The Engineering Economist, 50: 125–158

ISSN: 0013-791X print / 1547-2701 online
DOI: 10.1080/00137910590949887

Taylor & Francis
Taylor & Francis Group

MANAGING CAPACITY IN THE HIGH-TECH INDUSTRY: A REVIEW OF LITERATURE

S. David Wu

P.C. Rossin College of Engineering and Applied Science,
Lehigh University, Bethlehem, Pennsylvania, USA

Murat Erkoc

Department of Industrial Engineering, University of Miami,
Coral Gables, Florida, USA

Suleyman Karabuk

Department of Industrial Engineering, University of Oklahoma,
Norman, Oklahoma, USA

This article surveys a new generation of analytical tools for capacity planning and management, especially in high-tech industries such as semiconductors, electronics and bio-techs. The objectives of the article are to (1) identify fundamental theory driving current research in capacity management, (2) review emerging models in operations research, game theory, and economics that address strategic, tactical and operational decision models for high-tech capacity management, and (3) take an in-depth look at capacity-optimization models developed in the specific context of semiconductor manufacturing. The goal of this survey is to go beyond typical production-planning and capacity-management literature and to examine research that can potentially broaden capacity-planning research. For instance, we explore the role of option theory and real options in modeling

Address correspondence to S. David Wu, P.C. Rossin College of Engineering and Applied Science, Lehigh University, 19 Packard Laboratory, Bethlehem, PA 18015, USA. E-mail: sdw1@lehigh.edu

capacity decisions. We not only examine capacity-planning problems from the perspective of a particular firm, but also the interaction of capacity investment among supply chain partners. Not only are these issues increasingly important in the fast-changing high-tech environment, they draw on new tools from different disciplines and pose significant intellectual challenges. We also examine papers that represent the multifaceted nature of high-tech capacity planning, integrating capacity decisions with issues related to contracting, coordination, sourcing, and capacity configurations.

BACKGROUND AND INTRODUCTION

In high-tech industries such as semiconductor, consumer electronics, telecommunications and pharmaceutical, a firm's ability to manage capacity is arguably *the* most critical factor for its long-term success. Even in a stable economy, the demand for high-tech products is volatile and difficult to forecast; the rapid rate of technology innovation causes short product lifecycles, low production yield and, oftentimes, long production lead time, all of which hamper the firm's ability to respond to market changes. Uncertain economic times exacerbate these challenges. Whereas in an environment of sustained demand growth, firms might build inventory or hold excess capacity to buffer against demand variability, most are reluctant, or unable, to assume such financial risks in a downside market. Nevertheless, high-tech companies recognize that in order to sustain their customer base and to seize revenue opportunities, they must be able to manage successive technological innovations effectively, e.g., introducing high-margin innovative products at the right moment while maximizing the return-on-investment for older, more mature products. To do so, firms must structure capacities in their supply chain so that over time it is possible to respond to demand surge from new product introduction and market upside, and to absorb short-term decline due to technological migration and market downside.

The role of capacity management is even more important in industries in which capital equipment cost is high. For example, in the semiconductor industry, manufacturers are faced with astronomical capacity costs, long capacity lead times, high obsolescence rates and high demand volatility. A new semiconductor fab costs $1–4 billion to build, and the price for a single machine may be as high as $4–5 million. Moreover, the rapid technology innovation leads to short product lifecycles and thus to higher obsolescence rates and increased equipment usage costs. To make the situation worse, the demand variability during a particular quarter may peak above 80% of the average sales and, according to the Semiconductor Industry Association, the equipment procurement lead times are usually as long as

6–12 months. This means that the demand beyond the capacity lead times is highly uncertain. The above environment drives semiconductor manufacturers to adopt exceedingly conservative capacity expansion policies (Erkoc & Wu, 2004). However, in a fast-growing global market, the conservative capacity-expansion policy leads to severe shortfalls in service levels. In a recent survey of managers by the Fabless Semiconductor Association (FSA), the respondents emphasize significant concerns about whether the manufacturers (foundries) will be able to supply wafers to meet demand. The most common reason cited in the survey for missing forecasts is the lack of adequate capacity from the foundries rather than internal issues, such as the lack of a specific technology or process (Ridsdale, 2000).

A similar phenomenon is also observed in the biotech industry. Many bio-drugs on the market require such high volumes of manufacturing capacity that capacity is always in shortage. Recent surveys reveal that 50% of the contractors in this sector believe that the general capacity shortage is unlikely to improve soon; this is due to the current business practice that puts most of the capacity expansion risk on the manufacturer's shoulders (Fox et al., 2001). A typical manufacturing facility in the biotech sector costs between $200 million and $400 million to build, a significant risk to bear, especially when a drug has yet to receive regulatory approval. To address these challenges, many drug manufacturers have begun to work with their pharmaceutical and biotechnology customers to forge long-term commitments and co-investment plans. An emerging trend in the industry is that major drug developers begin to co-invest facilities and equipment with their manufacturing partners in exchange for guaranteed (capacity) slots.

