8: *To what extent enables (spatial) consumer flexibility electrification of car-sharing fleets?*

---

[5]The online planner is effectively trading-off efficiency and robustness.

To answer this question, we present a hybrid optimization model along the lines of a Generalized Assignment Problem (GAP) that combines EV charging constraints with both, scheduling constraints and non-overlap constraints. Our objective is the minimization of total cost of fleet operations using both technologies.

Beyond car-sharing as it is implemented already today, the advent of self-driving cars might render the concept of car-ownership obsolete to larger portions of urban population. In particular, it may strengthen the relevance of both, shared mobility approaches and solutions to corresponding economic coordination problems.

1.4 Structure

This section provides a short outline of the thesis structure (see Figure 1.1). Chapter 2 introduces the concept of flexibility and presents a broad literature overview in adjacent fields that the remainder of the thesis relies upon. Thereafter, the two main parts of the thesis follow.

Part II focuses on smart grids and online mechanism design. Chapter 3 presents the related work and fundamentals in the field of smart grids. Chapter 4 adapts the canonical online mechanism (Parkes, 2007) to the smart grid, including a rich evaluation in multiple dimensions, including market power, payments and allocations.[6] Thereafter, in Chapter 5, Online Mechanism Design (OMD) is again applied to a richer demand model with the main focus on allocative efficiency. Furthermore, special emphasis is placed on the value of more accurate information about the future.[7]

Part III examines the value of flexibility in car-sharing using empirical data and both, offline and online optimization approaches. In Chapter 6, the role of car-sharing in future mobility systems is introduced, along with specific information on the dataset used in the subsequent two chapters for evaluation purposes. Chapter 7 presents offline and online optimization models leveraging consumer flexibility for fleet size minimization (offline) or QoS maximization (online). In Chapter 8 we sound out the prospects of EVs in car-sharing.

Finally, Part IV summarizes the main results and presents an outlook into avenues of promising future research.

---

[6]An earlier version was circulated as Ströhle et al. (2012)

[7]An earlier version was presented at AAMAS, 2014 (Ströhle et al., 2014).

| Part I | 1 | Introduction |
| | 2 | On Flexibility |

	3	Smart Grid Fundamentals		6	Car-sharing Fundamentals	
Part II	4	Model-free Online Mechanism Design for Scheduling Preemptive Jobs		7	Leveraging Consumer Flexibility in Station-based Car-sharing	Part III
	5	Model-based Online Mechanism Design for Scheduling Non-Preemptive Jobs		8	Leveraging Consumer Flexibility in Electrified Car-Sharing	

| Part IV | 9 | Conclusion |

Figure 1.1: Structure

Chapter 2

On Flexibility

This chapter presents an introduction to flexibility in different domains and aims to outline the boundaries of the broad concept of flexibility. There exists a vast body of literature on flexibility in the microeconomics and management science literature alone, which renders the definition of flexibility a challenging task. Therefore, we focus on fields and aspects that we deem most relevant for this work.

Flexibility in general is considered valuable if uncertainty and change are present in the considered system. Its value can be derived from all levels of decision making, from operational to strategic.[1] We introduce the literature on flexibility in roughly chronological order and begin with the field of manufacturing, starting on the microscopic level with flexibility in scheduling, continuing with operational and strategic flexibilities. Thereafter, we review the literature on flexibility between multiple firms, i.e., in supply chains and supply networks and compile a comprehensive literature review of flexibility in Operations Management (OM). We pay particular attention to process and structural flexibility (Jordan and Graves, 1995) and finally arrive at strategic flexibility. Based on these prerequisites, we provide details on flexibility in the realm of electric power systems and car-sharing, the two domains of interest in this work. We conclude with a discussion of flexibility and its relationship to the work at hand.

2.1 Flexibility in Manufacturing

The general concept of flexibility in manufacturing systems has been studied intensively and along many dimensions in the literature of past decades (Sethi and Sethi, 1990; Browne et al., 1984; Mandelbaum, 1978). Firms perpetually face uncertainty with respect to market conditions and technological change. More specifically, decreasing lead times in production and shorter product life cycles due to competition require flexibility on multiple decision levels. A firms' ability to adapt to both, changing environments and fluctuating demand, is increasingly valuable and possibly critical for sustained success in the market place.

---

[1] "Flexibility is not just an adaptive response to an uncertain environment. It has a proactive function in creating uncertainties that competitors can not deal with." (Gerwin, 1993).

Flexibility, however, comes at a cost in the form of larger engineering efforts or more expensive machinery, and hence, determining the optimal amount of flexibility is an interesting question. While the drivers for the value of flexibility are coarsely understood, an exact and universally accepted definition of flexibility itself is lacking. This section provides an overview of the literature on flexibility in the manufacturing and adjacent domains.

Qualitative Flexibility Concepts   Gupta and Somers (1992) borrow from Sethi and Sethi (1990) to repeat that "manufacturing flexibility is a complex, multidimensional and difficult-to-synthesize concept." The central role of flexibility is acknowledged as it provides a "critical measure of total manufacturing performance." They conducted both literature review and a questionnaire survey with responses from 269 firms at the chief executive level to cast (manufacturing) flexibility into nine constructs. Those constructs were validated through additional data collected from 113 firms (Gupta and Somers, 1992), and include volume, programming, process, product and production, market, machine, routing, material handling, as well as expansion and market flexibilities (Gupta and Somers, 1992, Fig. 2).

A comprehensive overview of concepts and measurement of flexibility in manufacturing systems is given in the overview paper of Gupta and Goyal (1989). The authors of a review article on manufacturing flexibility (Beach et al., 2000), following the literature, define flexibility as "the ability of the system to quickly adjust to any change in relevant factors like product, process, loads and machine failure". A simpler and more general definition is given with "the ability to change or react with little penalty in time, effort, cost or performance" (Upton, 1994).

Beach et al. (2000) synthesize manufacturing flexibility from most of the literature presented so far and place special emphasis on its interplay with uncertainty "from a largely operational perspective." They interpret manufacturing flexibility to be valuable in both, proactive and reactive uses. The latter is more relevant on a operational level, while the former use of flexibility relates more closely to a firm's strategic objectives. Eventually, Beach et al. (2000) posit that "any measurement of flexibility must, because of its nature, be user or situation specific."

A manufacturing plant is considered flexible (volume flexibility) if average cost of goods produced is rather flat over the quantity produced (p. 291 Sethi and Sethi, 1990, referring to Stigler 1939). However, "flexibility will not be a free good". A plant designed to accommodate large volume uncertainty and thereby remain profitable will likely have higher per unit cost than a plant operating at nameplate capacity. As Sethi and Sethi put it, the value of flexibility increases in the variation (variance) of the market price of the good and the "ability to predict market prices before making output decisions". Hence, manufacturing flexibility is not an end in itself, but rather required in order to successfully face change and uncertainty on the demand side.[2]

---

[2] A compact flexibility overview in lean manufacturing is provided in Herrmann (2013, pp. 25–28).

Zelenović (1982) highlights the trade-off involving operational manufacturing flexibility and productivity. On the other side, Zelenović highlights the perception that flexibility in the production process allows employees to make decisions on their own, which would be impossible in a rigid, inflexible system. Hence, as he expresses it, flexibility aids in "raising the level of humanization of work."

From deduction and literature studies, however, the introductory remark in Sethi and Sethi (1990) must be emphasized: "Flexibility is a complex, multidimensional, and hard-to-capture concept." The authors show by means of a literature review that at least 50 different terms exist that all describe various types of manufacturing flexibility. The common denominator in all of them (Sethi and Sethi, 1990) seems to be: "With flexible manufacturing, it becomes possible to bring the efficiency of mass production to batch production of multiple products."

Formal Flexibility Measures   Kumar (1987) develops formal measures for manufacturing flexibility that must be acknowledged as interesting, simple and elegant. First, seven essential and eight desirable axioms are outlined. A good measure for manufacturing flexibility must (should, in the case of the desirable axioms) satisfy those axioms. Thereafter, Kumar introduces the well-known information-theoretic measure of entropy as already defined by Shannon in 1948 into the manufacturing context. The only drawback of this measure is its lack of parametrization possibilities. Subsequently, three additional measures are introduced to alleviate this shortcoming. Unfortunately, neither of them fulfills all requirements introduced before, but the different measures may be applicable depending on the situation. In particular, these measures can be tailored to the situation under consideration by adjusting the corresponding parameter values appropriately.

Brill and Mandelbaum (1989) first outline the literature on flexibility in manufacturing to continue with the introduction of formal definitions of flexibility for both, individual machines and groups of machines with respect to a specific task. Regarding the latter, they introduce optimistic and pessimistic measures of flexibility, and a wide array of measures from in-between those two. The notion of optimistic and pessimistic flexibility measures demonstrates the complementary character of flexibility and uncertainty.

Dimensions of Flexibility   Browne et al. (1984) characterize flexibility in the manufacturing context and aim to highlight when exactly a manufacturing system should be termed "flexible". In their description, a Flexible Manufacturing System (FMS) can encompass flexibility in eight different dimensions: First and foremost *machine* flexibility, i.e., the "ease of making the changes required to produce a *given* set of part types", *process* flexibility, i.e., extending machine flexibility to different materials and different ways of production, *product* flexibility, the "ability to changeover to produce a new (set of) product(s) very economically and quickly." *Routing* flexibility describes

a situation in which the production process actually or potentially involves different routes and is thus less susceptible to individual machine breakdowns. *Volume* flexibility relates to the "ability to operate a flexibility manufacturing system at different production volumes." *Expansion* flexibility is a measure on how easily and modularly a manufacturing system can be extended, which is intricately linked to the concept of routing flexibility. *Operation* flexibility concerns the possibility to reorder operations for each part type. By deliberately refraining from fixed process steps on predetermined machines, the decision which machine to use for each step can be based on system state and thus enhance performance. Finally, *production* flexibility incorporates all previously mentioned types of flexibility and is, following Browne et al. (1984) measured in the set of part types a FMS can produce. Notably, Browne et al. reach beyond the mere description of flexibility and propose tangible measures for each kind of flexibility. Those involve, for example, machine set-up times, changes in production rate in case of machine breakdowns, or minimum profitable production volume.

Ideally, a FMS incorporates all eight kinds of flexibility; however, depending on the situation, it might be economical to include only subsets thereof. Historically, geographic location and type of industry have had a large influence on the composition of flexibility types for specific manufacturing applications. Browne et al. outline a range of flexible systems, ranging from micro-flexibility on a machine level, to macro-flexibility on a multi-line level. Bernardes and Hanna (2009), referring to another study from 1989, for example, define manufacturing flexibility as "the quickness and ease with which plants can respond to changes in market conditions."

Sethi and Sethi (1990) heavily rely on preceding work (Browne et al., 1984), and identify eleven, types of flexibility. Moreover, they emphasize the importance of an "appropriate organizational structure" which can be interpreted as an advance to the later article by Jordan and Graves published in 1995 introducing and examining the concept of structural flexibility. The flexibility categories identified are identical to those by Browne et al. (1984), apart from material handling, program and market flexibility. Material handling flexibility seems to have been included in process flexibility, while program flexibility captures the reliability of a production process, i.e., for how long production can continue untended (without manual intervention). The latter, market flexibility, includes a marketing perspective and aims to include modifications to the product that can later, after production facilities have been put in place, be leveraged. Furthermore, it enables "frequent product changes". Interestingly, they find "microprocessor technology" to be essential to successfully implement flexibility in manufacturing, as it avoids sacrificing efficiency (Zelenović (1982) uses the term productivity to express the same construct). The problem of defining flexibility is exacerbated by the various temporal levels (operational, tactical, strategic) on which it is examined.

2.2 Scheduling Flexibility

Definition and Objective Pinedo (2012) defines scheduling as "a decision-making process that is used on a regular basis in many manufacturing and service industries. It deals with the allocation of resources to tasks over given time periods and its goal is to optimize one or more objectives."

The most common objective in scheduling is makespan minimization.[3] To this end, the tasks are assigned to resources such as workers and machinery, which perform the necessary steps of each task. In this process, flexibility may arise in two ways. First, the order of individual tasks to achieve the overall goal may be flexible, i.e., different sequences in which the project tasks are completed may be admissible. Second, resources may exhibit flexibility in that they can be assigned to more than a single task. For example, workers may be trained to perform more than a single task (Iravani et al., 2005). Such flexibilities can be leveraged to improve the desired objective, e.g., reduce the makespan, increase profits, reduce costs or improve reliability.

A rich overview of scheduling problems, algorithms and systems is presented in Pinedo (2012). The presented models are coarsely grouped into deterministic and stochastic and ordered with respect to their complexity. Because this work is concerned with scheduling tasks on multiple resources in parallel under uncertainty, the stochastic, parallel machine models are most closely related (Pinedo, 2012, Ch. 12,13).

Scheduling of Pre-emptive Jobs Federgruen and Groenevelt (1986) show that parallel machine scheduling problems can be efficiently solved using maximum-flow techniques if the jobs can be interrupted, i.e., if jobs are preemptive. Preemption effectively dissolves the combinatorial character of the original problem. In this setting, flexibility is a result of the jobs' preemption property. Their work helped in overcoming computational complexity in scheduling in a wide array of applications. Hence, as a larger number of schedules can be computed in the same time frame, the applicability of stochastic scheduling is broadened and more frequent re-planning in online scheduling is enabled.

On-line Scheduling Chang et al. (2000b) propose an on-line scheduling algorithm for the case when a model of future arrivals is available, i.e., the distribution, but not the actual realization of future arrivals is known. They apply their model to schedule network packets under transfer capacity constraints.

Bent and Van Hentenryck (2004) build on this work and trade-off economic efficiency for reduced computational complexity. In more detail, they substitute costly computation of the expected value of a schedule, used by Chang et al. (2000b), with

---

[3]Makespan is the difference between the date at which a set of tasks is completed and the start time of the project.

a faster consensus-based approach. Their algorithm is applicable in highly time-constrained settings, but suffers, as the authors admit, from elitism.[4]

Hentenryck and Bent (2009) give an overview of online scheduling where, in the spirit of both, Chang et al. (2000b) and Bent and Van Hentenryck (2004), discarding jobs is allowed. Their objective corresponds to maximizing the weight (or value) of the schedule. Jobs are assigned heterogeneous weights and which counts towards the objective if a job is completed in time, i.e., before its deadline.

Herroelen and Leus (2005), who focus on project scheduling under uncertainty, treat flexibility and robustness of schedules as closely related terms and simply consider a schedule as being flexible if it "can easily be repaired". However, the authors do not define "easily".

On-line Scheduling with Self-Interested Agents   The literature presented so far assumed that the planner deciding on the respective schedules has access to truthful information. If, however, self-interested agents submit information to the planner, as is the case in distributed systems, they may be willing to exploit the planner's decision by strategically manipulating information.[5] This is true if the agents can improve their payoff by misreporting. To avoid such inefficient outcomes, additional constraints need to be introduced to ensure monotonicity in allocation decisions (Parkes, 2007). Grid and cloud computing present a realistic use case for such concerns. Self-interested agents, armed with private information, pose serious incentive problems to conventional, off-the-shelf scheduling approaches.

Stößer et al. (2010) highlight the cost of manipulation by strategic agents in grid markets. The scheduling problem they consider is a GAP in which a large number of heterogeneous jobs need to be matched to heterogeneous machines (computational nodes). Additionally, jobs can be migrated between nodes of a grid, complicating both, problem formulation and computational complexity, but adding realism. Due to its complexity, the problem cannot be solved to optimality within the corresponding time constraints (due to the online character of the problem), and hence requires heuristic solution approaches. The main contribution of Stößer et al. (2010) is less geared towards improving the quality of the solutions found, but comprises a scalable allocation and pricing heuristic that preserves incentive compatibility on the buyer side, the main design desiderata.

---

[4]Basing a decision on its expected value may yield more fine-grained information than on a majority vote of scenarios of the future, as is the case in consensus-based approaches. For example, taking action a_1 may be the better choice in two out of three scenarios. However, the losses from choosing a_1 in the third scenario may be such that choosing a_2 actually has higher expected value. It is in such situations that expectation (Chang et al., 2000b) and consensus-based approaches (Bent and Van Hentenryck, 2004) diverge in their respective decisions.

[5]Consider a scenario where decisions are made based on urgency of individual agents. In the absence of a truth-enforcing payment rule (Parkes, 2007), each agent has an incentive to misreport his type to receive preferential treatment.

2.3 Supply Chain Flexibility

Recent research has shifted the focus from factory-level flexibility towards more strategic aspects, i.e., flexibility on the supply chain level (Stevenson and Spring, 2007). In more detail, a hierarchy of flexibilities in the supply chain is established, ranging from the shop-floor level over tactical and strategic level (plant and firm levels) to the network level, i.e., supply chain flexibilities. Irrespective of the broad literature context which the authors provide, they urge caution in measuring flexibility; for a measure to be be meaningful, it needs to be defined adequately (Stevenson and Spring, 2007, p.693). Measuring flexibility becomes increasingly complex, or virtually impossible, if the scope of the measurement not only includes the actual level of flexibility of a supply chain, but also difficult-to-quantify hypothetical flexibility potentials.

Information Sharing One focal point of the literature review is concerned with information sharing, involving Information Systems (IS's) spanning multiple firms (Stevenson and Spring, 2007). Firms are – at least theoretically – interested in information sharing across corporate boundaries if doing so is beneficial to the firm. Benefits of information sharing may comprise improved "mix, volume, and new product flexibility" (Stevenson and Spring, 2007, p. 695).

Shared information can include demand or sales forecasts, inventory levels, lead times, quality (to increase trust between partners), and further dimensions. While it is found that all the studies cited report positive results from information sharing, some critical questions, revolving around individual incentives to take part in information sharing schemes, remain unanswered.[6] In particular, the question on how the arising benefits are distributed between participating firms is central to the success of information sharing. Beyond attribution of success and distribution of benefits, the question of how much flexibility to provide at what node in the supply chain is not trivially answered. Moreover, IS's facilitating inter-firm information sharing might imply a more rigid, less agile supply chain structure. This can be explained by the cost involved in setting up inter-firm links supported by IS's, as White et al. (2005) observe in a case-study on supply chain agility. While they do not doubt the benefits of cross-firm integration, White et al. (2005) also note that integration and flexibility on the supply-chain level are inversely related due to the cost of linking separate supply chain entities.

Flexibility to Create or Mitigate Uncertainty The relationship between flexibility and uncertainty in the context of supply chains has received special attention in the academic literature. Both, Upton (1994) and White et al. (2005) consider flexibility as

---

[6]The inclined reader is referred to the chapter on supply chain coordination and information sharing in Cachon and Terwiesch (2009, p.377).

a means to overcome uncertainty. This again highlights the complementary character of flexibility and uncertainty. However, other studies highlight another interesting aspect of flexibility: Gerwin (1993), for example are concerned with the strategic creation of uncertainty directed at creating additional cost towards a firm's competitors. Thus, flexibility may not only be used to cope with uncertainty, but instead to create additional uncertainty for competitors.

2.4 Flexibility in Operations Management

Operations management, according to the mission statement of the Manufacturing & Service Operations Management journal[7] is concerned with "the development of enduring knowledge that can lead to more efficient and effective processes for the creation and delivery of goods and services." Simchi-Levi (2013) gives an excellent overview of interesting problems in the field over the last two decades. These problems range from vehicle routing over manufacturing problems to revenue management and queuing problems faced by online retailers and ATM operators. Notably, the overview by Simchi-Levi includes an accessible intuition to the corresponding solution approaches.

Differentiating Agility, Flexibility and Responsiveness Bernardes and Hanna (2009) summarize the various views on flexibility within the manufacturing and operations management literature, and make the differing perspectives on the concept of flexibility clearer. In particular, they differentiate the terms agility, flexibility and responsiveness. They find *flexibility* to be an inherent system property and operating characteristic and define it as the "ability of a system to change status within an existing configuration (of pre-established parameters)". *Agility*, on the other hand is understood by Bernardes and Hanna to describe the "ability of the system to rapidly reconfigure". *Responsiveness*, in turn, is defined as the "propensity for purposeful and timely behavior change in the presence of modulating stimuli." Hence, agility can be seen as an approach to organize a system, responsiveness as the "outcome" of a system's organization and operations, and flexibility to be a lower-level "operating" characteristic (Bernardes and Hanna, 2009).

Revenue Management Early work in the field of Revenue Management (RM) has assumed prices as given and managers were only concerned with protecting or opening capacity of different fare classes in airline and hotel operations. Later on, due to both, the possibility of instantaneous comparisons between competitors, and decreasing costs of price-changes (so-called re-labelling costs), price has become an equally important control variable. Simchi-Levi (2013) outline the problem of price-based rev-

---

[7]http://pubsonline.informs.org/journal/msom#

enue management for an online retailer in the form of a case study. Therein, price decisions are subject to the exploration-exploitation trade-off which is well known from reinforcement learning (Sutton and Barto, 1998): Demand for new products is initially unknown. Through price modifications, the seller learns about demand (exploration), and can henceforth set revenue-maximizing prices (exploitation). Historically, the airline industry has been at the forefront of revenue management (Talluri and Van Ryzin, 2005) using computer-aided price and capacity control. Ideally, revenue management succeeds in maximizing capacity utilization, especially in the presence of fluctuating demand. Beyond the airline industry, revenue management finds application in hotel and car-rental industries, but also grid- and cloud-computing. While the services sold in this manner are highly heterogeneous, they share common characteristics such as time, location, and above all else, perishability.

Flexible Products Gallego and Phillips (2004) present a study on flexible products. They define a flexible product (flexible from the perspective of the seller) as "a set of two or more alternatives serving the same market such that a purchaser of the flexible product will be assigned to one of the alternatives by the seller at a later date." They present slightly different flight dates on the same day as an example for flexible products. The buyer can choose from either of the fixed alternatives, the flexible product, or not buying at all. The benefit to the supplier lies in observing demand up to the decision time point and hence, more efficient allocation decisions via reduced demand uncertainty. Gallego and Phillips introduce examples for flexible products in the area of air cargo, online advertising, but also hotel operators with multiple properties in close proximity, and, the field of opaque fares (such as Priceline[8] and Hotwire [9]) in general. Marketing such flexible (or opaque) products requires trading-off extra revenue from a highly price-sensitive clientele with the cannibalization of high-valued full-fare demand.

2.5 Structural Flexibility

As we have seen in previous sections, flexibility is a topic of central importance in manufacturing, supply chain management OM, but also other areas that we have not covered. Jordan and Graves (1995) coin the concept of "structural flexibility" and show that the addition of even minor amounts of flexibility to a system can lead to results virtually equivalent to fully flexible systems. In the automobile industry, according to Jordan and Graves (1995), "investment and tooling decisions" related to the assignment of products to plants are made at least one year prior to the begin of production. However, sales forecasts at that time are still subject to uncertainty, hence, an

---

[8]www.priceline.com
[9]www.hotwire.com
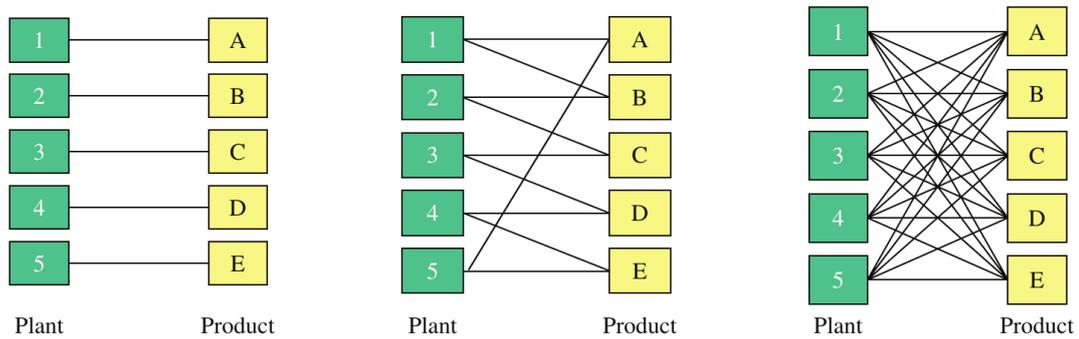
Figure 2.1: Process flexibility, from Simchi-Levi (2013)

inflexible system with a rigid 1:1 mapping of product to plant (Fig. 2.1, left side) would regularly experience underutilization or leave some demand unserved.

Flexibility vs. Capacity   This problem can be coped with through the introduction of process flexibility with respect to "product assignment decisions, i.e., decisions on which products are to be built at which plants or on which lines"(Jordan and Graves, 1995, p.577). Then, the question is "how much process flexibility is needed?" Jordan and Graves (1995) use (costly) flexibility in a sparse fashion, i.e., they propose adding a small amount of flexibility to each plant, such that a flexible plant is able to produce two products instead of a single one. This approach nicely contrasts with naïve, more costly approaches which would augment each plant with the flexibility to manufacture multiple products (illustrated in Fig. 2.1, right side). The decision which kind of flexibility to add to a specific plant is guided by the goal of constructing "flexibility chains". For maximum flexibility in such flexibility chains, it is important to "close the chain", i.e., to install capabilities regarding product *A* at plant 5 (Fig. 2.1, center). With such flexibility, production shifts through the chain from one plant to another become possible without installing the costly capability to produce every product at every plant (Cachon and Terwiesch, 2009).

If demand for specific products is uncertain, flexibility can act as a substitute for capacity (Jordan and Graves, 1995). Fig. 2.2 illustrates the relationship between expected sales and capacity utilization if demand for different products is negatively correlated.[10] The dotted lines indicate different capacity levels, the lower solid line illustrates the case of no-flexibility, i.e., dedicated plants, while the upper solid line represents the case of one-chain flexibility. Clearly, depending on the relative cost of flexibility and capacity, different system configurations are optimal. Flexibility, however, is most valuable in settings with balanced overall supply and demand, but where product-specific demand might vary. This corresponds to points *D* and *E* in Fig. 2.2. In the extreme cases of low capacity, high capacity utilization is achieved even without

---

[10]Jordan and Graves (1995) assume a correlation coefficient of −0.5.
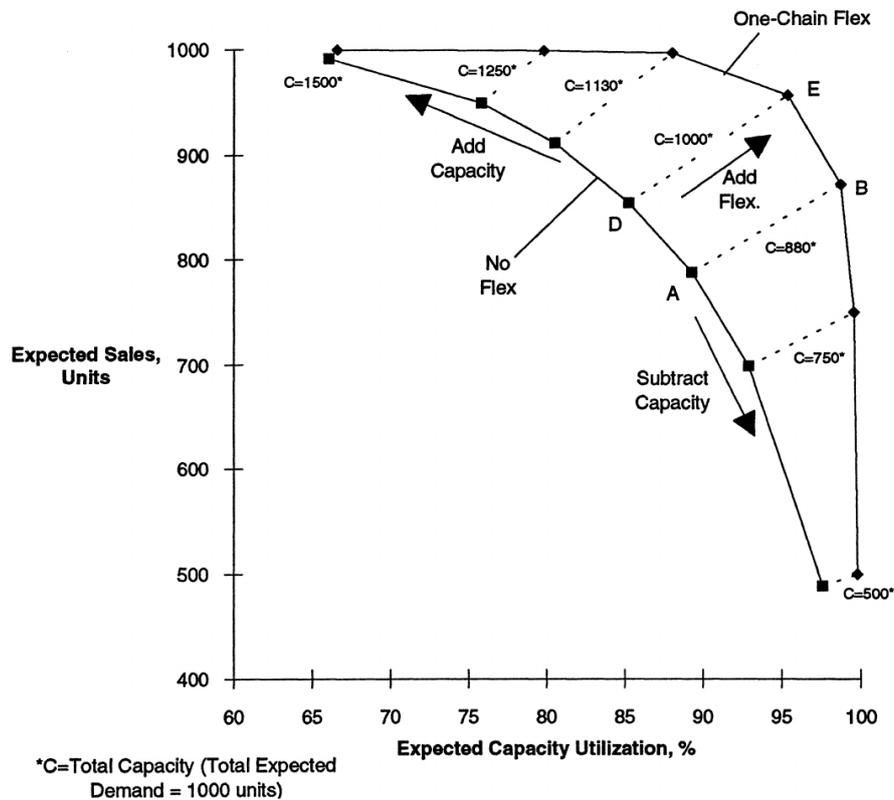
Figure 2.2: Gain from process flexibility under uncertain, negatively correlated demand for multiple products (Jordan and Graves, 1995). The largest improvement can be achieved in systems where quantity demanded is equal to capacity installed (D → E).

flexibility (lower right in Fig. 2.2). At excessive capacity, flexibility is of little value as well: Each plant (manufacturing a single product) is able to cope with fluctuations in demand by itself. Cachon and Terwiesch (2009) confirm this observation. Moreover, Graves and Tomlin (2003) show that one single "long" chain, i.e., a chain involving a large number of plants, is more valuable than several smaller, independent chains.

Formalization of Structural Flexibility Motivated by the above goal of creating more efficient processes and aiming to better comprehend the underlying essential components of structural flexibility, Iravani et al. (2005) build on the seminal work by Jordan and Graves and formalize the concept of structural flexibility. They aim at providing strategic insights for the design of manufacturing and service operations. They develop a variety of indices "that quantify the ability of a system to respond to variability in its environment." Compact information representation in the form of scalar indices representing structural flexibility (Iravani et al., 2005) enables improved managerial decisions at the strategic level regarding manufacturing processes. In more detail, structural flexibility in a graph of supply and demand vertices, differentiated by

the product that can be manufactured or consumed, is defined as the "number of non-overlapping paths a system can use to respond to a particular change in demand." The graph structure can be encoded into a structural flexibility matrix whose elements are obtained by solving the corresponding max flow problems between the set of *demand* vertices. Intuitively, the more connected the graph, the more flexibility can be expected from the manufacturing process. However, the authors stress the importance of intelligently adding flexibility to the corresponding process. Their results show that systems with lower flexibility can provide a superior response to changes in demand relative to a more connected system, if the edges in the graph are installed with structural flexibility in mind. The authors propose the Mean and the Eigenvalue indices and find that both are well-suited to express structure-inherent flexibility. These indices are attractive from a managerial point of view, as they allow the expression of flexibility as a single scalar. Hence, Iravani et al. show that flexibility with regard to system structure can indeed be measured and therefore be taken into account when deciding on system structure on a strategic level, i.e., at the time of system creation. In a way, their work is thus related to the entropic flexibility measures proposed by Kumar (1987) and introduced before. Both approaches feature a tractable formal definition of flexibility that is based on flexibility graphs.

2.6 Strategic Flexibility

On the one hand, flexibility may serve as a level to induce uncertainty for competing firms through corresponding investment decisions that retain product or volume flexibility for subsequent tactical and operational decisions. Besides obvious benefits in coping with uncertainty, flexibility can also have a strategic value that may open or close certain trajectories for a firm's future. For example, product (as well as plant configuration and plant location choices), can be made with the focus of gaining a long-term competitive advantage by allowing a greater set of options contingent on the decision today or simply to increase competitors' uncertainty (Sethi and Sethi, 1990; Beach et al., 2000).

Thereby, strategic decisions in the face of uncertainty may aim to increase flexibility, i.e., enlarge the space of states that are attainable after a certain decision is made. This section focuses on the latter, sequential decisions and, naturally arising thereof, real-options.

Sequential Decisions Sethi and Sethi, referring to Marschak and Nelson (1962), present an ordinal measure for flexibility in sequential / staged decisions. A first stage decision $a1$ is more flexible than another decision $a2$ if the set of states reachable as a consequence of choosing $a1$ is a superset of the set of states reachable as a consequence of choosing $a2$.

Benjaafar et al. (1995) give a detailed treatise on the strategic value of flexibility regarding manufacturing performance and formalize the concept of flexibility in (sequential) decision making under uncertainty. They attribute flexibility to an action that retains "future decision making freedom" (Benjaafar et al., 1995, p.444). Intuitively, flexibility can be understood as a property that "allows [decision makers] to change their minds once an initial action is taken." For ease of expositions Benjaafar et al. treat an artificial two-stage decision processes, which they claim "can [...] easily [be] extended to multi-stage situations." They posit the value of an action to be a function of initial uncertainty, initial estimate of expected value, expected future information, and expected future flexibility. In general, the authors show that "in situations where future system states cannot be easily predicted, control strategies should attempt to maximize future system flexibility." Remarkably, this is in line with the decision strategy proposed by (Petersen et al., 2012), which aim to maximize agility in the face of uncertainty and on-line decision making. It is also in line with the entropy-based flexibility measures of Kumar (1987).

Real Options Following the seminal work of Black and Scholes in 1973 on option pricing, there have been efforts to apply their method to value real-world flexibility. Academic research has, for example, modeled investment decisions under uncertainty (Dixit, 1994), giving rise to the notion of real options. Notably, Sethi and Sethi (1990) ignored an important aspect of strategic flexibility: The flexibility to defer investment or to even abandon a project. In contrast, Bengtsson (2001) present a broad overview at the intersection of manufacturing flexibility, in the sense of Sethi and Sethi (1990), and real option theory. Thereby they emphasize flexibility in the realm of project postponement and abandoning. The authors concede, however, that the particular type of strategic flexibility must remain rather simple (e.g., two-product firms) in order to allow for analytical solutions. A central question regarding the basic machine and routing flexibility remains unanswered, as it is not clear which type of distributions/underlying stochastic process should be assumed. This poses an important obstacle to practical application, as valuation results and corresponding decisions might be directly driven by flawed assumptions, and hence be highly unreliable. Extending previous work, Bengtsson and Olhager (2002) present a model to value product mix flexibility and evaluate it empirically through a case study. Clearly, results on the economic value of flexibility require careful conditioning on the model and domain under consideration. Based on the model, however, Bengtsson and Olhager (2002) are able to determine the optimal level of automation, involving automatic and semiautomatic manufacturing resources. Furthermore, they derive the value of extra capacity, which can provide additional (volume) flexibility.

2.7 Flexibility in Electric Power Systems

The electric power system has peculiar characteristics: First, both sides of the system must be balanced at all times to ensure safe system operation. However, as the cost of storing electrical energy is prohibitively high, supply (and demand) must be actively controlled and hence decided upon to achieve the system's goal. Second, the "transportation" of electrical energy from the location of generation to the final consumer is bound to the power grid. In order to retain a functioning system, flexibility at each stage can be leveraged to improve the economics of systems operation. Accordingly, the following sections introduce flexibility concepts on the sides of generation, grid operation and consumption.

Supply-Side Flexibility

The most fundamental constraint of power system operations concerns the equality of quantities supplied and demanded at all times. This constraint arises from the non-storability condition of electrical energy.[11] In power systems dominated by conventional generation, demand is forecast, while generation is controlled and adjusted to follow consumption, and thus a balanced power-system is achieved.

Dispatch Flexibility   Deciding which part of the generation park to activate with the goal of minimizing system cost while retaining safe and reliable system operations is a computationally hard problem. This so-called unit-commitment problem has historically received wide attention, as better solutions directly affected system efficiency (Happ, 1977).

After liberalization, however, the constraints for the individual firm have changed: Instead of ensuring safe and reliable system operation, the generating firm is required to meet contractual obligations. More recently, uncertainty from both, controllable loads and increasing volatility from renewable generators has been examined in the context of dispatch flexibility and unit commitment (Damousis et al., 2004; Bertsimas et al., 2013). Solving the unit-commitment problem to optimality is, for typical instance sizes challenging. Therefore, mixed-integer programs are complemented with meta-heuristic approaches in practice. While these approaches do not guarantee optimality of the solution, they are able to find good-enough solutions within acceptable time periods (Mantawy et al., 1999; Kazarlis et al., 1996). On shorter time scales (i.e., on the seconds to 15 minute level) deviations from a balanced system state are typically resolved through the dispatch of different types of ancillary services, differentiated by the time required for them to come online. These ancillary services are

---

[11]Electricity is storable in other forms of energy (potential, chemical, mechanical), but this comes at high cost.

typically provided for by conventional generation capacities.[12] To this end, the system operator can declare some generation capacities either as "must run" to provide the necessary control capacity, or, depending on the market design, generators taking part in the Day-Ahead Market (DAM).[13]

Ramping Flexibility   In the wake of increasing contributions of volatile and difficult-to-predict RES to the supply side, new challenges for safe and efficient system operations arise. First, ramping of conventional generators will be observed more frequently to compensate for fluctuations of RES. Second, conventional generators are replaced by their renewable counterparts, which depend on natural supply (wind, sunshine), reducing the control capabilities on the generation side of the system over time. The challenge in system operations will be (or already is) to provide acceptable quality of supply on decreasing amounts of control capacities. Projecting this trend into the more distant future, a paradigm shift on power system operation is necessary where the demand side is expected to take on a more active role. Otherwise, the promise of sustainability through integration of RES might turn out to be improvident.

Grid Flexibility   The power network has two original roles: First, providing a (physical) connection between the demand and supply sides. Second, enhancing supply reliability, system resilience, and overall economics of the power system through uncertainty and flexibility pooling. By means of uncertainty pooling (Cachon and Terwiesch, 2009), unexpected, random changes in demand by one consumer are, to a large extent, offset by other consumers' opposing changes, without requiring adjustment on the supply side.

   Flexibility on the grid level can be mostly categorized into volume flexibility, i.e., the power network can easily be operated at different levels of intensity. However, if the power network is equipped with active control equipment, its topology can be adjusted to reflect the current supply and demand situation by power switches and increase the system's transfer capacity. Such flexibility to disconnect individual lines and hence shape the network's morphology, allows to trade-off transfer capacity for power losses, hence providing an additional level to shape volume flexibility.

Demand-Side Flexibility

After introducing flexibility from a generation and grid perspective, we now turn our attention to demand-side flexibility. First, we provide a macro-perspective of economic flexibility potentials, continue with a taxonomy for different kinds of flexibility

---

[12]The inclined reader is referred to (Rebours et al., 2007a,b) for more detail on the technical and economic aspects of ancillary services.

[13]A comprehensive review of concepts of DAMs is provided in http://content.caiso.com/training/Day-Ahead_Market_Overview/index.html (accessed September 4, 2014)

and their interrelationship, to eventually arrive at portfolios of flexible loads.

Economic Potential of Flexible Loads   In order to quantify the economic benefits of DSM, standards that facilitate fair evaluation and comparison between competing approaches are required. However, as Strbac (2008) state "[...] a lack of methodologies for the quantification of cost and benefits" renders the task, at least so far, challenging. Nevertheless, Strbac present the (admittedly coarse) intuition that the value of DSM may be increasing in system load, i.e., higher in systems that are close to requiring system expansion. Still, the authors abstain from estimating monetary values.

A more specific use case concerning co-generation and heat pumps is examined in (Fehrenbach et al., 2014). Assuming a macro-perspective but relying on a bottom-up large-scale optimization model, the economic potential of virtual power plants consisting of micro-cogeneration plants, heat pumps and thermal storage for private households is estimated. They show that by combining "thermal and electric load management" significant flexibility potentials in Germany can be established, which in turn allow for better integration of RES into the power system. They identify two drivers for overall proliferation of heat-pumps (which are responsible for additional demand-side flexibility). First, the share of renewable generation in the power system, and second, the price of fossil fuels. Both factors are estimated to have a positive impact on the adoption of heat-pumps. Larger shares of RES reduce the cost of the main factor of production for heat-pumps. Micro-cogeneration, on the other hand, is negatively affected from higher fuel prices. Somewhat surprisingly, Fehrenbach et al. show that under their model assumptions, electrical heat pump capacity may account for up to 50% of peak system load in 2050. Furthermore, their results indicate that "the electrical capacities of heat pumps available for load management are consistently and significantly higher than those of micro-Combined Heat and Power (CHP) for flexible generation." This statement underlines the importance of demand side flexibility in future power systems relative to flexible generation.

Taxonomy for Flexibility   The term "demand-side flexibility" is mostly used in a context that relates to deferring loads in time (Strbac, 2008). Gottwalt et al. (2011), for example, outline to what extent home automation in the context of smart homes could provide load flexibility by basing operation decisions on price signals. Their earlier results indicate that the work of Fehrenbach et al. (2014) is directed towards the most promising part of demand side flexibility, i.e., the part offering the largest economic potential: Water and space heating. Both parts are excellent candidates for load deferral as most consumption takes place during nighttimes, i.e., there is virtually no loss in comfort from shifting the heating period to an earlier point in time, as the systems offer sufficient amounts of thermal storage.

In any case, demand side flexibility may be of heterogeneous quality. Compared to power systems exclusively controlled via generation capacities, classifying and quan-

tifying the quality of flexibility is a novel problem. Petersen et al. (2012) put it as follows:

> "In a traditional energy system, where flexibility is provided by power plants, the quality of a given flexible resource can be determined fairly unambiguously by activation time, length of reservation period and capacity. In a Smart Grid system, flexibility should be provided by flexible consumer appliances, which are not created for power management. Determining the quality and value of a flexible resource therefore becomes far more multifaceted, since additional performance constraints such as storage capacity, minimum runtime, temporal constraints (deadlines), ramp rates etc. must be considered."

In order to generically classify different qualities of flexibility, (Petersen et al., 2013) present a taxonomy for modeling demand side flexibility in smart grids. Depending on the device under examination (for example EVs or Heating, Ventilation and Air Conditioning), different qualities of flexibility can be provided to the system. Somewhat flexibly, Petersen et al. describe "flexibility as the ability to deviate from the plan". The main focus of their article is on establishing a taxonomy that relies on classifying devices, and therefrom arising flexibility potentials, according to operating regimes. Regimes are differentiated by power capacity, energy capacity, energy level at a specific deadline and minimum runtime. In particular, the authors derive three quality levels of demand-side flexibility, illustrated in Fig. 2.3 in the following order:

1. Buckets are the highest quality source of flexibility. They are constrained by only two restrictions: Energy and power capacity. HVAC as the prime example of this segment ensures indoor temperature to be within a predetermined energy level (indicated via dashed lines in Fig. 2.3). Power of heating and cooling is limited, hence the power capacity constraint.

2. Batteries are Buckets that are additionally constrained by a deadline and energy level. EVs are good examples as they require a certain State-of-Charge (SoC) at a certain time, in addition to being restricted in both dimension of energy and power.

3. Bakeries are Batteries with the additional constraint of having fixed power consumption over time and a fixed run time. The corresponding example relates to non-preemptive household appliances where the only decision, similar to Gottwalt et al. (2011) concerns the start date.

Flexibility Portfolios   Real-world power systems feature heterogeneous consumers (devices), and hence, a heterogeneous flexibility potential. To complicate matters further, the cost associated with leveraging flexibility might be heterogeneous, and be
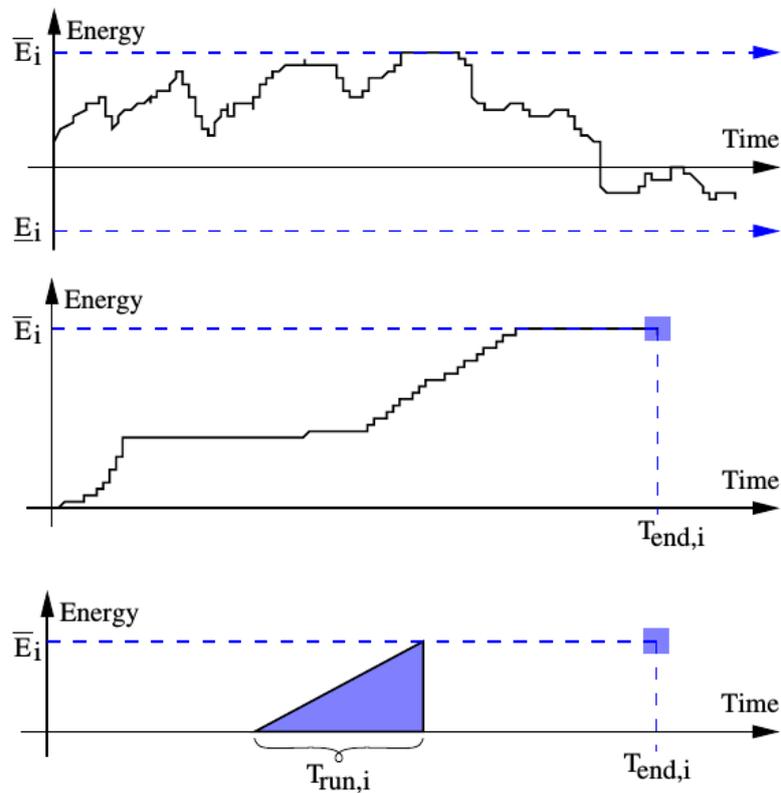
Figure 2.3: Buckets, Batteries, and Bakeries (Petersen et al., 2013, p.3)

only partially related to the quality flexibility, but rather with the amount of flexibility provided. In (Petersen et al., 2014), the authors strive for low waste of energy from RES and minimal use of conventional generation by means of different control algorithms for heterogeneous consumer portfolios. They show that one of the developed greedy algorithms (named "GRASPsorted") overall performs quite well in shaping the demand curve over time as close as possible to generation from RES. The authors stress the fact that algorithms that rely on fast sorting algorithms instead of solving more involved Mixed-Integer-Program (MIP) problems might yield sub-optimal solutions, but are very fast, and thus applicable to large-scale DSM scenarios. Nevertheless, for cases where there is only a limited amount of high-quality flexibility available, the tracking error increases considerably. Therefore, entities aiming to provide ancillary services based on flexible demand must ensure to have access to a well-balanced portfolio of sufficiently high quality.

Interestingly, the issue of incentives for consumers to provide their flexibility in the first place to a controlling entity is not elaborated on. Instead, it seems most authors believe that flexibility will be provisioned by consumers in an altruistic manner.

One exception is Gärttner et al. (2014); Gärttner et al. (2014) which focus on portfolio composition from the perspective of a utility. Selecting customers that offer the

appropriate kind of flexibility at an acceptable price, utilities can leverage consumer heterogeneity to their own advantage instead of procuring costly balancing power to offset fluctuations from RES. While the cost of consumer flexibility in their model is determined exogenously, it could perspectively be substituted by values derived from empirical customer research and hence provide valuable decision support to utilities in composing their customer portfolios.

He et al. (2013) take the temporal aspect of energy consumption into account when referring to demand side flexibility, but, in general, pursue a broader view than Petersen et al. (2013). Further they provide an engaging overview into the "matching between demand response services and contracts", differentiated by contract features and response requirements. As Fig. 2.4 illustrates, the load mix can be divided into storable and non-storable load. The non-storable part can further be separated into (non-) shiftable load, which can be further segmented into curtailable and non-curtailable load. At the very core of the illustration lays base load, that may not be modified in any way. Relative to the taxonomy provided by Petersen et al. (2013), storable load concerns both, Buckets and Batteries, while Bakeries are mostly identical to shiftable load. Interestingly, He et al. (2013) add the segment of curtailable load (i.e., "load that cannot be shifted without affecting the end-use service, but the service can be interrupted instantly") to the spectrum of flexibility.

Beyond the segmentation of demand according to flexibility, they concentrate on the segment of small private consumers for two reasons: Those are able to provide the level of decentralization that is most valuable in future power systems, and, in contrast to larger consumers such as manufacturing firms, they face more market barriers and transaction costs. Hence, such small, dispersed consumers are ideally suited to benefit from utilities' intermediary offerings and provide valuable flexibility to the system. He et al. showcase the need for diverse contract types to account for consumer heterogeneity. However, they acknowledge that distorted incentives for intermediaries might preclude the offering of a diverse range of contracts, limiting the potential of demand side flexibility.

2.8 Flexibility in Transportation Systems

Transportation systems oftentimes suffer from congestion, delay and breakdowns. However, in most systems different kinds of flexibility are present that allow to alleviate the negative repercussions. This section presents a brief overview and pointers into the literature for transportation system-related flexibilities.

Routing Flexibility   Routing flexibility allows the individual traveler to adjust initial travel plans in the wake of congestion or unexpected events. Through re-planning the traveler can avoid congested segments on his way and hence reduce overall travel time, often termed "latency" in the transportation literature. On a macroscopic level,

Load mix				
Storable load	Non-storable load			
	Shiftable load	Non-shiftable load		
		Curtailable demand	Non-curtailable load	
			Base load	
e.g.: EV charging, heating, cooling	e.g.: laundry, dryer, dish washer, stove	e.g.: lighting, computer, handyman tools, TV	e.g.: Appliance standby, burglary alarm	
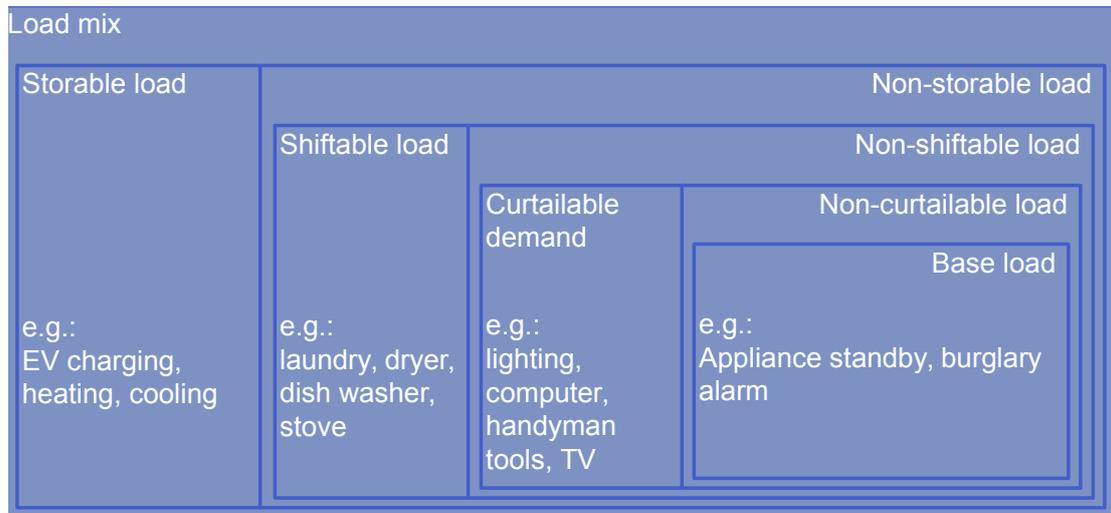
Figure 2.4: Load flexibility, adapted from He et al. (2013).

routing flexibility may be employed to reduce overall latency, i.e., increase social welfare.

Fundamental research concerned with exploiting routing flexibility can be traced back to the seminal traffic assignment problem in Beckmann et al. (1956). The authors considered a problem where the cost of congestion on an edge depends on the flow along that edge. In equilibrium, no vehicle can improve its travel times through unilaterally deviating from the chosen route.[14] Clearly, drivers on road networks are assumed flexible with respect to their route.

In logistics, routing flexibility may be leveraged through online updating of route plans, for example in Traveling Salesman Problems (TSPs), to better cope with unexpected changes to the network topology (e.g., accidents and construction sites) or congestion. As UPS has demonstrated through its widely known approach of avoiding left-hand turns for its delivery truck fleet, routing flexibility can be exploited in logistics to achieve better economic outcomes.[15]

Interestingly, the idea of pricing the use of a (public) good based on the negative externalities caused by usage were developed when William Vickrey was concerned with congestion pricing in transportation networks (Vickrey, 1955). More recently, Roughgarden (2005) has focused on the "price of anarchy", i.e., situations where selfish agents choose their travel paths in equilibrium, while there is assignment that achieves lower overall latency. The difference in social welfare between the optimal and the equilibrium assignment is termed the "price of anarchy". This line of work is closely related to the Braess Paradox, which states that the addition of an edge to a

---

[14]The inclined reader is referred to Sheffi (1985) for an excellent overview of transportation problems and equilibrium analysis.

[15]http://priceonomics.com/why-ups-trucks-dont-turn-left/

congested network may result in decreased overall performance (Braess, 1968).

Another strand of research in computer science, is concerned with fast route computation on continental-size road networks (Geisberger et al., 2008). Such algorithmic advances enable novel applications, i.e., ride-sharing. Here, the economic matching of drivers and riders demands for teaming travelers with similar routes within short time spans. In this spirit, Abraham et al. (2013), for example, present an efficient algorithm for finding alternative, reasonable paths in road networks, which may be suited for both, re-routing and ride-sharing.

Temporal Flexibility   Travelers may be flexible with respect to the time of departure and arrival of a certain trip. They may also accept a variety of travel durations, if longer travel times are compensated in another dimension. Further, there may be certain preferences regarding the expected travel time and the expected risk of choosing a specific travel option. Accordingly, travelers may be willing to accept longer travel times in exchange for more frequent service (which may constitute a proxy for reliability). Small (1982) employs time-dependent demand functions to model "scheduling of discrete activity directly at the individual level." He finds that there is considerable difference in the marginal rates of substitution of travel times and being late to work between white and blue collar workers. Thereby, blue collar workers are significantly more averse to being late than their white collar colleagues. Also, workers living in single households are less willing to give up leisure time to be at work on time. This seminal study has sparked a number of interesting research directions. One of them (Fosgerau and Karlström, 2010), estimates the value of reliability in public transport, which maps nicely into the previously mentioned trade-off between travel-time and travel frequency.

Multi-modal Mobility and Flexibility   Transportation systems offer various options for reaching one's destination. Depending on the distance and the origin-destination pair under examination, different mobility technologies are most appropriate. On the one hand side, flexibility with regard to the means of transportation may be used to increase transportation capacity in a multi-modal network. Additionally, it may also be used to enable certain origin-destination pairs, which, if flexibility was lacking, would be infeasible. Furthermore, such flexibility is the basic prerequisite for multi-modal mobility.

Orchestrating multiple modes of transportation to serve an origin-destination pair both creates flexibility for the traveler, but also requires flexibility on his behalf, i.e., the traveler must be willing to switch the means of transportation en route. While this may induce a cost in terms of time or inconvenience, it may also add numerous links to a space-time transportation network, hence augmenting its flexibility potential. In particular, dynamic reassignment between different means of transportation may enhance the reliability of the overall network. The potentially largest benefit from multi-

modal travel, compared to travel by car alone, may lie in more relaxed travel, reduced uncertainty regarding time of arrival and cost of travel, but also improved safety.

A critical building block in convincing travelers to use public transport options may be easily accessible and timely information on the current state of the system. This may include forecast arrival times, including detailed, yet accessible (probabilistic) information on future delays. Such information may foster consumers' perception that customers are indeed in charge and not at the mercy of the service provider. This is especially important for multi-modal mobility systems, where delays can quickly compound, negating the concept's positive potentials.