In this article, we review literature relevant to high-tech capacity planning and management. Using the competitive operational environment of the high-tech industry as the backdrop, we examine the impact of capacity from strategic as well as tactical and operational perspectives. The papers we review are not limited to applications in the context of high-tech industries but they *are*, in our opinion, representative of critical research ingredients in this area.

KEY RESEARCH ISSUES FOR CAPACITY PLANNING

We consider research issues for capacity planning from the strategic, tactical and operational levels. At the strategic level, capacity planning involves not only the firm's own capacity investment, but also its supply chain partners' investments. The capacity investment of one firm in the supply chain could have enormous impact on the performances of all upstream and downstream firms; thus, strategic interactions between two or more players need to be taken into account. In addition to monolithic models that employ tools such as expected utility theory and dynamic programming, the literature increasingly considers settings that model independent multiple decision makers in the context of supply chain management. Research in this area utilizes game theoretic models focusing on issues such as contracting, co-ordination and risk-sharing mechanisms.

At the tactical level, capacity planning focuses on capacity expansion tactics as related to the operational aspects of the firm. A comprehensive survey of the OM literature on the size, type, and timing of capacity investments is given by van Mieghem (2003). In this paper, we focus our examination of tactical capacity expansion literature on three key aspects. First, we review OM literature that incorporates characteristics of the high-tech industry in the traditional production, inventory and demand management models. Second, we review the growing literature on real options as related to high-tech capacity planning. In recent years, real options have become quite popular as a means for modeling capacity investment risks based on endogenous and exogenous factors. Third, we survey papers that examine risk sharing and vertical integration between suppliers and buyers through capacity reservation contracts.

At the operational level, capacity planning typically refers to decision support models developed for a specific operational environment. There is a significant literature for operational capacity planning in the semiconductor industry, which will be our main focus. We classify operational models in semiconductor manufacturing according to the level of detail that they capture and the length of the planning horizon that they consider.

STRATEGIC AND TACTICAL MODELS: GAME THEORETIC AND ECONOMIC ANALYSIS

Capacity Planning with Production, Inventory and Demand Management Perspectives

Demand uncertainty and the short product lifecycles of high-tech products are two key factors that influence capacity expansion models. High-tech manufacturers avoid carrying inventory due to high obsolescence rates. In fact, high-tech products are often treated as perishable goods. Inventory models developed in this context typically use news vendor or news vendor networks settings with single-period and stochastic demand. These models consider capacity investment by a single or multiple independent decision makers in a stationary environment; once capacity is built it stays unchanged during the planning horizon. The profit to the firm is modeled as a function of the capacity level, K, and the state of the world (e.g., realized demand), ξ. Extensions to news vendor models study multi-period settings with capacity investment adjustments over time. In this setting, capacity at time $t-1$, K_{t-1}, is adjusted to K_t at time t at some cost. If the investment is (fully or partially) reversible, then contraction (i.e., $K_t < K_{t-1}$) is possible. Otherwise, with irreversible investment the firm either expands or maintains its current capacity across periods.

Newsvendor-Style Models
This literature employs aggregate planning for the acquisition and allocation of resources to satisfy customer demand over a specific time period. The short product lifecycles in high-tech make such approach quite appropriate. A few papers in this category investigate capacity expansion and investment strategies jointly with inventory management and/or outsourcing policies. Bradley and Arntzen (1999), observe that firms achieve better financial results by optimizing their capacity and production/inventory decisions simultaneously. They demonstrate their result using a case study performed at an electronics firm. With the increasing pace of technological innovation and the increasing cost of manufacturing equipment, many OEMs are reluctant to respond to economic cycles by adjusting their own in-house capacity. Consequently, capacity outsourcing has become an integral part of capacity investment decisions (Mason et al., 2002). Atamturk and Hochbaum (2001), propose a four-way tradeoff among capacity, production, subcontracting, and inventory levels over a finite horizon. Kouvelis and Milner (2002) consider two-stage supply chains and analyze the impact of supply/demand uncertainty on capacity and outsourcing decisions. They conclude that greater supply uncertainty encourages vertical integration, because the OEMs have incentives to make investments in their suppliers to ensure reliable and continuous supply. In contrast, outsourcing becomes more attractive as uncertainty in demand increases.

Pindyck (1993) shows that demand uncertainty can discourage firms from capacity expansion when there is perfect competition; while Kulatilaka and Perotti (1998) show that higher uncertainty may increase the firm's incentive to invest when there is imperfect competition. Van Mieghem (1999) studies the trade off among capacity investment, production and subcontracting in a two-stage, two-player, two-market setting. He models the interactions of a manufacturer's and a subcontractor's decisions. He observes that the manufacturer subcontracts more (invests less on his own capacity) as the demand uncertainty increases, which induces the the subcontractor to invests more. Under a similar setting, Tan (2004) investigates capacity investment and pricing decisions for a manufacturer and a subcontractor with guaranteed availability. Van Mieghem and Dada (1999) examine the interplay among capacity, inventory and pricing decisions. The authors study the impact of the timing on these decisions and on the firm's profitability; they examine different market settings such as monopoly, oligopoly and perfect competition Other news vendor–like game theoretical models study sizing (Bernstein & DeCroix, 2004), timing (Ferguson, DeCroix, & Zipkin, 2002), and allocation (Karabuk & Wu, 2003; Mallik & Harker, 2004) of capacity under competitive settings.