One technology may be particularly helpful in introducing multi-modal transportation systems: The self-driving car (Thrun, 2010). Such vehicles may pose serious competition to established taxi services, on both, long and short distance travel routes. In particular, self-driving cars may consolidate the taxi and car-sharing businesses in the long run. Important past and current problems, such as the vehicle-routing problem and its numerous variations may enjoy increasing attention due to the corresponding mobility evolution.

2.9 Discussion

Flexibility was, is, and will be a concept difficult to define and hence measure due to its ambiguity. This chapter provides a broad outline of flexibility in various domains involving different classifications and measurement approaches. In this work it serves as a foundation regarding the treatment of flexibility in settings involving the Smart Grid and Car-Sharing. The concept of structural flexibility established by Jordan and Graves (1995) and subsequently enhanced and formalized by Iravani et al. (2005) is closely related to this work. They showed that small amounts of flexibility may be sufficient to achieve outcomes closely matching the results achievable in a fully flexible system. Applied to car-sharing this may imply, that few, appropriately selected flexible customers are sufficient to achieve close-to-optimal results. The same holds for the smart grid domain: In power systems dominated by RES not all demand must be flexible; instead, relatively small amounts of flexibility may again be sufficient to effectively integrate RES into the power system. Clearly, the value of flexibility in either application depends to what extent flexibility can be formed into "long" chains.

Compared to most established literature on the treatment of flexibility, the work at hand is specifically interested in highlighting the difference between online and offline decision making in the spirit of Chang et al. (2000b) and Bent and Van Hentenryck (2004). This separation allows us to quantify the value of information, i.e., the extent to which a good manager (planner) can improve operations. Beyond the optimization perspective, we are interested in designing mechanisms that make can be applied in the presence of self-interested agents with private information (this has been briefly introduced in section 2.2).

Part II

Smart Grid: Incentive Engineering for Flexible Consumers

Chapter 3

Smart Grid Fundamentals

Large-scale proliferation of RES and the associated decentralization of the power system introduces a novel, significant source of uncertainty into power system operations. Volatile and intermittent output from RES, in particular, poses a serious challenge to security of supply. To cope with this challenge, various measures to shape consumption according to generation have been proposed (Sioshansi, 1995; Strbac, 2008). These efforts mostly rely on the establishment of an appropriate information, communication and control infrastructure. The envisioned "Smart Grid" infrastructure enables robust economic control of large quantities of dispersed consumers, thus "activating" the demand side of power systems. Therewith, efficient integration of large amounts of RES may be achievable.

In this chapter, we establish the foundations for subsequently introduced economic coordination mechanisms in the smart grid, and provide both relevant and necessary background information concerning the smart grid.

The structure of this chapter is as follows. First, we introduce fundamentals of power system operations, including a brief overview of restructuring efforts in global key markets to gain a better understanding of how the current situation came into existence, and the current state of the electricity markets as well as technical constraints of grid operations. We continue with the smart grid concepts, followed by "smart markets", a conceptual separation introduced by the German Bundesnetzagentur to foster more precise communication.[1] Thereafter, motivated by the growing importance of RES in power systems, we present insights regarding the interplay of sustainability and uncertainty. Finally, we devote one section on DSM, which, as we will see later on, will play an essential role in economic coordination within smart grids. We conclude with a brief discussion.

---

[1] In their perception, the smart grid should be concerned with the network view, while the smart market side should be concerned with economic allocation and pricing decisions for which a functioning technical smart grid infrastructure is the necessary prerequisite.
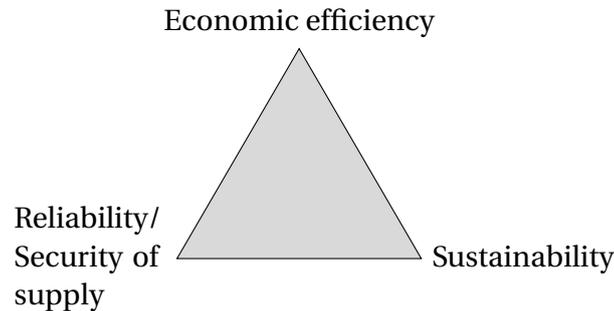
Figure 3.1: The energy trilemma. Power system design and operation must trade-off between ecological sustainability, economic efficiency, and reliability.

3.1 Power System Fundamentals

Germany's 2013 primary energy consumption, which includes the electric power sector, is dominated by oil (33.4%), natural gas (22.3%), anthracite (12.8%), lignite (11.7%), and nuclear fuel (7.6%). The share of renewables accounted for 11.5%; overall primary energy consumption amounted to 13,908 PJ[2]. While the electric power systems relies to a large extent on fossil fuels, it is in the midst of transformation, both physically, and economically, with RES gaining in importance. This section first introduces background information on technical constraints in power system operations. Thereafter, we present the restructuring process most power systems underwent within the last two decades, and which led to the emergence of a classic value chain in power systems. Finally, the current market architecture and traded products on European energy markets are introduced.

3.1.1 Technical Constraints in Power System Operations

Power system operations are subject to numerous technical constraints which ensure safe service provision and hence differentiate this sector from others. However, a number of constraints are actually soft constraints, i.e., slack, or flexibility, can be leveraged to achieve more favorable objectives in another dimension.

Frequency System frequency is determined by the speed at which rotating generation capacities are operated. System frequency amounts to $50\,Hz$ in European power systems, i.e., 50 phase cycles per second.[3] If quantities supplied exceed (fall short of) quantities demanded, frequency increases (decreases). Frequency is a global property

---

[2]http://www.ag-energiebilanzen.de/index.php?article_id=29&fileName=ageb_pressedienst_02_2014_jahresbericht.pdf accessed on 2014/09/01.

[3]North-American power systems are operated at $60\,Hz$.

of the power system, hence the entire power system is synchronized to the exact same frequency. To avoid damage to equipment and machinery connected to the power network, frequency must be held within narrow admissible corridors.

In case of deviations from the frequency set point, automatic control mechanisms based on negative feedback in the sense of control theory are executed and return the power system to the desired state. Most often, this involves re-adjusting the steam valves in thermal generation units. In cases where short term control measures are insufficient to achieve the desired effect, minute reserves are called upon, i.e., the system operator asks additional capacities to come on-line or to be shut down.[4]

Voltage   Voltage of an electrical current determines the amount of useful work that can be performed, i.e., it determines the power that is transferred via the current. To ensure efficient, and more importantly, safe operation of electrical equipment, voltage is kept within (compared to system frequency to somewhat less restricted) $\pm 10\%$ limits[5].

Voltage increases at the point of in-feed and decreases at the point of consumption. A major challenge in today's power networks relates to a lack of monitoring (and hence control), that is to be alleviated through smart grid adoption.

Line Limits and Power Flow   Formally, the flows of electric current can be modelled according to Kirchhoff's laws. Flows along the edges between two nodes A and B in a power network are split between connecting lines according to their respective conductance (Stoft, 2002).[6] This property sets the problem of modelling power flows apart from transportation / transshipment problems. Power flows between supply and consumption nodes in the power network are inversely proportional to the impedance of the connecting paths. The first consequence of this property is that not every location requires the same amount of generation for its demand to be satisfied, as losses vary between locations. Second, while aggregate network transfer capacity might be sufficient to cover demand at node B from supply at node A, the edge $\overline{CB}$ might be overloaded, requiring local supply at higher cost. Thus, the marginal cost of service differs depending on the node location in the graph of the power network.

Skånlund et al. (2013) analyzed historical price and power flow data for Germany and concluded

> "[...] that prices in Germany do not seem to reflect local balances sufficiently. [...] prices in northern Germany seem to be too high in some

---

[4]For a more detailed treatise into regulation services, see for example Eßer-Frey (2012, p.38ff) and Weidlich (2008, p.11ff)

[5]cf. DIN EN 50160:2011-02

[6]The European commission has commissioned a study on loop flows which is available at http://ec.europa.eu/energy/gas_electricity/studies/doc/electricity/201310_loop-flows_study.pdf, accessed on September 10, 2014.

Technology		Hard coal	Lignite	CCGT	GT
Load Gradient	$\%P_N$/min	1.5	1	2	8
Minimum load	$\%P_N$	40	60	50	50
Start-up times hot (< 8h)	h	3	6	1.5	0.1
Start-up times cold (> 48h)	h	10	10	4	0.1

Table 3.1: Ramping and minimum load constraints of selected conventional power generation technologies in today's power systems. P_N represents nameplate capacity.[8]

> instances, not only triggering "wrong" generation in Germany, but also triggering "wrong" cross-border flow and generation abroad, which potentially amplifies the challenges associated with unscheduled flows. "

Clearly, increasing RES capacities, in particular wind power generation capacities, pose a growing challenge for both, efficient grid operations and price determination due to lacking transmission capacities and volatile feed-in.

Ramping constraints   Power system flexibility is largely due to (conventional) generators' ramping flexibility, i.e., the ability to quickly adjust output up and down, according to system's needs. Different generation technologies exhibit various constraints on ramping. Typically, nuclear power plants are slow to adjust output, while gas turbines are considered to have most advanced ramping capabilities. Table 3.1 presents a detailed overview of ramping and minimum load constraints in today's power system, including historic generation capacities. Gas Turbines (GTs) can adapt their output by approximately 8% per minute, while power plants fueled by hard coal are constrained to 1.5% per minute. Compared to GT technology alone, utilizing waste heat from the gas turbines via Combined Cycle Gas Turbine (CCGT) technology achieves higher thermal efficiency, at the price of stricter ramping constraints. GT technology has also superior start up times, from either hot or cold state, of approximately 5 minutes.[7]

However, fast ramping capacities are usually more costly to operate than their slower-ramping peers. Hence, for efficient power system operations, the appropriate mix of both, low cost and flexible generation is important. With growing RES penetration, on the one hand, larger amounts of flexible generation are required for safe system operations. On the other hand, market prices are insufficient to cover the higher

---

[7]Values presented are representative for the installed generation park. Current state of technology for new installations is (slightly) better.

[8]http://www.effiziente-energiesysteme.de/fileadmin/user_upload/ PDF-Dokumente/Veranstaltungen/Workshop_Retrofit/3_SIEMENS_ Feldmueller.pdf

cost of flexible generation. This so called "missing money problem" (Cramton and Ockenfels, 2012) motivates the discussion about capacity markets to a large extent.

3.1.2 Liberalization

In the late 1980's governments began liberalization of the power sector in order to improve the sector's efficiency through competition. As large parts of the system were no longer considered natural monopolies, formerly vertically integrated entities underwent unbundling.[9]

Transmission continues to be considered a natural monopoly, while distribution, i.e., marketing and billing were liberalized. Consumers, similar to the telecommunications sector, were able to choose from a variety of competing retailers, often only differentiated by price and billing interval.

On this note, generation was liberalized, different market forms for exchanging energy established, and both, benefits and shortfalls of liberalized generation discovered.[10] As Sioshansi (2006) argues, borrowing from an article by Paul Joskow, "[...] in nearly all cases, the initial introduction of reforms has led to 'reform of the reforms' and in most cases to 'hybrid' markets, with significant challenges for policymakers." Clearly, separation of functions in the value chain that to this time were served by the same entity has oftentimes had unanticipated consequences. In some cases, for example Brazil, parts of Canada, and the Russian Federation initial public support waned in the aftermath of noticeable price increases. Beyond reversing political decisions, some European states have initially been rather reluctant to market competition, fearing the demise of strategically important energy champions among other reasons (Sioshansi, 2006). A prime example is the lack of creating additional cross-border transmission capacities to alleviate bottleneck situations. The general takeaway of the article is that "hybrid markets" have begun playing a significant role, as "textbook prescriptions" on how exactly markets should be implemented have led to disastrous failures. This is in line with an earlier remark by Sioshansi that "[...] deregulation is essentially a misnomer. No electricity market has been (or, in fact, can be) fully deregulated."

California Electricity Market Reform   The case of the California market reform has, due to its spectacular failure with rolling blackouts in 2000 and 2001, received special attention in both, the academic literature and general media. Besides problems from poorly implemented and monitored incentive schemes in the transmission network, issues related to the (mis-)use of market power led, among other consequences, to the

---

[9]Unbundling describes the separation of once vertically integrated (energy) companies into separate entities responsible for a single component of the value chain.

[10]For a compact representation of key electricity market reform in 15 world regions, the inclined reader is referred to Table 2 in Sioshansi (2006).

bankruptcy of Enron. Shortly before the deregulation of the California electricity market, large utilities had divested important parts of their generation portfolio (Bushnell, 2004). This, counterintuitively, increased in ownership concentration in southern California. However, Bushnell argues that market failure was only to a minor extent due to lackluster competition, but rather should be attributed to poor installation of long-term supply schemes, concentrating trading activities on short-term market venues, such as DAM.

Misuse of market power to increase profits is a recurring theme in the literature (Joskow, 2008), albeit not the only one. For instance, JP Morgan Ventures Energy Corporation (JPMVEC) were accused of fraudulently exploiting an incentive scheme between 2010 and 2012 installed by the California ISO (CAISO), which resulted in severe penalties and the eventual sale of the business. In more detail, JPMVEC exploited uplift payments "which provide additional compensation to generators when market revenues would not cover what is called the 'bid cost' of a resource the ISO has committed." As JPMVEC could reliably predict future congestion, it leveraged this information and profited from the necessary re-dispatch to alleviate congestion via excessive, bid-dependent uplift payments.[11]

Reform in the Russian Federation   The electricity market reforms in the Russian Federation are outlined in Gore et al. (2012). Unbeknown to most researchers in the field, this reform can be considered to be the worlds "most ambitions" by most metrics, including electricity usage and covered geographic area. Gore et al. are able to show that due to concentration in most market regions during times of peak demand (as measured via the Herfindahl-Hirshman Index), market power might pose a serious threat to efficient market outcomes and hence foster above-competitive prices. Other challenges in this (geographically) vast market concern congestion and lack of generation capacities in the European part of the country.[12]

Reform in Germany   As Ilg (2014, p.13f) notes, electricity market reform in Germany has mainly been driven by European regulation. Directive 96/92/EC meant the beginning of electricity market liberalization and placed special emphasis on introducing more competition through improved cross-border transmission capacities. This process was continued with directive 2003/54/EC whose main focus was on network access and opening of national markets, but also giving consumers greater freedom in choosing their supplier. The most recent directive, 2009/72/EC[13], is again concerned with increasing cross-border transmission capacities and in particular with

---

[11]http://www.ferc.gov/CalendarFiles/20130730080931-IN11-8-000.pdf   accessed on September 10, 2014

[12]These may have been partially addressed in the meantime through a round of mandatory investments by the wholesale generators.

[13]http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:
32009L0072&from=EN accessed on September 11, 2014
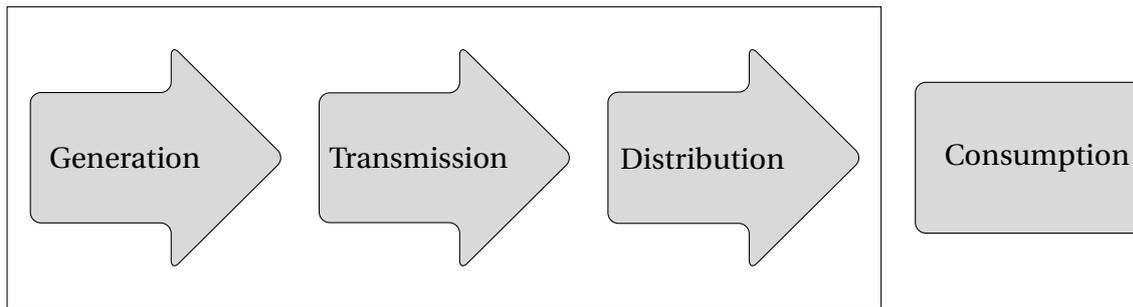
Figure 3.2: Post-liberalization, pre-Smart Grid Power System Value Chain

rolling network expansion planning with a long-term horizon (ten years). In Germany, this directive has been implemented in national law in August 2011 as part of the Energiewirtschaftsgesetz[14]. Currently, energy market reform in Europe has, according to the European Commission's energy strategy, five priorities:[15]

1. limiting energy use in Europe;

2. building a pan-European integrated energy market;

3. empowering consumers and achieving the highest level of safety and security;

4. extending Europe's leadership in the development of energy technology and innovation;

5. strengthening the external dimension of the EU energy market.

3.1.3 The Power System Value Chain

Before liberalization, the power sector was a strictly regulated natural monopoly and exhibited low innovation and high prices. Security of supply was considered crucial and, hence, fairly high. In order to overcome the drawbacks of monopolistic industries, electric power systems in the western world underwent liberalization, sometimes termed restructuring, beginning in the late 1980s. After liberalization, the electricity sector can conceptually be structured into a value chain consisting of generation, transmission, distribution and consumption (Fig. 3.2).

Generation   Historically, relatively few, high-powered generators are located in close proximity to population centers, their generation adapting to fluctuating demand. Demand in power systems varies with high regularity over the time of the day, over

---

[14]EnWG – Energy Industry Act

[15]http://europa.eu/legislation_summaries/energy/european_energy_policy/en0024_en.htm

the week and over the year; it exhibits so-called complex seasonalities.[16] Due to electric energy being a perishable good, fluctuating demand must be compensated via adjustments in generation (cf. Bertsimas et al., 2013). More recently, RES, such as wind, solar and biomass have gained an increasing share of the market, aggravating the control challenge.

Transmission   Transmission connects generation and consumption. We refer to the high-voltage echelon of the power system as transmission. However, the growing importance of RES, which are characterized by reduced power density, and are, thus, by definition decentralized (Smil, 2010), changes this situation.

Electrical power is not routable. In contrast to data packages on the internet, which can be routed along specific edges in the connection graph, electrical current flows over *all* edges connecting supply and demand (Schweppe et al., 1988; Keshav and Rosenberg, 2011). The physical laws of power flow have important implications regarding network congestion, which have been faced historically through locational pricing in the form of zonal or nodal pricing (Stoft, 2002).

Distribution   Distribution System Operators (DSOs) are responsible for supplying electrical energy to consumers via extensive networks at low voltage (400V to 22kV in Western Europe). Following liberalization in Germany, DSOs are required to provide infrastructure access to competing suppliers at a regulated price.

Historically, power distribution networks were built for unidirectional power flows. Due to a lack of deployed sensors and communication capabilities, network operators barely had information on the state of the network (Ipakchi and Albuyeh, 2009). To avoid capacity scarcity on the distribution level, (at least in Germany) excess distribution network capacity, e.g., lines and transformers, was erected.

However, with the advent of both, fluctuating, uncertain generation from RES and flexible, controllable loads, such as EVs, over-provisioning of capacity becomes increasingly costly and economically unsustainable (Schuller et al., 2014; Flath et al., 2013). Accordingly, recent research has focused on (local) economic coordination schemes to leverage demand-side flexibility instead of creating a "copper-plate" to accommodate increasing variation and uncertainty.

Consumption   In the past, electricity consumption has mostly been assumed inelastic, i.e., inflexible and non-controllable in the short run, hence the need to optimize generation. As transaction costs related to frequent metering dominated the benefits of a controllable demand side, prices were in practice mostly flat. Clearly, scarcity in time and space cannot be reflected by means of such non-informative prices.

---

[16] De Livera et al. (2011), for instance, presents impressive examples of complex seasonalities including non-standard calendar effects in the Turkish electricity market regarding both, price and quantity.

Smart grid proponents (Ipakchi and Albuyeh, 2009; Palensky and Dietrich, 2011) argue that smart grid adoption may significantly reduce transaction costs associated with adjusting behavior to scarcity signals, establishing the demand side as a control variable equally important as supply side control.

3.1.4 European Electricity Markets: Products and Structure

After the liberalization of the electricity sector, competition on the generation level commenced within newly designed electricity markets.

Market and product structure Energy trading can be separated into Over-the-counter (OTC) and exchange traded products. OTC offers high flexibility through integration of trading-partner specific clauses. Moreover, OTC trading allows for confidential trading. According to Judith et al. (2011), traded volume in OTC markets is significantly larger than on exchange-based markets, and is currently estimated at approximately 3000 TWh, while consumption amounts to approximately 600 TWh, both on an annual basis. Hence, electricity is bought and sold about five to six times before actual consumption takes place, via OTC contracts, alone.

Exchange trading, on the other hand, fosters price discovery and liquidity, but is restrained by the specific standardized trading products established by the exchange. Effectively, exchange-based trading provides price signals and benchmarking for OTC contracts and is thus relevant beyond the immediate trading that takes place on the platform.

The spot market is used to trade short-term contracts, e.g., contracts with delivery on the same or the next day. Typically, physical settlement is used. EPEXSpot, the result of a merger of German EEX and French PowerNEXT is the largest electricity spot market in Europe and provides standardized products for Germany, France, Austria and Switzerland on the DAM and intraday market. Monthly trading volume on the DAM has grown from roughly 6.5 TWh in 2005 to 22.5 TWh in 2014.[17]

The derivatives market offers the opportunity for hedging up to six years into the future using futures and options. Contract types include weekly, monthly, quarterly and yearly futures typically with financial settlement.[18]

Temporal market structure Traded products in exchange based trading are differentiated with respect to the time remaining until delivery. Schuller (2013, p.33) inte-

---

[17]http://www.ise.fraunhofer.de/de/downloads/pdf-files/data-nivc-/folien-electricity-spot-prices-and-production-data-in-germany-2014-engl.pdf, p.8, accessed September 11, 2014

[18]A detailed overview of European Energy Exchange (EEX)'s Contract Specifications is presented in   http://www.eex.com/blob/78218/d5899ea991b80d022df3b3ca048fb9d7/20140731-eex-contract-spezifications-0040c-e-final-pdf-data.pdf

grates trading venues' structure based on Judith et al. (2011) with the corresponding temporal aspects.

OTC trading typically takes place years and months before delivery. Trading on EEX' power derivatives market ranges from years to weeks before delivery. The day ahead market covers trading up to two weeks in advance of the delivery date, while intraday, takes place on the very same day, up to 45 minutes prior to delivery (Keles, 2013, p.12–16). Ancillary services are called upon with lead times of up to one hour, but usually on shorter notice. Ancillary services may be divided into primary and secondary reserve, as well as minute reserve power (Keles, 2013, p.16), in increasing order of lead time. The former must answer within 30 seconds of being requested, while secondary reserve may have a latency of up to five minutes. Short lead times require generation to already be on-line when it is called upon. This is in contrast to the minute reserve which guarantees delivery within 15 minutes and can hence be provided from non-spinning reserves.[19]

Capacity Markets   The successful proliferation of RES has had three major consequences for operators of conventional generation capacities. First, it led to reduced operating hours. Second, supply gluts during traditional high-price periods (midday peak periods) lowered prices, reducing economic attractiveness. Third, higher price volatility led to increasing cycle counts, which in turn have inflated maintenance and repair costs. All three developments together negatively affected profitability of conventional generation capacities and have led as far as generators threatening to shut down and mothball even newly erected gas-fired power plants. They argue that under current circumstances prices are insufficient to generate profits, or at least cover variable costs. Safe system operation, however, requires sufficient availability of reserve capacities, either to alleviate congestion in the power network, or to offset fluctuations in generation and consumption. The desire of generators to reduce fixed costs via capacity reductions is scrutinized by regulators and often denied.[20] Current market design, unfortunately, does not reward available capacity, relying on payments for energy only. To remedy this shortcoming, the incumbent generators have proposed establishing capacity markets with the goal of providing both, system security, and profits.

Interestingly, Cramton and Ockenfels (2012) indeed argue that capacity markets may improve economic coordination and functioning of power systems (which should not, given the financing of the study[21] by RWE – one of the large four incumbent generators in Germany – come as a surprise), but, at the same time urge great caution before introducing capacity markets.

---

[19]Detailed information on the procurement of ancillary services, for Germany, is provided at `https://www.regelleistung.net/`
[20]See,   for   example,   `https://www.enbw.com/unternehmen/presse/pressemitteilungen/presse-detailseite_51008.html`.
[21]The authors disclosed financing information in the article.

They present five major impediments to the successful introduction of capacity markets. First, political forces, playing a decisive role in market design decisions may foster theoretically flawed approaches, leading to potentially disastrous outcomes. Second, the time horizons underlying a rational adoption of capacity markets is on the order of decades, as they mainly concern long-term investment decisions while the political mindset in the aftermath of the decision to phase-out nuclear generation is more concerned with short-term solutions, such as avoiding blackouts. Third, capacity markets are a blunt knife in the face of political uncertainty, and might turn out to be a costly experiment. Fourth, a capacity market should include locational signals to avoid construction of generation capacities in congested locations. Finally, aligning the goals of a capacity market with existing subsidisation schemes for RES poses challenges that should be addressed "[. . .] before a capacity market is adopted". Cramton and Ockenfels prefer the creation of a "[. . .] stable and reliable political, and a sound market framework" (Cramton and Ockenfels, 2012, p.33) in order to reduce (political) uncertainty instead of patching the existing system with an additional capacity market. Furthermore, they propose, among other measures, improved, i.e., market-based RES integration and transmission expansion. Interestingly, they highlight the importance of addressing the far-reaching fundamental market failure in today's electric power systems: "[T]he absence of a robust demand side". Accordingly, once the mentioned issues are resolved to a satisfactory degree, capacity markets can "unfold [their] complementary value in assuring resource adequacy".

3.2 Smart Grids

Recent and future expected changes to the power grid require upgrades to the existing infrastructure in order for it to cope with the challenge of decentralization, efficiency, transparency, robustness, and sustainability. Upgrades towards the smart grid, according to Gellings (2009) may comprise "sensors, communications, computational ability and control in some form to enhance the overall functionality of the electric power delivery system." Ipakchi and Albuyeh (2009) argue for the transition to a smart grid to be successful, it must be evolutionary. The main drivers, according to Ipakchi and Albuyeh (2009), for smart grid adoption, may be summarized as follows: First, proliferation of intermittent resources on the transmission and distribution level and, second, adoption of electric vehicles. The smart grid is set to improve energy efficiency and to increase engagement of so far passive consumers (Farhangi, 2010). It aims towards establishing decentralized control and develop a robust infrastructure for a sustainable power system via pervasive control and monitoring. At the same time, it should help to reduce over-investment into additional generation and transmission capacity, guaranteeing higher cost-efficiency (Joskow, 2012). The smart grid is based on ubiquitous bidirectional information and communication technology, advanced metering and monitoring systems. Information is thus the backbone of

the smart grid. Automation activates the distribution network and taps into currently dormant flexibility potentials on the demand side. Ramchurn et al. (2012) argue that one main objective of future smart grids is the mitigation of uncertainty arising from fluctuating generation by means of flexible loads.

Coordination efforts in future power system require information technology and automation to play a key role in interconnecting the parts of the system at low transaction costs. Integration of volatile renewable energy sources and reduced operating cost are expected to drive technology adoption. The European Smart Grid Technology Platform (European Commission, 2012, p.27) defines the smart grid to be "an electricity network that can intelligently integrate the actions of all users connected to it – generators, consumers and those that do both – in order to efficiently deliver sustainable, economic and secure electricity supplies."

3.2.1 Smart Meters and Advanced Metering Infrastructure

Smart meters measure, store, and communicate consumer's consumption data in (relatively) high temporal resolution, i.e., in 15 minute or hourly intervals (Siano, 2014). Ideally, the roll-out of smart meters are accompanied by energy management or Demand Response (DR) applications. Advanced Metering Infrastructure (AMI) is considered an essential prerequisite to the roll-out of the smart grid and the next step following historically installed electromechanical meters and one-way automated meter reading (Farhangi, 2010, Fig. 6). It provides the basis for bi-directional communication and hence enables novel business models by means of controlling consumption. In particular, it foots on the roll-out of smart meters and enables all the concepts subsumed DSM. In effect, through analysis of logged data, operations, cost and customer service can be optimized. Furthermore, AMI through bidirectional communication, may provide timely feedback on outages and power quality, supporting further measures regarding grid automation (Siano, 2014).

3.2.2 Information and Communication Technology

Smart grid ICT standards are outlined in Gungor et al. (2011) where GSM, GPRS, 3G, WiMAX, PLC, and ZigBee are compared along data rate, coverage range, applications and known limitations. Depending on requirements, different technologies may be optimal. For instance, wired technologies are more costly to deploy on a large scale, but offer the possibility for improved "communications capacity, reliability and security." Wireless technologies are "constrained [by] bandwidth and security options", but offer reduced installation expenses. Communication requirements are outlined as well and comprise security, reliability and robustness, scalability, quality-managed communication between supplier and customer, to ensure low-latency communication for smart grid control purposes (Gungor et al., 2011).

The diversity of smart grid stakeholders requires exchange of information in an automated fashion over clearly defined and accepted interfaces. Gungor et al. (2011) provide an extensive overview of standardization efforts and existing norms that have come into existence through collaboration of various institutions and professional bodies (IEEE, ISO, NIST, ANSI, and others). Standards presented cover areas such as Home Area Networks (HANs), AMI, EVs, energy management systems and inter-control center communications (Gungor et al., 2011, Table 2). An alternative discussion of smart grid standards is provided in Farhangi (2010).

3.2.3 Smart Markets

The German Bundesnetzagentur, responsible for overseeing competition and regulation in network industries such as natural gas, electricity, telecommunications and mail, has issued a report in 2011 offering a definition and separation of terms concerning the "Smart Grid" on one side and "Smart Markets" on the other (Bundesnetzagentur, 2011). The Bundesnetzagentur defines "Smart Grids" as conventional power networks extended via information and control technology. In more detail (Bundesnetzagentur, 2011, p.11)

> "The conventional electricity network turns into a 'Smart Grid' through its extension with components regarding communication, measurement, control and automation technology. 'Smart' reflects both, capturing network state in real-time, and existence of means for controlling the network to fully utilize existing network capacity."

Hence, the "Smart Grid" is supposed to render power networks controllable and more reactive and therewith improve provision of transmission *capacity*. The power grid remains a natural monopoly and thus requires appropriate regulation. The definition of what belongs to the smart grid accordingly predefines which parts require regulatory oversight. The "Smart Market", in contrast, should, according to Bundesnetzagentur (2011), focus on trading of *energy* and abstract from capacity concerns. Innovation is expected to take place on the side of liberalized markets, where the forces of competition create a creative environment for economic experimentation.

This separation is partially in line with Bichler et al. (2010), who propose smart markets, i.e., algorithmically supported markets that enable novel forms of exchange, unlikely to happen in a traditional environment. One prime example of smart markets is the trading agent competition in the domain of supply chain management . Bichler et al. (2010) focus on combinatorial market design and outline both, perspective research avenues and hurdles to adoption of smart market adoption in reality.

Through smart markets, it may become possible to reach beyond optimization-based markets as proposed in Gallien and Wein (2005) and include aspects of game

theory, behavioral sciences, mechanism and organizational design (Bichler et al., 2010, p.688f). Smart markets are considered a key building block to encapsulate complexity and attain improved economic outcomes. However, transfer of theoretical results into practice often fails as the underlying assumptions required to apply game-theoretic solution concepts do not hold. Clearly, smart markets in the domain of electricity require smart grid infrastructure for monitoring, control and enforcement.

Deadline-differentiated pricing (Bitar and Xu, 2014) may be interpreted as a particular perspective on smart markets. The originally homogeneous good electricity may be separated into different quality classes and hence be priced separately. "The longer a consumer is willing to defer, the larger the reduction in price" (Bitar and Xu, 2014). The authors simplify preference expression to a three-dimensional type (price, quantity, deadline) possibly at the expense of efficiency that may be achieved through more complex type representation. The advantage of this approach lies in its simplicity, robustness and, most importantly, incentive compatibility for the demand side. Complexity is solely found on the supplier's side, who must decide on both, the pricing/quantity regime and execute an operating strategy, i.e., decide when to serve which demand. In this context, Woo et al. (2014) provide an in-depth review of electricity differentiation, including a reliability perspective, which may be employed to reach better operating, as well as system expansion planning decisions. Hence, differentiation may be considered an important building block of a reliable, sustainable and efficient power system with large shares of uncertain RES.

3.3 Sustainability, Uncertainty and Forecasting

RES render a power system more sustainable, but introduce uncertainty regarding future generation. Forecasting generation from RES yields higher accuracy, the larger the portfolio of generators under consideration. In order to cope with volatile generation in constrained power networks, however, either fast-ramping generators, a flexible demand side, or a combination of both is necessary.

3.3.1 Demand Forecasts

Forecasting demand has historically been an active field of research (cf. Taylor et al., 2006, and the references therein), as better forecasts translated into better optimization decisions, and more efficient systems operation. Gould et al. (2008) propose a novel method of forecasting aggregate power demand based on state-space models. Interestingly, their contribution is a novel way of comfortably including complex seasonalities. De Livera et al. (2011) extend the previous approach to also include multiple complex seasonalities (i.e., daily, weekly, and monthly periods) but is more robust to parameter overfitting. Moreover, the new method includes "non-integer" period, and, for the first time, dual-calendar effects. One of the application examples notably

includes Turkish electricity demand data over a period of nine years that exhibits dual-calendar behavior as well as the usual complex patters of electricity demand.

3.3.2 Wind Power Forecasts

Forecasting volatile output of RES is a rather novel field of research intricately linked to meteorology. With growing shares of sustainable energy sources, accurate forecasts are valuable and decisive in achieving satisfactory commitment decisions for conventional generation to cover residual load. Imprecise forecasting may lead to expensive re-scheduling decisions and, if market design allows, negative prices.

Giebel et al. (2011) provide an extensive literature review on short-term wind power prediction that spans more than 350 journal and conference papers, including physical and statistical forecast models, or hybrids thereof. Statistical model yields, by definition, the uncertainty associated with forecast results. Physical models require additional processing to include uncertainty in forecasts (Giebel et al., 2011, p.63). Clearly, longer forecasting horizon is associated with reduced forecasting accuracy (Giebel et al., 2011, Figures 4, 21). An interesting fact the authors mention is that "predictability" may influence wind farm siting decisions. Forecasting models can err in two dimensions, namely, level and phase. The former "misjudges the severity of the storm, while a phase error misplaces the onset and peak of the storm in time."(Giebel et al., 2011, p.8) Evaluation of forecasts' errors, i.e., essentially judging their quality is not trivial, with different measures, such as the popular Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) not reflecting different parts of the error information. Today's state of the art for comparing wind power forecasts includes bias, MAE, RMSE, coefficient of determination R^2, skill score (forecast error relative to a baseline forecasting model), and the histogram of the error distribution (Giebel et al., 2011, p.9). To conclude, including wind turbine shut-off events is of high importance to reduce the risk of outages during storms (Giebel et al., 2011, p.20,21). Pinson and Madsen (2012) employ Markov-chains and autoregressive approaches to model offshore wind power variation that explicitly include regime switches. Moreover, they allow for time-varying model coefficients. The authors highlight that offshore wind farms, in contrast to their on-shore peers, are usually densely packed, which yields highly fluctuating output and renders accurate forecasting a challenge. To account for uncertainty appropriately, they evaluate their model with respect to "point, interval and density forecasts" (Pinson and Madsen, 2012, p.4).

3.3.3 Solar Generation Forecasts

Reikard (2009) compare AutoRegressive Integrated Moving Average (ARIMA) models, transfer functions (which are ARIMA models augmented with causal variables), neural networks and hybrid models (ARIMA and neural network) on time scales from one to four hours on hourly data and from five to thirty minutes on high-frequency

data (minute basis). They find ARIMA models to yield best forecast accuracy. This is quite interesting, as adding further explanatory variables in the majority of cases deteriorates forecast accuracy traced back by the authors to variation in the causal variables themselves. However, with increasing sampling frequency of the underlying data, regression and neural net models may beat ARIMA models, as they are more suited to reflect short term fluctuations. The ARIMA models (Reikard, 2009, p.348), in contrast, owe their superior performance at lower data resolution to precisely reflecting the diurnal solar cycle. Marquez and Coimbra (2013) improve the very-short term (3-15 minutes) solar power generation forecast by explicitly taking cloud cover into account. To this end, they develop sky image processing techniques with minute resolution. Their article introduces the algorithmic details and processing steps to arrive at meaningful forecasts. They find their approach to outperform naïve persistence models, with the largest improvement in forecast accuracy at the 5-minute ahead period.

3.4 Management of Flexible Demand

3.4.1 Demand Side Management

The umbrella term DSM summarizes a variety of approaches to leverage demand flexibility potentials (see Palensky and Dietrich, 2011; Strbac, 2008). Generally speaking, DSM aims to *adapt* current consumption to current generation. In an early work, (Gedra and Varaiya, 1993) refer to DSM as "proposals [...] to reduce system load in an efficient manner". Clearly, in 1993, volatile RES played only a minor role in the power system. Accordingly, the concept of DSM was mostly concerned with reducing peak load and increasing infrastructure utilization, and only to a lesser extent with the efficient integration of RES into power system operations. However, in recent years, DSM is understood to be able to significantly assist in the integration of RES. (Vardakas et al., 2014), for instance, state "DSM [to] include all activities which target to the alteration of the consumer's demand profile in time and/or shape, to make it match the supply, while aiming at the efficient incorporation of renewable energy resources." The notion of "energy services" is central in this context: Schweppe et al. (1989) note, "[that] an end use device *uses* electric energy to provide a *service* to the customer." This *differentiated* view on electricity consumption paves the way for DSM approaches that adapt energy consumption to external signals such as availability of renewable generation, prices, system frequency or even temperature (Albadi and El-Saadany, 2008). A significant subfield of DSM is DR: According to Siano (2014, referring to further studies), "DR refers to 'changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized."'

As of today, DR implementations in practice are mostly limited to direct load control for industrial customers, heating control (night setback) and Time-of-Use (ToU) rate structures (Wang et al., 2010b). The benefits of DR are, according to Siano (2014), bill savings (for both, responding and non-responding customers), reliability benefits, prevention of market power exercise, improved choice for customers, system security through additional degrees of freedom to meet contingencies.[22] Spees and Lave (2007) as well as Siano (2014) argue that such ToU rate structures may render both, responsive and non-responsive consumers better off: The responsive consumers shift consumption to low-price periods, non-responsive consumers enjoy reduced rates during peak times. Going forward, smart homes and electric vehicle charging may emerge as novel flexible load types (Ramchurn et al., 2012; Hu et al., 2012). The emergence of these new flexibility potentials will also foster the notion of quality-of-service in power markets in lieu of more traditional DSM categorizations like "shifting" or "shedding": Retail electricity will no longer be a homogeneous good but will rather be differentiated with respect to reliability, delivery time or power quality (Faruqui et al., 2010).

According to the European Commission[23], incorporating demand side flexibility through measures of DSM into electricity markets is a win-win situation. First, flexibility on the demand side can substitute supply side flexibility. Hence, it can act as a replacement for both, storage and generation capacities. Second, by making the supply chain more efficient, lower energy costs and eventually better prices for consumers can be realized.

The Smart Grid, which, to a large extent, comprises of upgrading existing power system infrastructure on the distribution level with information and communication technology (Farhangi, 2010), will play a key role in enabling the transition towards a decentralized and ecologically more sustainable power systems architecture. It is furthermore expected that smart grid infrastructure will serve as an enabler of DSM, or DR, to achieve the goal of safe, sustainable and efficient power system operations.

To this end, a better understanding of the different types of demand side flexibility has been established in numerous recent publications (Albadi and El-Saadany, 2008; Gottwalt et al., 2011). The motivation behind gaining a better understanding of flexibility potentials is obvious: Through appropriate control mechanisms, stability on short time horizons and economic efficiency of the power system might be achievable at reduced cost compared to keeping costly flexible generation capacity online. Beyond better, i.e., more granular control of flexibility on either side of the system, economic incentives are required for flexibility potentials to be revealed to any decisive instance (Mohsenian-Rad et al., 2010).

---

[22](Flath et al., 2012, Appendix B) and Siano (2014) provide a compact overview of DSM benefits.
[23]http://ec.europa.eu/energy/gas_electricity/doc/com_2013_public_
intervention_swd07_en.pdf, accessed on 2014/09/03.
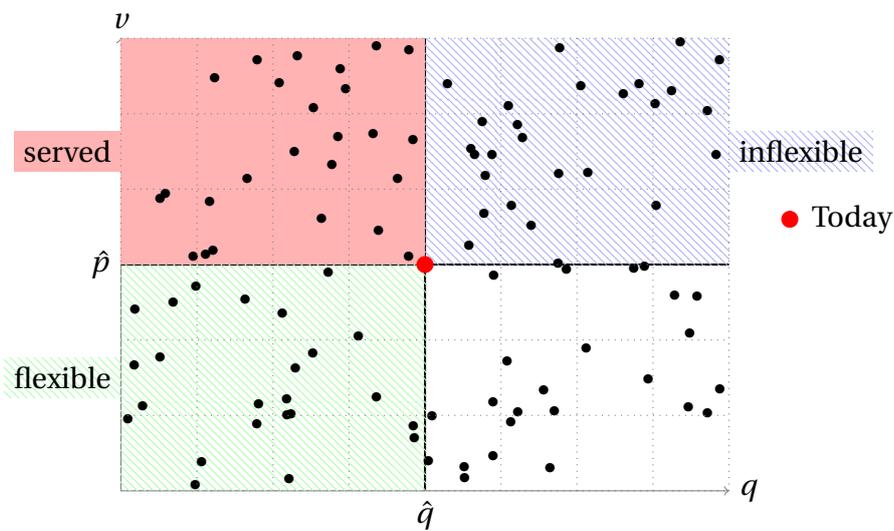
Figure 3.3: Heterogeneity in valuation and quality requirements for different services in the smart grid. Establishing heterogeneous quality classes opens up new applications for the homogeneous good electricity.

3.4.2 Scheduling of Flexible Loads

Quality Differentiation and Automation   Quality requirements for individual energy services are heterogeneous and vary by customer, device, time of day, and day of week. Such demand heterogeneity may be leveraged to align demand and supply in a smart grid. Fig. 3.3 abstractly illustrates the diapason of quality (q) and valuation (v) for an energy service. Current power systems, however, do not allow for differentiation by quality and valuation, but provide immediate service at a specific price per unit of energy $\hat{p}$, and a certain quality $\hat{q}$ (illustrated by the red dot in the center). "Quality" of energy services may comprise curtailing, deferral, interruption, and reliability in general. To avoid excessive transaction costs, a corresponding smart grid infrastructure is necessary to offer consumers a choice, which quality/reliability to subscribe to. Ideally, this differentiation should be possible on a very granular, i.e., on a per-device level. For behavioral changes to materialize, the associated benefits from more system-compliant behavior must be attributed, at least to some extent, to the stakeholder changing his behavior.[24]

Load Scheduling   Leveraging the potentials of flexible loads requires active dispatch of these assets. To this end, scheduling approaches for flexible, i.e., deferrable, loads have attracted significant research activity. Parvania and Fotuhi-Firuzabad (2010) schedule both flexible loads as well as decentral generation assets to minimize wholesale electricity costs. Using a mixed-integer programming framework, Sou et al. (2011)

---

[24]This is in contrast to DR, which historically let utilities collect a large fraction of benefits.

determine cost-minimizing power profiles which satisfy complex constraints such as non-interruptible and sequential operations, while Bosman et al. (2012) propose coordinated scheduling of combined heat and power plant (CHP) fleets. Relying on different scheduling routines, Subramanian et al. (2013) show that efficient demand side coordination can already be achieved with modest load flexibility endowments. Scott et al. (2013) present a comprehensive framework for scheduling different flexible loads in the presence of multiple sources of uncertainty. Recognizing the central connection of generation uncertainty and load deference, Papavasiliou and Oren (2014) develop solution techniques for large scale stochastic unit commitment problems in the presence of flexible loads. Varaiya et al. (2011) present "risk-limiting dispatch", a probabilistic approach to scheduling conventional generation in the presence of stochastic generation to eventually meet operating constraints regarding both, generation and transmission. While the authors interpret DR to increase stochasticity of the control problem, the demand side may as well be commissioned to alleviate uncertainty through appropriately engineered scheduling algorithms.

Petersen et al. (2013) present two heuristic algorithms, "predictive" and "agile" that control a virtual power plant of heterogeneously flexible load. The objective is to minimize residual load, defined as the absolute difference between generation from RES and scheduled demand. They find their heuristic (causal) agile control algorithm to yield suboptimal results, but to beat its predictive counterpart under short forecast horizons. The latter relies on integer programming, is therefore computationally more involved, and less applicable to real world large-scale flexible load dispatch problems. The agile algorithm, in contrast, leverages sorting to greedily dispatch the least-flexible devices and retain flexibility of the aggregate portfolio. The intuition to retain as much flexibility as possible within the unscheduled portfolio is reminiscent of the approach of Kumar (1987).

3.4.3 Pricing Regimes

Price-based incentives are intended to activate the demand side and hence assist in balancing supply and demand (Borenstein et al., 2002). The particular design decisions may follow a multitude of considerations, including complexity, risk (for consumers and utilities), and revenue considerations.

Constant rates, a constant price per unit of energy (Wh), is the most simple and common pricing regime for electric power is consumed. It is popular due to its low complexity, both for the consumer as well as with regard to metering complexity. Boiteux (1960) and Schweppe et al. (1988) present a detailed treatise on the pitfalls of constant energy prices (and argue in favor of real-time pricing). Their most significant downside is that scarcity is not reflected in prices and hence consumers have no incentive to adjust behavior to system state. Often, constant rate tariffs require the consumer to pay capacity-dependent connection charges.

Temporal Pricing   ToU differentiated prices vary over the day with at least two different prize zones, i.e., high and low price. Prices and zone lengths are published in advance and are static in nature, that is, they do not take dynamic system state into account. The goal of ToU is to defer load to low-load times, e.g., nighttimes, reduce overall peak load and increase system utilization factor. This regime adds consumer incentives, but clearly fails to dynamically reflect system state. Hence, while it has been successfully applied in the past (Borenstein et al., 2002, p.5), its successful application in power systems with strong penetration by RES is doubtful. Reference? Critical Peak Pricing (CPP) programs combine attributes from Real-Time Pricing (RTP) and interruptible programs. They feature time-varying prices as in ToU programs, augmented with an additional rate that can be called upon on short notice, but only a limited number of times, by the utility (Borenstein et al., 2002). Accordingly, CPP dominate ToU programs through better tracking of wholesale power prices, and hence lead to more efficient outcomes.

(Schweppe et al., 1980, 1988) comprehensively introduced the concept of RTP long before the advent of inexpensive communication means such as the internet. The authors argued that the power system's operating efficiency could be improved, capital investments reduced, and customers would be given a choice with respect to the reliability (we prefer to say quality) of purchased electricity. The utility benefits from reduced costs and improved capacity utilization as well, as some peak-load consumption is curtailed. Allcott (2011), however, demonstrate that the benefit of RTP to the consumer is rather small in their examined setting.

Locational Pricing   By adding a locational component prices can be further differentiated and include localized scarcity signals. This again may be important for both, investment and operation's decisions. (Bohn et al., 1984) argues in favor of locational pricing to foster efficient allocations, both, in the short and long run. "With spot pricing (a different price for each customer location at each moment), the utility can induce socially optimal behavior by each customer and avoid system overload without having to resort to collective or individual rationing schemes." [25]

3.4.4 Incentives and Mechanism Design in Smart Grids

Due to the distributed nature of smart grid systems (see Tanenbaum and Van Steen, 2002), the optimal dispatch as determined by a central scheduler cannot be directly implemented in practice. The information required to determine the schedule (i.e. willingness-to-pay and flexibility endowments, Fig. 3.3) are private information of individual system participants. Therefore, these inputs are not directly available to a central planner but rather need to be elicited from the distributed agents. Price-based

---

[25] We refer the inclined reader to Ilg (2014, Chapter 3.3) for an overview of the literature on electricity pricing.

DSM programs facilitate a decentralized decision paradigm by directly incentivizing customer behavior changes (Albadi and El-Saadany, 2008). However, purely price-based coordination (without feedback) typically cannot internalize hard operational constraints (Wang et al., 2010a). A typical phenomenon is the occurrence of new load spikes due to herding effects (Ramchurn et al., 2012). Using simultaneous bidding over a complete planning horizon, Mohsenian-Rad et al. (2010) show that competing distributed agents can converge to an optimal schedule in a distributed fashion.

However, every task needs to be scheduled, i.e. there is no consideration of quality-of-service differentiation. The combination of job admittance decisions, online decision-making and incentive-compatible pricing and allocation rules leads to the literature on online mechanism design: Friedman and Parkes (2003) outline the design challenges of such mechanisms and establishes the central assumptions like limited misreports. Ideally, algorithmic mechanisms must be designed such as to achieve monotonicity in their solutions (Parkes, 2007) and hence enable incentive compatibility.[26] Careful incentive design deserves special attention as the demand side may be actively engaged with system control. Here, poorly designed, widely deployed algorithms might lead to catastrophic system failures.

Vytelingum et al. (2010) propose a trading mechanism (and a set of trading strategies) for electricity in the presence of self-interest agents based on the Continuous Double Auction (CDA) that "degrades well" under increasingly constrained power network capacity.[27] Gerding et al. (2011) apply online mechanism design in the smart grid, addressing the coordination of plug-in hybrid vehicle charging decisions. They develop a model-free online mechanism for perishable goods in a discrete-time setting with multi-unit demand and decreasing marginal valuation. Focusing on non-decreasing valuations, Stein et al. (2012) introduce the concept of pre-commitment in combination with model-based online scheduling techniques to create an online mechanism for coordination of charging requests of pure electric vehicles. Gerding et al. (2013) propose a two-sided market for advance reservation EV charging relying on cost and availability information from the sellers and preferences regarding charging location and time by buyers. Again, to account for the dynamism in the model, the authors rely on principles of online mechanism design. The designed mechanisms allow trading-off price volatility and budget balance at high efficiency, while the benchmark mechanism suffers from low efficiency and a significant deficit.

---

[26] A wonderful exposition on market and incentive design is provided in Kalagnanam and Parkes (2004, Chapter 2).

[27] While the mechanism itself is not incentive-compatible, the underlying CDA has been shown to attain high efficiency.

Chapter 4

Model-free Online Mechanism Design for Scheduling Preemptive Jobs

4.1 Introduction

Today's energy system is experiencing transition from mainly fossil-fueled power generation towards greater shares of renewable energy sources (RES). While these new energy sources are clean and have quasi-zero marginal cost, they are highly fluctuating and feature a lower energy density than fossil fuels. This challenges conventional wisdom in power system planning: The future system topology will be much more decentralized with significant generation capacities being located on the distribution grid level. Furthermore, system operators need to be able to cope with much higher variability levels on the supply side. Recent experiences in power systems around the world suggest that traditional procedures will have to be adapted in light of these new challenges. For example, Subramanian et al. (2013) note, that "the current operating paradigm [...] works at modest penetration levels, but fails when 30% or more of total energy generation comes from renewables." Similar observations were made in Germany where the share of power generation from renewables increased between 2009 and 2013 from 15.9% to 23.4%. However, over the same period the number of days with exceptional RES shedding events in Germany's northeastern transmission grid increased from 4 to 142 — what used to be a rare event had become the norm.[1],[2]

Acknowledging the central role of the underlying stochastics of RES generation, Varaiya et al. (2011) claim that power grid operations will need to fundamentally change in a similar manner as the manufacturing industry did when it adopted IT-based just-in-time supply chain management. The research on demand side management (DSM) so far is generally based on variable, yet pre-announced, electricity prices. This resonates well with traditional forward power markets for conventional generators, but seems ill-suited to describe a RES-dominated power system which is characterized by quasi-zero marginal cost supply and limited reliability. A possible remedy is to introduce quality differentiation with respect to reliability and adopt

---

[1] www.50hertz.com/en/file/50Hertz-Almanac-2012-EN.pdf
[2] www.bdew.de

value-based pricing. Then, payments and realization of energy delivery are not specified a priori, but rather arise endogenously from heterogeneous customer valuations and flexibility endowments. Designing such a quality-differentiated market requires the explicit consideration of incentive-compatibility with respect to load flexibility.

We consider a geographically highly restricted smart grid market for direct and instant matching of flexible loads with volatile generation. We envision this market platform as a parallel, secondary marketplace for excess RES generation which can no longer be integrated into the primary wholesale market at reasonable costs. Consequently, demand side applications can choose which market to engage in, and highly flexible, non-mission-critical demand may be attracted by lower costs on the local flexibility market. Similar to Bitar and Low (2012), we apply a generic model to explore the general implications of such a secondary marketplace with differentiated service quality. Unlike most prior research, we do not make a distinction between shiftable and sheddable loads but consider demand flexibility on a job level. This means that individual jobs have freedom with respect to their execution time (shifting characteristic) and furthermore may not be executed at all (shedding characteristic). We complement this micro-founded demand model by an online mechanism which yields truthful revelation of jobs' preferences with respect to *both* value and temporal flexibility. Furthermore, the proposed mechanism is weakly budget-balanced under the mild assumption of homogeneous reservation prices on the supply side. Our analysis provides a quantitative assessment of the potentials to match renewable generation to a flexible demand side in a decentral smart grid. We find that, for a plausible range of the relevant parameters (generation capacity, demand side flexibility endowment), the cost of ensuring IC in a smart grid market is relatively small. This suggests that local matching of demand and supply can be organized in a decentral manner in the presence of a sufficiently flexible demand side.

This chapter addresses decentral coordination of flexible loads to balance increasing levels of generation uncertainty in the power system. Recently, demand side management and flexible loads have attracted increased research activity both from a technical as well as an economic perspective. Besides these directly related research strands, this chapter relates to the theory on (online) mechanism design and formal modeling of operational flexibility. Furthermore, we also connect to the literature on operational flexibility. Graphs play a central role in formally expressing flexibility. For instance, (Jordan and Graves, 1995) already demonstrated in their seminal paper that limited flexibility endowments can typically accrue a large share of maximal benefits achievable under full flexibility which may be represented through complete bipartite graphs. Borenstein (2000) presents precedence constraints and corresponding process flexibility in the manufacturing domain via directed acyclic graphs, enabling efficient representation of routing flexibility at various levels of detail. Kranton and Minehart (2001) focus on issues in strategic network formation. They use bipartite graphs to model trade relationships. The more links an agent establishes, the more flexible he becomes in his sourcing decisions. Chou et al. (2010) highlight the impor-

tance of separating between range and response in process flexibility, modeled via bipartite graphs. They conclude that in most cases improving response, i.e., the rate at which a system is capable of reacting to change, is more valuable than improving range, "the extent to which a system can adapt". More recently, Chou et al. (2011) compare "highly connected but sparse graphs"with complete graphs regarding their worst case performance over a large set of objective functions under uncertain and fluctuating demand. They propose an heuristic for the design of flexible process structures without information on uncertainty. This approach yields comparable efficiency relative to customized structures that rely on information about demand distributions. To summarize, graph theoretic approaches are common in modeling flexibility in economic systems. We follow this tradition by leveraging this technique to model load flexibility in smart grids.