Most of the above papers consider a single product and resource type. In reality, however, different products may share the same resource, and different types of resources may be needed to process a particular product. A line of literature considers the presence of multiple resources where decisions must be made for the configuration and selection of optimal resource types. This results in a network design problem. Few researchers tackle the capacity expansion problem with multiple products *and* multiple resources; however, there is growing interests in models that consider more general product/resource settings. In an early paper, Dixit (1997) discusses optimal investment policies in a two-resource setting. Harrison and van Mieghem (1999) propose a product-mix linear programming model and use its optimal shadow prices to extend the classical news vendor model to a so-called multidimensional news vendor solution with multiple resources and multiple product types, which leads to the well-known critical fractile solution. The critical fractile values balance overage costs with underage costs and are computed using shadow prices. Using the multidimensional news vendor approach, Van Mieghem (1998) studies a two-product setting in which the firm has the option to invest in two product-dedicated resources or one flexible resource that can process both products. The paper examines the impact of price, cost, demand uncertainty and demand correlations on the investment decisions. Later, van Mieghem and Rudi (2002) extend these models to news vendor networks that incorporate multiple products, multiple resources and multiple storage points. They observe that when demand is normally distributed, the optimal expected investment value is an increasing function of the demand vector and a decreasing function of any variance term. Bish and Wang (2004) also use news vendor networks to investigate investment policies for product-flexible versus dedicated resources in a two-product setting with correlated demand.

Multi-Period Models with Capacity Adjustments

Multi-period or dynamic capacity expansion models determine policies that specify the timing and scope of capacity adjustment so as to maximize the expected net present value of the firm's investment. These models seek answers for when and how much capacity to build in a dynamically changing environment. Capacity decisions are strongly influenced by the length of the planning horizon and the associated rate of depreciation, as well as by the cost of investment and the demand uncertainties during the aggregated planning period. Most papers in this area utilize stochastic dynamic programming models. Van Mieghem (2003) points out that there are three main challenges in capacity cost modeling: indivisibility, irreversibility and nonconvexity. Indivisibility implies lumpy capacity expansions, which are common in high-tech applications. Due to rapid technological innovation, almost all capacity investment in high-tech manufacturing is irreversible (in the sense that capacity expansion cannot be undone without significant cost). Irreversibility contributes to nonlinearity in capacity expansion decisions since it might prevent firms from making downward capacity adjustments to demand changes from one period to another, leading to 'no action' policies in some periods. Moreover, one may not be able to assume convex capacity costs due to fixed-cost, economies-of-scale, etc., adding another layer of complexity to the problem.

Research on dynamic capacity models with stochastic demand goes back to the seminal work of Manne (1961), in which he models demand growth using Brownian motion with positive drift for a single-resource system. The resulting regenerative process leads to uniform capacity increments that take place whenever the demand backlog goes beyond a threshold value. The timing of the expansion is modeled by the 'hitting time' of the Brownian motion. The author shows that by appropriately choosing the discount rate based on demand variance, a deterministic equivalent model can be built to solve the stochastic problem. In Eberly and van Mieghem (1997), the authors propose a more general continuous-time model that considers n resources and non-stationary demand. Interestingly, they show that optimal investment strategies follow a control limit policy, i.e., an ISD or invest/stay put/disinvest policy. At a given time, the ISD policy partitions the state space into various regions in an n-dimensional space based on current capacity and the state of the world. Each region and its boundaries specify the optimal investment policy for each resource that can be expanded, contracted or left in place. The authors are able to provide closed-form solutions when the uncertainty is modeled using a geometric Brownian motion. Assuming i.i.d. demand, a similar paper (Harrison & van Mieghem, 1999) applies the newsvendor networks approach to the above setting. For a similar problem with a single product setting, Narongwanich, Duenyas, and Birge (2002) consider the effect of indivisible capacity. They consider a manufacturing firm that introduces new (generations of) products in stochastic time intervals. The paper examines investment policies in dedicated systems that are reconfigurable for future capacity needs. They show that the optimal investment policy remains ISD, provided that all resources have identical adjustment sizes. In the case of non-identical adjustment sizes, the optimal policies are still ISD-like but require perturbations due to the 'lumpiness' of the capacity.

In many semiconductor fabs, the capacity level could have a substantial effect on the wafer processing costs. Consequently, there may be a strong dependency between production and capacity; this dependency needs to be taken into consideration in capacity planning problems. A simple cost model capturing this dependency is presented by Iwata and Wood (2002). Due to scale economies and short product lifecycles, when incorporating production costs as a function of capacity, capacity costs are non-convex. A few papers in semiconductor manufacturing tackle this phenomenon. An example is Dixit (1995), in which the author ties the production function to the capacity in a convex-concave fashion. Production first exhibits increasing return to scale followed by decreasing return to scale as a function of capacity, which leads to optimal investments that are lumpy. A similar approach is used by Murto, Nasakkala, and Keppo (2004) for an oligopoly market with competition. The authors investigate the tradeoff between the value of flexibility and economies of scale under competition by modeling the capacity investment using a Nash-game setting.