An earlier version of this chapter was circulated as Ströhle et al. (2012).

4.2 Scenario and Model Description

We consider a local market mechanism for continuous matching of discrete supply and demand jobs close to real-time. With "local" we refer to a setting where participants on both market sides are situated within spatial proximity of each other. Continuous matching, in contrast to fixed clearing intervals, fosters liquidity and efficiency on geographically bounded markets. The argument for shortening prediction horizons combined with frequent re-adaption of generation and consumption through intraday markets is reiterated in (Bitar et al., 2012). The presented market is used exclusively ex-ante to sell uncertain, unexpected and thus excess quantities of locally generated renewable energy to customers exhibiting different kinds of flexible demand. In general, electricity markets can be categorized according to their contracting time frame, geographical area, and market organization. Using the taxonomy presented by (Ramos Gutierrez et al., 2013; Botterud et al., 2009), the market (mechanism) is situated on the intraday market without leadtimes in a restricted geographical area. The way the market decides on allocation is closely related to the generalized assignment problem, which (in contrast to the ordinary assigment problem) allows for unused supply and unserved demand to be discarded with the problem remaining feasible. The major contribution in this work lies in solving this problem online and in the presence of strategic agents/jobs on the demand side.

4.2.1 Problem Description

In power markets with large shares of renewable generation capacity, the market design might allow for trading uncertain quantities of renewable generation on the regular market at a discounted price (due to its inherent uncertainty). Alternatively, only a secure equivalent of the uncertain generation can be traded on the regular mar-

ket, with excess quantities shedded or traded on specialized, complementary market venues. Our market mechanism serves as such a complementary market for energy available on short notice, where ad-hoc availability reflects local RES intermittency. Therefore, at some times excess supply will be available that, under traditional market rules (Botterud et al., 2009), would be shedded (thereby adversely effecting social welfare). Under the proposed market mechanism, however, excess renewable generation would be assigned to queued, flexible demand on short notice, enhancing social welfare. This contrasts with related literature, where load is mostly considered an exogenously given constraint and sophisticated RES forecasting is used to reduce uncertainty in unit commitment problems. As local RES forecasting is highly uncertain, we focus on providing incentives to the demand side that, combined with smart grid technologies, might provide large amounts of rather inexpensive flexibility for local supply-demand matching.

4.2.2 Demand Model

Demand is modeled as in the canonical demand model used in online mechanism design literature (Parkes, 2007). Each job has a type θ which consists of a tuple including its arrival time a, departure time d, temporal flexibility f, and the job's value v.

$$\theta = < a, d, f, v > \tag{4.1}$$

In our model, each job requires one unit of energy to be successfully completed. Jobs arrive at time a and remain available until their latest possible start date $a + f$, which coincides with their departure date d. If job j is assigned a unit of supply, j is considered served and both, the corresponding supply order and j are discarded from the order book. While this model is highly generic and cannot capture the peculiarities of quantities demanded, it lends itself to tractable analysis. Moreover, the generality of the model allows diverse demand such as energy storage, data centers, HVAC, agricultural pumping and so on (Bitar and Low, 2012) to be be described within this model.

The formulation of demand in discrete jobs is motivated by the idea of energy services (Schweppe et al., 1989): A consumer usually has limited information about the quantity of electrical energy consumed by a certain job, but she might be able to quantify utility she receives from a particular service that consumes electrical energy as its input. Additionally, the user may not be concerned when exactly a service is performed (in an unusable intermediate state) but when the service' end product becomes available. For simplicity it is assumed that the user's utility from a service is independent of when the service is provisioned, provided it is before the end of the job's deadline. Formally, if the job is allocated and thus matched between arrival and departure ($a_i \leq m_i \leq d_i$), resulting utility is the difference between valuation and pay-

ment, and zero otherwise.

$$u_i(m_i) = \begin{cases} v_i - p_i & \text{if} \quad a_i \leq m_i \leq d_i \\ 0 & \text{else} \end{cases} \tag{4.2}$$

4.2.3 Supply Model

In our supply model, offers are modeled as discrete energy "packets" of unit size matching the unitary energy demands of the jobs on the demand side. Supply jobs (offers) remain available in the supply queue only for a very limited timespan before they are matched or discarded (shedded). We denote the period during which the supply jobs are available for matching with α, and set it to a constant value for all supply jobs.

We instantiate our generic supply model with real-world generation data from a wind farm in Sotavento, Spain (cf. López et al., 2002). For the purpose of the discrete job model, we derive the arrival of supply jobs from hourly mean generation data $g(t)$. Briefly, wind generation data of the year 2012 are normalized by installed capacity C of the wind park, which we assume to equal the maximum mean hourly generation observed throughout the year, i.e. $C = \max_t(g(t))$. The resulting utilization levels of the wind park $g'(t) = g(t)/C$ are cumulated over time to form the continuous, monotonously increasing function G.

$$G(t) = \sum_{i \in \{t_0, \ldots, \lfloor t \rfloor\}} g'(i) + \frac{t - \lfloor t \rfloor}{\lceil t \rceil - \lfloor t \rfloor} \cdot g'(\lfloor t \rfloor) \tag{4.3}$$

G represents capacity-normalized generated energy in the period $[t_0, t]$. Different scaling factors regarding the supply side (i.e., installed capacities) are captured by the factor β in eq. (4.4). The arrival time of supply job k is then defined by a_k, the scaled inverse function of G.

$$a_k = \frac{1}{\beta} G^{-1}(k) \tag{4.4}$$

Parameter β is set such that total quantities demanded and supplied balance each other in the long run[3]. Based on empirical (hourly) measurements of wind speed and derived turbine output, we construct the arrival times of supply jobs. Intuitively, interarrival time of supply jobs is decreasing in wind speed. The generic supply model and the derivation of jobs' arrival times easily generalizes to fluctuating power sources other than wind.

---

[3]Average utilization of the Sotavento wind farm in 2012 was approximately 0.2, thus $\beta = \frac{1}{0.2} = 5$.

4.2.4 Load Flexibility Matching

Applied to the power systems domain this implies that not all demand must be served, and thus excess demand can be discarded, providing a lever to efficiently balance supply and demand. Clearly, this presents a major departure from conventional procedures in power systems operations. However, we argue that a radical change of the system structure warrants a radical change of operational paradigms.

Structures that formalize flexibility in assignment and scheduling problems are frequently formulated as bipartite graphs (sec. 2.5). As we are interested to what extent temporal flexibility improves achievable social welfare, temporal overlap between supply and demand jobs is encoded in bipartite graphs. The presence of edges indicates temporal overlap, i.e., possible assignments. Demand jobs are characterized by valuation, supply jobs by their reservation price r. The goal of maximizing social welfare (difference between valuations and reservation prices) is equal to finding the maximal weighted matching, where weights represent marginal contribution to social welfare from the corresponding assignment. The task is complicated by information on the graph becoming availabe only over time (online), and possibly being strategically misreported (valuation, edges). Starting from an offline formulation using full information, we address both challenges in the next section.
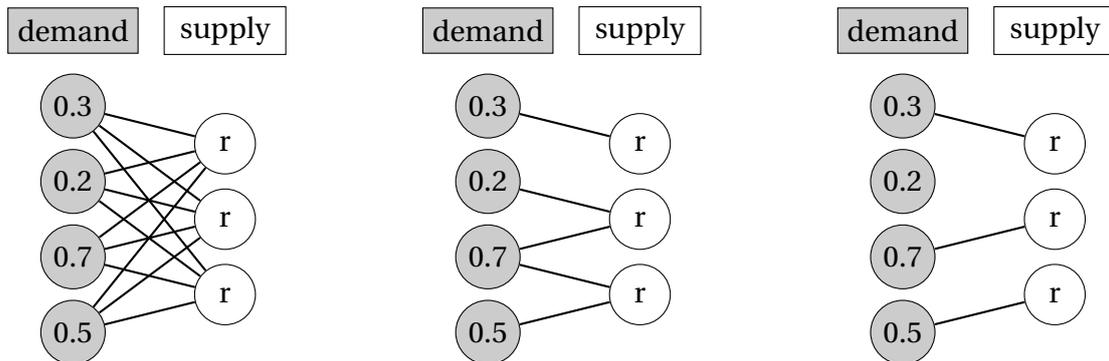
Figure 4.1: The generalized assignment problem as a bipartite graph. Edges between supply and demand jobs indicate temporal overlap. Demand jobs feature valuation v, while supply jobs are characterized by a (homogeneous) reservation price r.

4.3 Allocation Procedure – Planners and Mechanism

The task of any planner lies in identifying welfare-maximizing assignments. While off-the-shelf solutions are available for the offline case of full information, partial information on the graph in an online setting and strategic misreporting by self-interested agents (mechanism design) add interesting facets to the problem.

In power systems, continuously retaining the balance of supply and demand and providing near-constant power quality (with respect to frequency and voltage) constrains the decision space available to any coordinating instance. As the contribution of RES to total generation grows larger, keeping the necessary conventional spinning reserve capacities online to react to its unexpected fluctuations becomes an increasingly costly effort (Bitar et al., 2012). Therefore, we develop an approach that focuses on the demand side, mainly by introducing incentives to operationalize jobs' flexibility, which in turn reduces the cost of power system operations through improved integration of RES.

A clairvoyant (offline) planner serves as an offline benchmark that, given true information about consumers' preferences and suppliers' costs, computes welfare-maximizing allocations. For the online case we present a greedy model-free planner. However, such allocation designs suffer from incentive problems (Sioshansi et al., 2008) which we address through online mechanism design (Parkes, 2007). All three approaches ensure continuously balanced supply and demand, achieved by two means: First, unmatched supply jobs are discarded from the system at the end of their availability period d. Second, we assume that consumption rates of matched jobs are similar in a sense that individual disparities will not cause imbalances in the system during physical settlement.

4.3.1 Benchmark Offline Planner

Solving the NP-hard generalized assignment problem (Roth and Sotomayor, 1992, ch. 8) yields the maximally achievable social welfare under perfect knowledge if types θ are not misreported.[4] It thus constitutes an intuitive benchmark for alternative allocation procedures that might operate online or take incentives into account. Technically, our formulation deviates from standard formulation of the GAP as not every job on either side is compatible with all jobs on the opposite side, i.e., some jobs do not overlap temporally.

$$\max_x \sum_{j \in J} \sum_{i \in I} (v_j - v_i) \cdot x_{ij} \tag{4.5}$$

The optimization objective in Equation (4.5) is to maximize social welfare by selecting feasible matches between supply offers $i \in I$ and demand requests $j \in J$. The decision variables x_{ij} indicate whether supply job i and demand job j are matched. The optimization decisions constrained as follows: Each request and offer can only be

---

[4]We use Gurobi 5.6, a standard industry solver for optimization problems.

matched once at most (Eq. 4.6 and 4.7).

$$\sum_{j \in J} x_{ij} \le 1 \qquad\qquad \forall i \in I \qquad\qquad (4.6)$$

$$\sum_{i \in I} x_{ij} \le 1 \qquad\qquad \forall j \in J \qquad\qquad (4.7)$$

Constraint (4.8) ensures that allocation takes place only between those requests and offers that have some overlap in time. Correspondingly, if there is no such overlap, matching becomes impossible.

$$x_{ij} \in \{0, 1\} \qquad \forall (i, j) \in \{(i, j) \in I \times J | a_i \le a_j \le d_i \lor a_j \le d_i \le d_j\} \qquad (4.8)$$

Note that in contrast to the (ordinary) assignment problem, the number of jobs on one side of the market may differ from the corresponding number on the other market side, i.e., $|I| \ne |J|$ is possible in a *generalized* assignment problem.

4.3.2 Online Planner

Leaving the benchmark case behind, we are primarily interested in how well the generalized assignment problem can be solved in an online setting, i.e., under uncertainty regarding future arrivals of supply and demand jobs. For ease of exposition, we restrict ourselves to the model-free case, i.e., we assume no information on future arrivals of either kind of job. However, we argue that this is a reasonable assumption for the case of small, local markets, where forecasting is especially unreliable and might yield only little benefit.

The proposed online welfare maximization heuristic has the same goal as the offline planner, i.e., maximization of social welfare. However, in contrast to the offline planner, the *online planner* has no information on future release and due dates as well as valuations. Therefore, in order to utilize as much information as possible, supply and demand jobs are matched at the end of their respective active period, i.e., at the point in time when maximum information is available while making decisions is still possible (see Fig. 4.2). In our case of unit quantities, this reduces to checking whether a job is available on the other side of the order book on the considered job's due date.

When job j reaches its deadline $d_j = t$, virtual market clearing of all currently available jobs (demand and supply) $\Psi = \{j | j \in I \cup J \land j_a \le t \le j_d \land j_m = \emptyset\}$ is performed. To this end, first the corresponding matchings of a greedy (with respect to valuation) double auction is performed. If j is allocated in this *virtual* market clearing, i.e., $j_m = t$, it is effectively allocated and removed from Ψ.

Deciding as late as possible does not pose an optimal decision policy, but it reduces decision uncertainty in our model-free setting via postponement. A potential pitfall can be found in ignorance of situations, where an expiring supply job is allocated to
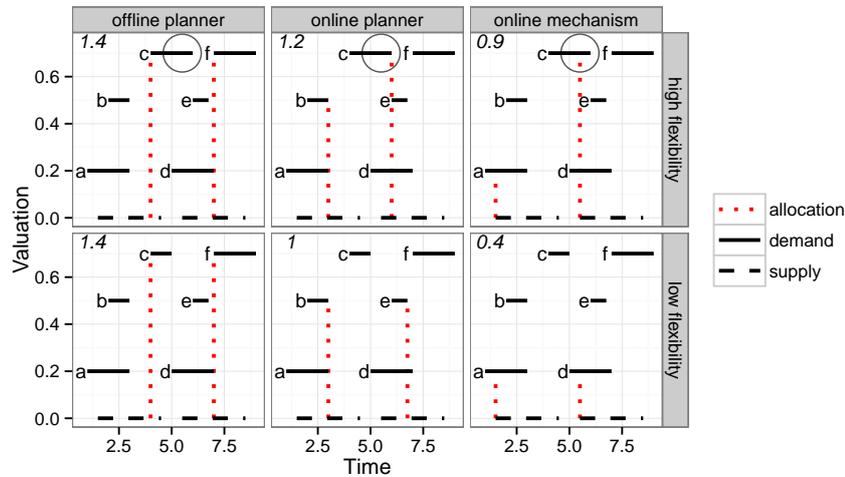
Figure 4.2: Comparing allocation decisions by the offline, online planner as well as the online mechanism. Increasing flexibility creates additional competition ('high' treatment - job *c* is more flexible) and lets the online planner and mechanism make better decisions. Resulting social welfare in top left corner of each subgraph.

the highest active demand job with a large remaining active period at the allocation date. A competing demand job with only slightly smaller valuation, but very little remaining active period is not allocated and remains eventually unallocated, likely reducing economic efficiency. A model of the future could improve decisions by the naïve, model-free online heuristic. Allocation decisions could be based on greedily computed opportunity cost of each allocation decision. If marginal social welfare from allocation were to exceed marginal cost, the respective job would be allocated. [5]

4.3.3 Online Mechanism

Reporting one's true type information under both planning approaches is not a dominant strategy: In some instances allocation might be facilitated by inflating flexibility, while in other instances the job is (in an online setting possibly repeatedly) pushed to later points in time for allocation, and eventually discarded as higher-valued demand jobs arrive. Thus, in order to prevent type manipulation by the jobs and induce truthfulness, allocation decisions by the planners must be both, modified to induce monotonicity in allocation decisions (Parkes, 2007) and supplemented with payments. Under a corresponding mechanism, single-unit demand consumers are incentivized to

---

[5] The computation of opportunity costs is straightforward, if future arrivals are known (as in the offline case). Under uncertainty, the future could be sampled from a model based on historical information. Then, the opportunity cost of a decision is the weighted average of the scenario-specific opportunity cost.

truthfully reveal their job valuation v and flexibility f on arrival a. The mechanism has the property of dominant-strategy incentive compatibility (DSIC). Suppliers, on the other hand, post generation offers at which they are willing to supply as soon as supply becomes available.[6] The mechanism continuously matches jobs on either side of the market based on participants' reported information, i.e., it computes allocations and payments. To qualify as an economic mechanism it should satisfy individual rationality (IR), incentive-compatibility (IC) and budget-balance (BB). However, pursuing these design goals does not permit maximization of economic efficiency any longer (cf. Myerson and Satterthwaite, 1983).

Allocation and pricing rule

To achieve incentive-compatibility in single-valued domains, both the allocation and pricing rule must take monotonous decisions (cf. Parkes, 2007). To this end, our online mechanism employs a greedy allocation rule, i.e., any newly incoming demand job is matched *on-arrival* to the lowest active offer. On the other hand, a newly arriving supply job is matched to the highest-valued active demand job, irrespective of their due dates and characteristics of competing orders.

If matching is not possible immediately, jobs are enqueued and either assigned at some later point in their active period or, in case of too fierce a competition on the respective side of the market, eventually discarded without assignment. Allocation and pricing decisions take place at two distinct points in time, i.e., payments are determined only at the end of the active period of the respective demand job (similar to Friedman and Parkes, 2003), while allocation decisions are made until the latest start date of the job. To avoid issues with jobs misreporting their departure date, access to the good (in case of successful allocation) is granted only at the reported due date of the job.[7]

Formally, the assignment decisions on either side of the order book are given by

$$i(j_i) = \begin{cases} \arg\min_{i \in I(j_i)}\{v_i\} & \text{if } v_i < v_j \\ \emptyset & else \end{cases} \tag{4.9}$$

$$j(i_j) = \begin{cases} \arg\max_{j \in J(i_j)}\{v_b\} & \text{if } v_i < v_j \\ \emptyset & else \end{cases} \tag{4.10}$$

where $I(j_i)$ is the set of supply jobs active on the *arrival* of demand job j_i and $J(i_j)$ is the set of demand requests active on *arrival* of supply request i_j. In case of equal valuations v, tie breaking is employed such that jobs with later deadlines, earlier ar-

---

[6]Supply is assumed non-strategic as it can easily be inspected ex-post and thus is not private information.

[7]Friedman and Parkes (2003) show that this prevents agents from inflating their flexibility endowments.

riving jobs and higher valued jobs receive preferential treatment. Under this greedy allocation rule, we can ensure incentive-compatibility ex-post by use of critical value payments (Nisan, 2007). [8]

Mechanism properties

The online mechanism achieves three out of the four desired economic properties: Individual rationality, incentive compatibility (regarding the demand side), and budget balance (under an additional weak assumption). We rely on the standard assumption of online mechanism design that early release and late due dates cannot be reported, (see Friedman and Parkes, 2003).

Proposition 1. *Participation in the online mechanism is individually rational for all jobs.*

Proof. Demand side jobs are either allocated or not: Due to the pricing rule, the job payment is always bounded by its reported valuation and thus a non-negative payoff in ensured. In case of non-allocation, the payment is zero and again, job payoff is non-negative.

Supply jobs (offers) are remunerated at a price greater or equal to marginal cost. □

Proposition 2. *The online mechanism is incentive-compatible with respect to job time reporting.*

Proof. Assume the true type of a job is characterized by a release date a and a due date d. We then consider two possible types of job time misreports — (i) reporting a later release $\hat{a} > a$ and (ii) reporting an earlier due date $\hat{d} < d$.

(i) If a job is not allocated under the true release report, he will neither be allocated under the misreported later release. When the job declares a release date later than the actual release, there can be three outcomes:

a) If the allocation under the true type takes place at $t > \hat{a}$, the job is allocated under the misreported release date $\hat{a}$ at the same price, or

b) if the original allocation took place in the interval $[a, \hat{a})$,

- the job may no longer be allocated (no matching arises over the new active period), or

- the job may be allocated at a price greater or equal the price under truth-telling.

Clearly, in none of these cases can the job improve its payoff.

---

[8]We provide a complete characterization of the payment rule in the appendix.

(ii) The same reasoning applies for the reported due date: Declaring an earlier due date cannot result in reduced payments, but may induce non-allocation (original allocation takes places in the interval $(\hat{d}, d]$) or payments greater or equal the payments under truth-telling (original allocation takes places in the interval $[a, \hat{d}]$).

Therefore, a job cannot profit from misreporting arrival and departure times which proves the proposition. □

Proposition 3. *The online mechanism is incentive-compatible with respect to valuation reports.*

Proof. As noted above, payments are chosen such that they correspond to the jobs' critical value. The combination of greedy allocation with critical values in our setting with single-item interesting sets ensures IC (Parkes, 2007). □

Proposition 4. *The online mechanism is budget-balanced under the restriction of homogeneous reservation prices.*

Proof. For the case of homogeneous reservation prices (all supply jobs are offered at the same price) the payments received from the demand side are exactly balanced by the payments made to the supply side. Thus, the mechanism does not require outside subsidies and is budget balanced in the strong sense, that is it neither runs a deficit nor generates a profit. □

Note that the mechanism's property of budget-balance does not hold under heterogeneous reservation prices: Consider a demand job i which is matched to a supply offer with reservation price p at time $t < d_i$. At a later point in time t', $t < t' < d_i$, a supply offer becomes available with $p' < p$. To retain the monotonicity property of the payment rule (and thus incentive compatibility with respect to demand job timing) the job's payment obtains as p' which is insufficient to cover the reservation price p. This problem arises from the online nature of our problem combined with heterogeneous reservation prices. However, we primarily focus on a setting where supply is provided by renewable generators for which marginal costs of supply close to zero seem warranted.[9] In summary, the proposed online mechanism satisfies the IR, IC, and BB requirement under mild restrictions. To assess the applicability of the proposed mechanism, we evaluate the achievable efficiency through numerical simulation studies.

---

[9]For completeness, we investigate the impact of (homogeneous, constant) non-zero reservation prices on welfare later on.

4.4 Evaluation

Evaluating the mechanism, we isolate the role of flexibility in the context of both, economic efficiency and market power, from individual jobs' incentives. To this end, results in sections 4.4.1 and 4.4.2 are based on homogeneous demand flexibility. The value of flexibility is examined in a QoS context in section 4.4.3, naturally relying on heterogeneous flexibility. In the following, we introduce the main parameters used throughout the economic evaluation of the proposed mechanism. The length of the period during which jobs on both sides of the market are available is central to the attainable quality of any assignment solution. For supply jobs we set this availability to the constant value of five minutes.[10] For demand jobs, flexibility is one of the central evaluation criteria and therefore varies with average values between one and thirty minutes.

In order to account for different market sizes (and respective liquidity), settings with 20 (small) and 120 (large market) demand jobs arriving per hour are evaluated for a period of 24 hours. Valuations of demand jobs are drawn from a uniform distribution, i.e., $v \sim U[0, 1]$. Furthermore, installed renewable generation capacity is set such that hypothetical yearly demand and supply would be balanced. Note that this still allows for instances of scarce or excess supply. As we are also interested in incentives for the supply side, (homogeneous) reservation prices are varied in steps of 0.1 between 0.0 and 0.9. Finally, our results are based on $N = 150$ simulation runs for each parameter combination.

4.4.1 Efficiency

The central question concerning local and dynamic matching of demand side flexibility using an online mechanism circles around the attainable (relative) social welfare. We investigate both the effect of the market structure (size and flexibility) as well as the allocation procedure. In comparison to the social welfare under the clairvoyant planner (Sec. 4.3.2), both, the online planner and mechanism, can by design achieve only reduced results. To structure our analysis, we consider two efficiency gaps: (i) The cost of information given by the efficiency gap between offline and online planner. (ii) The cost of decentralization, i.e., the welfare gap between online planner and online mechanism. Clearly, these welfare gaps can be engineered to be arbitrarily large. Yet, for an assessment of the practical relevance of local matching of variable generation, we apply the scenario described in Section 4.2 and discuss the effects of allocation procedure, market size and demand side flexibility on social welfare. Using numerical evaluation, we can characterize and assess the magnitude under non-adversarial set-

---

[10]Smaller values deteriorate economic efficiency slightly. Larger values would be more likely to violate the elecrical engineering constraint of balanced supply and demand in small markets.

tings.[11] Figure 4.3 depicts both normalized welfare (upper panel) as well as the welfare gaps from incomplete information and decentral decision-making.
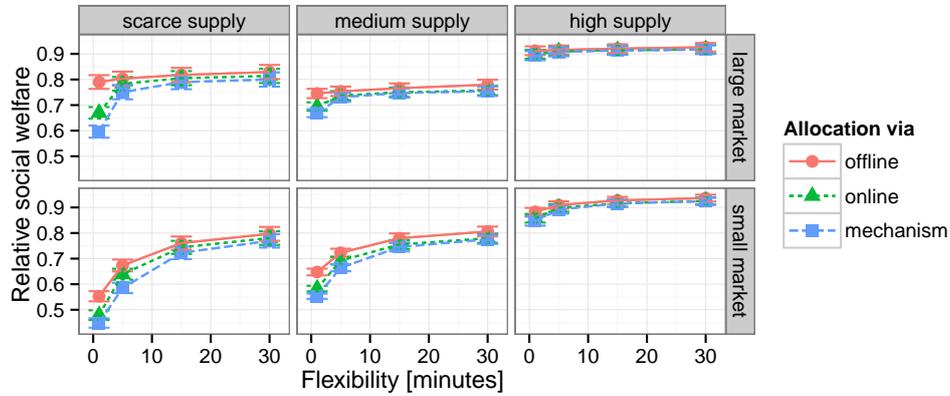
(a) Welfare over flexibility
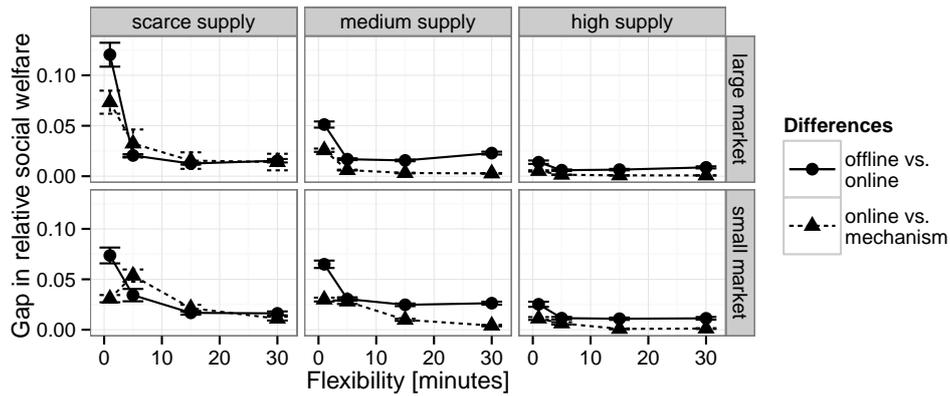
(b) Welfare gap over flexibility

Figure 4.3: Social welfare and welfare gap over demand flexibility with zero reservation price, differentiated by supply scenario (scarce, medium and high supply from RES) and market size (number of supply jobs per hour). Supply jobs are assumed to be available for five-minutes.

Market size and flexibility   Liquidity limitations are considered a central obstacle on the way towards more local, intraday electricity trading. Therefore, we are also interested in the effect of "localness" (i.e., liquidity or market size) on economic efficiency. Intuitively, the larger the market, the smaller the risk that orders remain un-

---

[11]Generation from RES can be assumed to be non-adversarial, i.e., driven by nature and not by a rational opponent.

matched due to a lack of temporal overlap between supply and demand jobs. However, physical grid limitations restrict possible market size. Interestingly, the analysis results indicate that demand flexibility can act as a substitute for market size: In the absence of demand side flexibility, small markets achieve the lowest relative welfare levels. For higher flexibility values, relative welfare approaches unity for all market sizes. At the same time, these results also suggest that in small markets larger flexibility endowments are necessary than in larger markets. In larger markets, already a minor amount of flexibility is sufficient to achieve near-maximum welfare. In small markets, marginal welfare from flexibility in low-flexibility scenarios is large, while the resulting welfare level is still significantly smaller relative to larger markets. More flexibility is required to secure significant total efficiency gains. This is a relevant finding for policy-makers who need to acknowledge that ensuring efficiency in heterogeneous market sizes will require heterogeneous flexibility endowments.

Effect of allocation procedure   The gap between the offline and the online planner is primarily governed by the level of demand side flexibility. Generally speaking, situations with more demand flexibility enable the online planner to make better decisions. This is due to the cost of poor allocation decisions being limited by a larger number of competing demand jobs. On the other hand, the cost of decentral decision-making is not only shaped by demand side flexibility, but also by market size and supply-demand imbalances, i.e., supply scenarios.

Interestingly, the effect of demand side flexibility on the price of IC is not monotonous: In the small market considered, the average gap between online planner and mechanism at zero flexibility is around 3%. For increasing demand side flexibility, the gap widens and reaches its peak of 6 % at five minutes of flexibility. For larger flexibility values this trend reverses and the gap again reduces to around 2 %. Furthermore, the larger the market, the quicker the welfare gap's decay over flexibility. Additionally, there is an interaction between flexibility and market size. For small markets with little demand side flexibility, the gap is small. Intuitively, this is a setting in which only few options are available and thus allocation decisions are easy to make. For larger markets the gap at zero flexibility is significantly larger and can exceed 12 % in the examined settings. It is in such settings (large market, low supply) where the value of flexibility is largest.

In summary, our results suggest that incentive-compatible coordination of flexible loads can be achieved at low cost (in terms of social welfare). Most notably, for non-zero flexibility levels, the welfare gap between the planning approaches and the IC mechanism never exceeds ten percent in total. Furthermore, for sufficiently flexible demand, market size becomes a factor of limited importance concerning the achievable social welfare.

4.4.2 Market Power

So far we assumed suppliers to post reservation prices equal to their marginal cost of generation, i.e., zero. We now depart from this assumption and examine the welfare effect of varying reservation prices for different demand flexibility levels, supply scenarios, and market size. Given the absence of clearing prices in the planner scenarios, this analysis solely focuses on social welfare as a result of the respective allocation decisions made by the mechanism. To assess the economic incentives for generators, we also consider the effect of non-zero reservation prices on supplier profits. This aids in understanding the economic sustainability of the proposed local marketplace.

Strategic suppliers will set reservation prices in a profit-maximizing manner. A departure from zero reservation prices will reduce the allocated volume as low-valued demand cannot be matched to available generation despite zero marginal costs. At the same time, this may increase allocation rates for higher valued demand. Consequently, welfare effects of strategic generator pricing are ambiguous.

Besides affecting the allocation results, strategic pricing of suppliers can also affect the mechanism's budget balance. This would be the case if suppliers were quoting heterogeneous reservation prices, as in the following example.

Example 1. *Suppose there are two jobs j_1, j_2 with consumption rates and total consumption $r_i = q_i = 1, \forall i \in \{1, 2\}$. Reservation price r is initially at zero. Both jobs are released in period t_0. j_1 is allocated immediately, while j_2 is to be allocated in period t_1. However, the reservation price is raised to $r > v_2$ in t_1. For efficiency reasons j_2 should no longer be allocated (as its value is less than the cost it incurs). However, in order to maintain incentive compatibility, it still must be allocated at the exact same price as the identical job j_1. Thus, IC in combination with varying reservation prices leads to settings in which BB is violated.*

To retain the property of budget balance, our strategic pricing analysis thus assumes homogeneous reservation prices.[12] Figure 4.4 illustrates the economic effects of strategic pricing for different supply scenarios as well as varying flexibility levels. Non-zero reservation prices mostly lead to reduced social welfare. The only exception are settings with scarce supply and little flexibility. Here, competition between demand jobs is particularly low and myopic decision making yields reduced social welfare. While supplier profit benefits from non-zero reservation prices, profit maximizing reservation prices exceed the welfare maximizing ones. The goal of increased supplier profits (and thus a more attractive market venue for suppliers) can be achieved either through the setting of reservation prices or through additional flexibility on the demand side. Interestingly, even in scenarios of excess supply, suppliers benefit from additional flexibility as it enforces competition on the demand side and therewith increases critical value payments. Furthermore, supplier profits are strictly positive

---

[12]Exploring which supplier motives and characteristics yield heterogeneous pricing decisions is an interesting opportunity for future research.
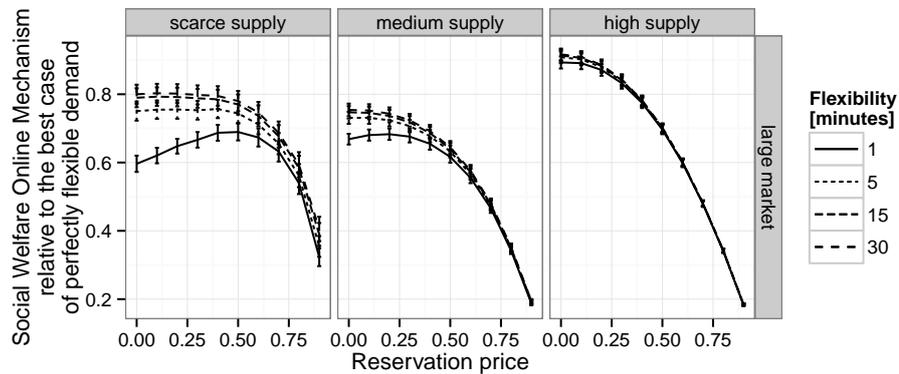
(a) Welfare over Reservation Prices.
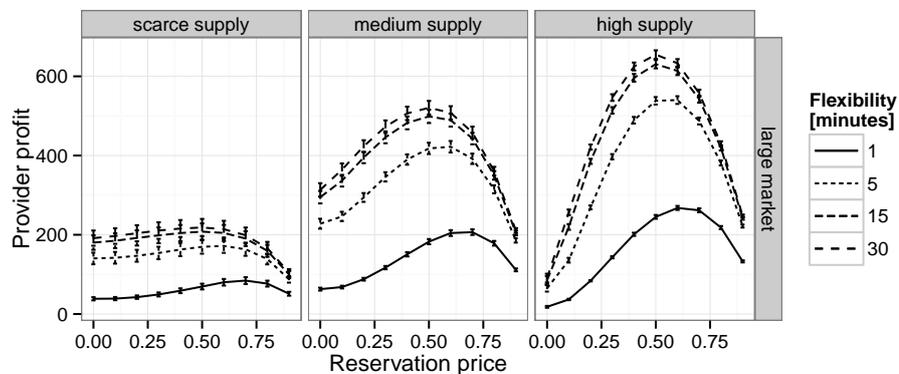
(b) Supplier Profits over Reservation Prices.

Figure 4.4: Trade-off between Profits and Efficiency

even under zero reservation prices. Naturally, increased competition on the demand side has positive effects on the profit of suppliers. Such competition can either be obtained through a larger, geographically more comprehensive market (which might require extra investment into physical assets such as transmission and distribution capacity) but also by means of a more flexible demand side. While the flexibilization of demand might convert so-far high-valued inflexible into flexible demand and thus lead to cannibalization of supplier revenues on the traditional primary energy market, it also increases competition on the presented market for excess renewable supply, raising revenues there, moderating the initially adverse effect for suppliers.

In addition to these desirable properties of the mechanism, it forms an interesting way of allowing suppliers to share quantity risk with consumers (this is in contrast to more traditional market designs where all demand needs to be served), similar to (Bitar and Low, 2012). Besides higher profits through increased flexibility, suppliers

benefit from reduced shedding of renewable supply in cases of excess supply.

4.4.3 Quality-of-Service and Individual Incentives

So far we assumed homogeneous flexibilities over all demand jobs. In this section, we focus on the effect of individual flexibility on probabilities of allocation and payments, and therefore introduce heterogeneous demand flexibility. More specifically, we draw flexibility f from an exponential distribution with parameter λ, i.e., $f \sim Exp(\lambda)$, where $\frac{1}{\lambda}$ amounts to the settings studied so far (1, 5, 15, 30 minutes). The resulting job heterogeneity allows for detailed insights into the interplay of job valuation and flexibility on payments and allocation probabilities. On a job level, we are particularly interested in the payments and how they relate to a job's flexibility. To this end we apply the notion of service levels and quality-of-service in the smart grid context (Oren and Smith, 1993). Different from most of the literature, were customers exogenously declare their preferred combination of service-level and price, we let service levels arise endogenously from jobs' valuation and flexibility.

The binary decision for a job to get served or not depends on system state (demand and supply) and job valuation v and flexibility f. The notion of a service level, i.e., the probability of successful allocation given its flexibility and valuation, can only be defined meaningfully for a set of similar jobs. With *similar*, we refer to jobs that exhibit akin valuations and flexibility, and define QoS as the probability of successful allocation for similar jobs given v and f, i.e., $QoS(v, f) = \mathscr{P}(\text{allocation}|v, f)$, a metric that is increasing in both valuation and flexibility.

Accordingly, job flexibility facilitates reliable service under unreliable supply regimes. Flexibility and valuation can act as both, complementary and substitutive attributes, while the value of flexibility for individual jobs is two-fold: Higher flexibility both increases service levels and reduces payments (critical value payments decrease monotonously in flexibility), thereby increasing job i's expected utility

$$E(u_i) = (v_i - p_i) * \mathbb{P}(\text{allocation}|\theta_i)$$

from allocation. Using this property, jobs can effectively choose in which currency, valuation or flexibility, to pay for reliable supply.

Fig. 4.5 illustrates aggregate results on the relationship between flexibility, valuation and expected utility at reservation price $r = 0$ and average jobs flexibility f of one minute.

The results indicate that market size (rows) has barely any influence on expected utility. On the other hand, the scarcer supply, the smaller the expected utility of a job. This is a natural consequence of critical value payments, which are increasing in competition on the demand side. For scenarios with rather limited supply, the substitutive character of flexibility and valuation becomes obvious, regardless of overall reduced utility due to increased competition on the demand side.
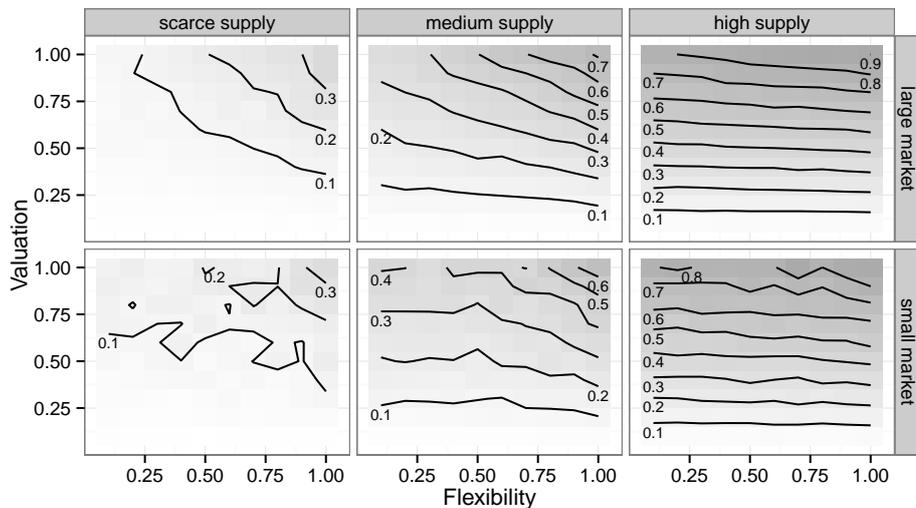
Figure 4.5: Contour plot of average job utility differentiated by valuation and flexibility quantiles. Expected utility increases in supply (columns from left to right). Market size (rows) has only litte influence on expected job utility. Through provision of flexibility, jobs increase expected utility. ($r = 0$, $f = 1$)

4.5 Discussion

Renewable energy sources are providing increasing amounts of energy to electrical power systems globally and introduce change to power systems on both short and long time scales. The uncertainty in generation from these energy sources poses a significant problem to (future) power grid operations. The use of even limited amounts of flexibility on the demand side can play a vital role to facilitate grid integration of fluctuating generation. Leveraging demand side flexibility may reduce network and storage expansion required for safe power system operations and thus assist the transition towards economically and ecologically more sustainable power systems.

Previous research has mostly focused on applying (online) optimization techniques to optimize the dispatch of flexible loads. However, we argue that for the vision of flexible demand in smart grids to materialize, incentives regarding the provision of flexibility on the level of the individual consumer must be taken into account: A future grid can greatly benefit from tapping into demand side flexibility by establishing incentive-compatible allocation and pricing rules. To this end, we present an online mechanism and apply it in a realistic smart grid setting with emphasis on local matching of flexible demand and uncertain supply from renewable generation. As information on demand side flexibility and its distribution over consumers (or even classes of devices) is scarce, our evaluation relies on simple, yet plausible assumptions regarding the demand side. For the supply side we instantiate a generic model with empirical

wind farm generation data.

Under this truthful mechanism, it is the dominant strategy for jobs to report their true type with respect to valuation and temporal flexibility, independent of other jobs' reports. This is because flexibility is valuable in two ways to the individual job: It reduces payments and increases the probability of allocation. Leaving an individual job's perspective, demand side flexibility is also valuable to the power system as a whole: We show that flexibility on the demand side reduces shedding of excess supply and thus increases the number of served jobs, positively affecting social welfare. Our evaluation indicates that in settings with reasonable flexibility levels, the price of information (i.e. the welfare gap between offline and online planners) is rather small (on average below 5%). Thus, we conclude that demand flexibility can substitute market size, rendering even small markets surprisingly efficient. While this is an encouraging result, we advance beyond online optimization and include incentives. Our results indicate that the price of incentive compatibility, or gap between the online planner and mechanism, respectively, is even smaller, and amounts in most settings to approximately 2%. Moreover, suppliers profit from flexibility as it increases competition and corresponding critical value payments, but also that it reduces the amount of shedded supply from renewable sources. In combination, this poses a powerful argument for fostering demand side flexibility as its (individual and global) value increases in the presence of greater levels of fluctuating generation.

For future work we see a number of worthwhile research avenues. The first concers the effect of different distributions of flexibility endowments over jobs on social welfare. What kind of general flexibility distributions are particularly valuable given different compositions of the generation side? Extending the presented model framework, a further interesting topic touches upon optimal investment in both renewable and conventional generation capacity on the one side and flexibility on the other. This question is also closely related to equilibrium prices on smart-grid markets with quality-differentiated products. Here, it would be interesting to see how matching on short-horizon markets affects outcomes on related markets. Finally, the trade-off between budget-balance and efficiency in conjunction with time-varying reservation prices is an interesting topic worthy of further exploration. As we have shown for small markets in low-flexibility settings, forgoing budget balance through non-zero reservation prices can ameliorate social welfare. There might be situations where abandoning budget balance in favor of higher efficiency is advantageous.

Chapter 5

Model-based Online Mechanism Design for Scheduling Non-Preemptive Jobs

5.1 Introduction

The availability of electrical energy from renewable sources such as wind and solar has been increasing rapidly in the past years, and is expected to significantly increase even further in the near future.[1] However, electricity generation from many renewable sources cannot be easily controlled and is often difficult to predict. Therefore, balancing demand and supply in the power system becomes increasingly challenging when solely relying on the ramping capabilities of conventional generators (e.g., CCGT). Alternatively, this problem can be addressed by introducing more flexibility on the demand side and allowing loads to be deferred. Thus, a major challenge in energy systems given both uncertain future demand as well as uncertain supply from RES lies in *online* scheduling of flexible loads.

To meet this challenge, we introduce and evaluate several novel algorithms for online scheduling of deferrable loads. These algorithms take into account probabilistic information about future supply and demand. Furthermore, we use a mechanism design approach to incentivize agents on the demand side to be truthful about their flexibility and the value of the loads. Specifically, we consider the problem of scheduling multiple non-preemptive loads (i.e., loads that, once started, cannot be interrupted) with a fixed load profile (i.e., power consumption rate of a load is an unalterable function of time from the starting point of the load) and uncertain supply from relatively cheap or even free renewable energy.

Our algorithms extend existing single machine scheduling algorithms called *expectation* (Chang et al., 2000a) and *consensus* (Hentenryck and Bent, 2009), which take uncertainty about future jobs into account. These algorithms generate scenarios from a probability distribution, and the scheduling problem is solved for each of these scenarios using an offline scheduling algorithm. Under the *expectation* algorithm deci-

---

[1]In Germany, 22.9% of electricity was supplied from renewable sources in 2012 (Sawin, 2013), up from 12% in 2006. The UK government has committed to ensure that 30% of electrical energy will be supplied from renewable sources by 2020 (DECC, 2011).

sions are based on the expected utility (over all scenarios) of scheduling a certain job immediately, whereas in the *consensus* approach, decisions are based on the number of votes by the scenarios to decide which job to schedule first. We extend both algorithms to not only cope with uncertainty on the demand side, but to also apply in settings with variable, uncertain supply where multiple, heterogeneous jobs, can run simultaneously. Furthermore, we use the concept of pre-commitment to convert these algorithms into truthful mechanisms.

Recently, *consensus* was applied in a smart grid setting with a specific focus on the context of energy allocation for the charging of electric vehicles (Stein et al., 2012). While that work addresses the problem of uncertain future demand, the model is restricted to deterministic future energy supply with constant marginal cost. We specifically focus on the problem of uncertain supply. More precisely, we assume that there are two sources of electricity: renewable supply, which is uncertain but free, and conventional generation, which is always available but costly. Another marked difference to Stein et al. (2012) is the consideration of non-preemptive loads. Stein et al. (2012) restrict their attention to preemptive loads, such as electric vehicles, which reduces the problem's complexity. However, in many real life settings jobs requiring electricity are non-preemptive (i.e., they cannot be easily interrupted and restarted).[2]

Our work is also related to Subramanian et al. (2012), who similarly consider the problem of scheduling deferrable loads. However, there are some important differences regarding the model and algorithms they consider. First, the predictive approach they employ, a simple point prediction of renewable supply, is incapable of reflecting uncertainty. Furthermore, their approach does not take (auto-)correlation of generation into consideration, which is common in renewable supply (e.g., if it is currently windy, it is likely that the following hour will be windy as well, similar with sunshine). In contrast, our *consensus* based approach considers several scenarios which are sampled from the distribution, and (auto-)correlation is taken into account when generating such scenarios. Another important difference is that they assume loads to be preemptive, as requiring a certain total amount of energy, and as being characterised by a flexible power consumption rate. In contrast, we assume non-preemptive loads with a fixed load profile and require that started loads need to be completed. Finally, unlike Subramanian et al. (2012) and similar to Stein et al. (2012), we assume that loads have monetary values, and our aim is to maximise the difference between the value of allocated loads and the cost of using non-renewable energy, i.e., to maximise social welfare.

Recently, researchers in the multi-agent community have begun looking at adapting online scheduling heuristics to deal with strategic agents. Such agents may misreport the value, arrival time or deadline for their jobs if, through such a misreport, they can get a better allocation or pay less. In order to ensure that an online schedul-

---

[2]Examples include washing machines, in a domestic setting, and a variety of heavy duty electrical machinery in an industrial setting.

ing heuristic is truthfully implementable with strategic agents a key criterion to be satisfied is *monotonicity* (Parkes, 2007): if a job has a type that is better in any of its dimensions than another (e.g., higher value, lower consumption rate, shorter length, earlier arrival or later deadline), and no worse in any other dimension, then its allocation must not be worse.

There are two main approaches to ensure monotonicity of online allocations. One approach is *output ironing* (Parkes and Duong, 2007; Constantin and Parkes, 2009), or cancelling that part of the allocation that breaks monotonicity constraints. While this is a principled approach, often a large part of the final allocation may need to be cancelled, and computing the ironing decisions can be intractable in realistically-sized settings. Another approach, which we follow here, is to partially pre-commit to serving jobs of sufficiently high value in the future, irrespective of future arrivals (Stein et al., 2012). This approach can limit efficiency, because it imposes additional constraints on future schedules. Yet, it has the advantage that it prevents strategic agents from misreporting.

Specifically, our main contributions are as follows:

- We consider, for the first time, the problem of online scheduling non-preemptive jobs and uncertain supply of resources.

- We present and compare several variants of two new algorithms for this setting: an extension of the *consensus* approach (Hentenryck and Bent, 2009) and an extension of the *expectation* approach (Chang et al., 2000a) to deal with both selecting *multiple* jobs at each time step and variable supply.

- We apply mechanism design to produce truthful variants of these algorithms by adopting the concept of pre-commitment.

This Chapter is structured as follows. First, we formalise the online scheduling problem. Then, we explain how we extend both *consensus* and the *expectation*-based approach to deal with possibly selecting multiple jobs at each time step, followed by a description of the issues and solutions to make these methods incentive-compatible. We conclude with an experimental validation and a discussion. The main parts of this chapter have been published in (Ströhle et al., 2014).

5.2 Problem Formulation

Our online scheduling problem with non-preemptive loads is characterised by demands of different values and requirements, which arrive online, and supply from two sources: an uncertain future amount of low-cost power from renewables, and controllable amounts of costly conventional generation. The decision of the scheduler concerns which of the incoming loads to schedule at what time with the goal

of maximizing expected social welfare (defined below). It should be noted that the offline version of the problem (with perfect knowledge of demand and supply) is already of combinatorial complexity, as finding the optimal schedule requires a combinatorial number of subsets to be evaluated and compared (Hentenryck and Bent, 2009). Hence, approximation algorithms are needed. Before we present the algorithms, we detail the formal definition of demand, supply, and the schedule.

5.2.1 Demand

We consider a setting where jobs arrive over a fixed finite time span (e.g., a day), modelled by a set of discrete time steps $T = \{1, \ldots, \mathcal{T}\}$. A *job* $j \in J$, where J is the set of all jobs, is characterised by a type $\langle v_j, r_j, l_j, a_j, d_j \rangle$, which comprises its value $v_j \in \mathbb{R}^+$, consumption rate $r_j \in \mathbb{R}^+$ (the amount of supply needed per time step), job length $l_j \in T$, arrival time $a_j \in T$, and departure time or deadline $d_j \in T$. Given this, we denote the total amount of energy required to serve a job as $q_j = l_j \cdot r_j$.[3] A job j cannot start before a_j, must end before d_j, and is non-preemptive, i.e., once it is started, it must continue running for l_j time steps (which we assume to be bounded by a constant maximum length). Thus, the latest start time for a job is $d_j - l_j$. A job's temporal flexibility can thus be expressed as $d_j - l_j - a_j$. We assume that the set of jobs J is not known a-priori, but is only revealed online as jobs arrive in the system. However, the scheduler can access samples of future jobs and their properties, e.g., valuation and length, and thus approximate the underlying probability distributions.

5.2.2 Supply

Electricity is supplied from two sources with different properties, renewable and conventional:

- Renewable sources, such as wind or solar power, are characterised by negligible marginal cost but also uncertain availability. For simplicity, we assume that costs for renewable energy are zero (although the algorithms carry over to settings where the marginal cost is constant, and can easily be generalised to other cost functions). Furthermore, we assume that the available renewable power is given by a stochastic process $X_T = (X_1, \ldots, X_{\mathcal{T}})$, whose realisation $x_t \sim X_t$ only becomes known at t. In order to allow for auto-correlated supply over time (which means that realisations at time t provide information about future supply), we encode the properties of this stochastic process via a Hidden Markov Model (HMM).

---

[3] For reasons of simplicity, we restrict our attention to jobs with constant power consumption. Generalization to more realistic power profiles is straightforward.

- Conventional generation, on the other hand, is characterised by unlimited output and a deterministic cost function c which is non-decreasing in the amount of power supplied at time t. In particular, we take these costs to be described by a linear function (constant marginal cost), i.e., $c(p) = b \cdot p, \quad b > 0$.

We acknowledge that supply is perishable, i.e., electrical energy that is not immediately consumed, cannot be stored and consumed in the future.

5.2.3 Schedule Determination

The solution to the online scheduling problem is a *schedule* $s = \langle s_1, s_2, \dots, s_\mathcal{T} \rangle$, which defines for every time step t a set of jobs $s_t \subseteq J$ to start at that time. A *feasible* schedule s must fulfill the following requirements , for all times $t \in T$ and for all jobs j started at time t, i.e., $j \in s_t$:

- j cannot start before its arrival, i.e., $t \geq a_j$,

- j must finish by its deadline, i.e., $t \leq d_j - l_j$,

- j can be started at most once, i.e., $\forall\, t, t' \in T : \text{if } j \in s_{t'} \text{ and } j \in s_t \text{ then } t = t'$.

We use $s = \langle \rangle$ to denote the empty schedule, i.e., where $s_t = \emptyset$ for all $t \in T$. Furthermore, given a schedule s and a time t, we denote the set of *running jobs* by $R_t(s) = \{j \mid j \in s_{t'}, t' \leq t < t' + l_j\}$. The net profit (or social welfare) $w(s)$ of a schedule s is then defined by the value of all scheduled jobs minus the cost of conventional generation. Formally,

$$w(s) = \sum_{t \in T} \left(\sum_{j \in s_t} v_j - c\Big(\max\Big\{ \sum_{j \in R_t(s)} r_j - x_t, 0 \Big\} \Big) \right) \tag{5.1}$$

which is the value we aim to maximise.

5.2.4 Strategic Behaviour

When designing the algorithms, we also need to consider strategic behaviour of demand side agents/jobs. Since the types of the jobs constitute private information, we would like to incentivize agents to reveal their types truthfully. Otherwise, agents could speculate and the scheduler might take suboptimal decisions based on incorrect/manipulated information. Specifically, the aim is to design a mechanism, i.e., a scheduling algorithm and corresponding payments, which is *dominant-strategy incentive-compatible*, i.e., reporting truthfully maximises an agent's utility, regardless of the behaviour of other agents. To guarantee this property, we assume that *each* job is owned by a different agent. Additionally, we require *individual rationality*, i.e., the required payment never exceeds a job's value, and is zero if a job is not run. In

Section 5.4 we return to these issues in detail, focusing first on the online scheduling problem.