Under a single resource, multi-agent setting, Armony and Plambeck (2003) consider a manufacturer that sells through two distributors. They use a queuing model to study the influence of double orders and cancellations on the manufacturer's capacity investment. At each distribution center, customers arrive according to a Poisson process demanding one unit at each time. If the end-item is out of stock at the distributor, in addition to the backorder the customer may order from another distributor with a probability, resulting in a double order. Moreover, the customer may cancel all outstanding orders after an exponentially distributed waiting time. They show that in this setting the manufacturer may overestimate the demand and the cancellation rates, and thus may over-invest in capacity. As such, the authors argue that the economic down-turn may not be the only factor to blame for the sizable write-offs witnessed in the semiconductor industry in recent years. Double orders placed by end customers through different distributors often leave semiconductor manufacturers with excess capacity as well. A caveat is that when the capacity cost is high and the manufacturer is unaware of double orders, she may still under-invest in capacity.

Taylor and Plambeck (2003) consider 'relational contracts' between a high-tech firm and a supplier. The firm periodically introduces innovative products, and the supplier needs to invest in capacity to produce components for the product. Relational (incomplete) contracts are needed when capacity investment must take place before the new product is fully defined. In other words, relational contracts specify informal agreements between parties about how they will behave. In this setting, the high-tech firm may promise to purchase at a price that reflects the cost of the capacity so that the supplier has incentive to increase capacity investment. Essentially, the value of future business relationships provides incentives for the firm to comply with the previously agreed price. Their analysis indicates that when capacity is expensive and demand is uncertain, the firm's preferred strategy is to commit to a specified unit-price only, but not to any quantity.

As mentioned earlier, most high-tech products have short product lifecycles that can be characterized by a single modal bell-shaped curve that represents the progression from initial ramp-up, to maturity, and then decline. The demands for these products can be viewed as a stochastically increasing function followed by a stochastically decreasing one, or a regime-switching model. The main challenge in these models is to capture the tradeoff between the capacity costs and the expected revenue from the product's demand over its life cycle. A few researchers consider regime-switching explicitly in their modeling (c.f., Angelus & Porteus, 2002; Angelus, Porteus, & Wood, 2000; Cakanyildirim & Roundy, 2002; Ho, Savin, & Terwiesh, 2002; Bollen, 1999). Angelus and Porteus (2002) study simultaneous capacity and production planning problem for a short lifecycle product where capacity can be reduced as well as expanded at exogenous costs. They first consider the case in which inventory carryover is not allowed, and thus demand in a given period must be satisfied by production, which is bounded by available capacity. They assume no lead-time for the capacity and no backlogs for the lost sales. They show that the optimal capacity plan can be reduced to a one-dimensional ISD policy that the authors refer to as the *target interval policy*. The target interval policy specifies a lower and upper capacity target. If the current capacity is below the lower target, capacity is increased to this target level. If the current capacity is above the upper target, capacity is disinvested and decreased to this target level. No action is taken if the capacity is in between. When there is no inventory carryover, the optimal capacity levels can be scheduled in advance, as they do not depend on the demand realizations. However, the same is not true when inventory carryover is allowed since optimal capacity at a given period depends on the current inventory level, which is a function of demand realizations in previous periods. In this case, it is shown that capacity and inventory are economic substitutes. They conclude that it is optimal to change the service level provided to customers across periods. Specifically, the optimal policy provides the lowest service level during the peak period and the highest service level when capacity is adjusted.

In a similar setting, Angelus, Porteus, and Wood (2000) consider only capacity additions, but capacity costs demonstrate economies of scale and non-negative lead-time. They show that an expansion policy that is analogous to the well known (s, S) inventory policy is optimal. Capacity is expanded to S if and only if the current level is below s. Cakanyildirim and Roundy (2002) expand the above setting to a multi-resource production process with lumpy capacity expansion/contraction. They propose a polynomial-time algorithm that applies bottleneck policies (BP) to optimally plan for capacity adjustments over time. Under BP it is always optimal to buy machines that are the same type as the bottleneck and to retire machines in the reverse order. Their algorithm determines the optimal clusters of machines to be installed and retired simultaneously and the timing of the adjustments. The authors illustrate their algorithm using real-life data provided by the SEMATECH databases. The timing of capacity expansion in this context is also studied by Huh and Roundy (2002) and Ryan (2004); however, they consider only stochastically growing demand. The former paper extends the model proposed by Cakanyildirim and Roundy (2002) to multi-product settings, whereas the latter incorporates lead-times for capacity expansion.

In product lifecycle models, the standard assumption is that lifecycle demand forecasts are given and independent from supply decisions; the impact of past sales on future demand is often ignored. To fill this gap, Ho, Savin, and Terwiesh (2002) incorporate capacity decision into the analysis of demand and sales dynamics in a supply-constrained, new-product diffusion model. The authors incorporate lifecycle demand into an optimal-control framework by modeling independently the *innovation dynamics* and the *interaction dynamics*; the latter refers to the interaction between customers (early adopters) who have purchased the product and potential customers who are not yet ready to adopt the product. The proposed model analyzes how much the firm should invest in capacity and when it should launch the new product in a Bass-like diffusion environment. The results suggest that it is often optimal for a firm to delay the product launch until after capacity is built; this allows the firm to pre-produce and build up initial inventories before entering the market. However, this is only true provided that the product diffusion does not occur before the product launch. As in the model proposed by Angelus and Porteus (2002), the initial inventory serves as a substitute for capacity. This strategy is particularly relevant for high-tech firms where physical capacity expansion is prohibitively expensive.

Bollen (1999) employs an option valuation framework to incorporate the stochastic lifecycle of a product into the firm's capacity investment and production planning problem, which is reviewed in the following subsection.

Capacity Investment through Option Valuation

Capacity investment in high-tech industry typically involves substantial cash exposures, volatile market demand and changing supply (technological) specifications. In this environment, the return on investment is highly uncertain; thus, the variability in returns is at least as important as the expected return. Moreover, most investment expenditures are irreversible and are considered sunk costs for the firm once invested. In general, capacity investment implicates considerable risk for high-tech firms. The traditional approach of maximizing net present value (NPV) when analyzing capacity investment is not sufficient to capture the managerial flexibility that exists throughout the capacity planning process (see Feinstein, 2002). A more accurate model should consider not only the purchasing and installation costs, but also the value of the options one could invest elsewhere (Dixit & Pindyck, 1994). Option theory provides a powerful tool to value risky investments through risk-neutral discounting and to incorporate risk without explicitly defined utility functions. Since the seminal paper by Black and Scholes (1973), option pricing has become a popular topic in finance. The enormous attention in literature on options is summarized in the comprehensive survey by Broadie and Detemple (2004). Although option theory has been primarily studied in finance, the potential benefits of real options have been recognized by researchers in operations management and engineering economics in recent years (see Miller & Park, 2002 for a recent survey). Real options theory is particularly relevant to capacity planning because it focuses on the combined importance of uncertainty and managerial decisions and it offers a dynamic view of firm's investment and operational decisions. Birge (2000) asserts that since operational decisions have the goal of maximizing value, the framework of real options can be used to evaluate decisions under risk. Making the connection between the effect of capacity on the firm and the pricing of a call option, he shows that risks can be incorporated into planning models through capacity adjustments. He proposes a model that integrates financial risk attitudes into a linear capacity investment problem.

In general, options theory has broad appeal to a variety of application areas in capacity and production planning. Examples include options to expand, options to defer production (Pindyck, 1988), options to abandon a project (Majd & Myers, 1990), options to wait or temporarily shut down production (McDonald & Siegel, 1985), and options to switch (van Mieghem, 1998). Johnson and Billington (2003) report that in recent years, high-tech companies are making use of real options to determine their investment and operating strategies, more specifically, the *timing* and *choice* of capacity adjustments. There is a growing espousal of real options models for high-tech capacity planning in the operations management literature. A broad description and discussion of real options can be found in Trigeorgis (1996) and Amran and Kulatilaka (1999).

Each unit of capacity provides the firm options to produce a certain quantity of the product throughout its lifecycle; such options are referred to as the *operating options*. The investment in capacity is the premium for the option, while the production cost corresponds to the exercise price. On the other hand, the firm usually has options to add more capacity, known as *growth options*. In general, options are early investments associated with firms' ability to expand in the future; the investment may be the acquisition of land (or access to) facility, technology, know-how or other resources. Following the legacy of option pricing, a majority of papers in this area employ geometric Brownian motion to model demand changes; as such, demands in future time periods can be modeled by a lognormal distribution. Under such a setting, Pindyck (1988) studies a capacity planning problem with irreversible investment. In this model, the value of the firm is determined by the present value of the expected flow of net gains generated by available capacity (i.e., the value of the operating options) plus the present value of additional profits that can be generated should the firm add more capacity in the future less the present value of the cost of capacity (i.e., the value of the growth options). Both valuations are functions of the current capacity level and the demand shift parameter. Pindyck (1988) shows that the firm's capacity choice is optimal when the present value of the expected cash flow from each marginal unit of capacity is equal to the total cost of that unit. The total cost is the procurement cost plus the opportunity cost of exercising the option to buy the unit. He models the demand at time t by $Q_t = \theta_t - \beta P_t$, where θ_t and P_t are the stochastic demand parameter and the price at time t, respectively. This particular way of modeling demand and price is quite common in this literature. The underlying assumption is that price adjusts instantly to balance supply and demand. Under this setting, the author concludes that uncertainty in demand increases the value of the firm's operating and growth options. Equivalently, the value of the unit capacity and the opportunity cost, both grows with uncertainty. However, since the increase in the latter is higher, the firm will end up holding less capacity when the uncertainty in future demand increases.

A notable case study that adopts Pindyck's approach in the context of semiconductor manufacturing is presented by Benavides, Duley, and Johnson (1999). The authors study the determination of the optimal scale, type and timing of IC manufacturing capacity expansion for a fab. Two distinct types of capacity, fixed and expandable, with different sizes are considered for a product whose demand changes stochastically over time. The analysis indicates that, under uncertainty, sequentially deployable fabs are economically more viable since they provide a growth option to the firm in addition to the operating option, which allows much of the fab's required capital investment to be delayed. The presented model considers only the growth phase of the demand for the product.