5.2.5 Offline Optimal: Clairvoyant Offline Scheduler

Online decision-making, i.e., under incomplete information, typically leads to suboptimal solutions. To illustrate the problem, consider the following example.

Example 2. *Consider a setting with two time steps t_1 and t_2 and two jobs $\{j_1, j_2\}$, each requiring exactly one unit of energy. The value of j_1's job is $v_1 = 7$, and it can run during either t_1 or t_2, while j_2 has a job which has value $v_2 = 5$, but only during t_1. At t_1 one unit of renewable supply is available with associated costs $c = 0$. For t_2, the scheduler* expects *one unit of renewable supply to be available, but there is a very small chance that this is not realised and then the alternative from conventional generation will be very expensive, i.e., $c = 10$. Moreover, the mechanism expects, with high probability, no further arrivals, but there is a small chance job j_3 with a high value $(7 < v_3 < 10)$ will enter the market at t_2. At t_1 the expected optimal allocation is to allocate j_2 (which has an earlier deadline), and postpone j_1. Now assume that j_3 enters the market at t_2. This leads to j_1 being discarded at t_2, and j_3 being allocated. With hindsight (i.e., offline optimal), however, it would have been better to discard j_2, allocate j_1 at t_1, and j_3 at t_2.*

Assuming full information on the realization of supply and demand (hindsight, or perfect foresight, respectively), resolution of the following mixed-integer progamming (MIP) formulation yields the offline-optimal benchmark schedule. In this formulation, the previously introduced schedule s is replaced by multiple decision variables: If job j is running at time t, $\alpha_{j,t} = 1$. Accordingly, $\beta_{j,t} = 1$, if job j is started at t, $\gamma_{j,t}$ encodes finishing of a job, $\phi_{j,t}$ describes whether j is served and thus included in the schedule, and $\kappa_{j,t}$ represents the consumed amount of supply by the respective job in period t.

Constraints   As indicated in Section 5.2.3, a valid schedule respects a number of constraints, which are expressed as follows.
First, any job can be started (β) and, accordingly, completed (γ) only once.

$$\sum_{t \in T} \beta_{j,t} \leq 1 \qquad\qquad \forall j \in J \qquad\qquad (5.2)$$

$$\sum_{t \in T} \gamma_{j,t} \leq 1 \qquad\qquad \forall j \in J \qquad\qquad (5.3)$$

Second, a job is active (α) if it has been started in a preceding or the current time step and is not yet completed. Furthermore, stopping a job is only possible, if it has been

active before.

$$\alpha_{j,t} = \beta_{j,t} \qquad\qquad \forall j \in J, \quad t = 1 \qquad (5.4)$$

$$\alpha_{j,t} = \alpha_{j,t-1} + \beta_{j,t} - \gamma_{j,t} \qquad\qquad \forall j \in J, \quad \forall t \in \{2,..,\mathcal{T}\} \qquad (5.5)$$

$$\gamma_{j,t} \leq \alpha_{j,t-1} \qquad\qquad \forall j \in J, \quad \forall t \in \{2,..,\mathcal{T}\} \qquad (5.6)$$

Fourth, starting a job obviously is only possible after its arrival. Fifth, for the job to be included in the schedule, it must be started before the latest possible start time and completed by the time of departure.

$$\sum_{t \in T} \beta_{j,t} \cdot t \geq a_j \cdot \phi_j \qquad\qquad \forall j \in J \qquad (5.7)$$

$$\sum_{t \in T} \beta_{j,t} \cdot t \leq (d_j - l_j) \cdot \phi_j \qquad\qquad \forall j \in J \qquad (5.8)$$

$$\sum_{t \in T} \gamma_{j,t} \cdot t \leq d_j \cdot \phi_j \qquad\qquad \forall j \in J \qquad (5.9)$$

Moreover, a job is characterized as being active over its entire length l. The job's rate of energy consumption κ must be greater or equal to the job's consumption rate r during the job's active time steps.

$$\phi_j \cdot l_j \leq \sum_{t \in T} \alpha_{j,t} \qquad\qquad \forall j \in J \qquad (5.10)$$

$$\kappa_{j,t} \geq \alpha_{j,t} \cdot r_j \qquad\qquad \forall j \in J, \forall t \in T \qquad (5.11)$$

Conventional generation can only be greater or equal to zero at any time, i.e., $x^{CG} \in \mathbb{R}^+$. Finally, aggregate supply from both sources must be equal or greater than total consumption of served jobs.

$$x_t^{CG} + x_t \geq \sum_{j \in J} \kappa_{j,t} \qquad\qquad \forall t \in T \qquad (5.12)$$

Objective Function   The objective function represents social welfare, the aggregate difference between sum of served jobs' valuation and the corresponding cost due to using conventional generation. Other objectives, i.e., maximizing the number of served jobs, minimizing the aggregate curtailment of renewable energy, or minimizing the use of conventional generation could substitute this objective. Formally, based on the notation introduced so far, the objective function can be expressed as follows:

$$w(\alpha, \beta, \gamma, \kappa, \phi) = \sum_{j \in J} v_j \cdot \phi_j - \sum_{t \in T} c \cdot x_t^{CG} \qquad (5.13)$$

This objective function is equivalent to Equation (5.1), but expressed in a different set of decision variables. We denote the objective value of the optimal solution to this

---
Algorithm 1: Greedy offline scheduler.

---
1 **Algorithm:** GREEDY-OFFLINE (J, x, s, t)

2 $J' \leftarrow \{j \in J \mid j \notin s, d_j \geq t + l_j\}$

3 **for** $j \in$ SORT(J') **do**

4 $\quad$ $t_{min}, c_{min} \leftarrow$ COSTMINIMALSTARTTIME(j, x, s, t)

5 $\quad$ **if** $c_{min} < v_j$ **then**

6 $\quad\quad$ $s_{t_{min}} \leftarrow s_{t_{min}} \cup \{j\}$

7 **return** s

---

program by w^*. We maximize the objective through appropriately setting the decision variables, i.e.,

$$w^* = \max_{\alpha, \beta, \gamma, \kappa, \phi} w \tag{5.14}$$

Due to the problem's combinatorial complexity this formulation is exclusively used for (economic) benchmarking purposes. In the online setting, where time-to-solution is critical, and, more importantly, there is incomplete information on future demand and supply, we will rely on greedy heuristics for efficient schedule construction.

5.3 Model-based Online Scheduling

In this section, we present our extensions of the online algorithms due to Hentenryck and Bent (2009). Similar to their work, our algorithms deal with uncertainty by sampling multiple future scenarios using an appropriate model of the system (in our case, sampled realisations of the future supply of renewable energy). Then, at each time step, an offline algorithm is used to solve each of these scenarios, and the resulting schedules are combined to yield the best decision to take in the current time step. However, unlike previous work, our algorithms are able to schedule multiple jobs per time step (rather than a single one), deal with uncertain future supply (rather than assuming this to be deterministic), and incorporate costs of exceeding the available supply (by using conventional generation).

As our algorithms rely on solving instances of an offline version of the scheduling problem, we first detail the corresponding algorithm in Section 5.3.1. Then, we provide a generic online algorithm that all our approaches follow, and conclude this section with two novel online scheduling algorithms: m-CONSENSUS, and m-EXPECTATION.

5.3.1 Offline Scheduling Algorithm

In the offline variant of our scheduling problem, we assume that all jobs J and the realization of the supply x are known in advance. However, finding an optimal sched-

ule even in this case is known to be computationally hard, as it is a generalisation of the NP-hard parallel machine scheduling problem Pinedo (2012). Following Dunke (2014), wo argues that solving scenarios – referred to as snapshots – to optimality, loses its "[..] efficacy once the situation [changes] upon arrival of new input elements" on one side, and aiming to maximize the information collected from the set of available realizations (Hentenryck and Bent, 2009, p.23) on the other, we use a greedy scheduling heuristic and refer to this as GREEDY-OFFLINE (Algorithm 1).

This algorithm has, besides the set of jobs to schedule J and renewable supply x two more arguments, s and t, which are used by our online algorithms later to encode past (and fixed) scheduling decisions (s) as well as the current time (t), which is the earliest time at which new jobs may be scheduled.[4] If not noted otherwise, the algorithm first sorts the available jobs J by decreasing value density v_j/q_j. Tie breaking is based on job size q with larger jobs receiving preferrential treatment.[5] For each job in this order, the function COSTMINIMALSTARTTIME then computes the starting time $t_{min} \in T, t_{min} \geq t$, that minimises the additional cost incurred by adding job j to s_t. If the net marginal welfare contribution from including j in the schedule is positive, i.e., the associated minimum cost, c_{min}, is less than the value of the job, the job is included in the schedule s. The scheduler thus effectively performs two operations for each job:

- First, deciding whether to include the respective job in the schedule, and if so

- at what time to schedule it.

For n jobs, the computational complexity of this scheduler is $\mathcal{O}\left(n\log n\right)$ for sorting (Skiena, 2011) , plus $\mathcal{O}\left(n\mathcal{T}\right)$ for finding the best start times for all jobs. In general, it can be bounded by $\mathcal{O}\left(n\mathcal{T}\right)$ because $\mathcal{T}$ is typically much larger than $\log n$.

5.3.2 Online Scheduling Algorithms

In online settings, new information is revealed over time, requiring sequential decision making. We consider two algorithms that are executed in the following context.

First, a set of $\mathcal{N}$ scenarios is created that serve as samples of the future, each representing one possible trajectory of the future. Technically, each scenario $i \in \{1,..,\mathcal{N}\}$ consists of the tuple $\langle J^i, x^i \rangle$, where J^i is a randomly sampled realisation of future *demand* which we refer to as the set of "virtual jobs", and x^i is a sampled realisation of *supply* based on a probabilistic supply model (or historical data).

Then, at every time point t, the online algorithm is invoked with the following arguments: the currently startable jobs (not scheduled so far) J_r, the realisation of the renewable supply x up until and including t, as well as the schedule so far. (Note that past scheduling decisions can have consequences beyond the current time point t,

---

[4]For now, these are set to $s = \langle \rangle$ and $t = 1$.

[5]Larger jobs are decided upon first, such that smaller jobs can be scheduled into possibly remaining gaps.

as jobs cannot be preempted, and the schedule can thus no longer be altered once a job is started.) To better incorporate the concept of limited foresight, the scenarios include future demand within the boundaries of horizon h, i.e.,

$$J^i(t) = \{j \mid j \in J^i \text{ where } t + 1 \leq a_j \leq t + h\}.$$

Supply is included in a similar fashion. However, jobs have length $l \geq 1$, and therefore, in order for the algorithm to make sensible acceptance decisions, the horizon regarding future supply h' is extended to $h' = h + \max\{l(j) \mid j \in J^i\}$. Future supply in scenario i can then be expressed as follows.

$$x^i(t) = \langle x_t, x^i_{t+1}, x^i_{t+2}, \ldots, x^i_{\min(t+h', \mathcal{T})} \rangle$$

The online algorithm returns the set of jobs to start next, s_t, which iteratively defines the full schedule s. We name the following two new online algorithms for selecting a set of non-preemptive jobs *multi-machine expectation* and *multi-machine consensus*.

Multi-Machine Expectation   Multi-machine m-EXPECTATION (Algorithm 2) relies on sampled scenarios and uses these to explicitly compute each job's marginal welfare contribution. m-EXPECTATION keeps starting jobs at time step t, until the best action is to no longer start jobs (line 13). To this end, in each scenario $i \in \{1, .., \mathcal{N}\}$ (line 4), $|J_r| + 1$ schedules are constructed: One for each $j \in J_r$, where j is started at the current time t and one additional schedule with no additional job being started. From this set of hypothetical schedules the job corresponding to the schedule with the largest marginal contribution to social welfare j^* is selected. Note that this can also be the empty job $\bot$. The selected job is then added to the schedule to be executed at the current time step t. As mentioned before, this procedure is repeated until the best action is to not add another job to the schedule s. This approach promises high efficiency, as the expected welfare directly represents the value we wish to maximize. However, it also incurs high computational cost, as we evaluate each scenario $|J_r| + 1$ times (once for each available job, and once to evaluate the case where no job is started). Accordingly, computational complexity of m-EXPECTATION (for a single time step t) is $\mathcal{O}(n^3 \mathcal{N} \mathcal{T})$. Following Hentenryck and Bent (2009), cubic complexity in the number of jobs under consideration in each time step prevents this algorithm from application in online domains with tight deadlines. By avoiding explicit computation of each scenarios' potential welfare, the following algorithm, m-CONSENSUS, succeeds in necessary complexity reduction.

Multi-Machine Consensus   The multi-machine *consensus* or m-CONSENSUS algorithm, is given in Algorithm 3. The algorithm solves the offline problem (line 5) for each scenario once and then schedules the job that is selected to be started immediately in the weighted largest number of scenarios (or none, if more scenarios do not

---

Algorithm 2: Schedule the jobs from J_r at t that have the highest added value.

---

1 **Algorithm:** m-EXPECTATION (J_r, x, s, t)

2 **repeat**

3   Reset counters f (with $-\epsilon$ for $f(\bot)$)

4   **foreach** *scenario* $\langle J^i, x^i \rangle$ **do**

5          $f(\bot) \leftarrow f(\bot) + w(\text{GREEDY-OFFLINE}(J^i(t) \cup J_r, x^i(t), s, t+1))$

6   **for** $j \in J_r$ **do**

7              $s' \leftarrow s$

8              $s'_t \leftarrow s'_t \cup \{j\}$

9              $f(j) \leftarrow f(j) + w(\text{GREEDY-OFFLINE}(J^i(t) \cup J_r \setminus \{j\}, x^i(t), s', t))$

10      $j^* \leftarrow \arg\max_{j \in J_r} f(j)$

11   **if** $j^* \neq \bot$ **then**

12          $J_r \leftarrow J_r \setminus \{j^*\}$; $s_t \leftarrow s_t \cup \{j^*\}$

13 **until** $j^* = \bot$

14 **return** s

---

start a new job). This is repeated iteratively, adding one additional job to the schedule at a time, until no more jobs are started. This repetition occurs at most n times (but usually much less frequently), [6] and so including the $\mathcal{O}(n\mathcal{T})$ per call to the offline algorithm, the computational complexity of m-CONSENSUS (for a single time step t) is $\mathcal{O}(n^2 \mathcal{N} \mathcal{T})$.

Compared with m-EXPECTATION, m-CONSENSUS is computationally more favorable, but might suffer from its elitism (Hentenryck and Bent, 2009): Decisions are based on votes, instead of the more decisive, but computationally more expensive social welfare criterion. For example, consider a setting with three scenarios: Consensus might schedule a job in two of the three scenarios, and accordingly, the job ends up being added to the eventual schedule. Expectation, on the other hand, has more fine-grained information and decides to reject the job, as it barely adds to the objective in the two scenarios, and considerably worsens the objective in the third scenario. In this example, the drawback of basing decisions on voting instead of the actual objective becomes evident.

5.3.3 Scenario Sampling

While the true distribution of future supply (or demand) is unknown, we assume the availability of a black box sampler, which returns independent scenarios from the distribution of future demand and supply. Hence, the true distribution can be approximated by sampling repeatedly.

Ideally, at each time point $t \in T$ these scenarios are updated to account for novel

---

[6] In addition, the offline scheduling problem gets smaller by 1 job in every iteration.

Algorithm 3: Schedule the jobs from J_r at t that occur in the most scenarios. Scheduling the empty job is denoted by $\perp$.

---

1 **Algorithm:** $m\text{-}\textsc{Consensus}\ (J_r, x, s, t)$

2 **repeat**

3   Reset counters f (with $-\epsilon$ for $f(\perp)$)

4   **foreach** *scenario* $\langle J^i, x^i \rangle$ **do**

5         $s' \leftarrow \textsc{Greedy-Offline}(J^i(t) \cup J_r, x^i(t), s, t)$

6   **if** $s'_t = s_t$ **then**

7             $f(\perp) \leftarrow f(\perp) + 1$

8   **else**

9   **for** $j \in J_r \cap s'_t$ **do**

10                $f(j) \leftarrow f(j) + 1$

11     $j^* \leftarrow \arg\max_{j \in J_r} f(j)$

12   **if** $j^* \neq \perp$ **then**

13         $J_r \leftarrow J_r \setminus \{j^*\}$; $s_t \leftarrow s_t \cup \{j^*\}$

14 **until** $j^* = \perp$

15 **return** s

---

information. However, due to the nature of the proposed mechanism, which irreversibly fixes decisions during the pre-commitment phase, updating scenarios poses specific challenges. In detail, the shorter the interval between decision time points (the higher the decision frequency), the more likely the occurence of constellations where a majority of scenarios are overly optimistic, i.e., characterized by high supply from RES and few and/or low-valued future jobs. As a consequence, the mechanism, relying on resampled scenarios, would decide to allocate all jobs active at time t, i.e., overcommit, crowd out future jobs with higher value, and hence achieve only reduced social welfare. The converse case with undercommittment eroding efficiency might also take place. The consequences, however, would be less severe: The corresponding decisions could be corrected at the next time step by committing to more jobs. To mitigate this undesirable re-sampling effect, we abstain from regular, rolling resampling and propose to sample only infrequently, i.e., once per day.

5.4 Online Mechanism Design

Classic optimization does not take incentives into account. We are, however, especially interested in creating allocation schemes, that render thruthtelling the dominant strategy for the individual agent/job. To this end, we depart from the scheduling paradigm and introduce an incentive-compatible (IC) mechanism. A mechanism can only be IC if allocation is monotonic, i.e., it must be impossible that a job reporting a lower type (i.e., a lower valuation, later arrival, earlier departure, or longer job require-

ment) is allocated instead of an job reporting a higher type (Parkes, 2007).

In order to ensure monotonicity, we take the approach first proposed by Stein et al. (2012). Here, the mechanism is split into two phases at each time step: pre-commitment and allocation. During the pre-commitment phase, the mechanism evaluates which jobs (might) contribute to a schedule's welfare given current commitments, as well as current and scenarios' future arrivals. If a job is deemed to increase social welfare, the mechanism commits to allocating supply to it before its deadline, regardless of the values and demands of future arrivals. Thus, an job has no incentive to misreport its type, e.g., via an earlier deadline or later arrival, because once pre-committed, the mechanism guarantees it to be allocated in the future.

During the allocation phase, the actual execution schedule is computed. The focus of this phase is on efficiency only, because incentives issues have already been dealt with during the pre-commitment phase. Note that, although for computing the allocation we can use the algorithms presented in the preceding section, all resulting schedules must respect the constraints from the pre-commitment decisions. So, for example, jobs which have been pre-committed and are flexible can be delayed, but once their deadline approaches, they must be scheduled, regardless of subsequent arrivals. In the following, we discuss these phases in separate sections.

5.4.1 Precommitment

In the pre-commitment phase, the mechanism needs to decide which of the jobs to commit from those that have already arrived in the system (J_r), and to what extent to retain spare (renewable) capacity for potential future arrivals. In order to ensure monotonicity (and hence IC), jobs that are precommitted must be scheduled before their deadline, regardless of future arrivals, thus precommitments may reduce flexibility of future allocations. In the worst case, i.e., if future arrivals are underestimated and future supply is overestimated, ensuring IC through precommitment may yield negative welfare (situations in which the cost of serving demand is more costly than the aggregated value of served demand). Hence, as the mechanism makes irreversible decisions, the quality of forecasting is of high importance to achieve high efficiency.

Example 3. *Assume the situation described in Example 2 and also assume that either the high-value job j_3 arrives or supply is not realised. Then, j_1 will not be allocated at t_2 and monotonicity of the allocation would be violated, because if j_1 reported an earlier deadline (i.e., only being available at t_1) he would always be allocated, whereas if he reports his availability throughout both t_1 and t_2 truthfully, then there is a chance of non-allocation. The mechanism will* precommit *to allocate j_1 at t_2, irrespective of the realisation of future supply or future arrivals. In these cases, with small probability, the allocation may be inefficient or the mechanism could make a loss, but, more importantly, monotonicity is guaranteed and incentive compatibility ensured.*

As discussed in Stein et al. (2012), in order to guarantee monotonicity, an additional *serialization* constraint needs to be imposed on the allocation in the pre-commitment phase. Jobs are first ordered by a monotonicity-respecting criteria, and the pre-commitment decision is taken by considering jobs sequentially, following this order. Specifically, possible orders that ensure monotonicity in this setting include: decreasing value, increasing length, increasing arrival time (i.e., earlier jobs first), decreasing deadline (i.e., later departures get priority), an increasing rate, as well as combinations of these, such as value density ($\frac{v}{r \cdot l} = \frac{v}{q}$). Tie-breaking rules must also use criteria that guarantee monotonicity.

Essentially, the mechanism considers each job, taken in this order, and considers whether it can fit in a schedule, i.e., increase social welfare of the resulting schedule, or not, given the previously pre-committed jobs. The procedure is formally defined in Algorithm 5: We use decreasing value density, and in case of ties, give preference to earlier arrival time (FCFS). Each unscheduled active job is pre-committed in this order, if the sum of scenarios in which the job is scheduled, is greater or equal to $\frac{\mathcal{N}}{2}$. The scheduling algorithm used is an adaption of the offline scheduler called Greedy-Offline-PC (Algorithm 4). This algorithm guarantees that all pre-committed jobs (denoted by argument P) are scheduled.

Theorem 1. *The allocation procedure defined in Algorithm 4 and 5 is monotonic, given an assumption of "no early arrivals or late departure"' misreports.*[7]

The proof of monotonicity by Stein et al. (2012) applies, with some modifications. Informally, the proof considers each dimension making up a job's type, and shows that a job's allocation is not worse than under another type which is identical in all dimensions, but is strictly worse in that dimension. So, for example, the allocation of a job with a given value, required amount of electricity and arrival time, but with a later deadline cannot be worse than the allocation of a job with exactly the same parameters, but reporting an earlier deadline.

5.4.2 Payments

Critical value payments are used to ensure truthfulness of the jobs. The critical value of a job in an online mechanism is the minimum value necessary for pre-commitment given the set of jobs active over the active period of the respective job j (Parkes, 2007, p.418). For a job j with value v_j which is pre-committed, its payment $p(j)$ is thus defined as follows.

$$p(j) = \min\{v_{j'} \mid j' \in \text{ schedule } s'\},$$

where s' is the schedule produced by the same algorithm in case j is replaced by j' (with value $v_{j'}$).[8] The payment thus is not just based on demand and supply at the

---

[7]This is a standard assumption in online mechanism design, cf. Friedman and Parkes (2003).

[8]Note that truthfulness (and individual rationality) entails $p(j) \leq v_j$.

Algorithm 4: GREEDY-OFFLINE-PC heuristically schedules *all* pre-committed jobs (P) and *value-adding* unscheduled future jobs (J').

---

1 Algorithm: GREEDY-OFFLINE-PC (P, J, x, s, t)

2 for $j \in$ SORT(P) // by decreasing job size, i.e., l*r

3 do

4 $t_{min}, c_{min} \leftarrow$ COSTMINIMALSTARTTIME(j, x, s, t)

5 $s_{t_{min}} \leftarrow s_{t_{min}} \cup \{j\}$

6 $J' \leftarrow \{j \in J \mid j \notin s, d_j \geq t + l_j\}$

7 for $j \in$ SORT(J')// by decreasing value density

8 do

9 $t_{min}, c_{min} \leftarrow$ COSTMINIMALSTARTTIME(j, x, s, t)

10   **if** $c_{min} < v_j$ **then**

11 $s_{t_{min}} \leftarrow s_{t_{min}} \cup \{j\}$

12 return s

---

moment of pre-commitment, but also at later times (until its latest starting time). This is required, since otherwise a job could report a later arrival time and reduce its payment. A consequence of this approach is that under specific conditions there is a chance that the received payments are not sufficient to cover the cost of conventional generation (e.g., in case of insufficient renewable supply, despite an optimistic forecast) incurred by the mechanism. The lack of budget balance can be mitigated through the use of reservation prices. Then, however, low valued jobs that could be served from renewable generation are not accepted, and, correspondingly, efficiency suffers. This dilemma cannot be remedied without harming the efficiency of the schedule.

5.4.3 Allocation

As long as all pre-committed jobs are scheduled, we are free to use any algorithm in the allocation phase, including the algorithms described in Section 5.3. To force scheduling of all jobs in P in Algorithm 6 , we replace the offline scheduler by GREEDY-OFFLINE-PC, and only select jobs from P to schedule at t.

While pre-commitment establishes constraints that render the mechanism IC, its introduction does not necessarily lead to worse results. This is illustrated in the following example.

Example 4. *We consider a situation without flexibility and compare the choices made by the mechanism to those by multi-machine* consensus *(the scheduler, Algorithm 3). Suppose $t = 0$, the following three active jobs have the current time as arrival time, a consumption rate of 1, and the following lengths l_j and values v_j. (The deadline then is exactly equal to the length l_j.)*

Algorithm 5: m-CONSENSUS-Precommitment pre-commits jobs that received votes of at least half of the scenarios.

---

1 **Algorithm:** m-CONSENSUS-Precommitment(J_r, x, s, t)

2 **for** $j \in$ SORT(J_r) **do**

3     $f \leftarrow 0$

4   **foreach** *scenario* $\langle J^i(t), x^i(t) \rangle$ **do**

5        $s' \leftarrow$ GREEDY-OFFLINE-PC($P, J^i(t) \cup J_r, x^i(t), s, t$)

6   **if** $j \in s'$ **then**

7           $f \leftarrow f + 1$

8   **if** $f \geq \frac{\mathcal{N}}{2}$ **then**

9        $P \leftarrow P \cup \{j\}$

10 **return** P

---

jobs						scenarios			
j	v_j	l_j	v_j/l_j	a_j		i	$t=0$	$t=1$	$t=2$
1	9	3	3	0		1	2	2	0
2	1	1	1	0		2	2	2	0
3	3	2	3/2	0		3	2	2	1
4	4	2	2	1		4	2	2	2

Suppose there are four scenarios, which all include the current supply of 2, and in some cases slightly different future supplies in $t = 2$. Additionally, each scenario includes a virtual job ($j = 4$) of value 4 and length 2, to be expected at $t = 1$. We assume the cost of conventional generation to be 10 per unit of power flow per period.

The decision for m-CONSENSUS is made by repeatedly scheduling all jobs in all scenarios (with a greedy heuristic, based on value density), and in each iteration starting the active job that occurs in most schedules. In this example, $j = 1$ is scheduled in scenarios 3 and 4, $j = 2$ is scheduled in scenario 1,2, and 4 (in the latter because the virtual job can then be included), and $j = 3$ is scheduled in scenario 1,2, and 3. Therefore jobs 2 and 3 are scheduled by m-CONSENSUS, for a total value of 4 and an expected value of 6 (there is a 50% chance that virtual job 4 can be executed).

The decision by m-CONSENSUS-Precommitment is done per job, heuristically ordered by value density. A job is pre-committed if at least half of the scenarios would schedule it. Job 1 meets this criterion and thus is committed first; job 3 then follows. This schedule has a value of 12, but a 50% chance of a cost of 10, which gives it an expected value of 7. This exceeds the value of the m-CONSENSUS schedule. Similar examples can be constructed as long as conventional generation is more expensive than any value density.

Algorithm 6: Schedule the pre-committed jobs P at t that occur in the largest number of scenarios, under the condition that they all are eventually allocated.

---

1 **Algorithm:** m-CONSENSUS-Allocation(P, x, s, t)

2 **repeat**

3   Reset counters f (with $-\epsilon$ for $f(\bot)$)

4   **foreach** *scenario* $\langle J^i(t), x^i(t) \rangle$ **do**

5 $s' \leftarrow$ GREEDY-OFFLINE-PC $(P, J^i(t), x^i(t), s, t)$

6   **if** $s'_t = s_t$ **then**

7 $f(\bot) \leftarrow f(\bot) + 1$

8   **else**

9   **for** $j \in P \cap s'_t$ **do**

10 $f(j) \leftarrow f(j) + 1$

11 $j^* \leftarrow \arg\max_{j \in P} f(j)$

12   **if** $j^* \neq \bot$ **then**

13 $P \leftarrow P \setminus \{j^*\}$

14 $s_t \leftarrow s_t \cup \{j^*\}$

15 **until** $j^* = \bot$

16 **return** s

5.5 Evaluation

This section presents the empirical evaluation of the algorithms. Our goal lays in quantification of the algorithms' and mechanisms' economic performance relative to a clairvoyant , and therefore optimal, scheduler.[9] The objective function is given in Equation 5.1. To this end, we vary

- the number of scenarios,

- the scenarios' horizon, and, most importantly,

- job flexibility.

Furthermore, by comparing algorithms and mechanisms, we quantify efficiency losses due to ensuring truthfulness via pre-commitment. The gap between scheduling algorithm and mechanism can be interpreted as an expected price of incentive compatibility.[10] Finally, we verify the computational complexity of the algorithms.

---

[9]We use Gurobi 5.6 to compute the offline-optimal benchmark with a 1% MIP gap.

[10] Due to the disadvantageous computational complexity of the presented Expectation-based algorithms, we abstain from evaluating them empirically and focus on the Consensus-based mechanism. We argue that for realistic problem sizes, Expectation-based approaches cannot be applied in the online setting.

5.5.1 Experimental Setup

We consider settings with $1, 5$ and 30 scenarios and limit the available uncertain information about the future by using horizons of $0, 1, 2$ and 4 periods for the demand side and, as explained in Section 5.3.2, correspondingly longer horizons on the supply side. Based on this setup, we are able to illustrate the effect of varying forecast horizon length and number of scenarios on the economic outcome. We use the models below to generate both the scenarios as well as the actual realisation independently.

Demand As introduced in Section 5.2.1, jobs $j \in J$ are characterised by $\langle v_j, r_j, l_j, a_j, d_j \rangle$. For simplicity, the consumption rate is set to $r_j = 1, \forall j \in J$ in the experiments. Job length l is sampled from a uniform distribution over $\{\underline{l}, \dots, \overline{l}\}$ with $\underline{l} = 1, \overline{l} = 4$. Job valuations v are drawn from a uniform distribution over the real interval $[\underline{v}, \overline{v}]$ with $\underline{v} = 1, \overline{v} = 10$.

The set of jobs in each instance (and the instances' corresponding scenarios) is created such as to closely follow a deterministic load curve over a time span of 24 periods as it is typically observed in electric power systems. Note that we use integer valued quantities demanded (eq. 5.15) by rounding a sinusoidal load profile to the closest integer value in each time step.

$$D(t) = \left\lfloor 5.5 + 3\sin(\frac{2\pi}{24} t) \right\rfloor \tag{5.15}$$

We first create a set of jobs including their valuation v and length l. Thereafter, iterating over all jobs in order of decreasing job size, we set their start dates such that aggregate consumption closely follows $D(t)$. So far, we have not considered the central control paramter job flexibility, which is defined as $f_j = d_j - a_j - l_j$. We set all jobs to be equally flexible, and vary flexibility between 0 (no flexibility case) and 5 time steps.

Renewable Supply To realistically model uncertain supply of renewable energy, we use publicly available historical wind data from the Sotavento wind farm in Galicia, Spain.[11] This wind farm consists of 24 turbines, with a combined output of up to 17.56MW. However, in order to scale the available supply in our experiments, we model only the wind speed and derive the corresponding power supply using a sigmoid power curve that is based on the installed turbine technology (Robu et al., 2012):

$$p_r(w_t) = C \cdot (1 + e^{6 - \frac{2}{3} w_t})^{-1} \tag{5.16}$$

---

[11]This data is available from www.sotaventogalicia.com, and we use hourly data from May 2008 to 2013.

where $p_r(w_t)$ is the available power from wind generation given the wind speed w_t at time t. Here, C is a factor that we use to scale supply and that corresponds to capacity of the installed wind generators. Specifically, in each experiment installed capacity C is scaled such that total supply from wind equals the total amount of energy demanded (Subramanian et al., 2012), i.e., $\sum_t p_{r,t} = \sum_{j \in J} q_j$. We scale supply to match aggregate demand, because the problem instances are most interesting, i.e., challenging to the algorithms, when there is almost sufficient renewable supply for all jobs. We use the wind data in two ways — to generate the actual supply available during that run, and second, to train a generative probabilistic model entering our scheduling mechanisms to generate new scenarios.[12] In more detail, we use a hidden Markov model (Juang and Rabiner, 1991) with ten hidden states as our generative model, as this yields good results in practice on the wind data.

Conventional Generation Conventional Generation (CG) as a source of reliable backup generation is necessary in order for the non-preemptive and pre-committed jobs to be served even in the case of an unanticipated shortfall of renewable generation. We employ the simplifying assumption that CG is characterised by constant marginal cost, i.e., $c_c(p_c) = b \cdot p_c$, and we set b to a value approximately 30% above average job value density. With this cost parameter, low-valued jobs should not be served if there is insufficient renewable generation. On the other hand, if there is some, but insufficient renewable generation to serve a job from renewable generation *fully*, social welfare benefits if the remaining part is served from CG instead of rejecting the job. As job flexibility increases, the amount of CG used by the algorithms should decrease.

5.5.2 Results

Our main experimental results are illustrated in Figure 5.1. We vary job flexibility between zero and five time steps (hours) and normalize social welfare by the result of offline-optimal under zero job flexibility. The reported results are the mean normalized social welfare and corresponding standard errors. We separate our results of 100 repetitions by the number of scenarios (columns) and horizon length. The solid line represents the mechanism's results. Our gold-standard, offline-optimal, is depicted in green. Figure 5.1 illustrates the following: In the presence of five hours of flexibility for all jobs, offline-optimal social welfare increases by approximately 21% compared to the zero flexibility case. However, without information on future demand (i.e., zero horizon length, top row), neither the schedulers nor the mechanism can take advantage of this flexibility welfare potential. If there is only a single scenario of future sup-

---

[12]The probabilistic model could also be leveraged to iteratively update scenario weights based on scenarios' likelihood as new information becomes available. In practice, however, with resampling forbidden, we weight scenarios uniformly.

Dimension		values
Job flexibility	f	[0,..,5]
Job valuation	v	[1,10]
Consumption rate	r	1
Job length	l	{1,2,3,4}
Number of experiments		100
Number of scenarios	$\mathcal{N}$	{1,5,30}
Scenario horizon	h	{0,1,2,4}
Specific cost of conventional generation	c	$1.3 \cdot \frac{\bar{v}}{\bar{q}}$

Table 5.1: Evaluation Parameters for Multi-unit Demand

ply available (which might not correspond to the actual realization), performance of the first-come first-serve scheduler even decreases over flexibility on the demand side (top left panel). Interestingly, increasing the number of scenarios (left to right), while retaining the zero horizon length on the demand side, does not improve the economic outcome.

However, adding even comparably short forecasting horizons (top to bottom) of one or two time steps (i.e., hours), the online schedulers can effectively harness demand side flexibility. Thereby, in case of a single scenario, the schedulers perform better than the mechanism. Both, the algorithms' and the mechanism's performance can be improved by increasing the number of scenarios. The mechanism's performance, although constrained by the incentive compatibility property, is on-par, and sometimes even better than the FCFS-based scheduling algorithm. This underlines the value of more advanced algorithms (Hentenryck and Bent, 2009).

To formulate the main results:

- First, social welfare initially increases with flexibility for all algorithms, given even a very short forecasting horizon. Furthermore, the mechanism achieves only slightly reduced welfare compared to the extended algorithms. Most notably, the average gap in social welfare, i.e., the cost of incentive compatibility in our settings does not exceed 4–5%. At flexibility equal to five periods, the online scheduler achieves results 15%-20% short of the offline optimal benchmark, depending on the number of scenarios and horizon. The corresponding mechanism achieves slightly lower values, with an additional reduction in social welfare in the range of 1%-5%. However, and importantly, these scheduling algorithms assume non-strategic, cooperative agents/jobs. An in-depth analysis of the results reveals that median performance of the mechanism is again nearly on par with the value-density based scheduler, and mostly better than the FCFS scheduler (Fig. 5.2). Evidently, average performance is skewed by outliers on the low end of the welfare spectrum. Notably, these outliers have an especially

strong presence for the mechanism, skewing average social welfare as the main performance metric.

- Second, a longer forecasting horizon (top to bottom, Fig. 5.1) can only very marginally improve economic performance of both, schedulers and mechanism. For the mechanism, maximum social welfare is achieved at horizons of one or two periods and remains constant beyond. Possibly, longer horizons provide stronger competition for currently active jobs, which may then end up without allocation. The forecasted jobs that never arrive, effectively pave the way for poor pre-commitment decisions. Regarding the number of scenarios, FCFS-based scheduling by design cannot benefit from a larger number of scenarios, while the mechanism and value-density based scheduler benefit significantly.

- Our results in Figure 5.3 show that computation time increases sub-linearly in job flexibility and linearly in the number of scenarios. The sub-linear increase of computation time in job flexibility can be attributed to decreasing marginal use of flexibility: The decision whether to include a more flexible job in the schedule is not repeated for all possible start dates, but only as long as the job is not included in a past schedule. Besides this interesting and useful result, the value-density based scheduling algorithm is the least expensive algorithm: For a horizon of 1 period, relying on 30 scenarios and assuming flexibility of five periods, construction of a schedule over 24 periods assuming the inputs from Tab. 5.1 takes 23 seconds with our implementation, i.e., ~ 1 second per period. The FCFS-based algorithm clocks in at slightly higher 25 seconds. The mechanism is more involved at about 48 seconds on average. The larger computational effort for the mechanism can be explained by the two-phase decision making, which approximately doubles runtime. Nevertheless, runtime in each time step is approximately 2 seconds, and thus acceptable for our online setting.

5.6 Conclusion

We extend the *Expectation* and *Consensus* algorithms to cope with multi-unit, non-preemptive demand under uncertainty on both, supply and demand side. Specifically, to deal with uncertain supply, we employ scenarios sampled from the underlying distribution. Furthermore, in order to apply the principles of these algorithms to settings with self-interested agents, we apply the pre-commitment approach to achieve monotonicity and thus incentive compatibility for the demand side. Finally, non-preemptive jobs and pre-commitments can only be scheduled if there are guarantees in terms of the availability of future supply. To deal with this problem, we consider two supply sources: low cost/free, but uncertain renewables, and costly but unlimited
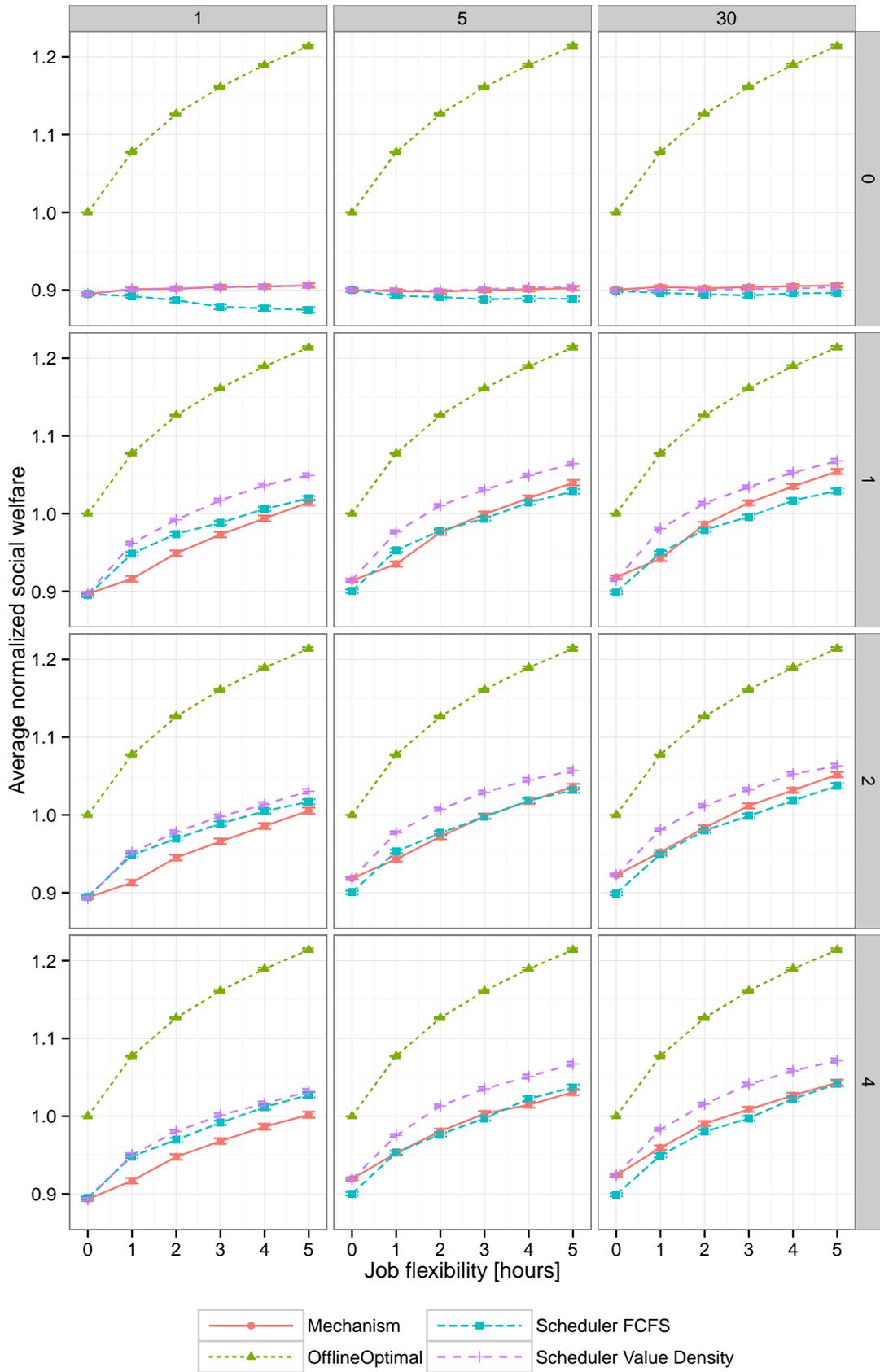
Figure 5.1: Average normalized social welfare and corresponding standard errors. Normalizing factor is the offline optimal given zero flexibility. Columns represent number of scenarios, rows indicate forecast horizon.
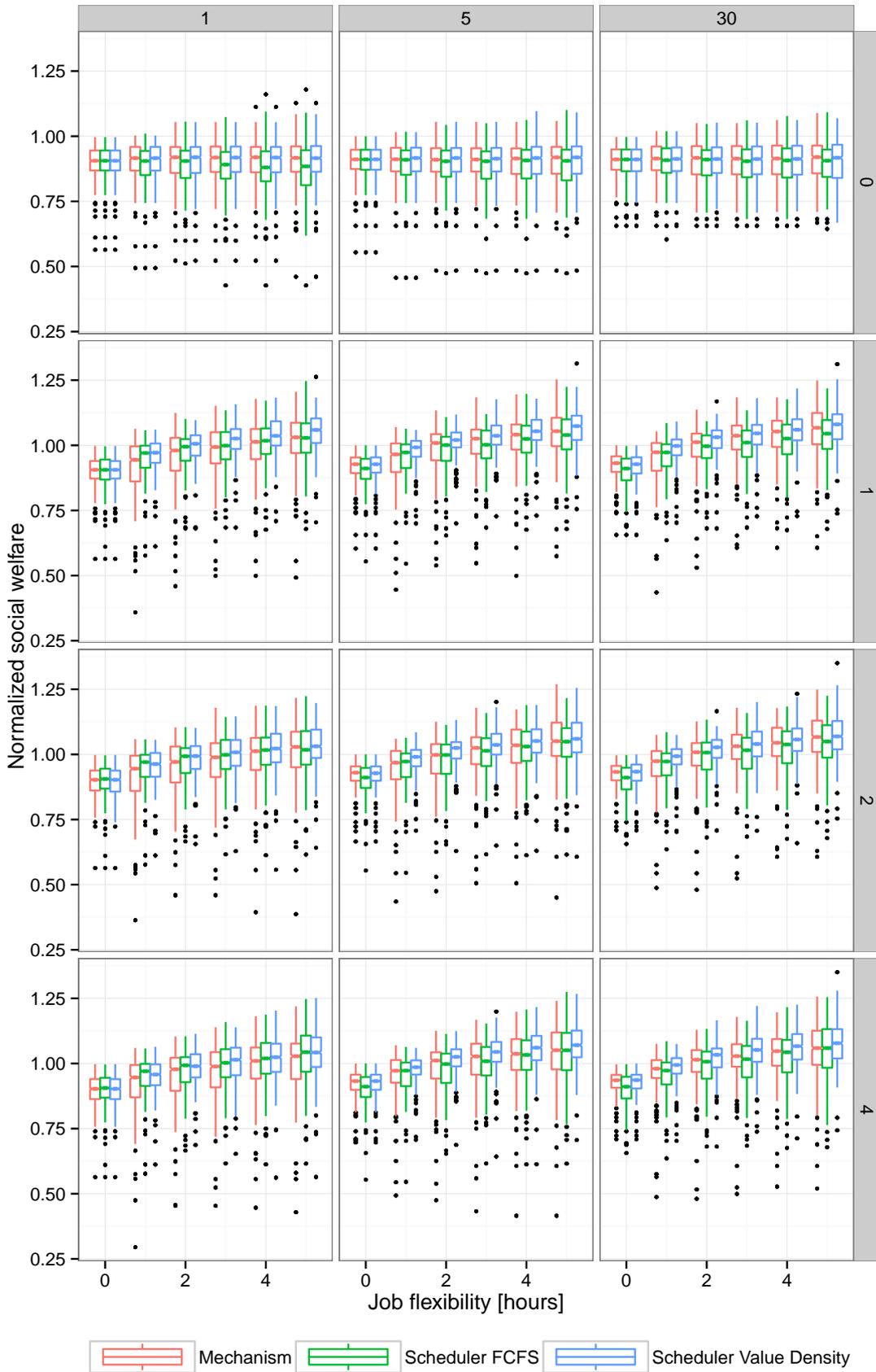
Figure 5.2: Normalized social welfare. Normalizing factor is the offline optimal given zero flexibility. Columns represent number of scenarios, rows indicate forecast horizon.
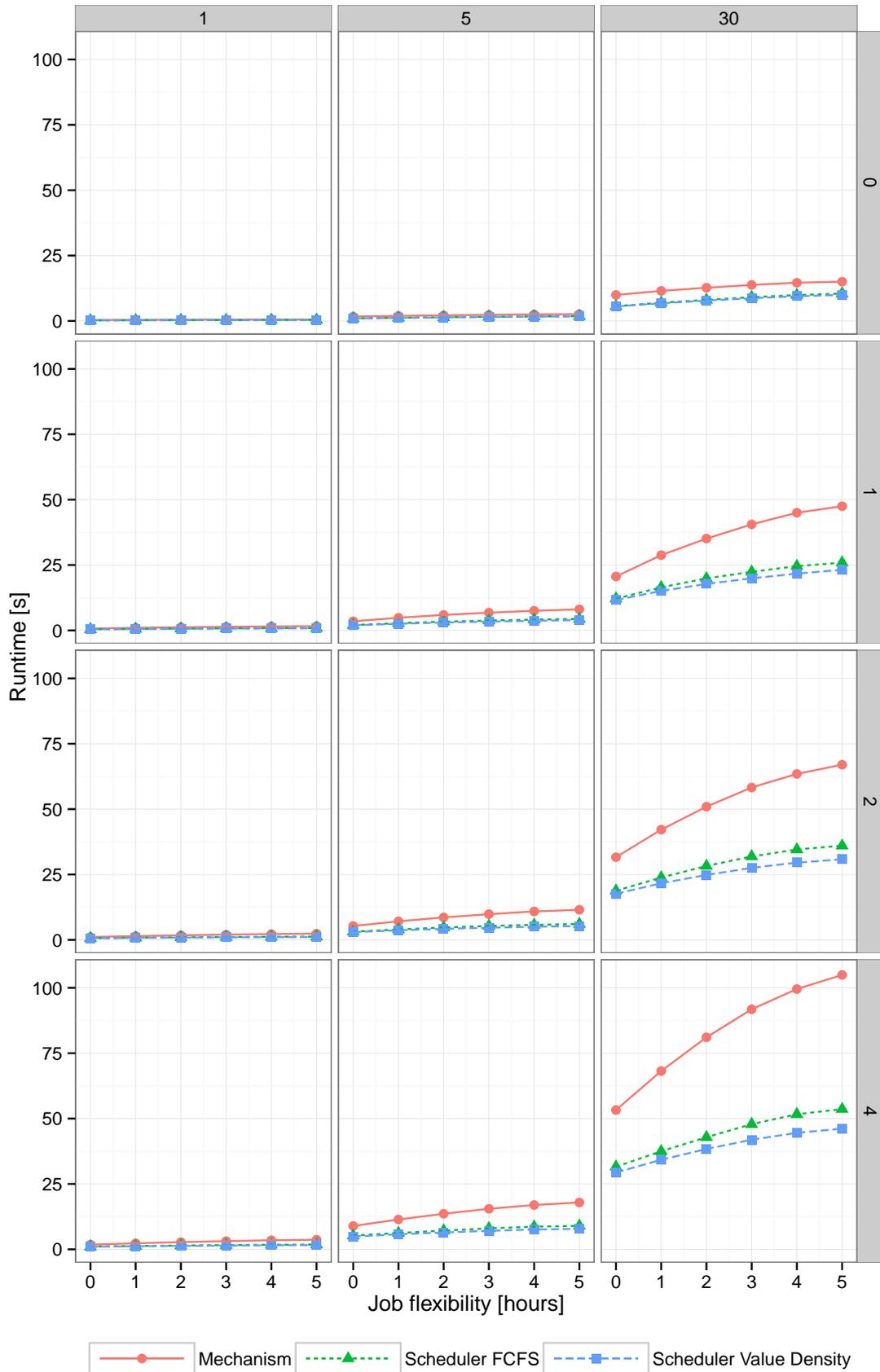
Figure 5.3: Runtime over flexibility. Columns represent number of scenarios, rows indicate forecast horizon.

conventional energy. This way jobs can be committed, even if supply is uncertain. The algorithms are designed to take into account the risk of using the costly alternative.

Our empirical evaluation shows a number of interesting results. First, as expected, social welfare increases in job flexibility. Second, model-free online schedulers and mechanism are easily improved using a model of the future involving rather moderate numbers of scenarios (to capture uncertainty) and comparably short horizons. This is an encouraging result, as the algorithms lend themselves to application in online settings with tight time constraints. Third, the cost of achieving truthfulness (i.e., by requiring pre-commitment) is very low and only approximately 4-5%. Economic performance of both, schedulers and mechanism could potentially be further improved using *expectation*-based approaches. The corresponding economic gain, however, would come at additional computational cost. We note that our evaluation of the mechanism is placed in the context of the electrical power domain. However, it could be applied to other settings where jobs are non-preemptive and there is a source of free (or cheap), expiring resources and costly backup supply.

In future work we intend to explore the trade-off between economic efficiency and budget deficits for the mechanism. Using the currently proposed mechanism, the system may run a loss when the payments it receives are low (e.g., due to lack of competition on the demand side), while incurring more than expected costs on the supply side. By under-committing, the mechanism could achieve budget balance (in expectation) at the cost of reduced efficiency as less of the free resource might be used.

Part III

Future Mobility Systems: Leveraging Consumer Flexibility

Chapter 6

Car-sharing Fundamentals

This chapter outlines the role of car-sharing in future mobility systems where the significance of inter-modal trips can be expected to increase. It also describes the different kinds of car-sharing that exist, presents an introduction into related literature including existing work in the car-sharing domain, but also covering related sharing concepts. Thereafter, we outline the various kinds of flexibility that might be valuable in the operations of car-sharing fleets and, finally, describe the data set that we base our evaluation on in the following chapters.

6.1 Future Mobility Systems

Multimodal Mobility   We employ the term *multimodal mobility* to describe passenger trips that make use of at least two modes of transport, such as train and car, or car and public transport in general. On a side note, the related term *intermodal transport* is used to describe the use of at least two different means of transport (e.g., truck and train) to transport goods (not passengers) from origin to destination. The combination of different modes of transport to satisfy travel demand is widely expected to gain in importance in future mobility systems. With mobility-related information becoming increasingly accessible, complex mobility chains and complex mobility services can be designed and offered. Such information access allows for the composition of both, ad-hoc and complex trips, potentially consisting of multiple legs and using different modes of transportation (multimodal transport). Car-sharing may be one of these modes, thereby to some extent relying on electric vehicles. The efficient integration of EVs, however, requires elaborate information and communication infrastructure beyond what would be necessary for Internal Combustion Engine (ICE) propelled vehicles.

Clearly, the sole availability of information does not render composition of complex trip patterns a self-fulfilling prophecy. Users' desire to select the best, or at least an appropriate combination of trip components from a plethora of options creates massive cognitive overhead. This complexity might create demand for Decision Support Systems (DSSs), as exemplified in Daimler's smart phone application "moovel". Internet-enabled DSSs might reach beyond the smart visualization of all available al-

ternatives, ordered, for example, greedily by one dimension. Instead, individual user's preferences could be taken into account when presenting information to the eventual decision maker, relieving him of unwanted clutter and thus assist in turning the vision of seamless multi-modal mobility into reality.

In shared mobility systems, economic coordination is of paramount importance to achieve acceptable economic outcomes. Except for the pathological case of excess transportation capacity, scarce capacity will have to be managed through efficient allocation decisions. In contrast to today's individual mobility sector, where decisions are made on longer time-scales, decisions granting access to mobility will be made more frequently but for shorter time horizons, adding complexity to decision making on the one hand, but also opening up new possibilities and associated economic potential.

Fig. 6.1 illustrates a simplified decision-making process where three alternatives $\{a_1, a_2, a_3\}$ are available. In the mobility domain, ownership decisions regarding a personal car are made once every couple of years, i.e., the user chooses one alternative (in red) and sticks with this far-reaching decision for a prolonged period. In the future, this might change to more frequent, short-term decisions, i.e., for each trip the appropriate mobility means is chosen. Hence, a larger number of small decisions might replace one single, important decision. With respect to Fig. 6.1 this implies that the time between decisions is reduced in a shared-mobility scenario. The individual user benefits from more appropriate offers, as the vehicle chosen for each trip is more appropriate than a vehicle that is required to satisfy a wider range of mobility requests ; on the other hand, he bears the additional cognitive overhead when making repeated decisions and might, depending on the design of the car-sharing system, be faced with the risk of lacking service. Interestingly, car-sharing also provides individual mobility to those that do not own a personal vehicle. For those users, car-sharing might be a welcome and highly valuable addition to their personal mobility portfolio and enable trips that otherwise would have been impossible to conduct.