Pindyck (1993) observes similar results concerning the effect of uncertainty on investment incentives; this research further considers firms in complete competition environments. In contrast to these results, Dangl (1999) observes that optimal capacity investment increases significantly with uncertainty if the firm has to fix capacity size at the time of installation for the entire life time of the facility. In such settings, the capacity is unexpandable once installed; as such, the author shows that it is optimal to delay the investment until high demand ranges even when uncertainty in demand is small.

In a different setting, Abel et al. (1996) investigates how options to expand capacity and options to disinvest affect the investment policies under uncertainty in a two-period model. They model these options as *call* and *put* (American) options respectively and examine their values and characterizations. In contrast to Pindyck (1993), their setting allows for partial irreversibility and thus put options to sell capital. In general, call (growth) options diminish the firm's incentive to invest since they add to the firm's value and are killed by investment. On the other hand, put (disinvestment) options increases the incentive to invest because it provides irreversibility. Consequently, the authors observe that an increase in uncertainty has an ambiguous impact on the investment incentives because it increases the value of both options. Similarly, Kulatilaka and Perotti (1998) conclude that for a setting in which the firm needs to make a decision on irreversible investment under imperfect competition, the impact of uncertainty on the value of the strategic growth options is context specific.

As pointed out earlier, high-tech products typically have a short lifecycle. Therefore a straightforward adoption of Black-Scholes formula in modeling capacity expansion may lead to inaccurate conclusions since the drift can change direction (Bowman & Moskowitz, 2001). Bollen (1999) recognizes the fact that simple stochastic processes may not accurately represent capacity investment in many important manufacturing sectors, such as the semiconductors and the pharmaceuticals, because they are characterized by well defined product lifecycles with bell-shaped growth patterns. He proposes a generalization of the real options valuation that explicitly incorporates stochastic product lifecycle into the firm's capacity investment and production planning problem. The product lifecycle is represented using a regime-switching process in which the planning horizon starts with a *growth regime* (increasing demand) and then switches stochastically to a *decay regime* (decreasing demand). He considers both the fixed (irreversible) capacity and the flexible (reversible) capacity. In the former case, the capacity must be fixed at the beginning of the project and cannot be adjusted throughout the lifecycle. The firm chooses the capacity investment that maximizes the project's NPV, given optimal production policies across periods, the cost of investment, the demand shift parameter and the discount factor. At the beginning of each period when it is possible to expand and contract capacity, the firm chooses the current capacity level based on the tradeoff between the cost of adjusting the capacity and the change in expected future profits. To model the stochastic demand in the two-regime context, the author employs the Wiener process; however, he considers different drift values for the demand parameter in the growth and decay regimes. While the drift is assumed to be positive in the growth regime, negative drift is used to model the decay regime. In a given period, the probability of switching from growth to decay is defined by a cumulative normal distribution function of the time elapsed since the beginning of the project. The author examines the sensitivity of the project and option values to demand uncertainty and the project lifecycle through a numerical analysis. He shows that by ignoring the product lifecycle, traditional approaches may undervalue the contraction option, by underestimating the probability that the demand may fall at some point in the future, and may overvalue the expansion option, by implicitly assuming that demand is expected to grow indefinitely.

Most capital investment decisions are made in multiple stages of the product lifecycle including R&D, product introduction and physical expansion. At each stage, one must decide whether to exercise the previously acquired option. This decision depends primarily on the valuation of the downstream options that will be created over the course of the product lifecycle. For example, in moving from the R&D stage into the product introduction stage, options on R&D investments must be exercised (made). This leads to the option of expanding product introduction, followed by the creation and exercise of subsequent physical expansion options. Therefore, a modeling approach that incorporates sequential investment decisions throughout the product lifecycle should employ *compound options*. The valuation of compound options is interdependent, i.e., exercising an upstream option generates a downstream real option. Therefore, the modeling of compound options requires the consideration for multiple sources of uncertainty. Herath and Park (2002) develop a compound real option valuation model assuming four sequential investment opportunities, namely, R&D investment, product introduction, the first expansion phase and the second expansion phase. In order to value compound options under multiple uncorrelated sources of uncertainty, the authors employ an extended version of the binomial lattice framework. Using Monte Carlo simulation they compare their approach with the traditional NPV method. The proposed model is novel in that it combines R&D investment decisions with physical capacity decisions in a real options framework. One of the main distinguishing characteristics of the high-tech industry is that R&D costs are mostly comparable to capacity costs even though they significantly affect the firm's overall value (Hicks, 1996).

Risk Sharing through Capacity Reservation

Traditional approaches in capacity investment assume that all the investment risks are absorbed by the firm who owns (builds) the capacity.