Choice Between Alternative Modes of Transport   Choosing between modes of transport is a classic example of discrete choice models. Consumers select from a finite set of alternatives (e.g., bus, train, taxi, car) the one alternative that maximizes their individual utility. Mode choices were the motivating problem for the eventual formation of discrete choice models based on random utility models These models were developed by McFadden and colleagues in the 1960s and 70s who received the Nobel Price in economics in 2000 for his contribution (McFadden, 2001). A seminal contribution to the development and application of discrete choice models was in forecasting the use of transportation alternatives in the period surrounding the introduction of Bay Area Transportation in the 1970s in California. Assuming utility-maximizing agents, observed user choices allow derivation of the values of different attributes of the choice process, e.g., the value of time, the value of spending time alone in a car vs.
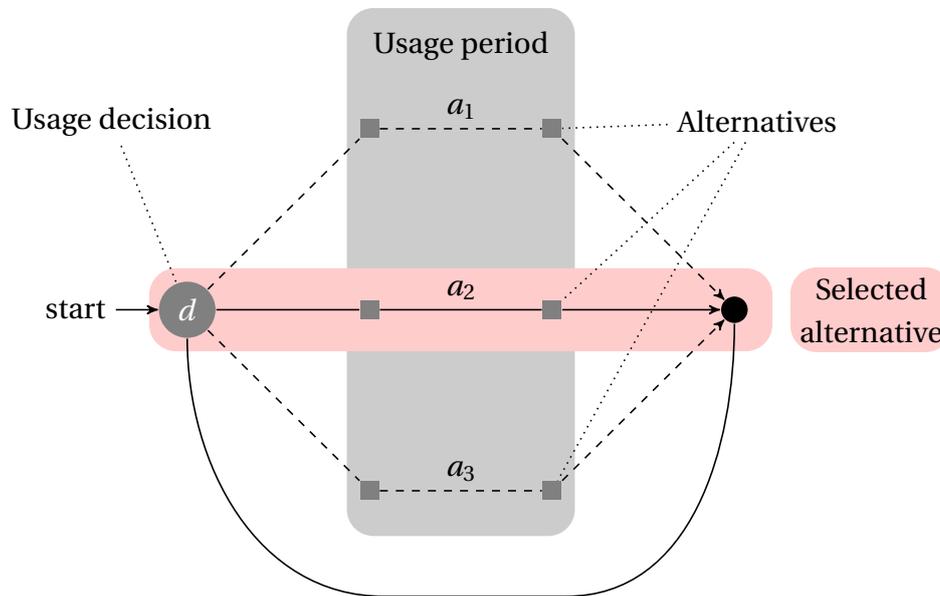
Figure 6.1: Alternatives and Decisions in Mobility Choices.

spending it in means of public transport, number of transfers, etc.

Historically, most trips in developed countries are conducted using a single mode of transport, most often using the automobile (Streit et al., 2013). While we do not expect this structure to change massively in the foreseeable future, car-sharing can function as an enabler of richer multimodal travel chains. This is especially true for the case of one-way car-sharing (e.g., car2go or DriveNow, to name currently popular choices). Accordingly, the availability of car-sharing offers might enable trips that otherwise would either not be undertaken or be undertaken via different modes. Ceteris paribus, adding one more accessible alternative to the menu of choices increases users' attainable utility. Clearly, the more accessible such novel mobility offers are relative to established means, the more likely users will choose them from their menu of mobility options.

We conjecture that the most important consequence of the introduction of car-sharing might be that families must no longer ensure that even the last bit of peak demand can be covered through their owned vehicle. Instead, one future possible development might be that only the base demand for individual mobility is covered via an owned vehicle. The remaining demand for individual mobility can then be covered through shared systems, of which station-based car-sharing might assume an important role.

Choice in Station-based Car-sharing   Besides choice concerning the mode of transportation, the users must make further decisions, if car-sharing is chosen to be a com-

ponent of his travel plans. In particular, he must choose from the set of vehicles and locations, and potentially adapt his desired reservation duration with the time the vehicle is available (at least in the case of station-based car-sharing with advance reservation). Thereby, the utility-maximizing alternative might not be available, hence, the customer must next-best options. Presumably, in some instances, the loss in utility from lacking availability might be negligible, and the consumers are virtually indifferent between alternatives. These are the cases, in which consumers can offer flexibility to the operator of the car-sharing fleet, without sacrificing their own utility. Furthermore, if the consumer is adverse to making detailed decisions, he might even welcome the recommendation of a vehicle that is close-enough to the actual utility-maximizing assisted by the reservation system. In some cases, trading-off complexity with slightly worse service can actually turn out to be beneficial for the user, especially if the presumed allocation decisions is bound to monetary incentives. In situations, where there are numerous alternatives available that are nearly on-par with each other, i.e., in dense station networks, the basic idea of hidden markets, as brought forward in (Seuken, 2010) poses an interesting avenue for further research.

Following the theory on random utility models we posit that consumers' revealed preference in the form of time, station and vehicle choice might be the best available option. This however, does not imply that consumers would not be willing to choose different alternatives, given the originally chosen ones were not available.

Reservation Process and Efficient Operations   Today's car-sharing reservation process typically is fully managed by the customer. For example, in typical station-based car-sharing operations, the customer decides, based on a highly transparent system state, exactly which vehicle to reserve for a clearly defined time span.[1]

The implicit assumption underlying human-computer interface design in car-sharing as in Fig. 6.2 seems to stem from the perception that car-sharing customers can indeed easily translate their preferences into well defined and strong choices (regarding type of vehicle, time, duration, and location). We argue that such reservation system design can have detrimental effects with respect to the efficiency of car-sharing operations, that could be avoided through better system design.

Fig. 6.3 illustrates the typical reservation process in station-based car-sharing. First, the user inserts the reservation from the set of available alternatives (i). Each reservation has a begin and end date (b and e). The period during which the vehicle is actually in use lasts from pick-up at b' to drop-off at time e'.

In addition to this standard process, users can modify their reservations during the course of the reservation, i.e., insert an earlier return date, or, if the vehicle is available has not been reserved by another user, a later return date as well. Due to penalties regarding late returns, users typically reserve the vehicle for a longer period than the

---

[1]In one-way rental systems, the user usually only selects a vehicle and drops it off at the desired location, *without* specifying the return date in advance.
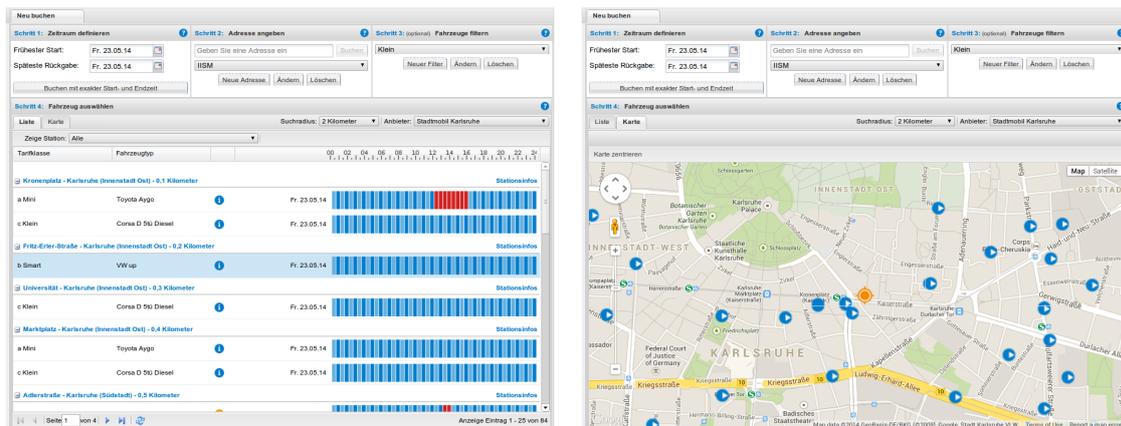
Figure 6.2: Typical customer reservation interface in station-based car-sharing. Customers can choose from the temporal view on the system (default) and a spatial view that reveals further information after clicking on the respective station.
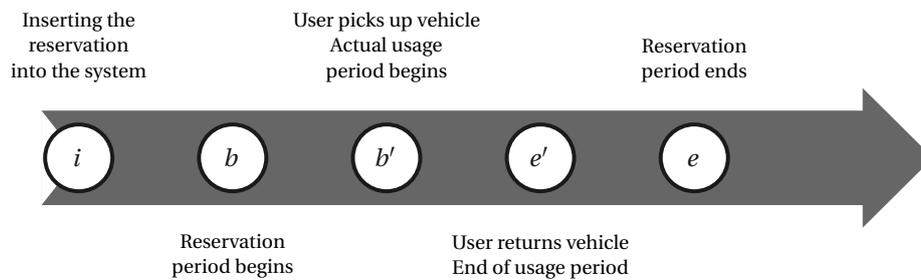
Figure 6.3: Reservation Process in station-based car-sharing.

actual usage period $b' - e'$. Hence, $e - b < e' - b'$ holds for most reservations.

Each individual reservation can have negative externalities on other users, mostly through reduced vehicle availability. For example, by placing a reservation on a particular vehicle, other users might no longer be able to fully serve their demand for mobility. Beyond crowding out other users, selfish users might place reservations in the system such that the time between reservations on the same vehicle is too short to accommodate an additional reservation. The user placing the reservation, however, can be nearly sure, not to be affected from potential late returns. System efficiency can be adversely affected by such reservation behavior.

Following these thoughts, the consequences of naively implemented transparency with respect to system state, as in today's reservation systems, are two-fold: First, it might prevent potential customers from participating (fearing the risk of low QoS) or, second, participating customers are forced to use other, less favorable modes of transport in order to satisfy their travel demands (due to a lack of vehicle availability). Consequently, the described reservation approach can lead to inefficient outcomes, with potential customers not joining the system and some customers being denied ser-

vice due to inefficient scheduling from distributed, self-interested customers. Clearly, there is potential in drafting alternative reservation schemes, however, they must be thoroughly evaluated before being rolled-out into the wild.

6.2 Shared Mobility Literature

The relevant related research contributions on car-sharing operations stem from both the literature on shared mobility systems as well as general insights from the operations management literature, especially on scheduling, fleet assignment as well as revenue and capacity management.

As noted before, there has been a recent and current trend towards an increasingly shared usage of mobility options. Consequently, the number of scholarly publications addressing strategic and operational challenges of such systems have recently become more numerous. Besides studies on car-sharing operations, there has been a great amount of research on bike sharing systems which face somewhat similar challenges. While not essentially considered a "sharing industry", car rental operations have in the past encountered and experienced a host of the challenges car-sharing operators are faced with as well. One particularly important problem in transportation research is the dial-a-ride problem, a special variant of the pickup and delivery problem (Savelsbergh and Sol, 1995). For future mobility systems, where autonomous (self-driving) vehicles could provide an important addition to the existing mobility spectrum, this problem and modifications thereof might receive renewed interest from both academia and industry. Clearly, shared mobility systems require economic coordination to manage and mitigate the adverse effects of scarcity on user acceptance. One point that has not been addressed in existing research on shared mobility systems concerns the design of incentives to truthfully reveal users' preferences. By basing economic allocation decisions on users' true preferences, efficiency can be improved. However, the trade-offs between economic benefits and privacy risks associated with the revelation of sensitive personal data must be addressed in design of the corresponding service designs; otherwise, user acceptance might suffer.

Car-sharing   Looking at typical car-sharing structures around the globe Barth and Shaheen (2002) establish a taxonomy for classifying such systems. System integration with public transit, the possibility of inter-nodal travel and the primary customer segment (residential vs. business) are identified as central aspects for classifying car sharing systems.

A large share of current car-sharing organizations do not facilitate inter-nodal travel (Shaheen and Cohen, 2007). Still, the optimization literature concerned with car-sharing has almost exclusively focused on mitigation strategies for the vehicle relocation challenge arising in systems with inter-nodal travel.

Kek et al. (2006) develop a simulation model to identify relocation techniques that can enhance service levels. Based on a data set from a local car-sharing company they demonstrate significant savings with respect to car lots, staffing levels and operational costs by using inventory balancing techniques. Kek et al. (2009) extend this work and develop a three-phase optimization-trend-simulation decision support system. This tool allows operators to determine a near-optimal manpower and operating parameters for the vehicle relocation problem. To the same end, Nair and Miller-Hooks (2011) propose a stochastic, mixed-integer program with joint chance constraints. This facilitates the generation of least-cost vehicle redistribution plans for shared-vehicle systems such that a proportion of all near-term demand scenarios are met. The potential of redistribution approaches is evaluated using a real-world application scenario in Singapore. Similarly, Correia and Antunes (2012) present an mixed-integer programming model to determine depot location in one-way car-sharing systems. The benefits of the approach are illustrated by means of a case study.

These research contributions focus on the operational side of car-sharing with real data sets being used to obtain representative booking streams for the model validation. Steininger et al. (1996) take a user-centered perspective and by means of a survey and a controlled experiment try to characterize car-sharing users and their adoption behavior. Morency et al. (2007) also aim at characterizing the behavior of car-sharing customers. Using data mining techniques on a transaction-level data set they are able to identify distinct customer clusters and day types. They argue that this type of analysis can help improve the efficiency of car-sharing operations.

Shaheen et al. (2012) conducted personal expert interviews to investigate the development of personal car-sharing, which they see as the next step in car-sharing. Relative to station-based car-sharing, personal car-sharing lowers the barriers to entry and usage, giving consumers greater choice in making mode-of-transportation decisions.

Bike-sharing   Another recent shared mobility trend is the establishment of bike-sharing systems in many cities around the globe. Given the lower cost of bicycles, these systems reach much larger scale than car-sharing operations. Typically, inter-nodal travel is the standard case in these systems and there has been a host of very recent papers on optimizing balancing and repositioning operations: Both Chemla et al. (2012) and Raviv et al. (2012) characterize the static bike repositioning problem and discuss different solution approaches. Raviv and Kolka (2013) introduce an inventory model to address the dynamic variant of this problem. They provide a numerical solution method as well as structural properties concerning the convexity of the model. Nair et al. (2013) present a quantitative analysis of a large-scale bicycle sharing system. They address several operational aspects (e.g., system characteristics, utilization patterns, and flow imbalances between stations) and present fleet-management strategies to deal with this asymmetry. Lin et al. (2011) provide a formal hub location inventory model formulation for the case of bike-sharing, where the design variables

reach as far as the creation of bicycle lanes between stations. They consider travel costs, bicycle inventory costs, facility costs (stations and lanes) as well as service levels in their objective.

Car Rental   Unlike car- and bike-sharing which can still be considered somewhat recent phenomena, car rental operations have been around for a fairly long time. Given this greater level of industry maturity, car rental operations is a well-explored research field. Many of the identified solution approaches could in principle be transferred to car sharing operations as well.

Focusing on a concrete decision support system implementation, Carroll and Grimes (1995) provide an overview of a rental car company yield management system which facilitates optimized control of availability and pricing for different car rental product combinations (car type, rental period, pickup and return location) over time. This helps to solve the closely-related problems of pricing, fleet planning, and fleet deployment. Savin et al. (2005) consider optimal rental capacity allocation policies. Using dynamic programming they determine properties of the optimal policy. Furthermore, they propose and evaluate a novel threshold heuristic. Finally, they consider the joint problem of fleet sizing and allocation. Similar to inter-nodal car-sharing systems, car rental companies also need to determine efficient vehicle relocation schemes. Fink and Reiners (2006) address this problem using a network flow optimization. They account for real-world constraints (e.g., country-wide network, partial substitutability across car types) and evaluate their approach using a real-world data set. Gans and Savin (2007) develop a stochastic control problem to address the optimal pricing and allocation approach when handling heterogeneous customer segments (fixed price contracts vs. walk-in demand).

Related problems in transportation research   One standard problem of transportation research is the Dial-a-Ride problem. Its objective is to "find vehicle routes and schedules for multiple users that specify pickup and drop-off requests between origins and destinations" (Cordeau and Laporte, 2007) while minimizing costs. The problem is due to its combinatorial characteristics computationally challenging and can be solved exactly for small problem instances only. Due to its importance in, for example, taxi and ambulance routing, a wide array of solution approaches has been developed and a large number of case studies on modifications of the original problem has been conducted (Jain and Hentenryck, 2011; Häll et al., 2009; Hanne et al., 2009).

Ride sharing, a research topic enjoying increased attention in the recent past (Kamar and Horvitz, 2009; Kleiner et al., 2011; Agatz et al., 2012; Coltin and Veloso, 2013) is also related to the Dial-a-ride problem. The major departure from the standard problem can be found in uncertain availability of drivers and different depot locations for each driver.

Electric mobility Electric mobility has recently received great attention, but its most prominent technology, Battery Electric Vehicle (BEV), suffers from limited range. Consequently, test users of BEVs have expressed range anxiety, the "concern, or even fear, of becoming stranded with a discharged battery in a limited-range vehicle, away from the electric infrastructure" (Eberle and von Helmolt, 2010). In the case of private ownership, the corresponding vehicle must be able to satisfy a wide spectrum of mobility requests of its user, e.g., ranging from the weekly short-distance shopping trip to the annual long-distance holiday trip. As a consequence, individuals are reluctant to forego the flexibility offered by conventionally fueled vehicles even if there are economic aspects that made the purchase and use of electric vehicles favorable (Steinhilber et al., 2013). Via separation of ownership and use, e.g., by means of car-sharing, Users can choose the most appropriate means to achieve their goals – but on a trip-by-trip basis. Hence, car-sharing might be able to play a significant role in fostering electrification of individual mobility in general and adoption of BEVs in particular. However, literature on electrified car-sharing so far is sparse. Some authors propose to provide a fleet of conventional vehicles for buyers of BEVs, allowing the owners to use this fleet for more extensive trips, instead of adding electric vehicles to the fleet (King et al., 2013). A study on the potential uptake of electric vehicles of 2011 provided disappointing results. However, these results (Doll et al., 2011) might be in large part due to poor assumptions and (poor) heuristics. In particular, the assignment of reservations to vehicles was following a pure FCFS approach, ignoring additionally available information. With the goal of the study in mind (better understanding of the economics of electrified car-sharing), the study's worst-case approach is difficult to comprehend.

6.3 Flexibility in Car-Sharing

Increasing popularity of various forms of car-sharing has paved the way to increasingly dense station networks of classic station-based car-sharing operators. A denser station network and an increasing number of available vehicles reduces the risk of remaining without service (i.e., without access to a vehicle) for the individual user, rendering car-sharing even more attractive. Traditionally, reservations requests are mostly inserted into the operator's system through the users themselves and beginning and end times are treated as hard constraints. However, consumers might be satisfied with slightly modified service relative to what they initially requested, but, in today's reservation system, are not able to express this flexibility approximately. In this work we focus on two types of consumer flexibility: Temporal and spatial flexibility. Each type can be used by the car-sharing operator to achieve more efficient outcomes, given consumers are willing to reveal their flexibility to the operator.

Temporal Flexibility   Users might have an earliest start date and latest end date for
their reservations that does not coincide with the reserved period, i.e., reservations
can be postponed or advanced in time. By shifting reservations in time within con-
sumers' flexibility limits as illustrated in the top right part of Fig. 6.4 , fleet size re-
ductions can be achieved while still serving all reservations. Instead of reducing fleet
capacity while still serving all reservations, temporal flexibility can be leveraged to in-
crease the number of served reservations while retaining original fleet capacity.
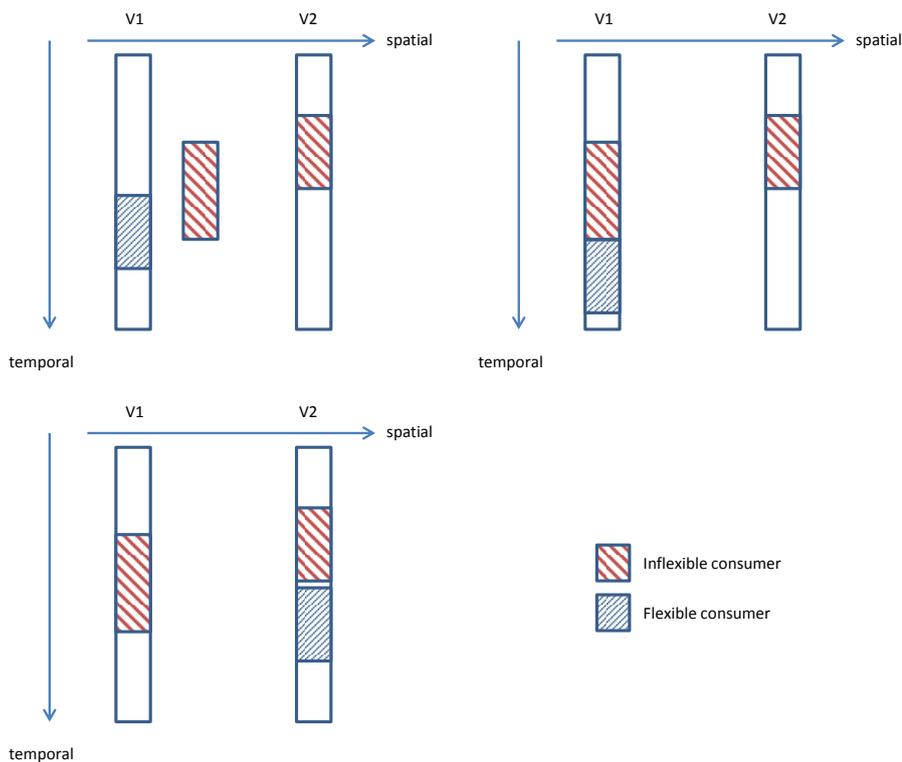
Figure 6.4: Spatial and temporal flexibility can be leveraged for more efficient fleet utilization

Spatial Flexibility   Consumers located in between two stations, might be rather in-
different, at which station their reservation is served. Hence, consumers might be
flexible regarding the location of their pick-up and drop-off station. This is especially
true as the station network available grows denser over time, reducing the associated
inconvenience from retrieving a vehicle at a neighboring station. Fig. 6.4 (bottom left)
illustrates in a minimal example how spatial flexibility can be leveraged to achieve
higher capacity utilization. By assigning a spatially flexible reservation to a vehicle

Data	Unit	Median	10% quantile	90% quantile
Reservations		62698		
Clients		7615		
Vehicles		206		
Distance per reservation	km	30	8	220
Reservation duration	h	4.50	1.77	16.5
Effective usage duration	h	3.63	1.08	14.82
Lead-time	h	3.42	0.16	76.53

Table 6.1: Car-sharing data set descriptives

located at a neighboring station, the newly arriving reservation can be served successfully.[2]

Vehicle Class Flexibility   Besides temporal and spatial flexibility, consumers might be flexible with respect to vehicle class. More specifically, a user might be satisfied with a smaller (larger) vehicle than originally reserved. For the provider in certain situations it might be beneficial to bump reservations to a higher vehicle class, (This might be troublesome with incentive compatibility.) instead of expanding the fleet in the lower vehicle class.

6.4 Data Selection

The car-sharing data set we employ in subsequent studies is parted into reservation stream and station data. The latter consists of the latitude/longitude information of each car-sharing station. The former, i.e., the reservation stream of 2012 contains approximately 63.000 reservations within city boundaries in the most important vehicle class "small", mostly comprised of Opel Corsas and Ford Fiestas. In more detail, each observation (reservation) comprises the following information: Anonymized user id, the date and time at which the reservation was inserted into the system, start of reservation, end of reservation, start of usage, end of usage, distance travelled, and station. Table 6.1 documents descriptive statistics of the reservation stream, comprising reservations of 7615 clients. Note that this data set includes reservations of the single vehicle class "Small".

Consumers can interact with the reservation system multiple times regarding a single reservation: Reservations can be inserted initially, modified, cancelled or ended before the end of the initially booked reservation period (case of early return). Un-

---

[2]In subsequent evaluations regarding the value of spatial flexibility, we use the Euclidean distance (see Section 7.2.4 for details).

fortunately, the data set available includes only the last interaction of the user with system and not the entire interaction history. As a consequence, the *insert* date is modified upon each interaction and loses its semantics. For example, in the case of early return, the *insert* date of a reservations will be later than the respective start date. Hence, computation of lead-times, the time between insertion and begin of a reservation, based on incorrect data yields insensible results.

Filters   We applied the following filters to the data: First, reservations beyond administrative city boundaries are discarded. This allows to focus on a smaller and more homogeneous data set, allowing for easier and more robust inference. Second, we focus on a single class of vehicles only, again to foster robustness and interpretability of results. Furthermore, we discard reservations with incomplete data. By focusing on a single vehicle class within city boundaries, approximately half of total reservation data is discarded. Further, approximately 10% of reservation data are removed due to incompleteness.

Duration and Distance of Reservations   Fig. 6.5 depicts the empirical distribution of reservation duration and travel distance. The mass of the distribution is in the lower left corner, indicating that smaller vehicles are mostly used for short periods and distances. Interestingly, there are numerous reservations where the use of traditional car rental offers would have been more economical to the user. We suspect that consumers prefer the convenience of pick-up and drop-off stations in proximity to their homes. Furthermore, they might be deriving value from knowing the vehicle type with certainty in advance[3]. Another explanation might be that the cost of spending time on picking up and dropping of vehicles at distant car-rental stations more than outweighs the additional cost from using the more convenient car-sharing offer.

Fig. 6.6 illustrates the distribution of reservation distances. Clearly, most reservations are rather short (also see Tab. 6.1), i.e., approximately 85% of reservations do not exceed a distance of 100 km. However, only about 35% of total distance driven is due to trips shorter than 100 km. The dashed line in Fig. 6.6 illustrates the share of total distance that could be served by a range-constrained vehicle. For example, to serve more than half of the distance traveled with a BEV, its range would have to exceed 170 km, assuming there were no other constraints.

Lead Time   The lead time with which customers are reserving vehicles is of special interest, as it reveal information how much ahead of time customers are reserving vehicles. Fig. 6.7 illustrates the overall distribution of reservation lead times.

Interestingly, approximately half of the reservations are booked with a lead time of four hours or *less*. Furthermore, in excess of 75% of reservations are inserted into the system within 24 hours before the beginning of the reserved period. Thus, assignment

---

[3]Except for rare cases, reservations are not reassigned to different stations or vehicle classes.
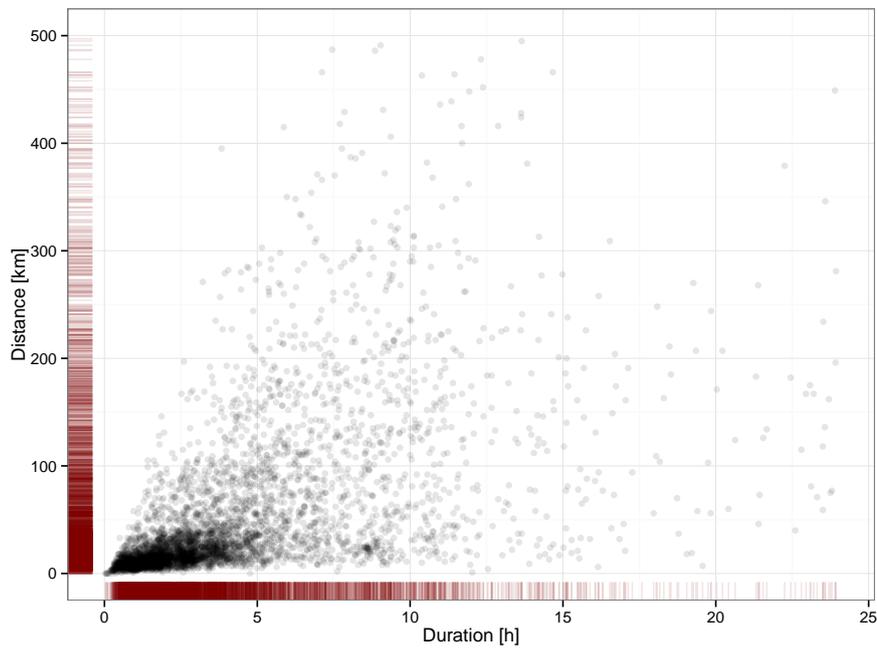
Figure 6.5: Sampled distribution of reservation distance and duration for vehicle class "small".

decisions are highly dynamic, and reservations enter the system mostly during a short time window before vehicle usage commences. Accordingly, one might suspect that reservations entering the system on short notice are rather flexible, either in time, space, or both, leaving room for optimization.

Fleet Utilization   An analysis and discussion concerning fleet utilization is available in Appendix B.
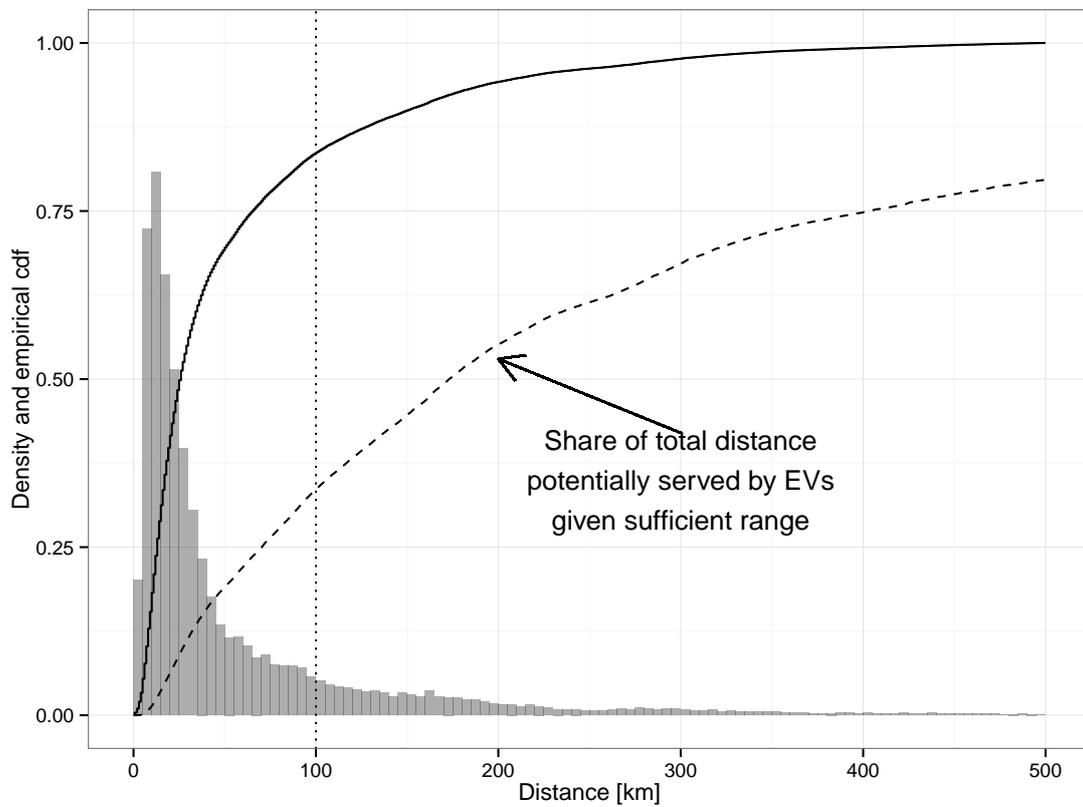
Figure 6.6: Distribution of reservation distances. Most trips are shorter than the typical EV range.
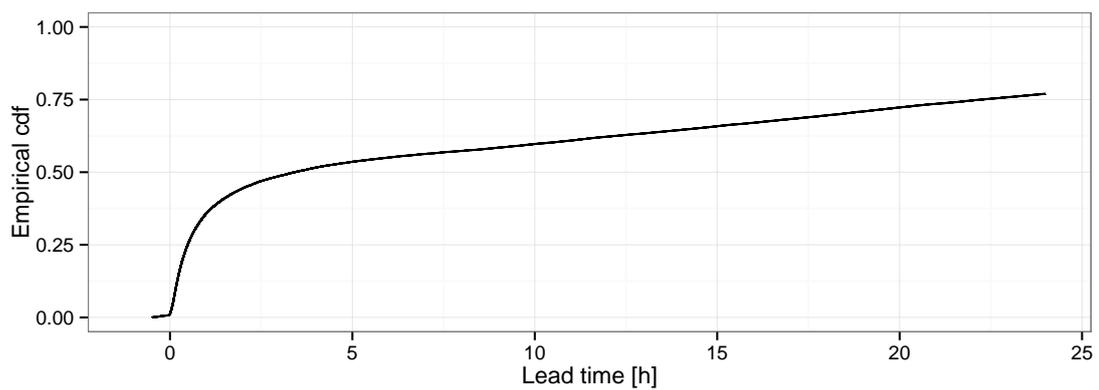
Figure 6.7: Distribution of reservation lead time. More than half of the reservations are booked with a lead time of four hours or less.

Chapter 7

Leveraging Consumer Flexibility in Station-based Car-sharing

This chapter is concerned with leveraging consumer flexibility to achieve improved economic outcomes in station-based car-sharing. Specifically, we examine to what extent consumer flexibility in space and time can be harnessed to increase fleet utilization or to improve QoS to consumers.

The intuition behind leveraging consumer flexibility is that of "opaque selling", a practice applied in revenue management for perishable assets (Fay, 2008; Jerath et al., 2010). In the hotel industry, for example, business guests might have a higher valuation for staying at a specific location (for the convenience of being close to their eventual destination). Tourists, on the other hand, might be willing to not know the exact location of their hotel in advance in exchange for a lower payment. By selling hotel rooms opaquely in a specific part of a city, but without information on the exact location, hotels can appeal to different market segments without cannibalizing their premium product. Effectively, the suppliers are able to enforce price-differentiation via quality-differentiated products. We follow the same idea of adding opacity to a service product and thus improve both capacity utilization, and revenues by attracting customer groups that otherwise would not use the service. In the case of car-sharing, opacity could translate into reservations on a city-district level instead of reserving a vehicle at a specific station.

The structure of this chapter is as follows. First, we introduce the problem definition and research questions, thereafter, we propose an offline optimization model that, serving all reservations, under perfect knowledge, minimizes fleet size. This model yields the value of consumer flexibility with respect to fleet size. Third, we introduce different online optimization approaches which aim to capture a large share of the value of flexibility in the online case, and present detailed results. Thereafter, we discuss the results and derive managerial implications for a car-sharing fleet operator. Finally, we conclude and provide an outlook into perspective future research avenues.

7.1 Problem Definition and Research Questions

Car-sharing operators are faced with the problem of trading-off fleet utilization and customer satisfaction. On the one hand, (prospective) customers value low utilization (high availability) of the vehicle fleet, while the operator is interested in high utilization in order to maximize profits, assuming that prices are given. As illustrated in Fig. 7.1, the outcome of economic coordination in car-sharing operations is determined by the following factors:

- Exogenously defined consumer flexibility determines the minimal fleet size necessary to serve all requests in the *offline* case.

- Fleet size and consumer flexibility are defined exogenously in the *online* case, while one of either lead time or QoS (dark circles) arises endogenously.

For ease of exposition we assume homogeneous fleets, i.e., both fixed and operating costs are identical for all vehicles.[1]

Offline   We posit that consumers are to some degree flexible in time and space, enabling ideas of the "opaque selling" kind (Gallego and Phillips, 2004). In the offline case, under perfect information, assignment decisions can be optimized within customers' flexibility constraints and higher fleet utilization may be achieved. Accordingly, the fleet size necessary to serve all reservations is minimized in the offline case, which also translates into minimal cost for the car-sharing operator.

The offline optimization reveals historically existing optimization potential. However, it is a rather theoretic construct that may serve as a baseline, but, due to uncertainty and imperfect forecasts, cannot be transferred to practice.

Online   To circumvent the shortcomings of offline optimization and outline an optimization alternative that may be applicable in practice, we introduce online optimization approaches that do not require information about the future. In the online case, the focus of the operator is on providing service to customers *after* the fleet size has been determined and thus can no longer be adapted to meet demand (or decrease costs). Moreover, there is no (or only limited information) on future demand in the online case.

We leverage an empirical dataset introduced in section 6.4 for our evaluation. If not noted otherwise, decisions taken by each of the online planners require access to retrospective data in the form of a reservation stream only, but can cope with a lack of foresight. This empirical reservation stream contains different aspects of reservation data, including

---

[1]We argue for this to be a reasonable simplification, as we restrict our attention to a single, homogeneous class of vehicles within the overall car-sharing fleet.

- the time when the reservation was entered into the system by the user

- beginning and

- end of the reservation period,

- actual beginning and

- end of the usage period,

- distance traveled.

The goal of the operator is maximization of the number of served reservations, given a stream of reservation data and assuming a specific, fixed fleet composition (number of vehicles, stations and stationing decisions) defined beforehand. Let QoS $\mathcal{Q}$ be defined as the ratio of served reservations and total amount of reservations in the reservation stream for the time frame under examination.

$$\mathcal{Q} = \frac{N_{\text{served}}}{N} \tag{7.1}$$

We posit the share of served reservations to be a reasonable proxy for the QoS the consumer eventually experiences.[2]

Research Question   For the purpose of increasing fleet utilization and QoS, we are interested in the value of consumer flexibility in station-based car-sharing.

As Fig. 7.1 indicates, flexibility may be a key factor regarding economic performance of car-sharing operations. In the offline case we focus on the interaction between fleet size and consumer flexibility (Fig. 7.1(a)).

In the online case, the situation is slightly more complicated: Three of the four levers that affect economic performance can be chosen exogenously, the fourth will arise endogenously. As indicated in Fig. 7.1(b), Quality-of-service, lead time and flexibility all affect the consumer, while the provider bears the cost of fleet provision. Note that by assuming consumers to be satisfied that allocation decisions are finalized only at the very latest possible moment before a reservation's period begins we take an optimistic approach to lead time. In practice, this might be unacceptable and decrease solution value. In order to quantify the value of different types of flexibility in the domain of car-sharing in detail, we pose the following questions.

Research Question 9 – OFFLINE VALUE OF FLEXIBILITY: *What is the value of spatial and temporal flexibility in terms of fleet size in the* offline *case?*

---

[2]For clarity and simplicity we restrict attention to a single evaluation metric. Still, fleet utilization (time or distance, or combination thereof) could pose another interesting metric for service quality.

(a) Offline
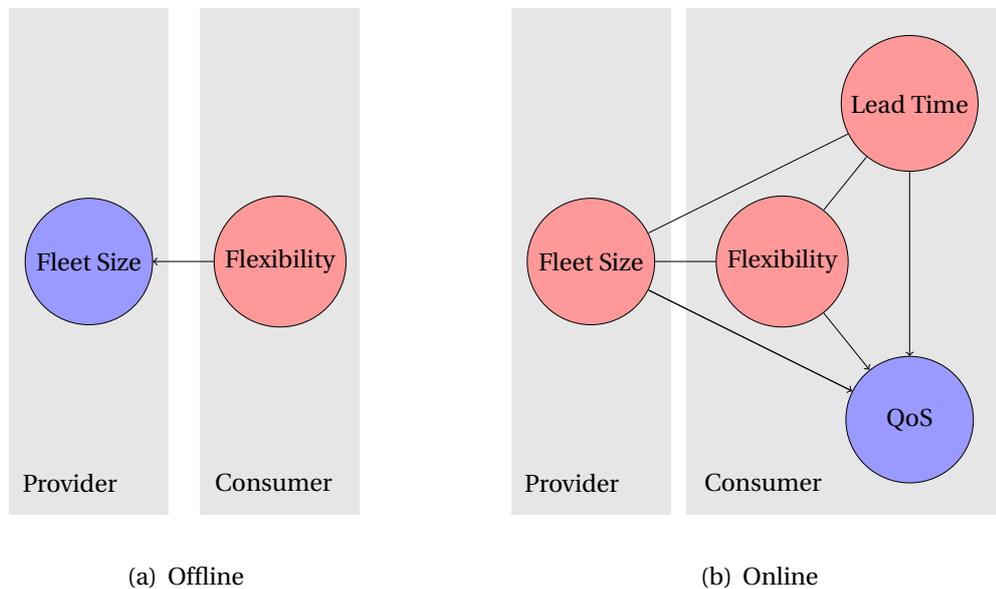
(b) Online

Figure 7.1: Levers for optimizing car-sharing operations. Exogenous variables are depicted in red, their endogenous counterparts in blue.

Research Question 10 – ONLINE VALUE OF SPATIAL FLEXIBILITY: *What is the value of spatial flexibility in the* online *case in terms of QoS?[3]*

We answer these questions based on empirical data, described in Section 6.4.

7.2 Offline Optimization

We introduce an offline optimization model that relies on perfect knowledge regarding reservations, i.e., has access to all reservations and is required to serve all reservations. We henceforth pursue minimization of total cost by means of fleet size minimization, subject to hard constraints on consumer flexibility as a valid optimization objective. Besides cost minimization, multi-criteria objectives involving both, operating costs and consumer inconvenience are promising objective candidates, but require careful weighting of the conflicting goals. Maximizing the number of served reservations (or distance/time covered) subject to an exogenously defined fleet is possible as well, but critically hinges on the exact specification of the fleet, i.e., the number of vehicles at each station. As we are interested in the value of customer flexibility, we let both size and configuration of the fleet arise as the result of optimization, while

---

[3]An alternative formulation could focus on the relationship between QoS and consumer inconvenience in the form of walkways.

introducing an upper bound on customer flexibility.

Solving the offline problem is possible in reality only ex-post (i.e., in hindsight) and thus, the value of the solution is limited concerning practical applications. However, reservation data exhibit recurring seasonalities over time. Therefore, an optimal fleet computed now may also be valuable in the future. Based on those patterns, computation of optimal fleets is valuable on strategic and tactical levels concerned with location or vehicle assignment planning.

Algorithmically, the problem of assigning reservations can be traced back to the GAP (Voudouris et al., 2010) which has the objective of profit-maximization.[4] In the special case in which any assignment of a reservation to a vehicle is equally costly, we are interested in minimizing the number of vehicles in use. This corresponds to the classic Bin-Packing Problem (Nemhauser and Wolsey, 1988).

7.2.1 Common Optimization Elements in Car-sharing

In our problem, multiple reservations must be assigned to a single vehicle in such a way that the resulting allocation is feasible, i.e., different reservations on the same vehicle do not overlap, each reservation is served by exactly one vehicle, and each vehicle in use serves at least one reservation.[5]

A reservation r is a tuple containing begin time, end time, and station, i.e., (t_b, t_e, σ). The set of all reservations is denoted as $\mathscr{R}$ and the set of vehicles as $\mathscr{V}$. The relation σ maps vehicles $\mathscr{V}$ and reservations $\mathscr{R}$ to the corresponding stations $\mathscr{S}$.

$$\sigma : \{\mathscr{V}\} \cup \{\mathscr{R}\} \mapsto \mathscr{S} \tag{7.2}$$

Decision Variables   Decisions concerning the assignment of reservation $r \in \mathscr{R}$ to vehicle $v \in \mathscr{V}$, are expressed via the binary decision variable $x_{r,v}$ (7.3), which is the decision variable of central importance.

$$x_{r,v} \in \{0,1\} \qquad\qquad \forall r \in \mathscr{R}, \forall v \in \mathscr{V} \tag{7.3}$$

First, we require all reservations $r \in \mathscr{R}$ in the system to be assigned to exactly one vehicle $v \in \mathscr{V}$, formally

$$\sum_{v \in \mathscr{V}} x_{r,v} = 1 \qquad\qquad \forall r \in \mathscr{R}. \tag{7.4}$$

---

[4]Assuming that any possible reservation-vehicle assignment decision is equally costly and the fleet was given exogenously, the problem could be further reduced to the Multiple-Knapsack Problem (Chekuri and Khanna, 2005).

[5]Otherwise it clearly should not be in use.

The second decision variable u incorporates the decision which vehicle to eventually include in the fleet. A vehicle v is in use if it is assigned to at least one reservation.

$$u_v \geq x_{r,v} \qquad\qquad \forall r \in \mathscr{R}, \forall v \in \mathcal{V} \qquad (7.5)$$

Fleet size N is defined by the number of vehicles in use.

$$\sum_{v \in V} u_v = N \qquad (7.6)$$

7.2.2 Modeling Temporal Flexibility

With complete information on all reservations, i.e., hindsight, vehicle-reservation assignment decisions can be optimized and the minimal fleet able to serve all reservations can be identified. In this section we develop the corresponding constraints of the mathematical optimization model that formalize customers' temporal flexibility, which is in turn leveraged to minimize both fleet size and total cost.

Decision Variable   Besides encoding reservation-vehicle assignment decisions in x, and usage of vehicle v via u_v, an additional decision variable θ is required to model temporal shifting of individual reservations. Furthermore, for auxiliary purposes we introduce the binary decision variable $\omega_{a,b}$ that encodes that a and b have temporal overlap, with b starting at a later point in time than a.

Objective Function   Minimal cost is achieved through fleet size minimization. Accordingly, we define the objective function to comprise of the number of vehicles used (7.6), which is then minimized via all three (assignment, vehicle in use, temporal shifting) decision variables, i.e.,

$$\min_{x,u,\theta} N. \qquad (7.7)$$

Note that the number of vehicles in use at zero flexibility can be extracted from a reservation stream of sufficient length in a straight-forward manner.

Temporal Flexibility Constraints   This model does not include spatial flexibility, hence, reservations must be served at the originally desired station. The station expressed in the original reservation r is indicated by σ_r. Accordingly, the following equality sets x to zero for infeasible reservation-vehicle combinations, i.e., combinations where the station of the vehicle and the reservation diverge. Formally,

$$x_{r,v} = 0 \qquad\qquad \forall \{(r,v) \in \mathscr{R} \times \mathcal{V} \mid \sigma_r \neq \sigma_v\} \qquad (7.8)$$

We denote the non-negative temporal shifting of an reservation r by θ_r. Overlap between two reservations r_1 and r_2 b in the case that r_1 has earlier begin time than r_2, i.e.,

$r_{1_{t_b}} < r_{2_{t_b}}$, is encoded in $\omega_{r_1,r_2} = 1$. Correspondingly, $\omega_{r_1,r_2} \neq \omega_{r_2,r_1}$ [6]. This is relevant as we only allow *postponement* of reservations. Shifting of reservations in time may not exceed a certain value, i.e.,maximum temporal shifting is constrained by $\bar{\theta}$.

$$0 \leq \theta_r \leq \bar{\theta} \qquad\qquad \forall r \in \mathscr{R} \qquad\qquad (7.9)$$

For analysis and interpretation we assume homogeneous temporal flexibility over all reservations. In an alternative problem formulation, temporal shifting may be elevated from being a constraint to being part of the objective function. However, doing so requires appropriate weights for customers' valuation of time, similar to Heinonen (2004), which we do not readily have access to. Therefore, we opt for including temporal shifting via constraints and forgo its inclusion in the objective.

In order to elegantly represent further constraints involving temporal overlap, we introduce the set of overlapping reservations $\mathscr{Z}$. It represents the set of two-tuples of reservations that are

1. placed at the same station, i.e., $\sigma_{r_1} = \sigma_{r_2}$,

2. have temporal overlap, i.e., $\omega_{r_1,r_2} = 1$ or $\omega_{r_2,r_1} = 1$, and

3. are distinct, i.e., $r_1 \neq r_2$.

Formally, $\mathscr{Z}$ is defined as follows:

$$\mathscr{Z} = \{(r_1, r_2) \in \mathscr{R} \times \mathscr{R} | (\sigma_{r_1} = \sigma_{r_2}) \wedge (\omega_{r_1,r_2} = 1 \vee \omega_{r_2,r_1} = 1) \wedge (r_1 \neq r_2)\} \qquad (7.10)$$

Intuitively, for each tuple $(r_1, r_2) \in \mathscr{Z}$, one of the two reservations needs to be shifted in order to ensure feasibility of the schedule. We constrain temporal shifting to be non-negative, i.e., reservations can only be postponed, but not brought forward. However, advancing reservations in time may be sensible in some settings.

Let $\underline{\theta}_{r_1,r_2}$ represent the required shifting of r_2 in order to accommodate both reservations, r_1 and r_2 on the same vehicle (cf. Fig. 7.2). If overlapping reservations r_1, r_2 are to be assigned to the same vehicle, the reservation with the later start time can be shifted at most by $0 \leq \theta \leq \bar{\theta}$. Reservations with temporal overlap $(r_1, r_2) \in \mathscr{Z}$ can be assigned to the same vehicle if one (or both) of the two reservations is shifted, i.e., postponed by at least as much as the difference in start and end time $\underline{\theta}_{r_1,r_2}$ (or $\underline{\theta}_{r_2,r_1}$) of the overlapping reservations.

Formally, we restrict overlaps between different reservations to be possible only in one direction, i.e., postponement. In order to assign two overlapping reservations $(r_1, r_2) \in \mathscr{Z}$ to the same vehicle v, at least one of the two (auxiliary) decision variables $\omega_{r_1,r_2}, \omega_{r_2,r_1}$ must be set to one (7.11).

---

[6] Except for pathological cases where beginning and end of both reservations r_1, r_2 are identical, and hence the exact definition is inconsequential.
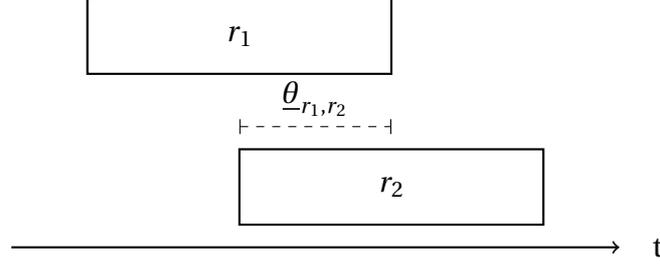
Figure 7.2: Illustration of temporally overlapping reservations r_1 and r_2. Shifting reservation r_2 by $\underline{\theta}_{r_1,r_2}$ allows to assign both reservations to the same vehicle. In this example $\omega_{r_1,r_2} = 1$, indicating the later start of reservation r_2 relative to r_1.

If there is temporal overlap, the relative shifting of reservations r_1 and r_2 (expressed via $\theta_{r_2} - \theta_{r_1}$) must at least be as large as the minimal required amount of shifting $\underline{\theta}_{r_1,r_2}$ (7.13) (and (7.14), for reasons of symmetry).

$$x_{r_1,v} + x_{r_2,v} \leq 1 + \omega_{r_1,r_2} + \omega_{r_2,r_1} \qquad \forall v \in V, \forall (r_1, r_2) \in \mathcal{Z} \qquad (7.11)$$

$$\omega_{r_2,r_1} + \omega_{r_1,r_2} \leq 1 \qquad \forall (r_1, r_2) \in \mathcal{Z} \qquad (7.12)$$

$$\underline{\theta}_{r_1,r_2} \cdot \omega_{r_1,r_2} \leq (\theta_{r_2} - \theta_{r_1}) \cdot \omega_{r_1,r_2} \qquad \forall (r_1, r_2) \in \mathcal{Z} \qquad (7.13)$$

$$\underline{\theta}_{r_2,r_1} \cdot \omega_{r_2,r_1} \leq (\theta_{r_1} - \theta_{r_2}) \cdot \omega_{r_2,r_1} \qquad \forall (r_1, r_2) \in \mathcal{Z} \qquad (7.14)$$

Applying this formulation, we restrict our attention to the cost incurred by the provider, while foregoing inclusion of costs on the consumers' side. Clearly, costs related to shifting reservations in time could be included in the objective, yielding a multi-criteria optimization problem. However, to come up with meaningful decisions via such an optimization approach, the weights of both sides need to be adequately specified – a task we leave for future research.

7.2.3 Modeling Spatial Flexibility

Let us now shift our attention from temporal to spatial consumer flexibility. We assume that some (not necessarily all) reservations are spatially flexible, i.e., flexible with respect to the station they are assigned to (in the spirit of "opaque selling" and "flexible products").

Technically, we add spatial flexibility to reservations by enlarging the set of stations $\mathscr{S}_r$ a reservation r can be assigned to. Accordingly, reservations are no longer to be served by vehicles located at their original station only, but also by vehicles located at adjacent stations within the defined perimeter of spatial flexibility.

Fig. 7.3 illustrates the potential of spatial flexibility based on real-world data. The more flexible a reservation, the larger the set of stations it can be assigned to besides
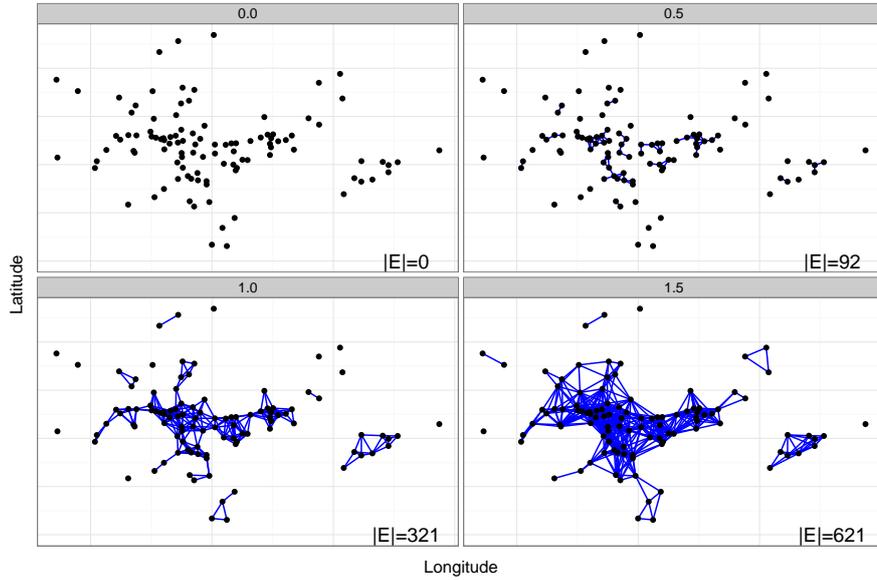
Figure 7.3: Increasing the solution space via spatial consumer flexibility (facet labels in km). Total flexibility potential is expressed in terms of valid inter-station edges.

the original station. The number of edges that arise between stations clearly increases in flexibility. Starting at zero edges, the number of edges $|E|$ increases to 92 at spatial flexibility of 0.5 km, 321 at 1.0 km and 621 at spatial flexibility of 1.5 km. Intuitively, the larger the number of edges between the stations, the more reservations can be moved to adjacent stations, potentially yielding better economic outcomes, but possibly also requiring increasing walkways from consumers.

The walkway w_r a consumer incurs when implementing reservation r is equal to the distance between the originally reserved pickup-station σ_r and the station of the assigned vehicle σ_v – conditional that r is assigned to vehicle v stationed within the flexibility perimeter of reservation r, i.e., $x_{r,v} = 1$. Formally,

$$\sum_{v \in V} d(\sigma_r, \sigma_v) \cdot x_{r,v} = w_r \qquad\qquad \forall r \in \mathcal{R} \qquad\qquad (7.15)$$

Assignments that violate the upper bound $\bar{w}$ on the admissible distance between original and actual station are excluded.

$$0 \leq w_r \leq \bar{w} \qquad\qquad \forall r \in \mathcal{R} \qquad\qquad (7.16)$$

Let the set of competing reservations $\mathcal{R}^{comp}$ comprise the reservations originating from neighboring stations that have temporal overlap (7.17). In contrast to the previous set of temporally overlapping reservations $\mathcal{Z}$, the set of competing reservations

$\mathscr{R}^{comp}$ includes reservations originating from adjacent stations, and is thus a superset of $\mathscr{Z}$. Formally, we define this set

$$\mathscr{R}^{comp} = \{(r_1, r_2) \in \mathscr{R} \times \mathscr{R} | d(\sigma_{r_1}, \sigma_{r_2}) \leq \bar{w} \wedge (\omega_{r_1, r_2} = 1 \vee \omega_{r_2, r_1} = 1) \wedge (r_1 \neq r_2)\} \quad (7.17)$$

Then, the no-overlap condition of two reservations on the same vehicle is expressed as

$$x_{r_1, v} + x_{r_2, v} \leq 1 \qquad\qquad \forall v \in V, \forall (r_1, r_2) \in \mathscr{R}^{comp} \qquad (7.18)$$

Based on sufficient empirical reservation data, the number of vehicles of each type at each station can be determined reliably. Let $|\mathscr{V}'(s)|$ encode the number of vehicles at station s based on empirical data. To coarsely retain the station size structure after optimization, we constrain the number of vehicles at each station not to exceed the number of vehicles stationed at station s by a factor of two.[7]

$$\mathscr{V}(s) = \{v \in \mathscr{V} | \sigma_v = s \wedge u_v > 0\} \qquad\qquad (7.19)$$

$$\mathscr{V}(s) \leq 2 \cdot |\mathscr{V}'(s)| \qquad\qquad \forall s \in \mathscr{S} \qquad (7.20)$$

This last constraint completes our MIP formulation.

7.2.4 Results

In this section we describe results from solving the optimization problems with the objective of minimizing fleet size under perfect information (offline). We begin with temporal flexibility, followed by the effect of spatial flexibility on fleet size.

Temporal Flexibility   The results in Fig. 7.4 show that temporal customer flexibility improves fleet utilization, or decreases minimal fleet size, respectively. For clarity, results for each week are connected by dashes; optimization leads to slightly different results for each week (mostly due to seasonal patterns spanning over the entire yearly in the reservation data), but the overall decreasing tendency is clearly visible. However, the effect of temporal flexibility is meager at best. For example, even at six hours of temporal flexibility, fleet size reductions are on average 6 %, i.e., virtually non-existent.

Initially, this is a surprising, yet plausible, discouraging result: Even elevated levels of temporal flexibility allow fleet size to be reduced only marginally. Nevertheless, this is a plausible result. Due to the fleet structure, which is characterized by a large

---

[7]By constraining station size, we forgo part of the optimization potential from pooling effects. However, by following a rather cautious approach involving a larger number of mid-sized stations instead of few large-sized stations, user acceptance can be expected to be higher.
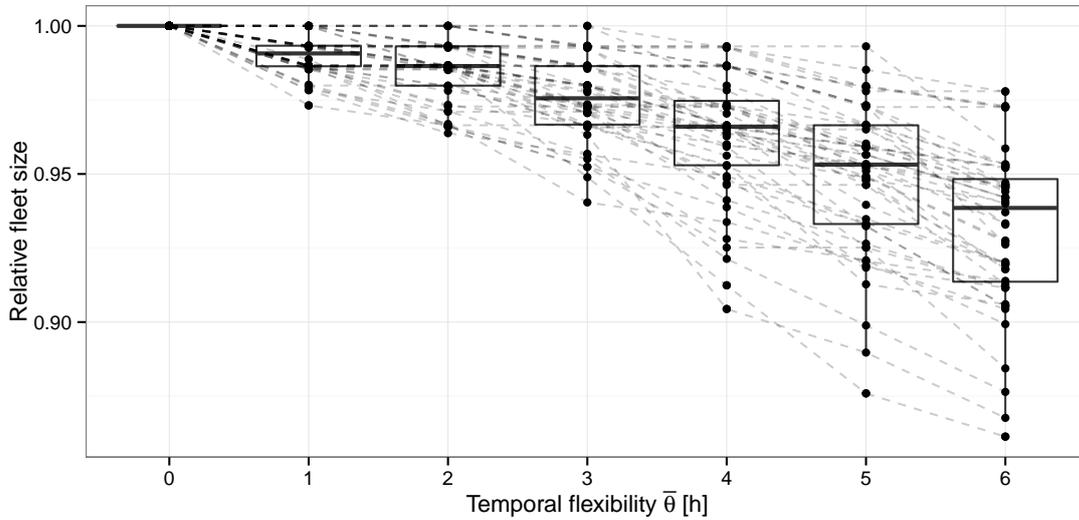
Figure 7.4: Fleet reduction over temporal consumer flexibility. Reservations of one week have been used. Sole use of temporal flexibility does not lead to significant fleet reductions.

number of small stations, and peaking demand at one single day of the week, temporal flexibility is only of little utility to the provider.[8] Under fleets with fewer, but larger stations, temporal flexibility might yield greater value, as it may enable more effective pooling of demand. However, under current fleet configurations, temporal flexibility is virtually useless for fleet optimization purposes.

Spatial Flexibility In contrast to the small gain from temporal flexibility, its spatial counterpart may yield significant improvements. Fig. 7.5 illustrates relative optimal fleet size at levels of 0, 0.5, 1.0, and 1.5 km spatial consumer flexibility (airline distance[9]). For example, assuming a flexibility level of 1.5 km fleet size can be reduced by up to 30% and approximately 18% on average.[10] Vice versa, under the given fleet, a similar amount of reservations could additionally be served assuming spatial customer flexibility of 1.5 km. Correspondingly, as all reservations are served in the described optimization model, fleet utilization increases by means of reducing fleet size.

The results so far highlight the importance of spatial customer flexibility in car-sharing operations. However, a conclusive test of flexibility's value requires an eval-

---

[8]Assume a station with two vehicles. With large probability at least two reservations are characterized by having temporal overlap in excess of customers' temporal flexibility. Consequently, neither of the vehicles can be decommissioned; the positive aspects of temporal flexibility cannot unfold.

[9] We compute airline distance in the Euclidean space as follows: $d(p_1, p_2) = \frac{40.000}{360}\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

[10]Note that limited station size has adverse effects on economic performance. If station size could be modified beyond current station sizes, further pooling of reservations could be achieved. This, however, comes at the price of consumer acceptance.
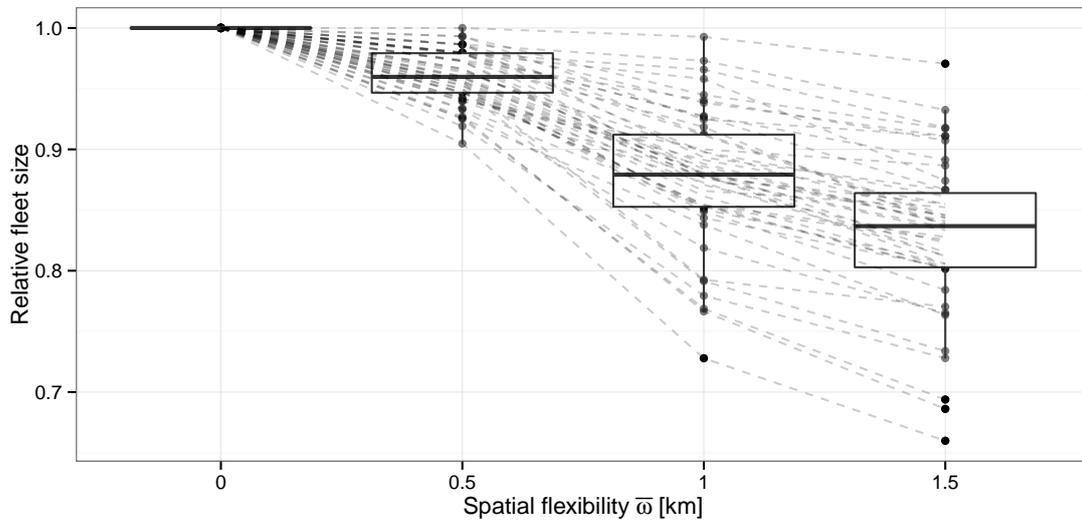
Figure 7.5: Relative size of optimized fleets over consumers' spatial flexibility. For example, assuming spatial flexibility of 1.5 km (airline distance), approximately 16% of the fleet could be retired.

uation in the online case, where access to future information is limited.

7.3 Online Optimization

In realistic, i.e., online, scenarios, the decision making process is complicated by imperfect information, with decisions being made as time unfolds and new information is revealed. However, there is only little information available on future arrival of reservations. In particular, reservations that arrive in the future cannot be included in assignment decisions that are taken now. In different words, decisions are made as time unfolds and new information is revealed. Based on the encouraging results in the preceding section on offline optimization, we focus on spatial flexibility, ignoring the potential of temporal flexibility in the online case. Naturally, the online attainable objective value, compared to offline optimization, is reduced. Fortunately, some knowledge about the future is readily available from the system state in the car-sharing domain. This allows for the construction and evaluation of different reservation-vehicle assignment policies.

First, reservations can be assigned to vehicles in the temporal order they arrive in the system (FCFS), with decisions being made at the very moment a reservation enters the system. The advantage of such an approach lies in its low complexity and ease of implementation on the one hand, and fairness and transparency as well as immediate feedback to the consumer on the other. The disadvantage of such a rather simple

approach might lie in low assignment success expressed in terms of QoS, and inflated consumer walkways.

Alternatively, assignment decisions can be postponed to the point in time when the reservation starts (or shortly before, leaving enough time to guide the consumer to the assigned station). By doing so, more information can be leveraged and better outcomes can be achieved.

Fundamentally, one can differentiate between designs that inform the customer about the acceptance of a reservation and the exact vehicle assignment immediately (on entering the reservation into the system), and those designs where acceptance and assignment decisions are communicated at different points in time. The former may enjoy higher user acceptance due to simplicity, the latter, however, may yield better performance by allowing the system operator to re-plan and re-optimize as more information about the future becomes available.[11] Obviously, leveraging information on currently active reservations whose vehicle assignment so far has not been decided upon, can foster improvements in both, fleet utilization and QoS. The later the customer needs to be informed about the eventual vehicle to be assigned for his reservation, the higher the bound for aggregate system performance be. However, customer acceptance may limit how late exactly feedback can be given to the customer.

7.3.1 First-Come First-Served

A FCFS approach requires no information on the future. Earlier arrival of a reservation in the system (relative to the time of desired execution) has two-fold consequences. First, it increases the probability of allocation as remaining capacity diminishes over time, and second, in case of assignment at a neighboring station, the distance between desired and assigned station may be reduced. On the one hand, consumers have an incentive to enter information concerning their reservation intention as early as possible, on the other hand, earlier information revelation may conflict with uncertainty on the level of the individual consumer.

Upon arrival of a reservation, the FCFS algorithm traverses the list of available vehicles, sorted by distance from the originally desired station σ_r in ascending order and assigns the reservation under consideration to the closest (i.e., first) available vehicle. Due to its simplicity and similarity to today's assignment process, we choose FCFS to be the benchmark for later evaluations. Admittedly, this simple heuristic may deliver relatively poor results (low share of served requests, inflated walkway distances) when vehicles are scarce, as externalities of individual assignment decisions are ignored.

---

[11]Besides immediate and late feedback approaches, re-planning is of interest as well. The idea here is to immediate return information on the assigned vehicle. Changing the assigned vehicle (and more importantly, the corresponding station) later on results in a penalty. Hence, the planner will only implement changes regarding re-planning into the assignment decisions, if doing so yields significant benefits.

7.3.2 Least-Utilized Station First

Our next approximation heuristic, Least Utilized Station First (LUSF) requires no information on the future, but leverages historic data in order to achieve better QoS. The heuristic assigns, as its name implies, reservation r greedily to the historically least-utilized station within the flexibility constraints of the reservation's originally desired station σ_r. On the positive side, QoS is maximized if current and past observations are correlated, i.e., expose the same or similar patters in time and space. However, on the negative side, mechanically assigning reservations to least-utilized stations may induce (i) longer-than-necessary walkways, and (ii) reduced utilization of the initially preferred stations. Nevertheless, we argue, and this is in-line with the literature on machine scheduling (Jain and Elmaraghy, 1997; Aspnes et al., 1997) that assigning newly arriving reservations to the historically least utilized station poses a promising approach and often performs very well in practice. This is especially true, if QoS is defined in terms of reservations served, ignoring walkways, as in our case.

7.3.3 Leveraging Advance Information

Customers mostly reserve vehicles some time in advance of using them (Fig. 6.7). If the assignment decision is not made immediately upon arrival of the reservation in the system (as is the case in FCFS), some information on future vehicle usage is readily available via look-ahead and can be leveraged to improve assignment decisions for reservations that are due at the current point in time (cf. Dunke, 2014). In contrast to the heuristics in sections 7.3.1 and 7.3.2, GREEDY indeed leverages additionally information available on future reservations. GREEDY tries to maximize the (expected) number of reservations served while retaining reasonable walkways for consumers. For instance, if assigning a reservation evokes unreasonably large aggregate walkways (i.e., exceeding the upper bound on spatial consumer flexibility $\bar{\omega}$), the corresponding reservation remains without allocation. At its core, it relies on a greedy offline assignment algorithm that is executed at each time step $t \in \mathcal{T}$, i.e., the points in time when assignment decisions need to be completed.[12]

GREEDYOFFLINE   Both the GAP and the Multiple Knapsack Problem are NP-hard. Determining the optimal assignment decisions given a homogeneous fleet of cars on the one hand and a set of reservations on the other can be reduced to the Multiple Knapsack Problem and hence is also NP-hard. We avoid the complexity associated with determining the optimal schedule at each time step $t \in \mathcal{T}$ by introducing an offline heuristic which we name GREEDYOFFLINE (Alg. 7).[13] It takes as input a schedule

---

[12]For instance, half-hourly time blocks used in car-sharing may serve as decision times.

[13]Only for very small problem sizes can the optimal solution be computed within reasonable time limits. For medium and large-sized problems we approximate the optimal schedule using a greedy heuristic and thus trade-off solution quality for a complexity reduction.

s, the fleet $\mathcal{V}$, and a set of reservations $\mathcal{R}$.

In order to assign reservations to available vehicles, the algorithm iterates over the sorted list of reservations $\mathcal{R}$ in the first step(line 1). To this end, sorting reservations by their length (in decreasing order) has empirically yielded good results. Initially, the empty vehicle (also: non-allocation decision) represented by $\perp$ is assigned to v^*. In the next step (line 3), we iterate over all vehicles in the fleet that are within the admissible spatial distance from the reservation's originally desired station. Thereby, the set of vehicles is sorted by increasing distance from the originally desired station σ_r (of reservation r). If the vehicle under consideration v *isavailable* to serve reservation r given schedule s, v^* is assigned v and GREEDYOFFLINE breaks the loop begun in line 3; thereafter, r is added to the schedule of vehicle v^* (line 8). Otherwise, if the vehicle under consideration is not available to serve r given schedule s, we iterate through all potential vehicles in increasing order of their distance to r. If no vehicle is available, i.e., $v^* = \perp$ still holds in the final iteration, r is discarded from the schedule.

Computational complexity for GREEDYOFFLINE is $\log(|\mathcal{R}|) + |\mathcal{R}| \cdot (\log(|\mathcal{V}|) + |\mathcal{V}|)$, including sorting of both, reservations and vehicles, and assuming that the subroutine *isavailable* requires constant time (e.g., via hash maps). However, complexity is typically much less since the size of the set of vehicles $\mathcal{V}$ can be limited by spatial flexibility to neighboring stations only. Furthermore, for most reservations, only a small subset of vehicles need to be iterated over before an available vehicle is found. In total, complexity is defined by the product of reservations and fleet size, i.e., $\mathcal{O}(|\mathcal{R}||\mathcal{V}|)$.

Algorithm 7: GREEDYOFFLINE

Input: $s, \mathcal{V}, \mathcal{R}$
Output: A schedule s

1   **for** $r \in sorted(\{\mathcal{R}\})$ **do**
2      $v^* = \perp$
3   **for** $v \in sorted(\{v \in \mathcal{V} \mid d(\sigma_v, \sigma_r) \leq \bar{w}\})$ **do**
4   **if** $isavailable(s, r, v)$ **then**
5              $v^* \leftarrow v$
6   break
7   **if** $v^* \neq \perp$ **then**
8          $s(v^*) \leftarrow s(v^*) \cup r$
9   **return** s

Greedy   The GREEDY algorithm is given in Algorithm 8, and relies on GREEDYOFFLINE to quickly compute schedules based on look-ahead information (Dunke, 2014) available at the time of invocation.

GREEDY is invoked at every decision time step and takes as inputs the current schedule s, the fleet of vehicles $\mathcal{V}$, reservations currently being active $\mathcal{R}_t$, i.e., the set of

---

Algorithm 8: GREEDY

Input: $s, \mathcal{V}, \mathcal{R}_t, \mathcal{R}_{t^+}$

Output: A schedule s

1   **for** $r \in sorted(\{\mathcal{R}_t\})$ **do**
2   Reset $f(v) = 0 \;\; \forall v \in \mathcal{V}$
3     $s' \leftarrow s$
4     $s' \leftarrow \text{GREEDYOFFLINE}(s', \mathcal{V}, \{\mathcal{R}_t\} \cup \{\mathcal{R}_{t^+}\} \setminus \{r\})$
5     $f(\perp) \leftarrow \Omega(s') + \pi$
6   **for** $v \in \{v \in \mathcal{V} \mid d(\sigma_v, \sigma_r) \leq \bar{w}\}$ **do**
7       $s' \leftarrow s$
8       $s'(v) \leftarrow s'(v) \cup \{r\}$
9       $s' \leftarrow \text{GREEDYOFFLINE}(s', \mathcal{V}, \{\mathcal{R}_t\} \cup \{\mathcal{R}_{t^+}\} \setminus \{r\})$
10      $f(v) \leftarrow \Omega(s')$
11    $v^* = \arg\min_v f(v)$
12   **if** $v^* \neq \perp$ **then**
13      $s(v^*) \leftarrow s(v^*) \cup \{r\}$
14 **return** s

---

reservations that require assignment decisions during the current invocation, as well as future reservations that have already been entered into the system $\mathcal{R}_{t^+}$ and will only be served in upcoming periods (look-ahead). The algorithm returns a schedule s after each invocation.

First, we iterate over the set of reservations requiring assignment decisions in the current invocation of GREEDY ($r \in \mathcal{R}_t$), sorted in decreasing order of the duration of the reserved period (line 1). GREEDY caches solution quality of competing assignment decisions internally via an appropriate data structure f. We reset the dictionary f (line 2), and assign the current schedule s to a temporary schedule s'. Next (line 4), the temporary schedule based on all available information except for the reservation currently under consideration r, is computed. The quality of the corresponding solution, the sum of consumer walkways as derived by $\Omega(s')$ and a penalty π for the potential failure to allocate a reservation to a vehicle, is stored in $f(\perp)$, where $\perp$ represents the no-allocation (alternatively: empty vehicle) case.

The same procedure, but including r, is repeated from line 6 on. Note that sorting the set of vehicles at this point with respect to distance can be omitted, as we need to evaluate all admissible reservation-vehicle decisions and thus sorting will not yield performance improvements. The difference in each iteration of the for-loop concerns the assignment decision of r and v, i.e., we begin the construction of a temporary schedule s' via the inclusion of r, assigned to a different vehicle in each iteration (line 8).

Solution quality is measured in aggregate walkways, i.e.,

$$\Omega(s') = \sum_{v \in \mathcal{V}} \sum_{r \in s'(v)} d(\sigma_v, \sigma_r)$$

and the algorithm greedily selects the vehicle v^* with the highest solution quality (line 11) for the assignment decision in line 13.

Hence, if the assignment of a reservation to a vehicle causes excessive walkways (in future assignment decisions), it may be advantageous to discard that particular reservation and instead include other reservations. Furthermore, in order to avoid the pathological case in which maximal solution quality would be achieved by not allocating any reservation at all, we penalize discarding a reservation by a certain amount of artificial walkways (line 11) expressed via π. [14]

Clearly, overall complexity of GREEDY (including GREEDYOFFLINE) is higher than for both, FCFS and the LUSF heuristic. In more detail, GREEDY calls GREEDYOFFLINE $|\mathcal{R}_t| \cdot (|\mathcal{V}| + 1)$ times, accordingly, overall complexity amounts to $\mathcal{O}\left(|\mathcal{R}'|^2 \cdot |\mathcal{V}|^2\right)$, where $\mathcal{R}'$ includes currently active and future reservation, i.e., $\mathcal{R}' = \{\mathcal{R}_t\} \cup \{\mathcal{R}_{t^+}\}$. [15]

7.3.4 Results

Compared to the offline setting, we adapt the evaluation context as follows in the online setting. Instead of minimizing fleet size while serving all reservations as we did before, we now consider the problem of maximizing the number of served reservations given a particular, fixed fleet. This fleet is the result of the offline optimization problem (sec. 7.2.3) leveraging spatial flexibility. The fleet under consideration is held constant, and was reduced by approximately 14% relative to the no-flexibility setting.[16] We chose this particular summer week as it required a rather extensive fleet to serve demand. Hence, due to the relatively large fleet size, satisfactory performance in the online setting could be expected.

We argue that retaining the same fleet for all evaluations both fosters comparability and interpretability of results. Furthermore, it nicely fits with the situation a car-sharing provider faces: The reservation stream changes over time, while the fleet is fixed for an extended period of time. In the following, we describe the results in more detail. Fig. 7.6 illustrates both, average weekly walkways and QoS, including the corresponding standard errors, on the ordinate and spatial consumer flexibility on the abscissa, differentiated by (left to right) algorithm in use. Note that all subsequent

---

[14]We set $\pi = \bar{w}$, indicating that the scheduler should prefer discarding reservations that cause excessive consumer walkways.

[15]Again, actual complexity is much lower, as the admissible set of vehicles is usually only a fraction of the whole fleet.

[16] In more detail, this fleet is based on data of the reservation stream of calendar week 33/2012 assuming spatial flexibility of $\bar{w} = 1.5$ for all reservations. Selecting reservation data of a different week as input to the optimization, we expect quantitatively different, but qualitatively similar results.
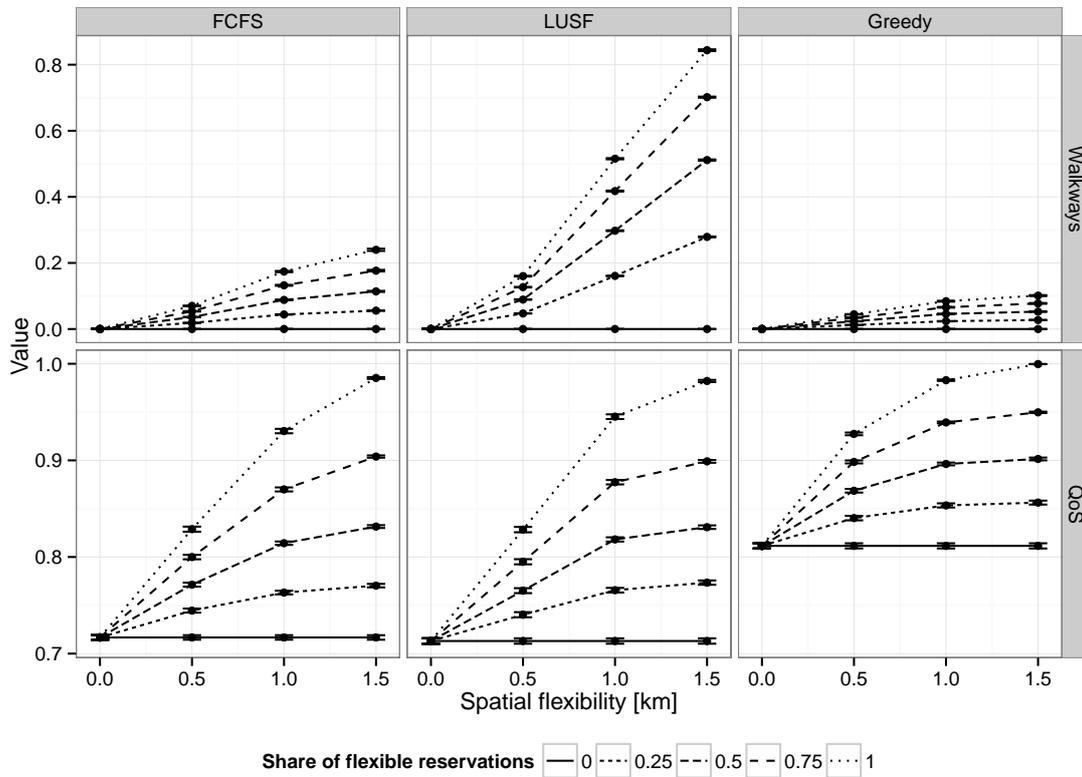
Figure 7.6: QoS and walkways differentiated by algorithm and consumer flexibility.

results were achieved under one single fleet configuration. This fleet was derived by assuming consumer flexibility to be 1.5km for all reservations in a given week. Accordingly, the size of this optimized fleet is, relative to the actual fleet in use, reduced. Clearly, walkways are zero if consumers exhibit zero flexibility for all algorithms while the associated QoS is low (approximately only 70%).

Our general observations are as follows: First, and following intuition, attainable service levels are increasing in the share of flexible consumers. Second, FCFS and LUSF exhibit similar performance regarding QoS overall. Depending on the share of flexible reservations, LUSF and FCFS achieve up to 98% of served reservations. However, this positive result is achieved at the cost of significant consumer inconvenience in the form of relatively extensive walkways. LUSF, for example, on average requires walkways close to 0.8km per served reservation in order to achieve the described high QoS. The last-introduced GREEDY algorithm, in contrast, succeeds in serving all reservations under the assumption of high spatial flexibility and the assumption of all reservations to be flexible (compared to only a fraction being flexible). It also shows superior performance with respect to QoS and consumer inconvenience at lower levels of flexibility. More specifically, GREEDY serves virtually all reservations at average
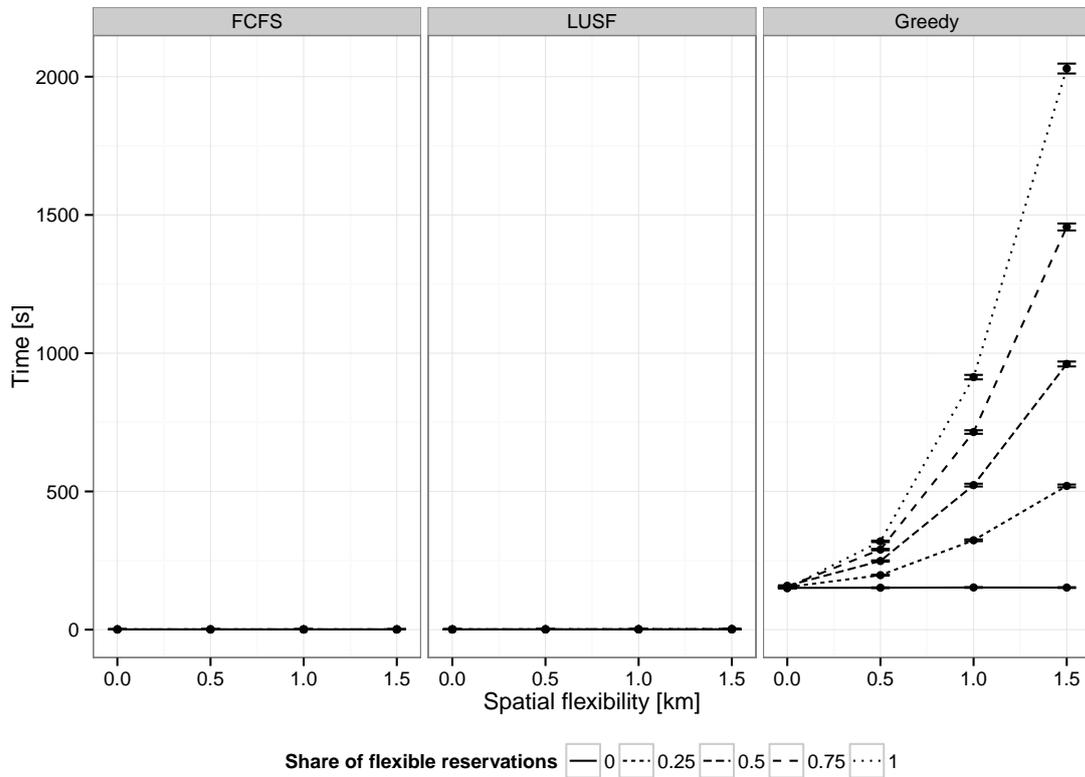
Figure 7.7: Empirical computational complexity

normalized walkways at values slightly in excess of 100m. Hence, GREEDY, leveraging available information, makes better decisions and delivers superior economic performance. The downside to such performance lies in the reliance on all available information and the associated cost of additional computation. Our observations regarding the algorithms' runtime confirm the theoretical complexity analysis. Fig. 7.7 presents average empirical computation time (and the associated standard errors) for computing the assignments for one week of reservations. As we do not vary the look-ahead length, spatial flexibility translates into larger available fleets for the corresponding reservation. Computational effort, as presented at the end of Section 7.3.3 grows supralinearly in fleets size. This is confirmed in Fig. 7.7. Interestingly, GREEDY requires longer run-times than its benchmarks even if spatial flexibility is absent. This effect can be attributed to GREEDY leveraging pooling flexibility present on the level of individual stations (cf. leftmost panel of Fig. 7.8), outperforming both FCFS and LUSF, which return decisions near-instantaneously. GREEDY requires, on current hardware, up to one second to return the assignment decision for a single reservation. Runtime of the GREEDY algorithm increases in consumer flexibility and the number of reser-
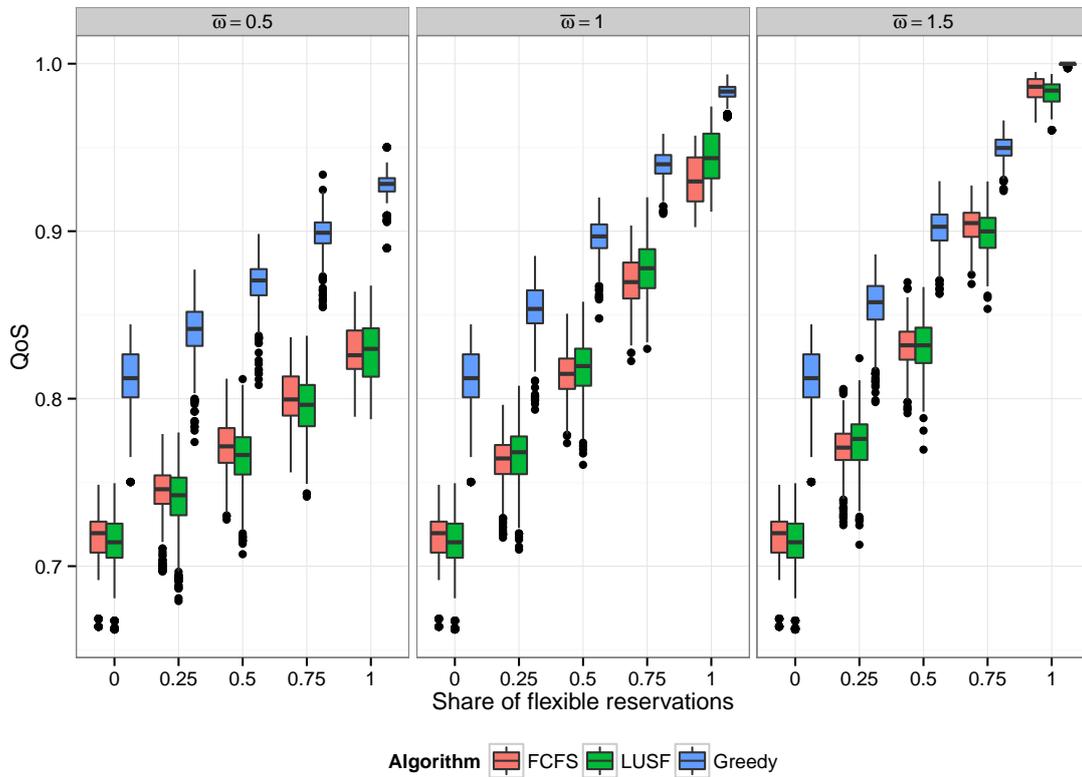
Figure 7.8: QoS differentiated by the share of flexible reservations and their spatial flexibility. Service quality is increasing in flexibility. Greedy assignment leverages all available information and thus achieves better QoS. Note that higher QoS does not entail increasing consumer inconvenience under the GREEDY algorithm (Fig. 7.9).

vations in the system at the time of decision.[17] By reducing the set of reservations that is considered when deciding upon assignments, run time can be reduced at the expense of economic performance. Altogether, the economic outcomes are encouraging: Leveraging advance reservation data in the system via appropriate algorithms such as GREEDY, may render significant fleet size reductions possible in practice.

Let us now turn towards inter-algorithm performance comparison, as illustrated in Figures 7.8 and 7.9. Both figures illustrate the range of values for QoS and consumer walkways, differentiated by share of flexible reservations (abscissa), upper bound of spatial flexibility (panels) and the algorithm in use (color coded). Overall, FCFS and LUSF yield comparable QoS. LUSF, by design, clearly yields highest walkways. FCFS, in contrast, avoids excessive walkways. For instance, the median walkway per served reservation at spatial flexibility of $\bar{\omega} = 1.5$ is 0.8 km for LUSF; FCFS requires only

---

[17]We deem such response times to be acceptable in the context of web applications. However, should the need arise, there is certainly still room for optimization.
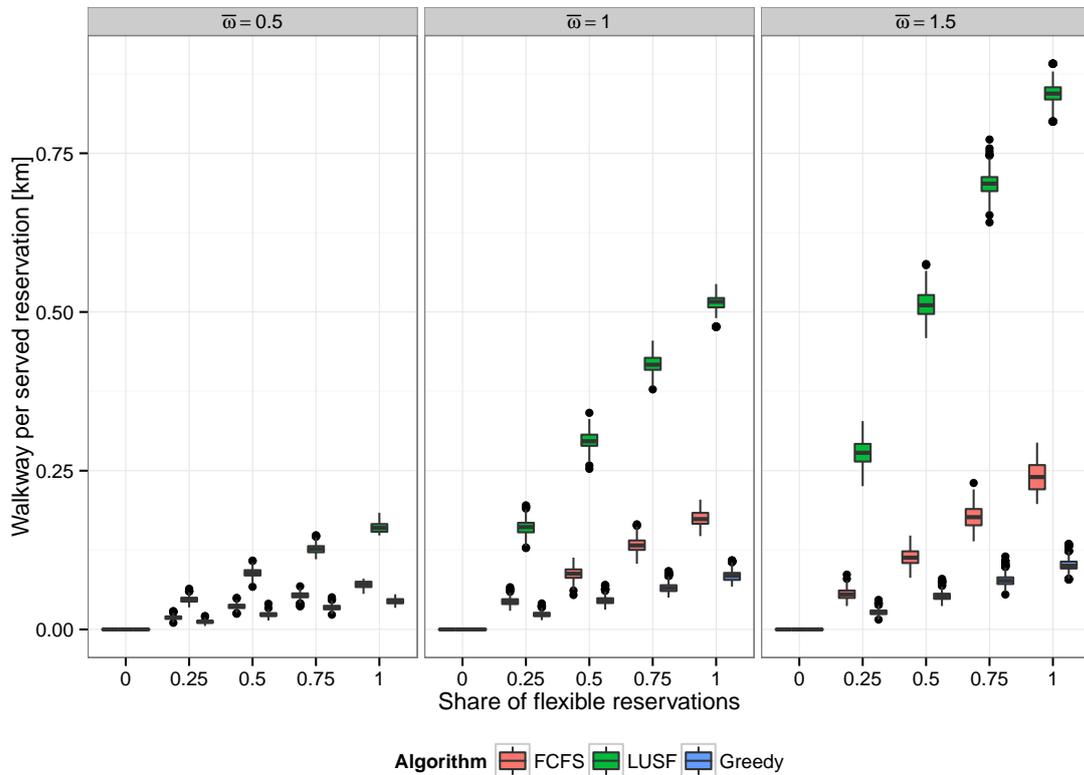
Figure 7.9: Walkways at different levels of spatial flexibility and share of flexible reservations. GREEDY achieves lower walkways per served reservation than its naïve competitors.

0.23 km, a difference of more than 200%. GREEDY beats both its competitors, LUSF and FCFS. Average walkways are only 0.1km for the corresponding parameter combination, an impressive 8-fold decrease over LUSF. This result is of high importance, as consumer convenience is a critical factor upon which the success and acceptance of algorithmically modified reservation-vehicle assignment decisions hinges critically.

In general, our results indicate that reducing fleet size comes at reasonable costs in terms of QoS and consumer inconvenience. We thus conclude that customer flexibility can be leveraged not only theoretically under perfect information (offline), but also in the practically relevant online case with access only to imperfect information.

Accompanying higher utilization and smaller fleet size, Fig. 7.10 illustrates the change in vehicle utilization over flexibility, with the algorithms color coded. Again, panels indicate reservations' spacial flexibility while the share of flexible reservations is given on the abscissa. Adding spatial flexibility increases the number of served reservations; as the number of vehicles is fixed, fleet utilization increases in parallel. In line with the results on served reservations, GREEDY yields highest fleet utilization. This result is especially pronounced for low levels of spatial flexibility. In these situa-
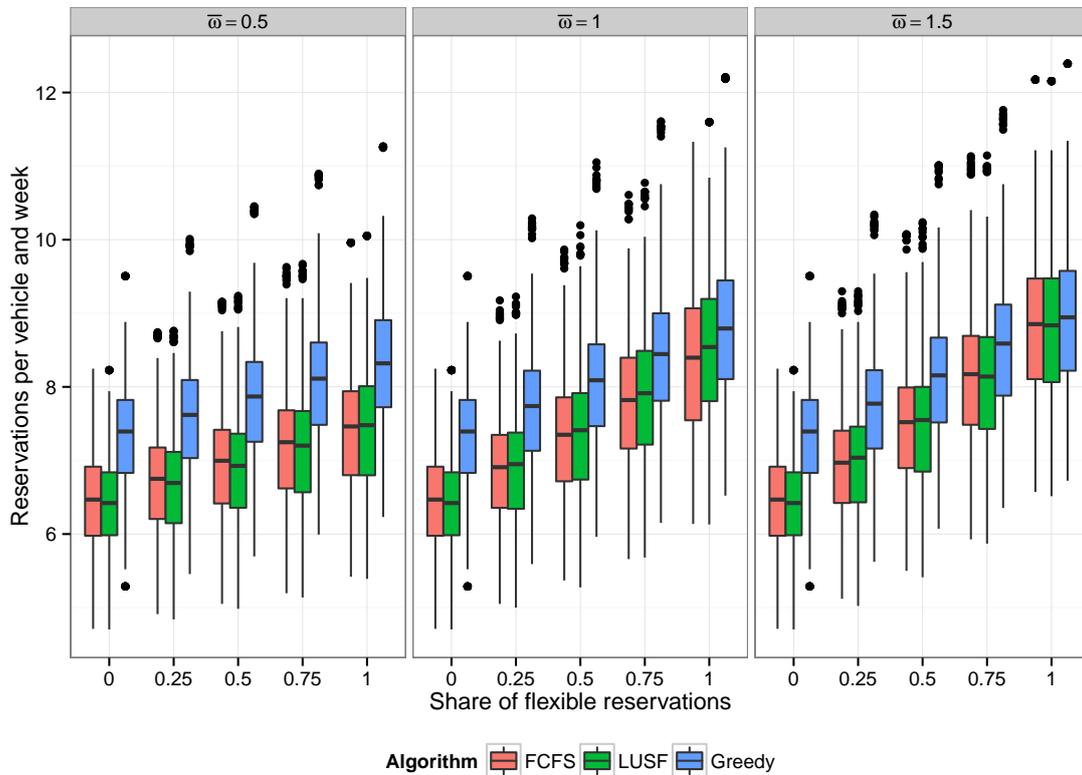
Figure 7.10: Served reservations per vehicle and week over share of flexible reservations differentiated by maximum spatial flexibility per reservation (panels). Flexibility increases fleet utilization. Similar to previous results, GREEDY can leverage flexibility better than its competitors.

tions, careful evaluation of assignment decisions is especially worthwhile.

7.4 Discussion

Our results on fleet size reduction by means of consumer flexibility are based on an *offline* optimization problem (Section 7.2.3), which assumes perfect knowledge of all reservations. In the more realistic *online* setting not all reservations can be served due to the reduced fleet, and lack of information about the future. Nevertheless, Section 7.3.4 documents the realistic potential in the online case. Here, it is interesting to see that the economic potential of customer flexibility can be leveraged effectively by intelligently assigning reservations to vehicles. On the one hand, this allows to significantly reduce fleet sizes in the first place. On the other hand, our results illustrate that service quality to consumers is not degraded tangibly under reduced fleets, po-

tentially fostering user acceptance of the proposed fleet adaptations.

The online optimization models presented so far rely on different information sets. FCFS and LUSF only require information on reservations that must be served at the respective point in time. GREEDY, in contrast, requires information on reservations that have been entered into the system, but have not been decided upon, yet. Fortunately, in the car-sharing domain, consumers make their intended actions known to the system before-hand. Advanced reservation provides a valuable information source that can immediately be used to the advantage of the fleet operator. Thus, relying on available information *only* a la GREEDY seems not to be affecting performance in a pronounced way. A challenge that we have left for future research concerns the inclusion of uncertain information regarding future reservations. The main challenge, identified in extensive testing, is to either develop accurate models of future reservations that include complex spatio-temporal relationships or, simply rely on past reservations for predicting the future (known as "back casting"). The issue of appropriately blending both, scenarios of uncertain future (virtual) reservations and deterministic reservations that have already arrived in the system, poses a special challenge in this context. In particular, the corresponding algorithm must be designed such that the inclusion of uncertain future (i.e., virtual) reservations in the decision making process improves outcomes. In particular, special emphasis should be placed on avoiding the preemption of actual reservations.

Adaptation of Optimal Fleets   As mentioned before, our results should be taken with caution: The reservation information on which the reference fleet is optimized in the *offline* case is a snapshot of demand for a very limited period of time. Depending on which period to optimize the fleet structure for, i.e., a period of low/high demand, results on QoS and walkways might differ. Nevertheless, the car-sharing provider has the opportunity to adapt fleet size and stationing decisions over time, controlling the adverse effect of demand variation on system performance. By doing so, he can make granular decisions regarding the trade-off between fleet utilization and the inconvenience faced by consumers. In other words, leeway on the strategic and tactical decision level provides the car-sharing operator with the opportunity to continuously balance short term (e.g., profit) and long term (growth and market share) goals.

Parking Spaces and Capital Expenditures   Finally, by leveraging consumer flexibility intelligently, better service and vehicle availability can be provided to consumers. The implications might be far reaching: Increased fleet utilization effectively allows to circumvent the historically most significant bottleneck to operations – the lack of parking spaces. Accordingly, smarter assignment decisions and higher fleet utilization provide a means to instantaneous growth without major capital expenditures for operators. This reason alone might serve as a powerful incentive to re-think current practices.

Flexibility regarding Vehicle Class   In the work so far, we have deliberately ignored one additional lever for car-sharing operations optimization, i.e., customer flexibility with respect to vehicle class. For fleet operators offering only one class of vehicle, this restriction is clearly not of interest. However, in typical station-based car-sharing settings, multiple vehicle classes are offered. On the one hand side, vehicle-class flexibility could provide another means to achieve more efficient vehicle assignments (reducing fixed costs due to reduced fleet size at the expense of increased variable costs). This is especially attractive from an operation's point of view if vehicle usage was complementary during the course of the week, i.e., high utilization of larger vehicles throughout the week, i.e., for business purposes such as corporate travel, and high utilization of smaller vehicles on the weekend, i.e., for shopping purposes, or vice versa. On the other hand, leveraging inter-class vehicle flexibility might introduce more pronounced incentive issues if self-interested customers are assumed.[18] Without appropriate allocation and payment rules, customers might try to gamble the system by strategically misreporting their true preferences, i.e., reserving a "small" vehicle when in fact a "large" vehicle is what the customer is speculating on. The cost of fuel is included in the reservation price. Hence, upgrading reservations to higher vehicle classes is, from the perspective of the operator (relative to classical car-rental) costly.

Nevertheless, through appropriately designed incentives, flexibility regarding vehicle class can provide an additional lever to improve the economics of car-sharing. In any case, the design of economic mechanisms (not only in car-sharing) with the goal of improving upon existing, possibly simpler, mechanisms, must anticipate strategic consumer behavior and either limit or deal with it in such a manner that reporting one's true preferences becomes a dominant strategy (Parkes, 2007).

Retrospective Data   The work at hand is limited in a sense that it relies on reservations that have actually taken place only (based on retrospective data), but does not take into account intended reservations that were never carried out (e.g., due to a lack of available vehicles). The observability of "out-of-stock" situations (Verbeke et al., 1998) as well as consumer-reaction to it is a common problem in retailing and OM, hence, we do not focus on challenges associated with it. However, we suspect that in the context of car-sharing and web-based user interfaces, data on consumer interaction with the reservation interface could yield valuable insights in how consumers cope with lack of vehicle availability, e.g., postponing their reservation to a later time, switching vehicle classes, or switching to nearby stations where the originally desired instance of mobility can be provisioned successfully. The resulting insights could assist in making efficient fleet expansion decisions on the strategic decision level as well.

---

[18]With car-sharing becoming main-stream, incentive-compatibility of allocation mechanisms based on private information should receive additional attention, as the average customer may increasingly be self-interested, and less attached to the original goal of sustainable transportation.

Beyond out-of-stock situations, we also do not deal with the influence of differing lead time on consumer decisions and the operational flexibility arising from different levels of lead time (Hua et al., 2010).

Imprecise Data   Furthermore, one drawback of the data used in this study concerns the lack of actual consumer location (home, work, etc.), or desired pickup location data, respectively. Instead, in all of our analyses, we rely on the location of the station at which the customer chose to reserve a vehicle (revealed preferences). As a consequence of this imprecise data, larger (or smaller) amounts of consumer flexibility might be necessary to achieve the described economic outcomes. While we suspect these deviations not to be material, our results should nevertheless be treated with caution.

7.5 Extensions and Managerial Implications

Building upon our results in an established car-sharing environment characterized by a dense station network, we claim that (mostly spatial) consumer flexibility should be incentivized and utilized in order to achieve more efficient reservation-vehicle assignments. Our results serve as a first step towards broader adoption of car-sharing.

We formulate three recommendations to the car-sharing operator to partly utilize the described potential:

Adaptation of the Reservation Interface   One avenue for improvement might be as simple as adapting the reservation interface. Instead of letting the consumer make assignment decisions, the least detrimental (with respect to system performance) assignment given the current state of the system (including advance reservations) could be recommended. If the consumer is not content with the recommendation, she can always choose to deviate from the recommendation and choose a more convenient, i.e., situated in closer proximity, vehicle. Thus we are not proposing to force the system-optimal choices on the customer, but give her a (slightly) biased initial choice upon which she can improve at the expense of system efficiency.

Implementation of a Reservation-Vehicle Assignment (Software) Layer   Conditional on the introduction of an additional software layer, customers no longer make decisions only by themselves, but rather submit abstract reservation requests (time frame, vehicle class, preferred station, e.g., "small vehicle in the city center") that are efficiently assigned to the vehicle fleet by an online planner. Following the idea of "opaque" selling, the downside of such an additional software layer is that the customer loses control over the details of the specific assignment. Thus this variant could be more invasive than the adaption of the reservation interface. Hence, this approach

requires a careful design in order not to harm customer acceptance of car-sharing in general.

Introduction of an Augmented Payment Structure   A third, and more involved way of leveraging customer flexibility in car-sharing could be to base payments on the amount of flexibility the consumer offers. The payments could be designed to account for flexibility either in a static way or in a dynamic way. In the former, monthly member fees could be adapted with respect to the customer class the customer self-selected. Flexible consumers could for example be rewarded through reduced monthly membership fees. In the latter case the change in payment due to flexibility could be based on time of day (similar to time-of-use rates in modern electric power systems), or, more advanced, on real-time demand for vehicles in immediate and neighboring areas. Potentially, the payment of a consumer could be determined by the externalities she is causing to the system, i.e., in terms of crowded-out/postponed competing reservations. Here, special attention must be payed to ensure that consumers are not encouraged to gamble the system, as this might lead to reduced system efficiency through low utilization and reliability.

7.6 Conclusion and Outlook

Based on our results we conclude that moderate levels of spatial consumer flexibility in car-sharing are sufficient to significantly reduce fleet size (e.g., between $10\% - 25\%$ at spatial flexibility of $\bar{w} = 1.5km$). High QoS can be achieved even in the online case ($> 95\%$) at massively reduced fleet sizes, while retaining reasonable average consumer walkways (≤ 0.2km). Here, the choice of algorithm attains high importance, as it is the deciding factor to achieve both, high QoS and low walkways. Temporal flexibility, on the other hand, is virtually useless in order to foster higher fleet utilization in car-sharing. Furthermore, we posit that periodic re-adjustments of fleet size might attain an economically more favorable balance of supply and demand under varying seasonal demand patterns (Ehrenthal et al., 2014).

   We are of the opinion that the most interesting question for future research, in different directions and research communities in the domain of (traditional) car-sharing, are as follows:

- The question of user-interface adaption and associated user acceptance alone poses a wide array of interesting research questions. Most notably, leveraging consumer flexibility will not come at zero cost. Therefore, it will be interesting to explore consumers' willingness to provide flexibility and the associated reimbursement required.

- A further extension concerns the impact of varying lead times on solution quality. The rational here is as follows: The later a user must be informed about

the eventual vehicle assignment decision of her reservation, the larger the solution space that can be used by the car-sharing operator. The gain from notifying users at a later time must be carefully weighted against acceptance issues that might arise from such late-notification schemes. One possible variation could involve immediate notification of the vehicle after entering the reservation intention into the system. However, this initial decision could be altered to account for unexpected reservations entering the system at a later time. By re-optimizing the schedule via re-assignments of reservations to vehicles, economically preferable outcomes can be achieved. Clearly, to avoid nervous schedules, parameters involving the number and associated penalty of re-optimization must be set carefully.

- The questions how to exactly design incentives requires careful design and evaluation. The allocation mechanism could (and in our opinion should) be adapted in order to achieve incentive compatibility, i.e., for users it should be the best strategy to reveal their (potentially complex) private information with respect to flexibility and valuation. To enable such a mechanism, the objective would have to be modified to, for example, include social welfare or provider profit, departing from treating reservations uniformly. The modified mechanism would require users to reveal information regarding their willingness to pay for service at a specific location, time and vehicle class, and thus introduce additional complexity. Again, any such change should be carefully designed and tested, before considering a broad roll-out.

The vision of self-driving cars (Thrun et al., 2006; Thrun, 2010) might lead to fundamental repercussions with car-ownership as well as operations of car-sharing providers. The introduction of self-driving cars, might further enhance the car-sharing experience for consumers and reduce the importance of consumers' spatial flexibility regarding more efficient operations. However, the central issue of economics, i.e., scarcity and the efficient allocation of goods and services, will continue to pose interesting economic problems. Data-driven decision-making may be of high value in propelling this idea from a vision state towards an actually sustainable addition to individual transportation. On this path, classic research questions concerning the related dial-a-ride problem (Cordeau and Laporte, 2003) and related questions might receive renewed attention: Where should vehicles be stationed? Which vehicle to use to serve a specific request? Which routes should be used? What is the optimal fleet size given a projection of demand? What is the value of being able to relocate vehicles? Beyond the questions concerning the assignment of single reservations to vehicles, interest might be sufficiently sparked to also cover questions concerning the related domain of ride-sharing, i.e., which riders to pick up on what path.

Irrespective of technological advances that might enable self driving cars and other innovations, the associated economic questions of interest will change but nevertheless can be expected to retain importance.

Chapter 8

Leveraging Consumer Flexibility in Electrified Car-Sharing

Car-sharing operations based on ICE-propelled vehicles constitute a first step towards more sustainable, urban transportation. Further improvements may be found in (locally) emission-free fleets of EVs that pose both an ecologically sustainable and economically viable complement to existing means of urban passenger transportation.

In this chapter we focus on the role EVs may assume in car-sharing fleets. In more detail, we examine the trilemma that a designer of electrified car-sharing systems faces: It involves trading-off required consumer flexibility, QoS and operating costs. To this end, we employ a subset of the empirical dataset introduced in Chapter 6 and examine the economics of electrified car-sharing in the presence of spatial consumer flexibility.

This chapter is structured as follows: We first describe the scenario, then we formulate the problem as modification of the Multi-Knapsack Problem (MKP) (Nemhauser and Wolsey, 1988), present evaluation results and discuss them, and provide a brief outlook into promising future research.