However, in high-tech supply chains, in which capacity is capital intensive, products have short lifecycles and demand uncertainty is high, on one hand the component manufacturers (the suppliers) often adopt an exceedingly conservative capacity expansion policy; this reduces their downside risk at the expense of upside potentials. Consequently, their downstream buyers (e.g., OEM manufacturers) may not have adequate supplies to fill the market orders. On the other hand, the buyers will avoid making *firm* commitments to the suppliers on their future purchases due to high uncertainty. However, to ensure higher availability, the buyers might be willing to share risk by sharing *partial* liability for the capacity as long as it is economically justified. To achieve this, risk-sharing mechanisms that create proper economic incentives must be developed. There is a growing literature about supply chain *coordinating contracts* that describe mechanisms that align the incentives of supply chain partners via risk/profit sharing. An excellent review of coordinating contracts is provided by Cachon (2003).

Capacity reservation contracts are an increasingly popular way to model the allocation of risks across suppliers and buyers in high-tech supply chains. Similar to the research surveyed earlier, work on these contracts regards capacity as an option to be exercised in the future to produce needed goods. In fact, they are known as *option contracts*. Typical settings consist of one supplier and one buyer interacting with each other in two phases. In the first phase, a *reservation contract* specifying a reservation fee (option price), r, an execution fee, e, and a reservation quantity, Q, is agreed upon by both parties. At this stage, the demand is unknown and usually represented by a probability distribution function. While the reservation fee is immediately payable, the exercise fee is due when the option is exercised (after demand uncertainty is resolved). Based on the reservation fee, the buyer chooses Q, which is matched by the supplier's capacity. In the second phase, the buyer decides on the exercise amount and pays the exercise fee after observing the realized demand. By appropriately choosing the contract parameters, both parties can improve their expected profit. Capacity reservation contracts are applications of *call* option models since each reserved capacity gives the buyer the right to purchase in the future. On the other hand, capacity coordination can be also achieved by *buy-back contracts* which correspond to *put* options in finance. The buyer pays for the capacity in full upfront, but she has the right to return the unsold products to the supplier for a price that is usually below the wholesale price. It should be noted that the pricing of options in this context is different than the pricing of financial options and those discussed in the previous section where the prices are based on no-arbitrage and Black and Scholes principles. Instead, in capacity reservation contracts, the trading parties determine their actions and valuations in accordance with their incentives and strategic interactions.

The strategic interactions between the supplier and the buyer are usually modeled using a game-theoretic approach based on *Stackelberg* games. Stackelberg games, also known as leader-follower games, are used to model the competitive behaviors of independent players when they act in sequence. Typically, one of the players (the leader) acts first, and the other (the follower) reacts. In capacity reservation contracts either the supplier or the buyer can be a leader depending on each one's respective market power. For example, in a supply chain with a powerful OEM as the buyer and a contract manufacturer as the supplier, it is more realistic to model the buyer as the leader. As another example, the OEM could be the supplier producing a custom designed telecommunication IC device for another OEM and thus has enough power to act first. Depending on who is leading the channel, the terms for reservation can be determined either by the supplier or the buyer. Other types of games can be also incorporated into the model. For example, Nash games can be used to model simultaneous actions of multiple buyers (suppliers) competing for the supplier's capacity (buyer's orders). For readers unfamiliar with game-theory, Fudenberg and Tirole (1991) and Allprantis and Ckakrabarti (2000) provide useful backgrounds. In what follows, we review papers that study capacity reservation contracts using the above described framework as their basic settings. We refer the reader to Kleindorfer and Wu (2003) and Spialer (2003) for further reading on this subject.

Research on capacity reservation contracts can be categorized into two groups based on how they motivate the buyer's incentives for reserving capacity. Capacity reservations are motivated by either 1) reducing potential cost through early commitments, or 2) ensuring availability during demand upsides. Quite a few papers that fall in the first group show that cost reduction through capacity reservation can be realized through advanced contracting and early commitments. This line of research typically considers multiple ordering opportunities in which the buyer has the option of committing to an order quantity in advance and then purchasing additional quantities at a higher cost (e.g., spot market price) after demand information is updated. Related papers include Brown and Lee (1998), Serel, Dada, and Moskowitz (2001), Bonser and Wu (2001), Wu, Kleindorfer, and Zhang (2002), Spinler, Huchzermeier, and Kleindorfer (2002) and Wu and Kleindorfer (2003). Brown and Lee (1998) study capacity reservations in the context of semiconductor manufacturing; in particular, they discuss 'pay-to-delay' capacity reservation contracts. The buyer provides an initial forecast and a contract consisting of both firm commitments and capacity options. The main incentive for the buyer is to minimize procurement cost by committing earlier and taking advantage of discounts offered by the supplier. It is typically assumed that the supplier always has sufficient capacity to offer. The authors focus their analysis on the buyer's perspective and derive optimal policies for the buyer only. Serel and Dada (2001) extend the capacity reservation problem to a multi-period setting with stationary demand. In their model, the buyer contracts a certain number of products from the supplier for each period by paying a reduced rate. As such the supplier guarantees the delivery of the buyer's order up to the contracted amount. Bonser and Wu (2001) propose a similar multi-period setting, in which future supplies are secured by a long-term contract and spot market purchases. To minimize procurement costs, the buyer must fulfill long-term contract commitments to avoid an 'underlife' penalty while at the same time taking advantage of spot price fluctuations.