8.1 Problem Formulation

Optimizing car-sharing fleets comprised entirely of conventionally propelled vehicles, we aimed, in the previous chapter, to minimize fleet size, which poses a good proxy for minimizing total cost.[1] To this end, we adapted the bin-packing problem and, hence, minimized the number of bins (vehicles) to serve all reservation requests.

In the wake of EVs, however, modifications to the problem formulation become necessary, as not all (sequences of) reservations are compatible with the constraints imposed by EV technology (time required for re-charging, sufficient SoC to complete trips). On the one hand, the introduction of EVs can facilitate more competitive fleet operations through reduced operational expenses. On the other hand, the operator

---

[1]Assuming a single vehicle technology and the constraint that any reservation must be served, finding a cost-minimal fleet is equivalent to minimizing fleet size.
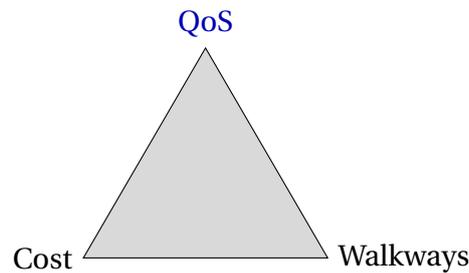
Figure 8.1: Decision trilemma in electrified car-sharing

may find itself with a less potent fleet, no longer able to fully serve peak demand, and hence, be forced to discard some reservations. The arising decision trilemma of a car-sharing fleet operator, involving cost, QoS, and walkways, is depicted in Fig. 8.1.

If consumers are flexible, fleet operators may be able to achieve both, improved utilization, which entails lower cost, through appropriate flexibility dispatch, and high QoS. In this vein, EVs may pose an interesting complement to fleets in shared mobility settings. Accordingly, we aim to answer the following question in this chapter:

Research Question 11: *To what extent does spatial consumer flexibility enable economically efficient electrification of car-sharing fleets?*

In order to determine the cost-minimal fleet serving all reservations in the reservation stream, consisting of both, conventional and electric vehicles, a computationally involved GAP (Cattrysse and Van Wassenhove, 1992; Roth and Sotomayor, 1992) would have to be solved. The solution comprises decisions on

- the number of vehicles in use,

- the type of each vehicle in use (conventional or electric),

- the stationing of each vehicle,

- the assignment of reservations to vehicles, as well as

- when to charge each EV.

The problem of finding the optimal solution to a GAP is NP-complete (Nemhauser and Wolsey, 1988). Solving the problem to optimality, irrespective of recent advances in MIP, is impractical for realistically sized instances, especially under higher levels of spatial flexibility. Therefore, we refrain from describing this model in detail and turn to a different formulation in which we aim to separate the problem into its strategic, tactical, and operative layers, and apply heuristics on the top two layers to reduce

computational complexity. The three layers we will refer to in the following are as follows:

1. Strategic – the number and types of vehicles in use,

2. Tactical – the vehicle-station assignment decision, and

3. Operational – the reservation-vehicle assignment decisions as well as the charging decisions.

By means of this separation, we trade-off optimality for reduced complexity. while still retaining meaningful, and valuable results for realistically sized problem instances.

8.2 Reservation Assignment as a Multi-Knapsack Problem

If strategic and tactical fleet composition decisions are taken care of beforehand, the remaining reservation-vehicle assignment problem clearly constitutes a problem of reduced complexity. Each vehicle can then be interpreted as a *knapsack* and the planner seeks to maximize the number of reservations placed into multiple knapsacks by appropriate assignment decisions. Accordingly, we model the assignment of reservations to vehicles as a MKP. Optimizing the corresponding performance criterion, the optimal solution yields information on reservations-vehicle assignment decisions leveraging spatial consumer flexibility. Usually, the problem is formulated with the objective of profit maximization. However, the profit is not easily defined in our setting; we, hence, require a different measure for meaningful optimization, and opt for a multi-objective criterion including three, in our opinion highly relevant, criteria (Fig. 8.1) of QoS ($\mathscr{Q}$), consumer walkways ($\mathscr{W}$) and cost ($\mathscr{C}$).

8.2.1 Fleet Determination

In order to compute reservation-vehicle assignments in the MKP, fleet configurations need to be established first. Vehicles constitute the essential constraints of the problem, serving as knapsacks in the MKP. Before solving the problem, we derive plausible fleet sizes and configurations, starting with the minimally required fleet under the assumption of zero spatial customer flexibility. Thereafter, we reduce fleet size and increase EVs' share (Fig. 8.2).

Initial Fleet   First, we determine the minimally sized fleet consisting of conventional vehicles only. This fleet is able to serve all reservations in the absence of consumer

Electrified fleet   Smaller, electrified fleet

substitute   substitute
conventional   conventional
with EVs   with EVs

Initial fleet ⟶ Smaller fleet
reduce fleet size
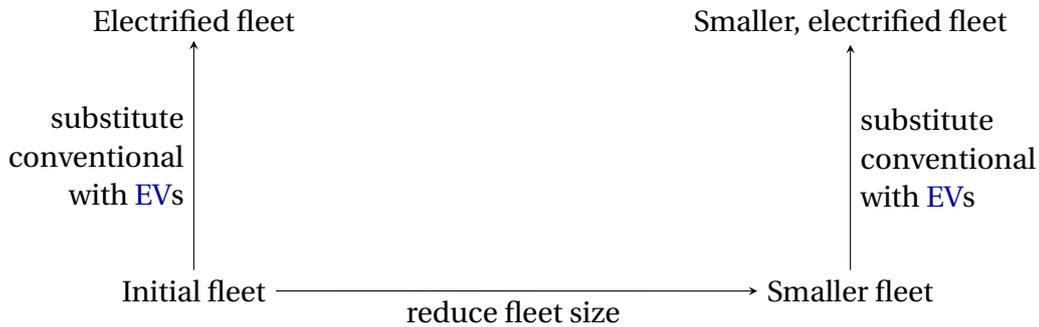
Figure 8.2: Determination of mixed fleets on the strategic decision level.

flexibility. Fleet information comprises number and stationing of vehicles. In our subsequent economic evaluations, we examine the implications of substituting conventional vehicles with EVs. This initial fleet serves as the starting point to the derivation of the following fleets.

Substituting Conventional by Electric Vehicles   In a first modification step, parts of the existing fleet need to be converted into EVs (Fig. 8.2). To this end, the question arises, which particular vehicles to "convert" to EVs. Interestingly, the requirement to convert an integer number of vehicles from conventional to electric, yields a problem similar to the problem of assigning parliamentary seats to parties. Therefore, we leverage the Sainte-Laguë Algorithm (Alg. 9) to convert the desired share of vehicles from the initial fleet into electric vehicles.

The total number of EVs (Ξ) can be interpreted as the number of seats in the parliament; stations then are the equivalent of parties (P), while EVs at a particular station are the members of the corresponding fraction. Votes per party (Θ) coincide with the number of vehicles originally stationed at the corresponding station. Hence, all fleet adaptation decisions are based on original fleet and station size.

Our choice of the Sainte-Laguë algorithm stems from its property of "proportionality" (Lijphart, 2003), i.e., the number of seats assigned to a party is considered closer to proportionality than under different regimes, such as d'Hondt.[2]

Fleet Size Reduction and EV Substitution   As illustrated in Fig. 8.2, besides pure substitution of conventional vehicles by EVs, we are interested in the (joint) effect of fleet size reduction and EV substitution. Both, reduction and substitution are performed according to Alg. 9. Thereby, fleet size is reduced in a first step. In a second run of the algorithm, the number of EVs at each station is determined. Thereby, the com-

---

[2]The algorithm has received growing attention in worldwide parliamentary elections, including German federal elections. For instance, it is used to determine the number of fractions' seats in the lower house of the German federal parliament.

Algorithm 9: SAINTE-LAGUË; Computes seats per party θ_p, $\forall p \in P$ from votes per party Θ and total number of available seats in the parliament Ξ. Used to assign parliamentary seats to parties in municipal, state, and federal elections.

   Input: P, Θ, Ξ
   Output: θ
1 $\theta_p \leftarrow 0 \quad \forall p \in P$
2 **while** $\sum_{p \in P} \theta(p) < \Xi$ **do**
3   **for** $p \in P$ **do**
4        $q_p = \Theta_p / (2\theta_p + 1)$
5     $p^* = \arg\max_p q$
6     $\theta_{p^*} = \theta_{p^*} + 1$
7 **return** θ

---

puted, intermediate number of vehicles per station serves as input to EV stationing decisions.

8.2.2 Decision Variables

The optimization program decides on the assignment of reservations $r \in \mathcal{R}$ to vehicles $v \in \mathcal{V}$. To this end, we again make use of the formulation established in Section 7.2.3 regarding the decision variable $x_{r,v}$. For each problem instance, the number, technology and stationing of vehicles of the fleet is determined exogenously. For the sake of more accessible modeling, we partition the set of vehicles $\mathcal{V}$ into disjoint sets of electric and conventional vehicles, i.e., $\mathcal{V} = \mathcal{V}_e \cup \mathcal{V}_c$, where $\mathcal{V}_e \cap \mathcal{V}_c = \emptyset$. Inspired by the problem formulation in (Schuller et al., 2014; Flath et al., 2013) we partition time into $\mathcal{T}$ discrete periods of equal length. Beyond assignment decisions, further charging decisions are formally expressed via the continuous decision variable $\phi_{v,t} \in [0, 1]$. Note that we express charging in terms (of fractions) of battery capacity C. However, vehicle charging is restricted to times when the vehicle is at its station, modeled via the binary variable $\kappa \in \{0, 1\}$. Furthermore, the battery is discharged during driving, which we model via the continuous decision variable $\gamma \in [0, 1]$, again in terms of capacity C.

8.2.3 Objective

Assigning a well-defined profit measure to individual reservations is not a straightforward task. On the one hand, consumers value vehicle availability. On the other hand, excessive fleet provisioning incurs high costs, contrary to the operator's interest. Pure cost minimization, however, leads to low service quality and might in the long term deter consumers from using the service at all.

There are, however, several possible ways to formulate the objective function in the multi-objective case (Marler and Arora, 2004). One possibility for removing di-

mensionality lies in normalizing the objective functions. If the planner's preferences are known, multi-objective optimization problems can be converted into ordinary, single-objective optimization problems, involving, for instance, weighted sums.

In order to circumvent additional complexity, we apply the classic weighted sum method (Marler and Arora, 2004). The individual criteria are

- QoS (share of reservations served), expressed via $\mathcal{Q}$,

- cost to the operator (monetary), formally $\mathcal{C}$, and

- cost to the consumer (in terms of walkways), formally $\mathcal{W}$.

As we would like to explore the trade-off between provider cost, consumer inconvenience and QoS (see also Sec. 8.2), this approach arguably constitutes a reasonable choice. We believe that reasonable assumptions regarding the relative weights of criteria yields tractable, valuable insights into the economics of electrified car-sharing. Hence, the objective function can be formulated as

$$\max_{x,\phi,\kappa,\gamma} \Gamma = \alpha \cdot \mathcal{Q} - \beta \mathcal{W} - \mathcal{C} \tag{8.1}$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$ are the corresponding exogenously defined weights regarding QoS and walkways.

8.2.4 Constraints

In order to obtain valid solutions, the decision variables must be constrained appropriately.

Reservation-Vehicle Assignment   Assignment decisions are binary, eq. (7.3). Each reservation may be assigned to a single vehicle, only, as each reservation is served only once, or not at all. Note the inequality, which differentiates this constraint from (7.4).

$$\sum_{v \in \mathcal{V}} x_{r,v} \leq 1 \qquad \forall r \in \mathcal{R} \tag{8.2}$$

Further, a reservation can only be served by vehicles at the desired and adjacent stations, i.e., stations within the perimeter of spatial flexibility, i.e., $d(\sigma_r, \sigma_v) \leq \bar{w}$, where the distance between two points is denoted by d. Correspondingly, *other* reservations are made impossible via the following constraint.[3]

$$x_{r,v} = 0 \qquad \forall (r,v) \in \{(\mathcal{R}, \mathcal{V}) | d(\sigma_r, \sigma_v) > \bar{w}\} \tag{8.3}$$

---

[3]For efficient implementation, it may be advisable not to create the corresponding variables in the first place. We choose this presentation for ease exposition.

Section 7.2.3 includes an alternative formulation for valid assignment decisions leveraging the idea of sets of competing reservations ($\mathcal{R}^{comp}$).

Scheduling   In order to achieve valid schedules, only a *single* reservation may be served by a vehicle at any time. This constraint is formalized in inequality (7.18).

Electric Vehicle Charging   If EVs are present in the fleet under consideration, i.e., $\{\mathcal{V}_e \neq \emptyset\}$, recharging between reservations is required. A vehicle can only be charged during times of inactivity, i.e., when it is not in use. Due to vehicles being shared, trips (reservations) are not tied to specific vehicles. Therefore, the assignment of any reservation r to vehicle v during time t renders the vehicle non-chargeable ($\kappa_{r,v} = 0$). For modeling purposes, let B be a matrix of dimensions $|\mathcal{R}| \times |\mathcal{T}|$. An element $b_{r,t}$ of this matrix assumes value one, if reservation r is active during time t, and zero otherwise. Accordingly,

$$\kappa_{v,t} = \sum_{r \in \mathcal{R}} 1 - (b_{r,t} \cdot x_{r,v}) \qquad \forall v \in \mathcal{V}_e, \forall t \in \{1,..,\mathcal{T}\} \tag{8.4}$$

This vehicle-availability condition constrains possible charging patterns. Furthermore, the maximum possible charge-rate c_{max} as well as SoC from the previous period constrain EV charging and require adequate modeling.

$$\phi_{v,t} \leq c_{max} \cdot \kappa_{v,t} \qquad\qquad\qquad \forall v \in \mathcal{V}_e, \forall t \in \{1,..,\mathcal{T}\} \tag{8.5}$$

$$\phi_{v,t} \leq (1 - SOC_{v,t-1}) \cdot \kappa_{v,t} \qquad\qquad \forall v \in \mathcal{V}_e, \forall t \in \{2,..,\mathcal{T}\} \tag{8.6}$$

Discharging of electric vehicles' batteries takes place only during driving.[4] Given the set of reservations $\mathcal{R}$ and the characteristics of a specific vehicle technology (e.g., Table 8.1) the (average) amount of energy required for driving purposes in each period $\underline{R}$, in terms of battery capacity C, can be computed in a straightforward manner. Distance of reservation r is encoded in d_r. Electric vehicles' fuel efficiency in $\frac{kWh}{km}$ is denoted via η_e.

$$\underline{R}_{r,t} = \frac{1}{C} \cdot \frac{b_{r,t}}{\sum_t B_{r,t}} \cdot d_r \cdot \eta_e \tag{8.7}$$

Equation 8.8 constrains the amount of energy discharged from vehicle v in period t to a value no less than the amount of energy required for reservation r, iff r is assigned to v (hence the product constraint). Discharge of vehicle v at time t is denoted $\gamma_{v,t}$.

$$\gamma_{v,t} = \sum_{r \in \mathcal{R}} \underline{R}_{r,t} \cdot x_{r,v} \qquad \forall v \in \mathcal{V}_e, \forall t \in \{1,..,\mathcal{T}\} \tag{8.8}$$

---

[4]We do not explicitly allow for feeding energy from a vehicle back into the power grid. However, under time-varying electricity prices, this may pose an interesting extension.

To achieve continuity over time regarding SoC, we introduce the following constraint.

$$SOC_t = SOC_{t-1} + \phi_t - \gamma_t \qquad \forall\, t \in \{2,..,\mathcal{T}\} \tag{8.9}$$

Furthermore, SoC must be constrained to technically valid values.

$$0 \le SOC_t \le 1 \qquad \forall\, t \in \{1,..,\mathcal{T}\} \tag{8.10}$$

For easier comparability, we require a fully charged battery at the initial and final period under consideration.[5]

$$SOC_{v,1} = 1 \qquad \forall\, v \in \mathcal{V} \tag{8.11}$$

$$SOC_{v,T} = 1 \qquad \forall\, v \in \mathcal{V} \tag{8.12}$$

Walkways, QoS, and Cost   Flexible consumers may be required to walk for a certain distance to the assigned vehicle. This walkway is equals the distance between the initially chosen station and the allocated station and is modelled as follows.

$$w_r = \sum_{v \in \mathcal{V}} d(\sigma_r, \sigma_v) \cdot x_{r,v} \qquad \forall\, r \in \mathcal{R} \tag{8.13}$$

Aggregate walkways are the sum of walkways over all reservations.

$$\mathcal{W} = \sum_{r \in \mathcal{R}} w_r \tag{8.14}$$

The achieved QoS is modelled via the share of served reservation.

$$\mathcal{Q} = \frac{1}{|\mathcal{R}|} \sum_{r \in \mathcal{R}} \sum_{v \in \mathcal{V}} x_{r,v} \tag{8.15}$$

Total cost of service provision by the operator is jointly determined by fixed and variable costs. Fixed costs are mainly due to the depreciation of the investment charged to the period under consideration ($\mathcal{H}$). To this end, P_e and P_c denote the price of a new electric or conventional vehicle, respectively. Variable costs arise from operation of the vehicle and are directly related to assignment decisions. The price of gasoline (per litre) is denoted p_g, that of electricity p_e. Fuel efficiency of conventional vehicles is represented via η_g.

---

[5]Note that this may lead to artifacts in charging patters towards the end of the period under consideration. However, as this is not the main focus of this study, we accept this shortcoming.

$$\mathscr{C} = \sum_{t \in \{1,...,\mathscr{T}\}} \sum_{r \in \mathscr{R}} \sum_{v \in \mathscr{V}_e} (x_{r,v} \cdot \underline{R}_{r,t} \cdot p_e)$$
$$+ \sum_{r \in \mathscr{R}} \sum_{v \in \mathscr{V}_c} (x_{r,v} \cdot d_r \cdot \eta_g \cdot p_g) \tag{8.16}$$
$$+ \frac{1}{\mathscr{H}} \cdot (|\{\mathscr{V}_e\}| \cdot P_e + |\{\mathscr{V}_c\}| \cdot P_c)$$

8.3 Evaluation

This section first presents and provides the rationale for the chosen input parameters. Thereafter, optimization results are presented and discussed.

8.3.1 Input Parameters

Input to the optimization are the technical characteristics of the Nissan Leaf Visia, a compact-class EV available on the market for purchase (Table 8.1). We assume a depreciation period of six years for both, conventional and electric vehicle. Range is a limiting factor only for the electric vehicle, hence the missing value for the conventional model. In order to compensate for overly optimistic fuel efficiency by manufacturers, we deviate from the official technical data on fuel efficiency and assume 6.8 l/100km, instead of the manufacturing data listed in Tab. 8.1, roughly a 40% increase. On the other hand, we assume fast-charging to be possible; at $c_{max} = 1$, the vehicle's battery is fully recharged within one hour. The price of gasoline is assumed to be 1.55 EUR/l, the price of electricity is set to 0.22 EUR/kWh.

Furthermore, we set the weight assigned to QoS in the objective function (8.16) to be $\alpha = 200$ per reservation, while one km of walkways is weighted at $\beta = 10$. We assume all consumers to exhibit homogeneous spatial flexibility.[6] Furthermore, to avoid excessive solving times, we set the MIP-gap to one percent.

The reservation stream, as well as the station locations, are taken from one week in May 2012[7] from city center stations. Reservations' distance and duration heterogeneity is illustrated in Fig. 8.3.

Fig. 8.3(a) presents a standard empirical Cumulative Distribution Function (CDF) (upper black line) and a distance-weighted CDF (lower blue line). While more than 80% of reservations are used to cover distances smaller than 100 km (Fig. 8.3(a)), half of total distance driven is due to long-distance reservations, i.e., reservations with distances exceeding 200 km.

---

[6]This is in contrast to the previous section. We do not, however, expect the introduction of heterogeneous flexibility to qualitatively alter our results

[7]The dataset contains all active reservations (including partially active ones) during the time of May 7 and May 14, 2012

Dimension	unit	notation	Nissan Leaf Visia[a]	Nissan Note Visia[b]
Range	km		160	
Battery capacity	kWh	C	24	
Charging speed	$\frac{kW}{kWh}$	c_{max}	0.125	
Specific consumption		η	$0.125\frac{\text{kWh}}{\text{km}}$	$4.8\frac{l}{100\text{km}}$
Price	€	P	29 690	13 990
Depreciation horizon	a	$\mathcal{H}$	6	6

Table 8.1: Vehicle specifications based on (Nissan, 2014). For added realism, we deviate from the charging speed and fuel consumption data as inputs to the optimization, see the text for details.

---

[a]http://www.nissan.de/DE/de/vehicle/electric-vehicles/leaf/prices-and-equipment/prices-and-specifications.html, last accessed October 2014

[b]http://www.nissan.de/DE/de/vehicle/city-cars/note/prices-and-equipment/prices-and-specifications.html, last accessed October 2014

Fig. 8.3(b) similarly illustrates reservations' duration heterogeneity. More than 90% of reservations are shorter than 12 hours, illustrated by the upper black line.The median reservation duration of this particular dataset is slightly below 3 h. Long-duration reservations, i.e., exceeding periods of 8 hours, are responsible for approximately half of temporal fleet utilization. Fig. 8.3(c) presents both reservation distance and duration jointly.

Furthermore, we rely on the historically established station network as the underlying spatial structure to the fleet of vehicles. Hypothetical fleets that serve our evaluation needs are determined via the Sainte-Laguë algorithm, introduced in Section 8.2.1. For example, the number of vehicles in the fleet is varied between 70 and 100% of the necessary fleet size in the no-flexibility case. Moreover, we examine the impact of different EV penetration levels, ranging between purely conventional fleets on one side and including up to 30% electric vehicles on the other. For instance, if fleet size is not reduced, and an EV-level of 30% is assumed, fleet size will be 64 vehicles, including 19 electric vehicles. Further details are documented in Table 8.2.

8.3.2 Optimization Results

We first present results regarding the objective function's components, i.e., cost $\mathcal{C}$, walkways $\mathcal{W}$, and QoS $\mathcal{Q}$. In a second step, implications on fleet utilization and details of the assignment decisions are revealed.

(a) Empirical cdf of reserva-
tions' distance

(b) Empirical cdf of reserva-
tions' duration

(c) Reservations' distance over
duration
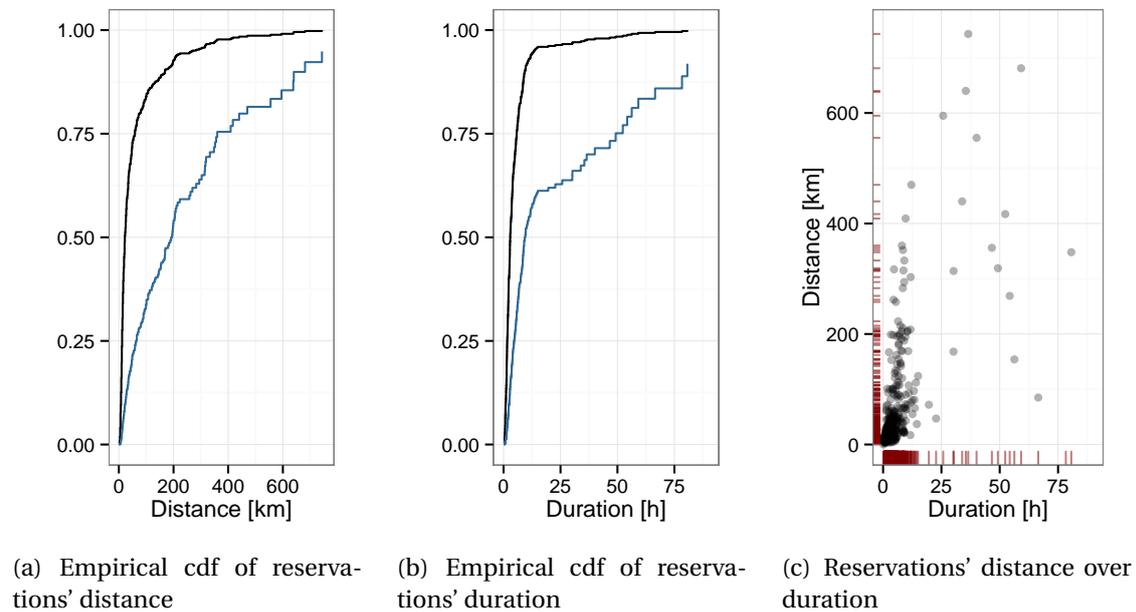
Figure 8.3: Distance and duration characteristics of evaluation data

Cost, QoS, and Walkways   Our result illustrations are faceted by consumers' spatial flexibility $\bar{w}$, varying EV penetration levels use of different line types.

Fig. 8.4 illustrates the cost of fleet operations, differentiated into fixed, variable and total costs. Clearly, variable costs are decreasing in EV penetration level due to relatively low operating cost of EVs. The opposite, however, is true for fixed costs. Moreover, variable costs exhibit a slight decreasing tendency in fleet size F for $\bar{w} \in \{0.5, 1\}$. However, as EVs are more expensive to purchase, increasing the associated depreciation amount, the effect of lower operating cost is overcompensated, yielding increasing total cost in EV penetration level.

Total costs are also increasing in spatial consumer flexibility. While counterintuitive at first sight, spatial flexibility – given our objective – allows a larger set of reservations to be served (thus increasing QoS), correspondingly increasing variable costs (but also payments to the provider). Interestingly, the difference in total cost between a purely conventional fleet and a fleet featuring approximately 30% EVs does not exceed 10% under the selected input parameters.

Fig. 8.5 illustrates an evaluation dimension not explicitly included in the objective, i.e., relative driving distance served. At zero spatial flexibility ($\bar{w} = 0$), higher EV penetration translates into lower service quality, i.e., some reservations can no longer be served. At $\bar{w} = 1$, there is virtually no difference between different fleet compositions. While less than 70% of total distance requested is served under zero flexibility (left facet) only, this value increases to approximately 95% in the case of $\bar{w} = 1$ (right facet).

Fleet size	relative F	70%	80%	90%	100%
	absolute	44	51	57	64
EV-level	0%	0	0	0	0
	10%	4	5	5	6
	20%	8	10	11	12
	30%	13	15	17	19

Table 8.2: Fleet composition, number of electric vehicles

Judging from these results, the influence of spatial consumer flexibility appears more pronounced than effects that can be traced back to fleet modifications.

Fig. 8.6 depicts QoS over fleet size. Again, we find only little influence of fleet size on this performance criterion, while spatial flexibility is the decisive lever for achieving high QoS. For instance, at $\bar{w} = 0$, only approximately 85% of reservations are served, whereas EV penetration slightly reduces performance. At $\bar{w} = 1$, service levels in excess of 98% are consistently achieved. These results indicate that spatial flexibility is a valuable lever to control customer satisfaction and fleet utilization in the case of electrified car-sharing.

Fig. 8.7 illustrates walkways per served reservation, i.e., the additional cost incurred by a flexible consumer. Clearly, at $\bar{w} = 0$ walkways must be zero. At $\bar{w} = 0.5$, average walkways per reservation fluctuate around 0.08 km, with EV penetration yielding slightly higher average walkways, and walkways of 0.16 km at $\bar{w} = 1$. Accordingly, only a small fraction of the assumed flexibility potential is actually employed. Note that these essentially "flat" curves of walkways over fleet size may in large part be due to the chosen combination of weights α and β. If the sole objective was cost-minimization, and walkways were not penalized, larger fleets may foster reduced walkways. In the current formulation, however, all "profitable" reservations, i.e., reservations that augment the objective value, are served. The penalty arising from walkways in the objective does not change the mechanics of the problem, especially if this penalty, weighted through β, is homogeneous over all reservations.

Assignment Decision Details   Fig. 8.8 illustrates the distance covered by technology.

Clearly, the bulk of total distance traveled is assigned to conventional technology in our solution. Furthermore, the share of distance traveled electrically is increasing in EV penetration level and customer flexibility. Combining the insights from Fig. 8.8 with the data in Table 8.2, weekly average driving distance of EVs after optimization ranges between 400 and 550 km, well within the operating range of EVs. For instance, in the case of $\bar{w} = 1$ and no reduction in fleet size (top right facet), around one third of total distance traveled is served electrically.

Fig. 8.9 pursues a different perspective in that it illustrates the number of served
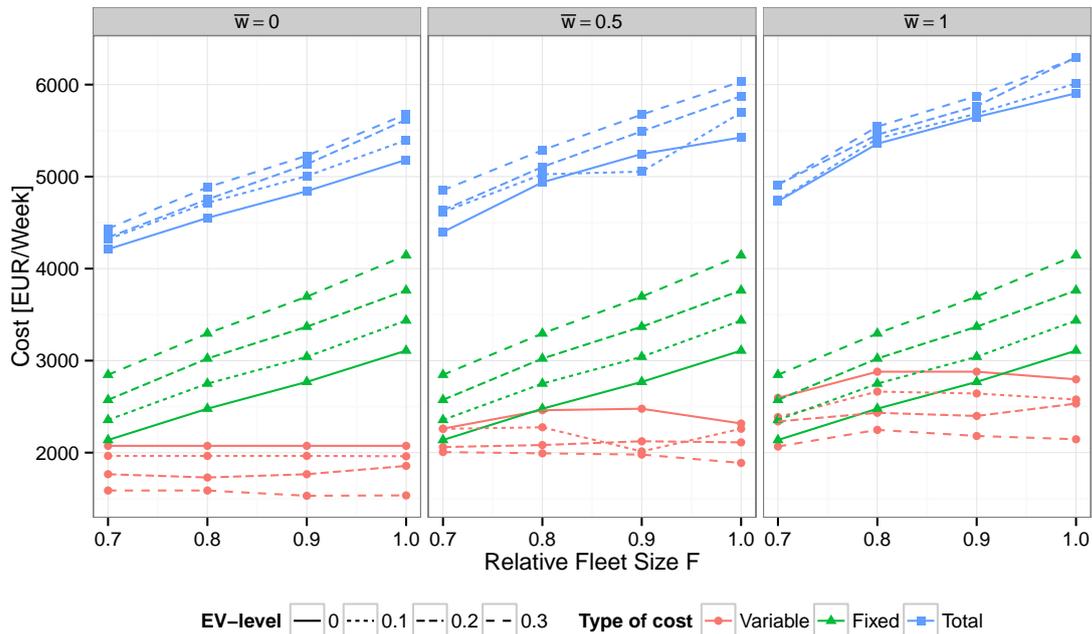
Figure 8.4: Fixed, variable, and total cost $\mathscr{C}$ over fleet size for different EV penetration levels and fleet sizes

reservations per vehicle. Under spatial flexibility, EVs serve more reservations per vehicle than their conventional counterparts. In the most extreme case, 14 reservations per week are served per electric vehicle, while conventional vehicles are only assigned 9 reservations.[8]

Fig. 8.10 illustrates the total number of reservations assigned to vehicles of either technology. Naturally, a larger share of EVs enlarges the share of reservations served electrically. For instance, around 50% of reservations are served electrically under fleets of the original size, 30% EVs, and $\bar{w} = 1$.

8.4 Discussion

Our optimization results indicate that, under quite conservative assumptions regarding the cost of EVs, up to half of reservations could be served electrically, thereby only marginally increasing costs to the provider. This finding stands in contrast to results reported in the literature so far, which are more pessimistic about the use of electric vehicles in car-sharing (cf. Doll et al., 2011). More optimistic results could be obtained, if the cost of EVs was to further decrease. Nevertheless, we find consumer flexibility

---

[8]Case of reduced fleet size $F = 0.7$, $\bar{w} = 1$, and 10% share of EVs.
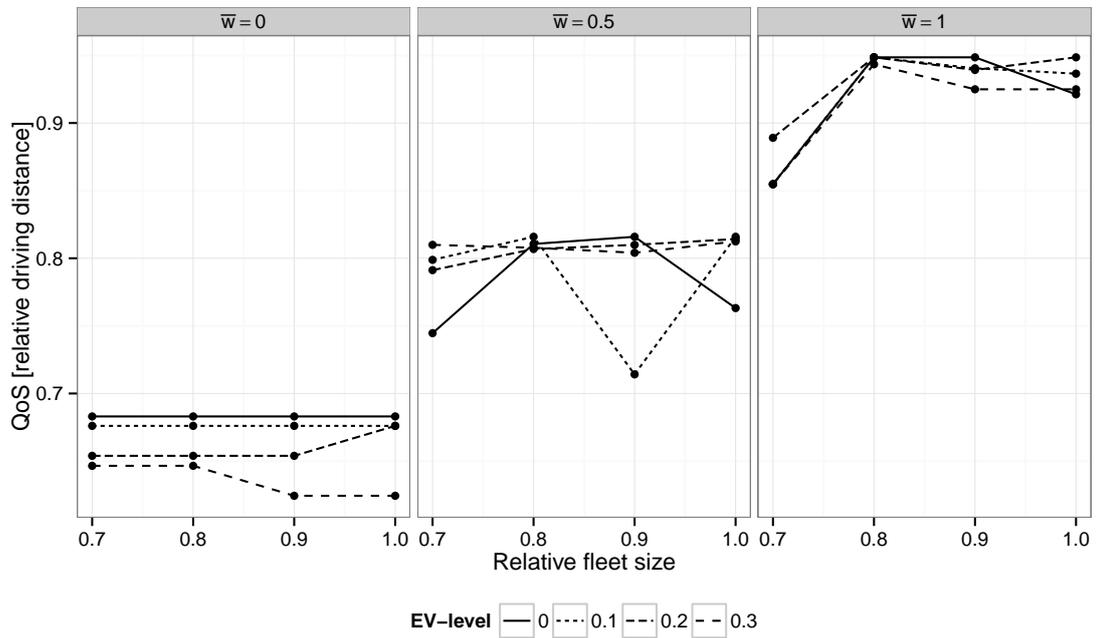
Figure 8.5: QoS in terms of distance driven over fleet size, differentiated by EV penetration level and fleet size
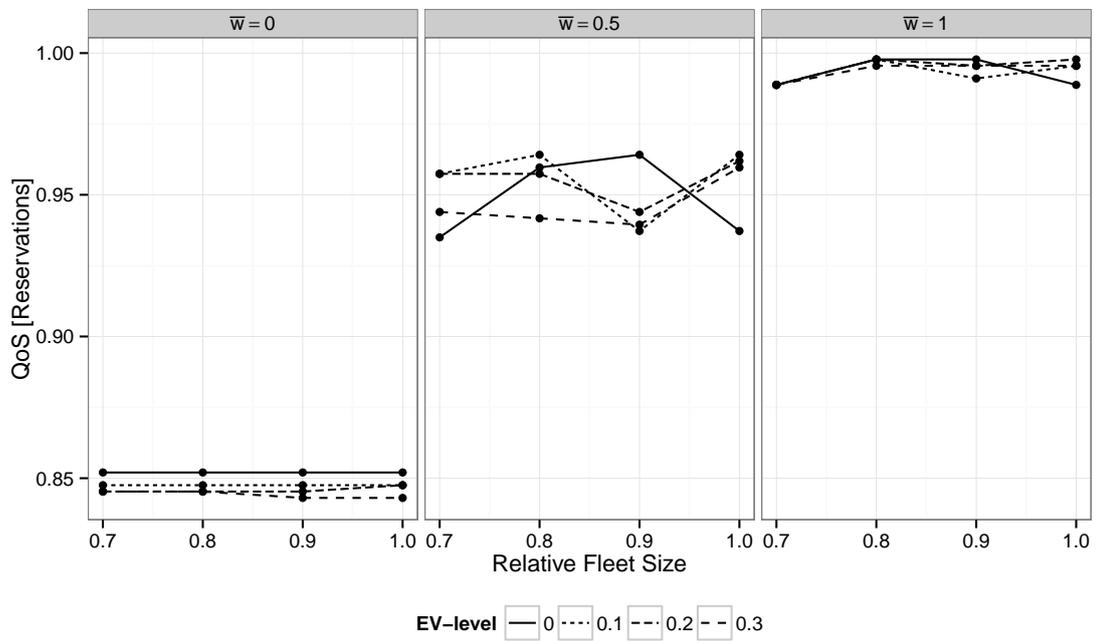
Figure 8.6: QoS over fleet size, differentiated by EV penetration level and fleet size
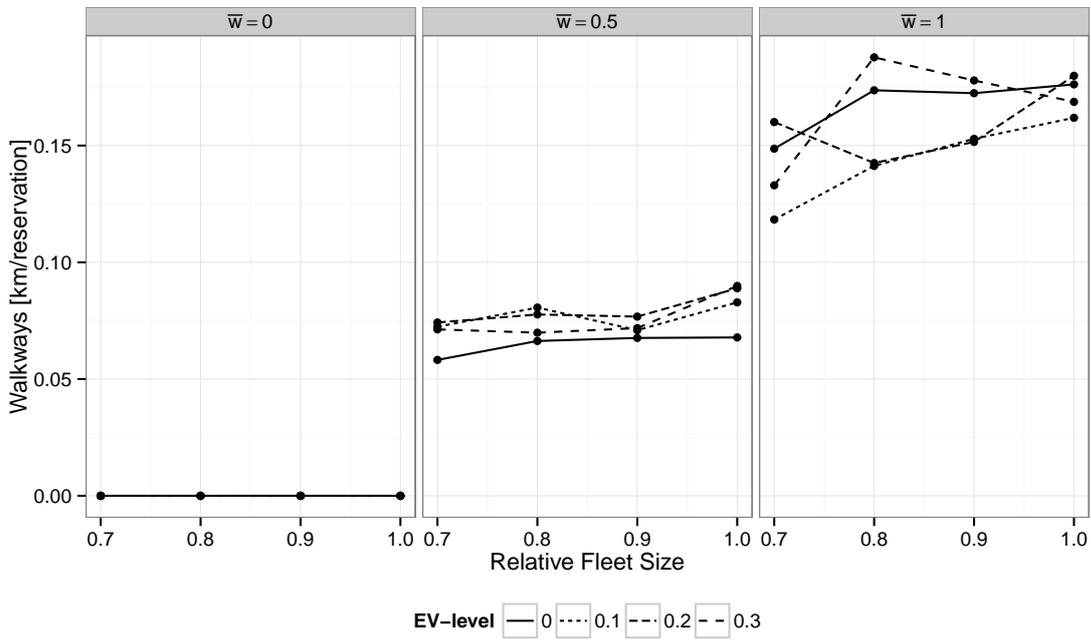
Figure 8.7: Walkways over fleet size, differentiated by EV penetration level and fleet size
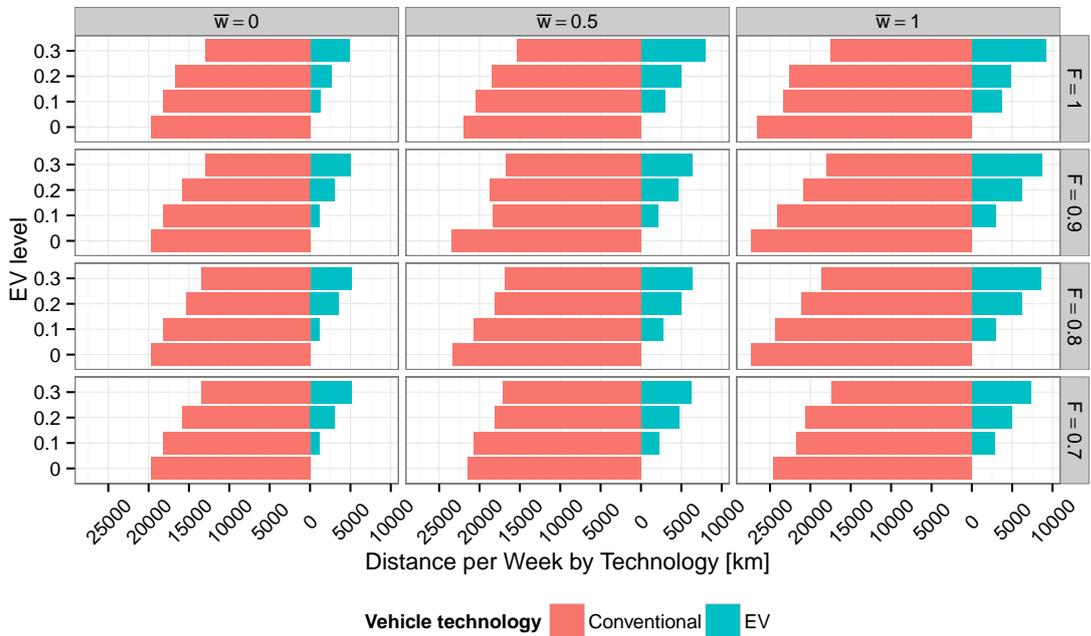

Figure 8.8: Driving distance covered by technology, EV penetration level and fleet size
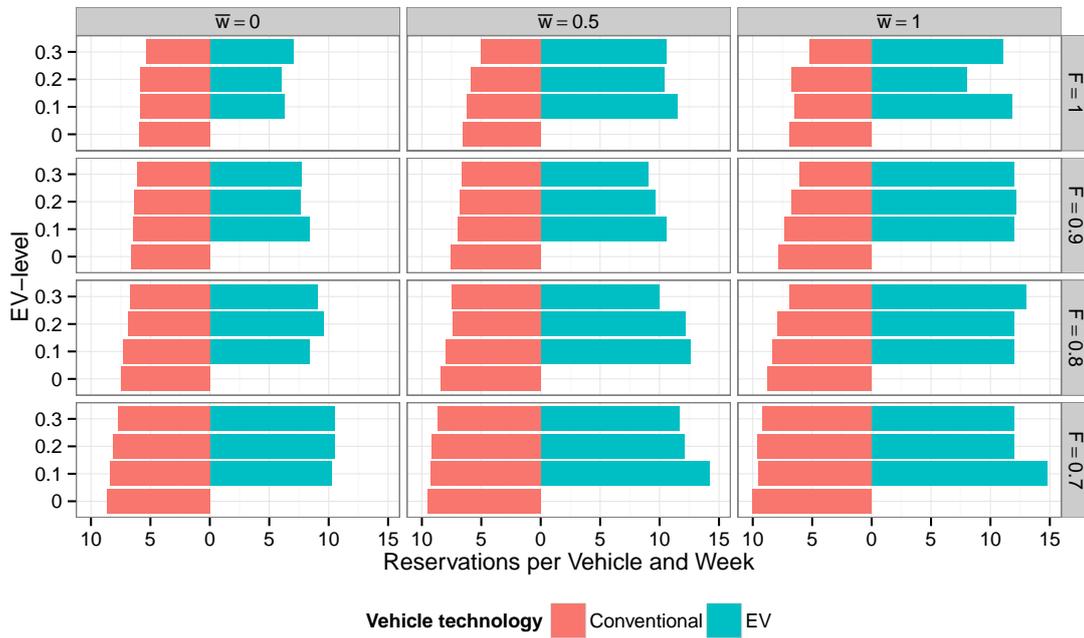
Figure 8.9: Vehicle utilization by technology, EV penetration level, and fleet size
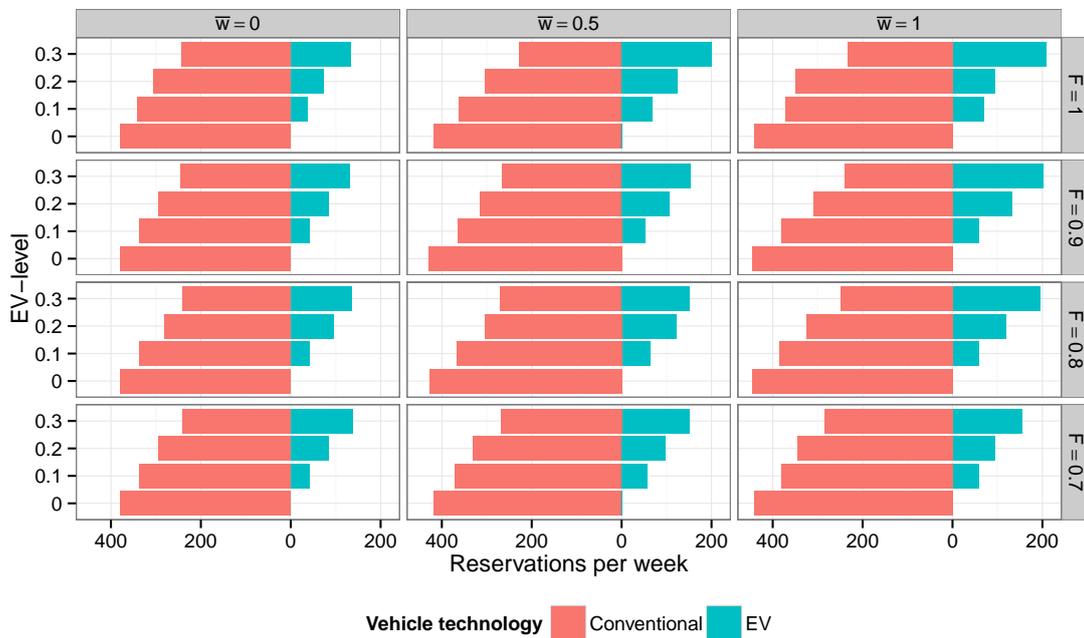
Figure 8.10: Reservations by technology

to foster electrification to a surprising extent. Accordingly, car-sharing might provide a near-ideal starting ground for the widespread adoption of electric mobility: First, the distribution of driving distances (Fig. 8.3(a)) maps nicely into the technical constraints of EVs. Second, it allows consumers to gain first-hand, risk-free experience with this new technology.

Car-sharing and EVs – Complementary Goods? In summary, we find EVs in car-sharing fleets not to be competitive currently. Interestingly, range restrictions of the adolescent EV technology do not pose effective obstacles to EVs adoption in car-sharing fleets. Rather, high capital expenditures[9] render EVs an inferior alternative to conventional vehicles. Given the weekly vehicle utilization in terms of distance described in Section 8.3.2, and if we assume aggregate daily distances driven to be distributed approximately uniform over the week (cf. Fig. B.1), range and recharging limitations do not pose major constraints to EV adoption. Accordingly, we posit the presented results not to be driven by the assumed – relatively high – battery charging rate c_{max}.

Two conflicting factors regarding the evaluation should be noted: On the one hand, both the fleet composition and stationing decision used as inputs to the optimization model may not be optimal, rendering our results rather conservative. On the other hand, the offline planner has access to relevant information for the entire period under consideration, yielding rather optimistic results. In realistic situations, uncertainty with respect to future reservations may deteriorate economic performance.

In this study we do not provide for a detailed variation regarding vehicle parameters. Nevertheless, we suspect that decreasing vehicle costs, possibly due to learning effects may pivot the overall situation in the near future. Employing vehicles with smaller batteries may provide a viable path in order to decrease vehicle costs. Presumably, but this is subject to further examination, smaller batteries may be able to serve similar parts of reservations.[10]

Appropriateness of the Optimization Objective In the definition of our objective function, we linearly incorporate cost, walkways and QoS. However, the distribution of walkways may be bimodal, raising the question of fairness on the one side, and compensation and incentives on the other. A more uniform distribution of walkways over consumers may be achieved by including individual walkways quadratically in the objective. Furthermore, the driving distance of served trips enters the objective only via the cost term.Long (costly) trips affect the overall objective adversely, hence, the optimal solution does not include them. Clearly, this is a problem of our objective

---

[9]We ignore interest on capital employed in our simple model.

[10]Determining the optimal technical parameters of EVs in car-sharing applications may pose a promising research endeavour.

formulation and deserves further exploration and, possibly, more appropriate formulations.

Imprecise Distance and Flexibility Assumptions Importantly, we lack access to the users' most desired geographic starting point regarding each reservation. Instead, we rely on the (possibly erroneous) assumption that the chosen station equals the most desired pick-up location, and consumers are, centered at this point in space, equally flexible in each direction. Additionally, customer flexibility so-far has been based on airline-distance only, which is clearly an abstraction from reality. A technically more involved model could rely on actual, real-world distances.

8.5 Outlook

This section aims to provide a brief overview of promising research directions in the field of electrified car-sharing.

Solution Technique and Optimal Fleets Due to the complexity associated with deriving cost-minimal fleets from reservation data, we formulated the problem at hand as a MKP. The field of metaheuristics (cf. Glover and Kochenberger, 2003; Raidl, 2006), however, provides a rich set of applicable solution approaches that may successfully assist in overcoming complexity of the related GAP. Alternatively, better decisions on the strategic and tactical level, i.e., number and type of vehicles, as well as their stationing, may be achieved via application of different centrality measures from graph-theory and social-network analysis (Freeman, 1978; Borgatti, 2005).

Modeling the Demand Process Future work will be concerned with modeling and simulating synthetic reservation streams that exhibit the main characteristics of the empirical demand data. To this end, capturing latent relationships in reservation data over time and space via graphical models (Koller and Friedman, 2009) may provide further insights. Model-based online optimization may in particular benefit from a synthetic model of reservation streams.

Online Optimization Spatial flexibility may be a potent substitute for both battery capacity and maximum charging rate. On the one hand, reduced vehicle cost may foster car-sharing electrification. On the other hand, poor decision making may become more costly as reducing battery capacity effectively removes part of the operator's flexibility. In the case of online optimization, spatial flexibility may be crucial in enabling acceptable QoS-levels.

Mechanism Design Optimization approaches are unable to introduce appropriate incentives to consumers regarding the revelation of private information, such as driving distances and spatial flexibility. Online Mechanism Design may render truthful information revelation the best strategy for consumers and thus lead to more informed decisions, improved efficiency and, potentially, proliferation of (electrified) car-sharing schemes.

Part IV

Finale

Chapter 9

Conclusion

Demand-side flexibility in both the smart grid and car-sharing domains, is the topic overarching this thesis. We present optimization and mechanism design-based approaches that leverage consumer flexibility to achieve more efficient outcomes. In the smart grid settings we focus on temporal flexibility, i.e., shifting and shedding, to better integrate volatile supply from RES and enable more sustainable power systems. Flexibility revelation by consumers, however, requires the establishment of appropriate incentive schemes in the first place. To this end, we propose incentive-compatible online mechanisms that achieve high efficiency. In the domain of car-sharing, we examine to what extent (mostly spatial) consumer flexibility may be employed to improve fleet utilization. Higher fleet utilization, in turn, may enable efficient electrification of car-sharing fleets. This work, hence, positions consumer flexibility as a powerful lever to align both, power and mobility systems. This conclusion summarizes the main findings and limitations of the presented work and provides a brief overview of prospective research avenues.

9.1 Summary and Contribution

Smart Grid Demand-side flexibility and demand-side management have been topics of considerable interest in smart grid research (cf. Strbac, 2008, and the references therein). So far, however, most research lacks the decisive factor – incentives. Rather, engaging consumers in DSM is mostly motivated by utilities' expected savings due to improved operations and reduced investment. Consumers, on the other hand, are mostly assumed to be willing to accept centrally-controlled allocation decisions ignorant of individuals' preferences.[1]

Part II of this thesis closes this gap and extends the state-of-the-art with respect to incentive design for consumers in the smart grid. In more detail, we establish simulation models for shiftable and sheddable demand in order to derive the value of flexibility under uncertainty, and, more importantly, with incentives taken into account.

Chapter 4 introduces a model of single-unit shiftable and sheddable demand and stochastic supply from RES. For the corresponding (offline and online) planners and

---

[1] Gerding et al. (2011); Stein et al. (2012) on EV charging coordination are relevant exceptions.

an incentive-compatible online mechanism, the model provides the necessary inputs. Although the demand model and the corresponding mechanism rely on a set of strong assumptions, the results remain interesting and encouraging. On the macroeconomic level, we are able to demonstrate the price of IC in terms of welfare to be rather small. Demand-side flexibility, i.e., shifting and shedding, may hence pose an important lever to facilitate economically efficient integration of stochastic supply from RES into the power system. Moreover, even small amounts of flexibility can significantly improve integration of RES. On the microeconomic level, we show that, ceteris paribus, more flexible loads enjoy higher allocation probabilities. Critical-value payments are monotonously decreasing in flexibility, completing incentives for flexibility provisioning.

Chapter 5 extends the demand model to allow for shiftable and sheddable multi-unit demand (jobs) that may not be interrupted after having been started, i.e., are *non-preemptive*. Different to the setting described before, we introduce costly conventional generation to support the case that jobs have been started, but supply from RES *only* is insufficient. Again, we establish the upper bound of the value of flexibility by means of an offline planner; it serves as a benchmark for online planners and mechanisms. Here, our contribution lies in the adaptation of existing packet scheduling algorithms (Chang et al., 2000a; Bent and Van Hentenryck, 2004) to allow for uncertain supply and demand. In the discrete time setting employed, allocation of *multiple* jobs at each point in time constitutes an additional contribution. Furthermore, we provide a Dominant Strategy Incentive Compatible (DSIC) online mechanisms that relies upon the Consensus idea of (Bent and Van Hentenryck, 2004) for allocation decisions, which we name, following Gerding et al. (2011), the *pre-commitment* stage. In the subsequent second stage we find Consensus to yield high economic efficiency.

Car-sharing   Part III explores the value of flexibility in car-sharing. In particular, we employ both, temporal and spatial characteristics, to reduce fleet size and, hence, raise fleet utilization. We formulate and solve a modified version of the bin-packing problem (Nemhauser and Wolsey, 1988), packing reservations on vehicles. Reservations, however, can only be assigned to a subset of all vehicles, i.e., either to vehicles at exactly the same station only (in the case of temporal flexibility), or to the vehicles at the originally desired station, as well as neighbouring stations (in the case of spatial flexibility). By enlarging the set of admissible vehicles, we are able to significantly reduce the overall number of used vehicles. Interestingly, temporal flexibility barely reduces fleet size for reasonable values of temporal flexibility. Spatial flexibility, on the other hand, is highly valuable and allows fleet size reductions of up to 30% during selected weeks, under spatial flexibility of 1.5km (airline distance). Median fleet size reduction amounted to 16% with spatial flexibility set to 1.5km, while ninety percent of all observations ranged between reductions of 8 and 26%. Note that only few consumers were required to provide their full flexibility potential, while most were served

at their desired station.

Besides uncovering the fleet reduction potential, we are particularly interested in more realistic, i.e., online settings. To this end, we hold the fleet of vehicles fixed and make allocation decisions online, with limited information on future reservations.[2] In more detail, we design an online planning algorithm incorporating look-ahead (cf. Dunke, 2014) on future reservations, to make informed assignment decisions. This approach clearly outperforms two naïve assignment schemes, both in terms of walkways and QoS/capacity utilization. Although computational effort is more pronounced, our novel algorithm yields results in typcial settings fast enough for practical application. Accordingly, we consider our algorithm to be a valuable tool for car-sharing operators interested in improving fleet utilization.

Improved fleet utilization in combination with a large share of short distance reservations draws our attention towards (partial) electrification of car-sharing fleets. In Chapter 8 we present a Multi-Knapsack optimization problem with additional EV-charging constraints. We systematically vary fleet size and composition and solve the corresponding Multi-Knapsack problem. Under the examined parameters, our results indicate that EVs are not (yet) economically viable for car-sharing operations. However, the additional cost of EV adoption to operators appears rather small. Reduced EV costs may render electrification of car-sharing fleets economically efficient in the foreseeable future.

9.2 Outlook

This thesis provides novel insights into the use and value of demand side flexibility in future power systems and car-sharing. Research is never complete, therefore, we identify drawbacks of the work so far and provide an overview into prospective research avenues.

Smart Grids   Chapters 4 and 5 have been concerned with the engineering of incentives in the smart grid. However, this work abstracts from power network constraints. Future research in the field of online mechanism design in smart grids should take those constraints into account (cf. Vytelingum et al., 2010). In this vein, dynamic adaptation of market size to current and expected supply and demand constellations may be an interesting research direction.

In the setting described in Chapter 5 we trade-off budget-balance for efficiency. A more detailed understanding of the solution space, however, may be valuable in assisting the design of market places a la Weinhardt et al. (2003) that are sustainable in the long run. Hence, we deem the trade-off between economic efficiency and budget balance (cf. Gershkov and Moldovanu, 2009) worthy of further research.

---

[2]Reservations are typically entered into the system before the begin of the corresponding usage period. Hence, some information on future reservations is available in advance.

One question we have abstracted from in this work concerns the handling of shedded demand. So far, we assume shedded demand to be removed from the system in its entirety. However, other approaches may be imagineable as well. Stochastic models and queuing theory may offer interesting methodological approaches to further explore this question.

Throughout this thesis, we assumed distinct agents for each job submitted to the mechanisms. In reality, however, agents may submit multiple jobs, representing a multitude of devices, to the mechanism. Then, mechanism design must ensure that revelation of one's private information regarding flexibility on a job level remains the dominant strategy (or at least the best strategy in expectation), and that by submitting additional jobs, allocation decisions cannot be manipulated to the benefit of an individual agent.

From a regulatory perspective it may be worthwhile to evaluate the economics of flexibility and efficiency in more detail. Thereby, it may be interesting to determine the conditions under which either flexibility or system efficiency should be expanded. In particular, it may be interesting to examine to what extent presence of flexibility may lead to counterintuitive results such as increasing use of emission-intensive base-load technologies (lignite and coal), and diminishing value of fast ramping capacities.

Finally, understanding and elicitation of preferences require more attention. For instance, leveraging both historical and smartphone sensor data may enable forecasting energy service consumption on the user level with high accuracy, hence, relieving the user to a large extent of the mundane task of specifying temporal flexibility. The challenge of accurate preference elicitation may require an interdisciplinary approach involving methods from machine learning, statistics, operations research, and microeconomics. In order to repeatedly make accurate decisions and efficiently integrate user feedback, methods from the field of active learning (c.f. Tong and Koller, 2002; Shann and Seuken, 2013) may prove valuable.

Car-sharing   Future work with the objective of improving the operational efficiency of car-sharing fleets may focus on two directions. Inclusion of uncertain information about future reservations may further improve online decision making and also be relevant to online mechanism design approaches.

Supply can, in contrast to power systems with significant shares of RES, be assumed fixed. Hence, methods from RM may find useful application in this field. Regardless of the direction pursued, a better understanding of consumers' preferences regarding the provision of flexibility is necessary. So far, empirical literature on car-sharing user behavior is, to the best of our knowledge, lacking. Methods from experimental economics, such as surveys along the lines of discrete choice theory, should therefore be employed to develop a fine-grained understanding of consumer behavior and flexibility. In particular, we deem a more precise estimate of the disutility from flexibility

provision in general, and the trade-offs between spatial and temporal flexibility, in particular, valuable.

Beyond consumer flexibility in non-electrified car-sharing, online planning and mechanism design approaches may receive more attention in electrified car-sharing. Here, according to our offline evaluation, consumer flexibility facilitates electrification of shared vehicle fleets. Hence, the provision of appropriate incentives in online settings may turn out decisive to (i) foster sustainability in mobility systems and (ii) to eventually link both, electric power and mobility systems.

In summary, the integration of consumer flexibility in smart grids and mobility systems on one side, and provision of appropriate incentives to consumers on the other, raises a number of interesting research questions. This thesis points out the benefits of an activated demand side in both domains. However, the presented OMD-based approaches may turn out to be of disproportionate complexity to consumers. Hence, simpler approaches involving (static) product and price differentiation on the basis of consumer segmentation may be more attractive and thus easier to introduce into retail markets. It will be interesting to see, to what extent a reduction in complexity negatively affects economic efficiency. At this point, we leave the task of exploring the efficiency-complexity trade-off for future research.

Appendix A

Online Mechanism Design Details

This appendix provides the technical details of the model-free online mechanism presented in Chapter 4.

Mechanism Properties

- **Online decision making.** Decision must be made online, i.e., as time is progressing and no information on future events can be assumed.

- **Incentive compatibility**[1] (IC). In an IC mechanism, it is a (weakly) dominant strategy for any job to reveal its true type to the mechanism. This basic property is of paramount importance, as only based on true job types can the allocative/economic efficiency of a mechanism be determined. Incentive compatibility in all dimensions for demand jobs must be ensured to avoid strategic actions, as such behavior can have adverse effects on social welfare.

- **Budget-balance** (BB). A budget-balanced mechanism requires neither external subsidies (notion of weak BB) nor does it accumulate payments (if BB in the strong sense), i.e., $\sum_i p_i = 0$

- **Individual rationality** (IR). Through participation, the job can only improve its utility. IR is achieved by making jobs pay only for goods in case of allocation.

Payment rule: A critical value approach

For the definition of a payment rule that renders the described allocation rule IC, we establish the sets of directly and indirectly competing demand jobs as well as non-competing supply jobs. These sets need to be defined to determine a demand job's critical value, which ensures IC (Nisan, 2007, cf. p. 229).

---

[1]Incentive compatibility is also referred to as truth-telling, truthfulness, or strategy proofness.

Directly Competing Demand Jobs   We name the set of demand jobs $DCD(x)$ that are active (both matched and unmatched) at *allocation time m* of demand job *x directly competing* with *x*.

$$DCD(x) = \left\{ j \in J \setminus \{x\} | a(j) \le m(x) \le d(j) \right\} \tag{A.1}$$

Note that directly competing demand jobs can eventually end up non-allocated.

Indirectly Competing Demand Jobs   We name a demand job *j indirectly competing* (ICD) with demand jobs *x* if it is being *allocated* during the active period of *x*; in order for a request *j* to be ICD, it does not have to be directly competing (at allocation time) with *x*.

$$ICD(x) = \left\{ j \in J \setminus \{x\} | (a(x) \le m(b) \le d(x)) \right\} \tag{A.2}$$

We will see shortly, how the valuation of such indirectly competing demand jobs becomes relevant for determining the payment associated with the allocation of *x*. Indirectly and directly competing demand jobs differ in two aspects: Eventual allocation (ICD *are* allocated) and time of competition (DCD are competing *at* allocation time, while IDC might be competing earlier or later). Accordingly, an element of $DCD(x)$ may also belong to the set of $ICD(x)$, i.e. the two sets are not necessarily subsets of each other.

Non-competing Supply Jobs   We name a supply job *non-competing* (NCS) if it is either non-allocated or only allocated after its release date (*i*) and if it is active at some point during the active period of *x*. The set of non-competing supply jobs is formally defined by

$$NCS(x) = \{ i \in I | (\nexists m(i) \vee a(x) < m(i)) \wedge AnyOverlap(x, i) \} \tag{A.3}$$

where

$$AnyOverlap(x, i) = a(i) < d(x) \wedge d(i) > a(x) \tag{A.4}$$

Furthermore, we define an order on these sets of jobs, where the following holds:

$$b_i < b_j \equiv (v(b_i) < v(b_j)) \quad \forall b_i, b_j \in B \quad \forall b_i, b_j \in A \tag{A.5}$$

The price a job eventually pays for successful allocation is formally determined by the job's critical value:

$$p(x) = v(\min\{\max(DCD(x)), \min(ICD(x)), \min(NCS(x))\}) \tag{A.6}$$

More specifically, the individual components are given by:

- the highest valuation of directly competing demand jobs at allocation time $v(\max(DCD(x)))$ (this is always less than the valuation of the allocated request),

- the lowest valuation of demand jobs allocated during the active period of the respective job x, i.e., $v(\min(ICD(x)))$,

- the lowest reservation price observed during the active period if the demand job has remained without any competing job for any duration of the respective supply job during its active period, i.e., $v(\min(NCS(x)))$.

The critical value's monotonicity renders the mechanism incentive compatible. Accordingly, the more flexible a demand job (i.e., the longer the active period between release and due date), the lower the payment.

Appendix B

Car-sharing Fleet Utilization

A fraction of approximately 10 percent of the fleet's small vehicles is in use at any given time, forming the base-load for the car-sharing system. Peak-load can easily be observed once per week with distinct peaks in the usage patterns. However, taking a macroscopic view on aggregate fleet utilization over the period of a whole calendar year renders recognition of robust patterns a difficult exercise. Accordingly, to derive a better understanding on the existence and nature of patterns, we embrace a micro-perspective on aggregate reservation behavior.

Fig. B.1 decomposes fleet utilization by month (each box represents one month) and by week (lines in different colors represent different weeks). At this level of detail, patterns can be identified visually . First, there is a pronounced 24-hour pattern with peak utilization during daytimes and low utilization at night. Second, the largest weekly peak can be found on Saturdays, followed by Sundays. The largest exception can be found in December, around Christmas time. Generally, peaks in fleet utilization can be traced back to public and school holidays. In particular, a larger-than-usual share of vehicles is used overnight, presumably to visit family, friends and relatives during Christmas-holidays without returning the vehicles at night to their stations. Maximal fleet utilization is observed on the Pentecost weekend at the end of May and the beginning of August, which coincided with the end of the summer holidays in the state of Baden-Wuerttemberg. Hence, from visual inspection of Fig. B.1 we can infer that recurring utilization patterns do in fact exist. Recognizing these regular patters, they might be useful for better economic decision making. Here, the interesting part is that knowledge of these patters can effectively mitigate uncertainty on future fleet utilization, at least on a macroscopic level.

The untypical reservation behavior around Christmas time stands out clearly from the rest of the data.

In Fig. B.2, fleet utilization by station is depicted. There are stations for which fleet utilization assumes values in excess of 40%. The low utilization of some other stations (i.e., < 20%) clearly demonstrates that there is potential in optimizing the current reservation scheme.
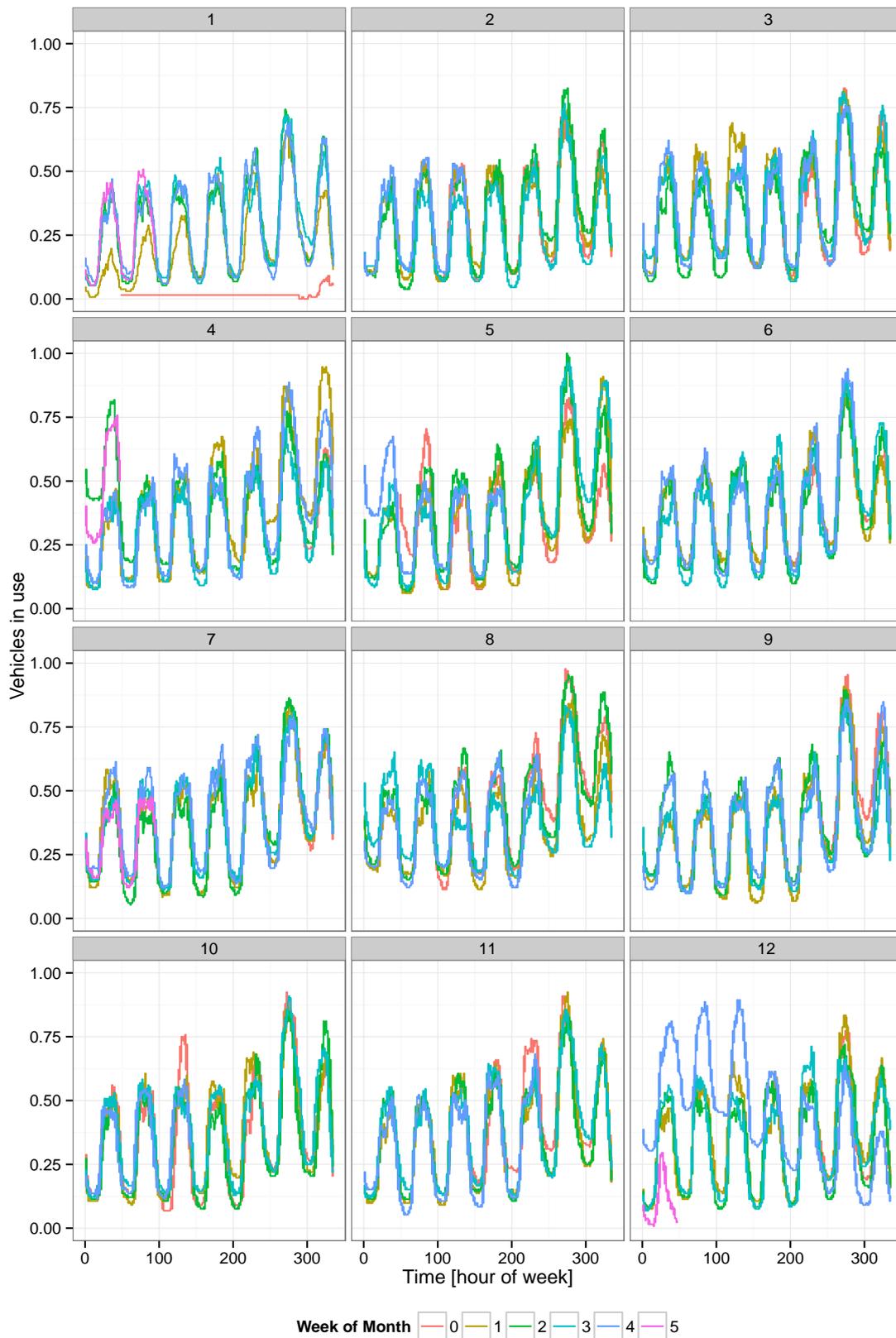
Figure B.1: Seasonalities of car-sharing usage: Significant daily and weekly usage patterns are present.
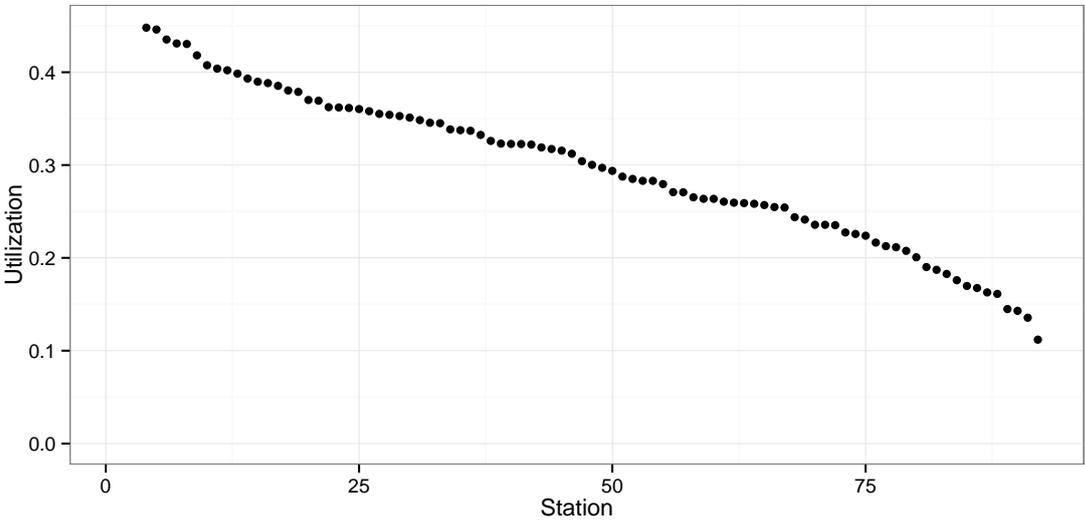
Figure B.2: Utilization of different stations for the small vehicle class

Bibliography

Abraham, I., D. Delling, A. V. Goldberg, and R. F. Werneck (2013). Alternative routes in road networks. *J. Exp. Algorithmics 18*, 1.3:1.1–1.3:1.17.

Agatz, N., A. Erera, M. Savelsbergh, and X. Wang (2012, December). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research 223*(2), 295–303.

Albadi, M. and E. El-Saadany (2008). A summary of demand response in electricity markets. *Electric Power Systems Research 78*(11), 1989–1996.

Allcott, H. (2011). Rethinking real-time electricity pricing. *Resource and Energy Economics 33*(4), 820–842.

Amin, S. and B. Wollenberg (2005). Toward a smart grid: power delivery for the 21st century. *IEEE Power and Energy Magazine 3*(5), 34–41.

Aspnes, J., Y. Azar, A. Fiat, S. Plotkin, and O. Waarts (1997, May). On-line routing of virtual circuits with applications to load balancing and machine scheduling. *J. ACM 44*(3), 486–504.

Banal-Estañol, A. and A. R. Micola (2009). Composition of electricity generation portfolios, pivotal dynamics, and market prices. *Management Science 55*(11), 1813–1831.

Barth, M. and S. A. Shaheen (2002). Shared-use vehicle systems: Framework for classifying carsharing, station cars, and combined approaches. *Transportation Research Record: Journal of the Transportation Research Board 1791*(1), 105–112.

Beach, R., A. P. Muhlemann, D. H. R. Price, A. Paterson, and J. A. Sharp (2000). A review of manufacturing flexibility. *European Journal of Operational Research 122*(1), 41–57.

Beckmann, M. J., C. B. McGuire, and C. Winsten (1956). *Studies in the economics of transportation.* New Haven: Yale University Press.

Bengtsson, J. (2001). Manufacturing flexibility and real options: A review. *International Journal of Production Economics 74*(1–3), 213–224.

Bengtsson, J. and J. Olhager (2002). Valuation of product-mix flexibility using real options. *International Journal of Production Economics 78*(1), 13–28.

Benjaafar, S., T. L. Morin, and J. J. Talavage (1995). The strategic value of flexibility in sequential decision making. *European Journal of Operational Research 82*(3), 438–457.

Bent, R. and P. Van Hentenryck (2004). The value of consensus in online stochastic scheduling. In *ICAPS*, Volume 4, pp. 219–226. 00040.

Bent, R. and P. Van Hentenryck (2005). Online stochastic optimization without distributions. In *ICAPS*, Volume 5, pp. 171–180.

Bernardes, E. S. and M. D. Hanna (2009). A theoretical review of flexibility, agility and responsiveness in the operations management literature. *International Journal of Operations & Production Management 29*(1), 30–53.

Bertsimas, D., E. Litvinov, X. Sun, J. Zhao, and T. Zheng (2013). Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems 28*(1), 52–63.

Bichler, M., A. Gupta, and W. Ketter (2010). Research commentary—Designing smart markets. *Information Systems Research 21*(4), 688–699.

Bitar, E. and S. Low (2012). Deadline differentiated pricing of deferrable electric power service. In *51st Annual Conference on Decision and Control (CDC)*, pp. 4991 –4997. IEEE.

Bitar, E., R. Rajagopal, P. Khargonekar, K. Poolla, and P. Varaiya (2012). Bringing wind energy to market. *IEEE Transactions on Power Systems 27*(3), 1225–1235.

Bitar, E. and Y. Xu (2014). Deadline differentiated pricing of delay-tolerant demand. *arXiv:1407.1601 [cs, math]*.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *The journal of political economy*, 637–654.

Bohn, R. E., M. C. Caramanis, and F. C. Schweppe (1984). Optimal pricing in electrical networks over space and time. *The RAND Journal of Economics 15*(3), 360–376.

Boiteux, M. (1960). Peak-load pricing. *The Journal of Business 33*(2), 157–179.

Borenstein, D. (2000). A directed acyclic graph representation of routing manufacturing flexibility. *European Journal of Operational Research 127*(1), 78–93.

Borenstein, S., M. Jaske, and A. Rosenfeld (2002). Dynamic pricing, advanced metering, and demand response in electricity markets.

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks 27*(1), 55–71.

Bosman, M., V. Bakker, A. Molderink, J. Hurink, and G. Smit (2012). Planning the production of a fleet of domestic combined heat and power generators. *European Journal of Operational Research 216*(1), 140–151.

Botterud, A., J. Wang, C. Monteiro, and V. Miranda (2009). Wind power forecasting and electricity market operations. In *Proceedings of the 32nd IAEE International Conference, San Francisco, USA.*

Braess, D. (1968). Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung 12*(1), 258–268.

Brill, P. H. and M. Mandelbaum (1989). On measures of flexibility in manufacturing systems. *International Journal of Production Research 27*(5), 747–756.

Browne, J., D. Dubois, K. Rathmill, S. P. Sethi, and K. E. Stecke (1984). Classification of flexible manufacturing systems. *The FMS magazine 2*(2), 114–117.

Bundesnetzagentur (2011). „Smart Grid" und „Smart Market" – Eckpunktepapier der Bundesnetzagentur zu den Aspekten des sich verändernden Energieversorgungssystems. Technical report, Bonn.

Bushnell, J. (2004). California's electricity crisis: A market apart? *Energy Policy 32*(9), 1045–1052.

Cachon, G. and C. Terwiesch (2009). *Matching supply with demand : An introduction to operations management* (2. ed.). Boston [u.a.]: McGraw-Hill Irwin.

Carroll, W. J. and R. C. Grimes (1995). Evolutionary change in product management: Experiences in the car rental industry. *Interfaces 25*(5), 84–104.

Cattrysse, D. G. and L. N. Van Wassenhove (1992). A survey of algorithms for the generalized assignment problem. *European Journal of Operational Research 60*(3), 260–272.

Chang, H., R. Givan, and E. Chong (2000a). On-line scheduling via sampling. In *ICAPS*, pp. 62–71.

Chang, H., R. Givan, and E. Chong (2000b). On-line scheduling via sampling. *Artificial Intelligence Planning and Scheduling (AIPS) (Breckenridge, Colorado, 2000)*, 62–71.

Chassin, D. P. and L. Kiesling (2008). Decentralized coordination through digital technology, dynamic pricing, and customer-driven control: The gridwise testbed demonstration project. *The Electricity Journal 21*(8), 51–59.

Chekuri, C. and S. Khanna (2005). A polynomial time approximation scheme for the multiple knapsack problem. *SIAM Journal on Computing 35*(3), 713–728.

Chemla, D., F. Meunier, and R. Wolfler Calvo (2012). Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization*.

Chou, M. C., G. A. Chua, and C.-P. Teo (2010). On range and response: Dimensions of process flexibility. *European Journal of Operational Research 207*(2), 711–724.

Chou, M. C., G. A. Chua, C.-P. Teo, and H. Zheng (2011). Process flexibility revisited: The graph expander and its applications. *Operations research 59*(5), 1090–1105.

Coltin, B. J. and M. Veloso (2013). Towards ridesharing with passenger transfers. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pp. 1299–1300.

Conejo, A. J., J. Contreras, and M. A. Plazas (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting 21*(3), 435–462.

Constantin, F. and D. Parkes (2009). Self-correcting sampling-based dynamic multi-unit auctions. In *ACM EC*, pp. 89–98.

Cordeau, J.-F. and G. Laporte (2003, June). The dial-a-ride problem (DARP): Variants, modeling issues and algorithms. *Quarterly Journal of the Belgian, French and Italian Operations Research Societies 1*(2), 89–101.

Cordeau, J.-F. and G. Laporte (2007). The dial-a-ride problem: models and algorithms. *Annals of Operations Research 153*(1), 29–46.

Correia, G. H. d. A. and A. P. Antunes (2012). Optimization approach to depot location and trip selection in one-way carsharing systems. *Transportation Research Part E: Logistics and Transportation Review 48*(1), 233–247.

Cramton, P. and A. Ockenfels (2012). Economics and design of capacity markets for the power sector. *Zeitschrift für Energiewirtschaft 36*(2), 113–134.

Damousis, I., a. Bakirtzis, and P. Dokopoulos (2004). A solution to the unit-commitment problem using integer-coded genetic algorithm. *IEEE Transactions on Power Systems 19*(2), 1165–1172.

Dash, R. K., N. R. Jennings, and D. C. Parkes (2003). Computational-mechanism design: A call to arms. *IEEE intelligent systems 18*(6), 40–47.

De Livera, A. M., R. J. Hyndman, and R. D. Snyder (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association 106*(496), 1513–1527.

DECC (2011). UK renewable energy roadmap. Technical report, UK Department of Energy and Climate Change.

Dixit, A. K. (1994). *Investment Under Uncertainty.* Princeton University Press.

Doll, C., M. Gutmann, and M. Wietschel (2011). Integration von elektrofahrzeugen in carsharing-flotten-simulation anhand realer fahrprofile. *Fraunhofer ISI.*

Dunke, F. (2014). *Online Optimization with Lookahead.* Ph. D. thesis, Karlsruhe Institute of Technology, Karlsruhe.

Eberle, U. and R. von Helmolt (2010). Sustainable transportation based on electric vehicle concepts: a brief overview. *Energy & Environmental Science 3*(6), 689–699.

Ehrenthal, J. C. F., D. Honhon, and T. Van Woensel (2014, October). Demand seasonality in retail inventory management. *European Journal of Operational Research 238*(2), 527–539.

European Commission (2012). European technology platform SmartGrids: Smartgrids SRA 2035 strategic research agenda - update of the smartgrids SRA 2007 for the needs by the year 2035.

Eßer-Frey, A. (2012). *Analyzing the regional long-term development of the German power system using a nodal pricing approach.* Ph. D. thesis, Karlsruhe Institute of Technology, Karlsruhe.

Farhangi, H. (2010). The path of the smart grid. *IEEE Power and Energy Magazine 8*(1), 18 –28.

Faruqui, A., R. Hledik, and S. Sergici (2010). Rethinking prices. *Public Utilities Fortnightly 148*(1), 30–39.

Fay, S. (2008). Selling an opaque product through an intermediary: The case of disguising one's product. *Journal of Retailing 84*(1), 59–75.

Federgruen, A. and H. Groenevelt (1986). Preemptive scheduling of uniform machines by ordinary network flow techniques. *Management Science 32*(3), 341–349.

Fehrenbach, D., E. Merkel, R. McKenna, U. Karl, and W. Fichtner (2014). On the economic potential for electric load management in the german residential heating sector – an optimising energy system model approach. *Energy 71*, 263–276.

Fink, A. and T. Reiners (2006). Modeling and solving the short-term car rental logistics problem. *Transportation Research Part E: Logistics and Transportation Review 42*(4), 272–292.

Flath, C., J. Ilg, and C. Weinhardt (2012). Decision support for electric vehicle charging. *AMCIS 2012 Proceedings*.

Flath, C. M. (2013). *Flexible demand in smart grids : Modeling and coordination / von Christoph Michael Flath*. Ph. D. thesis, Karlsruhe Institute of Technology, Karlsruhe.

Flath, C. M., J. P. Ilg, S. Gottwalt, H. Schmeck, and C. Weinhardt (2013). Improving electric vehicle charging coordination through area pricing. *Transportation Science*.

Fosgerau, M. and A. Karlström (2010). The value of reliability. *Transportation Research Part B: Methodological 44*(1), 38–49.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks 1*(3), 215–239.

Friedman, E. J. and D. C. Parkes (2003). Pricing WiFi at starbucks: issues in online mechanism design. In *Proceedings of the 4th ACM conference on Electronic commerce*, pp. 240–241.

Gallego, G. and R. Phillips (2004). Revenue management of flexible products. *Manufacturing & Service Operations Management 6*(4), 321–337.

Gallien, J. and L. M. Wein (2005). A smart market for industrial procurement with capacity constraints. *Management Science 51*(1), 76–91.

Gans, N. and S. Savin (2007). Pricing and capacity rationing for rentals with uncertain durations. *Management Science 53*(3), 390–407.

Gärttner, J., C. M. Flath, and C. Weinhardt (2014). Load shifting, interrupting or both? Customer portfolio composition in demand side management. In *Proceedings of the 11th Conference on Computational Management Science*. Springer. Lecture Notes in Economics and Mathematical Systems.

Gedra, T. and P. Varaiya (1993). Markets and pricing for interruptible electric power. *IEEE Transactions on Power Systems 8*(1), 122–128.

Geisberger, R., P. Sanders, D. Schultes, and D. Delling (2008). Contraction hierarchies: Faster and simpler hierarchical routing in road networks. *Experimental Algorithms*, 319–333. 00273.

Gellings, C. W. (2009). *The Smart Grid: Enabling Energy Efficiency and Demand Response*. The Fairmont Press, Inc.

Gerding, E., V. Robu, S. Stein, D. Parkes, A. Rogers, and N. Jennings (2011). Online mechanism design for electric vehicle charging. In *The Tenth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2011)*, pp. 811–818.

Gerding, E. H., S. Stein, V. Robu, D. Zhao, and N. R. Jennings (2013). Two-sided online markets for electric vehicle charging. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS '13, Richland, SC, pp. 989–996.

Gershkov, A. and B. Moldovanu (2009). Dynamic revenue maximization with heterogeneous objects: A mechanism design approach. *American Economic Journal: Microeconomics 1*(2), 168–198.

Gerwin, D. (1993). Manufacturing flexibility: A strategic perspective. *Management Science 39*(4), 395–410.

Giebel, G., R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl (2011). The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, ANEMOS. plus.

Glover, F. and G. A. Kochenberger (2003). *Handbook of Metaheuristics*. Springer Science & Business Media.

Gore, O., S. Viljainen, M. Makkonen, and D. Kuleshov (2012). Russian electricity market reform: Deregulation or re-regulation? *Energy Policy 41*, 676–685. 00011.

Gottwalt, S., W. Ketter, C. Block, J. Collins, and C. Weinhardt (2011). Demand side management—A simulation of household behavior under variable prices. *Energy Policy 39*(12), 8163–8174.

Gould, P. G., A. B. Koehler, J. K. Ord, R. D. Snyder, R. J. Hyndman, and F. Vahid-Araghi (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research 191*(1), 207–222.

Graves, S. C. and B. T. Tomlin (2003). Process flexibility in supply chains. *Management Science 49*(7), 907–919.

Gungor, V., D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. Hancke (2011). Smart grid technologies: Communication technologies and standards. *IEEE Transactions on Industrial Informatics 7*(4), 529–539.

Gupta, Y. P. and S. Goyal (1989). Flexibility of manufacturing systems: Concepts and measurements. *European Journal of Operational Research 43*(2), 119–135.

Gupta, Y. P. and T. M. Somers (1992). The measurement of manufacturing flexibility. *European Journal of Operational Research 60*(2), 166–182.

Gärttner, J., C. M. Flath, and C. Weinhardt (2014). Designing demand response portfolios. In *1st International Workshop on Demand Response*.

Hanne, T., T. Melo, and S. Nickel (2009). Bringing robustness to patient flow management through optimized patient transports in hospitals. *Interfaces 39*(3), 241–255.

Happ, H. (1977). Optimal power dispatch; a comprehensive survey. *IEEE Transactions on Power Apparatus and Systems 96*(3), 841–854.

He, X., N. Keyaerts, I. Azevedo, L. Meeus, L. Hancher, and J.-M. Glachant (2013). How to engage consumers in demand response: A contract perspective. *Utilities Policy 27*, 108–122.

Heinonen, K. (2004). Reconceptualizing customer perceived value: the value of time and place. *Managing Service Quality 14*(2/3), 205–215.

Hentenryck, P. and R. Bent (2009). *Online stochastic combinatorial optimization.* The MIT Press.

Herrmann, K. (2013). *Technologische und organisatorische Systembewertung und -gestaltung spanender Fertigungslinien nach den Prinzipien der schlanken Produktion.* Paderborn: Universitätsbibliothek.

Herroelen, W. and R. Leus (2005, September). Project scheduling under uncertainty: Survey and research potentials. *European Journal of Operational Research 165*(2), 289–306.

Hu, M., J. D. Weir, and T. Wu (2012). Decentralized operation strategies for an integrated building energy system using a memetic algorithm. *European Journal of Operational Research 217*(1), 185–197.

Hua, G., S. Wang, and T. C. E. Cheng (2010). Price and lead time decisions in dual-channel supply chains. *European Journal of Operational Research 205*(1), 113–126.

Hyndman, R. J. and G. Athanasopoulos (2013, October). *Forecasting: principles and practice.* OTexts.

Häll, C. H., H. Andersson, J. T. Lundgren, and P. Värbrand (2009). The integrated dial-a-ride problem. *Public Transport 1*(1), 39–54.

Ilg, J. P. (2014). *Coordination in power systems for efficient grid utilization.* Ph. D. thesis, Karlsruhe Institute of Technology, Karlsruhe.

Ipakchi, A. and F. Albuyeh (2009). Grid of the future. *IEEE Power and Energy Magazine 7*(2), 52–62.

Iravani, S. M., M. P. V. Oyen, and K. T. Sims (2005). Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Science 51*(2), 151–166.

Jain, A. K. and H. Elmaraghy (1997). Production scheduling/rescheduling in flexible manufacturing. *International Journal of Production Research 35*(1), 281–309.

Jain, S. and P. v. Hentenryck (2011). Large neighborhood search for dial-a-ride problems. In J. Lee (Ed.), *Principles and Practice of Constraint Programming – CP 2011*, Number 6876 in Lecture Notes in Computer Science, pp. 400–413. Springer.

Jerath, K., S. Netessine, and S. K. Veeraraghavan (2010). Revenue management with strategic customers: Last-minute selling and opaque selling. *Management Science 56*(3), 430–448.

Jordan, W. C. and S. C. Graves (1995). Principles on the benefits of manufacturing process flexibility. *Management Science 41*(4), 577–594.

Joskow, P. (2008). Lessons learned from electricity market liberalization. *The Energy Journal 29*(2), 9–42.

Joskow, P. L. (2012). Creating a smarter U.S. electricity grid. *The Journal of Economic Perspectives 26*(1), 29–47.

Juang, B. H. and L. R. Rabiner (1991). Hidden markov models for speech recognition. *Technometrics 33*(3), 251–272.

Judith, D., G. Meeßen, J. Hartog, F. Engelsing, F. Simonis, and L. Locher (2011). Sektoruntersuchung Stromerzeugung und -großhandel. Abschlussbericht (B10-9/09), Bundeskartellamt.

Kalagnanam, J. and D. C. Parkes (2004). Auctions, bidding and exchange design. In D. Simchi-Levi, S. D. Wu, and Z.-J. Shen (Eds.), *Handbook of Quantitative Supply Chain Analysis*, Number 74 in International Series in Operations Research & Management Science, pp. 143–212. Springer US.

Kamar, E. and E. Horvitz (2009). Collaboration and shared plans in the open world: Studies of ridesharing. In *IJCAI*, Volume 9, pp. 187.

Kazarlis, S., A. Bakirtzis, and V. Petridis (1996). A genetic algorithm solution to the unit commitment problem. *IEEE Transactions on Power Systems 11*(1), 83–92.

Kek, A. G., R. L. Cheu, and M. L. Chor (2006). Relocation simulation model for multiple-station shared-use vehicle systems. *Transportation Research Record: Journal of the Transportation Research Board 1986*(1), 81–88.

Kek, A. G., R. L. Cheu, Q. Meng, and C. H. Fung (2009). A decision support system for vehicle relocation operations in carsharing systems. *Transportation Research Part E: Logistics and Transportation Review 45*(1), 149–158.

Keles, D. (2013). *Uncertainties in energy markets and their consideration in energy storage evaluation.* Ph. D. thesis, Karlsruhe Institute of Technology, Karlsruhe.

Keles, D., M. Genoese, D. Möst, and W. Fichtner (2012). Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Economics 34*(4), 1012–1032.

Keshav, S. and C. Rosenberg (2011). How internet concepts and technologies can help green and smarten the electrical grid. *ACM SIGCOMM Computer Communication Review 41*(1), 109–114.

King, C., W. Griggs, F. Wirth, and R. Shorten (2013). Using a car sharing model to alleviate electric vehicle range anxiety. In *The 16th Yale Workshop on Adaptive and Learning Systems.*

Kleiner, A., B. Nebel, V. A. Ziparo, and A. Srl (2011). A mechanism for dynamic ride sharing based on parallel auctions. In *IJCAI*, Volume 11, pp. 266–272.

Koller, D. and N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques.* MIT Press. 02017.

Kranton, R. E. and D. F. Minehart (2001). A theory of buyer-seller networks. *American Economic Review 91*(3), 485–508.

Kuhnimhof, T., R. Buehler, M. Wirtz, and D. Kalinowska (2012). Travel trends among young adults in germany: Increasing multimodality and declining car use for men. *Journal of Transport Geography 24*, 443–450.

Kumar, V. (1987). Entropic measures of manufacturing flexibility. *International Journal of Production Research 25*(7), 957–966.

Kwon, P. S. and P. Østergaard (2014, December). Assessment and evaluation of flexible demand in a danish future energy scenario. *Applied Energy 134*, 309–320.

Lijphart, A. (2003). Degrees of proportionality of proportional representation formulas. In *Electoral Laws and Their Political Consequences*, Number 1 in Agathon series on representation. Algora Publishing.

Lin, J.-R., T.-H. Yang, and Y.-C. Chang (2011). A hub location inventory model for bicycle sharing system design: Formulation and solution. *Computers & Industrial Engineering.*

López, J., A. Dorado, J. Álvarez, A. Feijóo, C. Carrillo, J. Cidrás, and E. Menéndez (2002). The Sotavento experimental wind park. In *Global Windpower Conference, Paris, April 2002.*

Mandelbaum, M. (1978). *Flexibility in decision making: an exploration and unification.* University of Toronto.

Mantawy, A. H., Y. Abdel-Magid, and S. Selim (1999). Integrating genetic algorithms, tabu search, and simulated annealing for the unit commitment problem. *IEEE Transactions on Power Systems 14*(3), 829–836.

Marler, R. T. and J. S. Arora (2004). Survey of multi-objective optimization methods for engineering. *Structural and multidisciplinary optimization 26*(6), 369–395.

Marquez, R. and C. F. M. Coimbra (2013, May). Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy 91*, 327–336.

Marschak, T. and R. Nelson (1962). Flexibility, uncertainty, and economic theory. *Metroeconomica 14*(1-2-3), 42–58.

McFadden, D. (2001). Economic choices. *American Economic Review*, 351–378.

Mohsenian-Rad, A., V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia (2010). Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid. *IEEE Transactions on Smart Grid 1*(3), 320 –331.

Morency, C., M. Trépanier, B. Agard, B. Martin, and J. Quashie (2007). Car sharing system: what transaction datasets reveal on users' behaviors. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pp. 284–289. IEEE.

Myerson, R. B. and M. A. Satterthwaite (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory 29*(2), 265–281.

Nair, R. and E. Miller-Hooks (2011). Fleet management for vehicle sharing operations. *Transportation Science 45*(4), 524–540.

Nair, R., E. Miller-Hooks, R. C. Hampshire, and A. Busic (2013). Large-scale vehicle sharing systems: Analysis of velib'. *International Journal of Sustainable Transportation 7*(1), 85–106.

Nemhauser, G. L. and L. A. Wolsey (1988). *Integer and combinatorial optimization*, Volume 18. Wiley New York.

Nisan, N. (2007). Introduction to mechanism design (for computer scientists). In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (Eds.), *Algorithmic Game Theory*, pp. 209–241.

Nissan (2014). Preise und Technische Daten - NISSAN LEAF - Elektroautos von NISSAN.

Nobis, C. (2006, January). Carsharing as key contribution to multimodal and sustainable mobility behavior: Carsharing in germany. *Transportation Research Record: Journal of the Transportation Research Board 1986*, 89–97.

Nobis, C. (2007, January). Multimodality: Facets and causes of sustainable mobility behavior. *Transportation Research Record: Journal of the Transportation Research Board 2010*, 35–44.

Nolden, C., M. Schönfelder, A. Eßer-Frey, V. Bertsch, and W. Fichtner (2013). Network constraints in techno-economic energy system models: Towards more accurate modeling of power flows in long-term energy system models. *Energy Systems 4*(3), 267–287.

Oren, S. S. and S. A. Smith (Eds.) (1993). *Service opportunities for electric utilities: Creating differentiated products*. Boston: Kluwer Academic Publishers.

Palensky, P. and D. Dietrich (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. *Industrial Informatics, IEEE Transactions on 7*(3), 381–388.

Papavasiliou, A. and S. Oren (2014). Large-scale integration of deferrable demand and renewable energy sources. *Power Systems, IEEE Transactions on 29*(1), 489–499.

Parkes, D. and Q. Duong (2007). An ironing-based approach to adaptive online mechanism design in single-valued domains. In *AAAI*, pp. 94–101.

Parkes, D. C. (2007). Online mechanisms. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (Eds.), *Algorithmic Game Theory*, pp. 411–439.

Parvania, M. and M. Fotuhi-Firuzabad (2010). Demand response scheduling by stochastic SCUC. *IEEE Transactions on Smart Grid 1*(1), 89–98.

Petersen, M., L. Hansen, J. Bendtsen, K. Edlund, and J. Stoustrup. (2014). Heuristic optimization for the discrete virtual power plant dispatch problem. *IEEE Transactions on Smart Grid*.

Petersen, M., L. H. Hansen, and T. Mølbak (2012). Exploring the value of flexibility: A smart grid discussion. pp. 43–48.

Petersen, M. K., K. Edlund, L. H. Hansen, J. Bendtsen, and J. Stoustrup (2013). A taxonomy for modeling flexibility and a computationally efficient algorithm for dispatch in smart grids. In *American Control Conference (ACC)*, pp. 1150–1156.

Pinedo, M. (2012). *Scheduling: Theory, Algorithms, and Systems*. Springer.

Pinson, P. and H. Madsen (2012). Adaptive modelling and forecasting of offshore wind power fluctuations with markov-switching autoregressive models. *Journal of Forecasting 31*(4), 281–313.

Raidl, G. R. (2006). A unified view on hybrid metaheuristics. In F. Almeida, M. J. B. Aguilera, C. Blum, J. M. M. Vega, M. P. Pérez, A. Roli, and M. Sampels (Eds.), *Hybrid Metaheuristics*, Number 4030 in Lecture Notes in Computer Science, pp. 1–12. Springer Berlin Heidelberg.

Ramchurn, S., P. Vytelingum, A. Rogers, and N. Jennings (2012). Putting the "Smarts" into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM*.

Ramos Gutierrez, A. I., C. De Jonghe, D. Six, and R. Belmans (2013). Demand response within current electricity wholesale market design. *Working paper*.

Raviv, T. and O. Kolka (2013). Optimal inventory management of a bike-sharing station. *IIE Transactions*.

Raviv, T., M. Tzur, and I. A. Forma (2012). Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics*, 1–43.

Rebours, Y. G., D. S. Kirschen, M. Trotignon, and S. Rossignol (2007a). A survey of frequency and voltage control ancillary services—part i: Technical features. *Power Systems, IEEE Transactions on 22*(1), 350–357.

Rebours, Y. G., D. S. Kirschen, M. Trotignon, and S. Rossignol (2007b). A survey of frequency and voltage control ancillary services—part II: economic features. *Power Systems, IEEE Transactions on 22*(1), 358–366.

Reikard, G. (2009, March). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy 83*(3), 342–349.

Robu, V., R. Kota, G. Chalkiadakis, A. Rogers, and N. R. Jennings (2012). Cooperative virtual power plant formation using scoring rules. AAMAS '12, pp. 1165–1166.

Roth, A. E. (2002, July). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica 70*(4), 1341–1378.

Roth, A. E. and M. A. O. Sotomayor (1992). *Two-sided matching: A study in game-theoretic modeling and analysis*, Volume 18. Cambridge Univ Press.

Roughgarden, T. (2005). *Selfish Routing and the Price of Anarchy*. MIT Press.

Savelsbergh, M. W. and M. Sol (1995). The general pickup and delivery problem.
  Transportation science 29(1), 17–29.

Savin, S. V., M. A. Cohen, N. Gans, and Z. Katalan (2005). Capacity management in
  rental businesses with two customer bases. *Operations Research 53*(4), 617–631.

Sawin, J. L. (2013). Renewables 2013, global status report. Technical report, REN 21.

Schuller, A. (2013). *Electric Vehicle Charging Coordination - Economics of Renewable
  Energy Integration.* Ph. D. thesis, Karlsruhe Institute of Technology, Karlsruhe.

Schuller, A., B. Dietz, C. Flath, and C. Weinhardt (2014). Charging strategies for battery
  electric vehicles: Economic benchmark and V2G potential. *IEEE Transactions on
  Power Systems 29*(5), 2014–2022.

Schweppe, F., B. Daryanian, and R. Tabors (1989). Algorithms for a spot price respond-
  ing residential load controller. *IEEE Transactions on Power Systems 4*(2), 507 –516.

Schweppe, F. C., M. C. Caramanis, R. D. Tabors, and R. E. Bohn (1988). *Spot pricing of
  electricity.* Boston u.a.: Kluwer Academic.

Schweppe, F. C., R. D. Tabors, J. L. Kirtley, H. R. Outhred, F. H. Pickel, and A. J. Cox
  (1980). Homeostatic utility control. *IEEE Transactions on Power Apparatus and
  Systems PAS-99*(3), 1151–1163.

Scott, P., S. Thiébaux, M. Briel, and P. Hentenryck (2013). Residential demand re-
  sponse under uncertainty. In C. Schulte (Ed.), *Principles and Practice of Constraint
  Programming,* Volume 8124 of *Lecture Notes in Computer Science,* pp. 645–660.
  Springer Berlin Heidelberg.

Sethi, A. K. and S. P. Sethi (1990). Flexibility in manufacturing: A survey. *International
  Journal of Flexible Manufacturing Systems 2*(4), 289–328.

Seuken, S. (2010). Hidden market design. In *Proceedings of the 9th International
  Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1,* pp.
  1661–1662.

Shaheen, S. A. and A. P. Cohen (2007). Growth in worldwide carsharing: An interna-
  tional comparison. *Transportation Research Record: Journal of the Transportation
  Research Board 1992*(1), 81–89.

Shaheen, S. A., M. A. Mallery, and K. J. Kingsley (2012). Personal vehicle sharing ser-
  vices in north america. *Research in Transportation Business & Management 3,* 71–
  81.

Shann, M. and S. Seuken (2013). An active learning approach to home heating in the smart grid. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI'13, Beijing, China, pp. 2892–2899. AAAI Press.

Sheffi, Y. (1985). Urban transportation networks: Equilibrium analysis with mathematical programming methods.

Siano, P. (2014). Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews 30*, 461–478.

Simchi-Levi, D. (2013). OM forum—OM research: From problem-driven to data-driven research. *Manufacturing & Service Operations Management 16*(1), 2–10.

Sioshansi, F. P. (1995). Demand-side management: The third wave. *Energy Policy 23*(2), 111–114.

Sioshansi, F. P. (2006). Electricity market reform: What have we learned? what have we gained? *The Electricity Journal 19*(9), 70–83.

Sioshansi, R., S. Oren, and R. O'Neill (2008). The cost of anarchy in self-commitment based electricity markets. *Competitive Electricity Markets: Design, Implementation and Performance.*

Skiena, S. S. (2011). *The Algorithm Design Manual* (Auflage: 2nd ed. 2008 ed.). London: Springer.

Skånlund, A. B., A. v. Schemde, B. Tennbakk, G. Gravdehaug, and R. Grøndahl (2013). Loop flows – final advice. Technical Report TE-2013-36.

Small, K. (1982). The scheduling of consumer activities: work trips. *The American Economic Review 72*(3), 467–479.

Smil, V. (2010). *Energy Transitions: History, Requirements, Prospects.* New York: Praeger Publishers.

Sou, K. C., J. Weimer, H. Sandberg, and K. H. Johansson (2011). Scheduling smart home appliances using mixed integer linear programming. In *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), 2011*, pp. 5144–5149. IEEE.

Spees, K. and L. Lave (2007). Demand response and electricity market efficiency. *The Electricity Journal 20*(3), 69–85.

Stein, S., E. Gerding, V. Robu, and N. R. Jennings (2012). A model-based online mechanism with pre-commitment and its application to electric vehicle charging. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '12, Richland, SC, pp. 669–676.

Steinhilber, S., P. Wells, and S. Thankappan (2013). Socio-technical inertia: Understanding the barriers to electric vehicles. *Energy Policy 60*, 531–539.

Steininger, K., C. Vogl, and R. Zettl (1996). Car-sharing organizations: The size of the market segment and revealed change in mobility behavior. *Transport Policy 3*(4), 177–185.

Stevenson, M. and M. Spring (2007). Flexibility from a supply chain perspective: definition and review. *International Journal of Operations & Production Management 27*(7), 685–713.

Stoft, S. (2002). *Power System Economics: Designing Markets for Electricity* (1. ed.). John Wiley & Sons.

Strbac, G. (2008). Demand side management: Benefits and challenges. *Energy Policy 36*(12), 4419–4426.

Streit, T., B. Chlond, P. Vortisch, M. Kagerbauer, C. Weiss, and D. Zumkeller (2013). Deutsches Mobilitätspanel (MOP) - Wissenschaftliche Begleitung und Auswertungen Bericht 2012/2013: Alltagsmobilität und Fahrleistungen. Technical report, Institut für Verkehrswesen, KIT.

Ströhle, P., E. Gerding, M. de Weerdt, S. Stein, and V. Robu (2014). Online mechanism design for scheduling non-preemptive jobs under uncertain supply and demand. In *13th International Conference on Autonomous Agents and Multi-Agent Systems*.

Ströhle, P., C. M. Flath, and C. Weinhardt (2012). An online mechanism for incorporating demand side flexibility into the smart grid. Working Paper.

Stößer, J., D. Neumann, and C. Weinhardt (2010). Market-based pricing in grids: On strategic manipulation and computational cost. *European Journal of Operational Research (EJOR) 203*(2), 464–475.

Subramanian, A., M. Garcia, A. Dominguez-Garcia, D. Callaway, K. Poolla, and P. Varaiya (2012). Real-time scheduling of deferrable electric loads. In *American Control Conference*, pp. 3643–3650.

Subramanian, A., M. J. Garcia, D. S. Callaway, K. Poolla, and P. Varaiya (2013). Real-time scheduling of distributed resources. *Smart Grid, IEEE Transactions on 4*(4), 2122–2130.

Sutton, R. S. and A. G. Barto (1998). *Introduction to Reinforcement Learning* (1st ed.). Cambridge, MA, USA: MIT Press.

Talluri, K. T. and G. J. Van Ryzin (2005). *The theory and practice of revenue management*, Volume 68. Springer.

Tanenbaum, A. S. and M. Van Steen (2002). *Distributed systems*, Volume 2. Prentice Hall.

Tang, C. and B. Tomlin (2008). The power of flexibility for mitigating supply chain risks. *International Journal of Production Economics 116*(1), 12–27.

Taylor, J. W., L. M. de Menezes, and P. E. McSharry (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting 22*(1), 1–16.

Thrun, S. (2010, April). Toward robotic cars. *Commun. ACM 53*(4), 99–106.

Thrun, S., M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, and others (2006). Stanley: The robot that won the DARPA grand challenge. *Journal of field Robotics 23*(9), 661–692.

Tong, S. and D. Koller (2002, March). Support vector machine active learning with applications to text classification. *J. Mach. Learn. Res. 2*, 45–66.

Upton, D. (1994). The management of manufacturing flexibility. *California management review 36*(2), 72–89.

Varaiya, P., F. Wu, and J. Bialek (2011). Smart operation of smart grid: Risk-limiting dispatch. *Proceedings of the IEEE 99*(1), 40–57.

Vardakas, J., N. Zorba, and C. Verikoukis (2014). A survey on demand response programs in smart grids: Pricing methods and optimization algorithms. *IEEE Communications Surveys Tutorials*.

Verbeke, W., P. Farris, and R. Thurik (1998, December). Consumer response to the preferred brand out-of-stock situation. *European Journal of Marketing 32*(11/12), 1008–1028.

Vickrey, W. (1955, May). Some implications of marginal cost pricing for public utilities. *The American Economic Review 45*(2), 605–620. 00138.

Voudouris, C., E. P. Tsang, and A. Alsheddy (2010). *Guided local search.* Springer.

Vytelingum, P., S. D. Ramchurn, T. D. Voice, A. Rogers, and N. R. Jennings (2010). Trading agents for the smart electricity grid. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1*, pp. 897–904.

Wang, J., S. Kennedy, and J. Kirtley (2010a). A new wholesale bidding mechanism for enhanced demand response in smart grids. In *Innovative Smart Grid Technologies (ISGT), 2010*, pp. 1–8.

Wang, J., S. Kennedy, and J. Kirtley (2010b). Optimization of time-based rates in forward energy markets. In *Energy Market (EEM), 2010 7th International Conference on the European*, pp. 1–7.

Weidlich, A. (2008). *Engineering Interrelated Electricity Markets: An Agent-Based Computational Approach.* Springer.

Weinhardt, C. (2012). Engineering smart retail electricity markets. *IEEE Smart Grid Newsletter*.

Weinhardt, C., C. Holtmann, and D. Neumann (2003). Market-engineering. *Wirtschaftsinformatik 45*(6), 635–640.

White, A., E. M. Daniel, and M. Mohdzain (2005). The role of emergent information technologies and systems in enabling supply chain agility. *International Journal of Information Management 25*(5), 396–410.

Woo, C. K., P. Sreedharan, J. Hargreaves, F. Kahrl, J. Wang, and I. Horowitz (2014). A review of electricity product differentiation. *Applied Energy 114*, 262–272.

Zelenović, D. M. (1982). Flexibility—a condition for effective production systems. *International Journal of Production Research 20*(3), 319–337.