Wu, Kleindorfer, and Zhang (2002) consider a supplier-lead channel in which both the supplier and the buyer have access to spot markets to sell or to buy unstorable goods. However, uncertainty in future spot market prices creates incentives for both parties to embark on a long-term capacity reservation contract in which the supplier acts first and determines her reservation cost and her exercise fee by anticipating how the buyer will react. Next, the buyer chooses the reservation amount. In this model the only source of uncertainty is the spot market price. The buyer will not decide how many of her options to exercise until she observes the realized spot market price. Spinler et al. (2002) extend this framework by incorporating uncertainty into the buyer's future demand and the seller's future marginal costs. They show that under option contracts both parties are better off compared to other market schemes. In this environment, the buyer's demand for options depends on the correlation between buyer demand and spot price. Motivated by emerging B2B exchanges, Wu and Kleindorfer (2003) extend the model in Wu, Kleindorfer, and Zhang (2002), they examine the situation where suppliers compete to provide capacity for a single buyer. They investigate the optimal portfolios of contracting and spot market transactions for the buyer and the suppliers, and determine the market equilibrium pricing strategies. Interestingly, the authors observe that competition in options markets improves overall efficiency in contrast to forward contracts.

The second group of papers investigates capacity reservation contracts that are motivated by ensuring availability during market upsides. Papers in this category include Cachon and Lariviere (2001), Jin and Wu (2001), Barnes-Schuster, Bassok, and Anupindi (2002), Burnetas and Ritchken (2003), Tomlin (2003), Ozer and Wei (2003), Cheng et al. (2003) and Erkoc and Wu (2004) Cachon and Lariviere (2001) and Tomlin (2003) focus on buyer-lead models and investigate forced and voluntary compliance regimes. The former paper examines capacity contracting in the context of supplier-buyer forecast coordination. The buyer provides an initial forecast and a contract consisting of firm commitments and capacity options. After the supplier builds capacity, the buyer places an order based on the up-to-date forecast. They show that although supply-chain coordination can be achieved through option contracts in the full information case, it is only possible when compliance is forced. They conclude that in the absence of forced compliance, higher supplier capacity cannot be induced. Tomlin (2003) enhances this approach by introducing an intermediate compliance regime which he refers to as *partial* compliance. He shows that under nonlinear price-only contracts, options do increase the supplier's capacity. However, full coordination is not necessarily achieved. On the other hand, Erkoc and Wu (2002) show that in a supplier-lead channel the supplier will always have incentive to signal the buyer that she will be fully compliant by offering a noncompliance penalty scheme that won't be disputed by the buyer. Thus, they prove that coordination can still be achieved under voluntary compliance regime in a supplier-leading channel.

Barnes-Schuster, Bassok, and Anupindi (2002) study a general case in which the buyer both places firm orders and also purchases options under a two-period setting with correlated demands. The buyer faces uncertain demand in both periods. She places firm commitments for both periods and an optional quantity for the second period. The buyer has the option of carrying inventory from the first period to the second. She utilizes her first-period demand to update the forecast for the second period's demand and then exercises her options based on the updated information at the beginning of the second period. The authors examine optimal ordering policies and their implications on supply chain coordination.

Jin and Wu (2001) study capacity coordination under exogenous wholesale price by utilizing take-or-pay contracts set by the supplier in a high-tech manufacturing supply chain. In their setting, a per unit penalty for unused capacity is charged to the buyer only if the utilized portion of the reserved capacity falls below a certain threshold. The contract specifies both the penalty and the threshold. The assumption of exogenous wholesale price is a realistic one in high-tech industry where long before the negotiation on capacity reservation, the buyer (mostly an OEM) would have entered an agreement with the supplier to jointly develop the technology (known as the "design-win" phase). At this time, the supplier would assess the expected demand based on limited market information, and negotiate the (wholesale) pricing. A direct adjustment on the wholesale price is generally avoided due to buyer resistance. However, corrections through side payments or fees are usually possible and more practical. While under the endogenous wholesale price case, capacity reservation with options can always provide win-win solutions for both the supplier and the buyer, entering such an agreement is not necessarily a viable strategy for the trading parties under exogenous prices. In such an environment, the supplier would expand her capacity based on her assessment on the demand with or without reservation. Knowing this, the buyer would only make reservation when she fears that the supplier's capacity choice is not sufficient to fulfill her revenue potentials.

In this context, Erkoc and Wu (2002) investigate the supplier's and the buyer's incentives to (or not to) enter into a reservation contract in the first place. In their model, the buyer pays a reservation fee for each unit of reservation. The reservation fee is later deducted from the wholesale price for each utilized capacity unit. When the capacity is costly, the authors observe that capacity reservation contracts become more appealing to both players as uncertainty increases. They propose partial deduction and cost sharing contracts that can achieve coordination when the wholesale price is exogenous and examine the impact of the capacity cost structure (*i.e.*, linear vs. nonlinear) on the supplier's contract selection decision. The authors extend this model to a multiple competing buyer setting in Erkoc and Wu (2004). In this setting, in addition to interaction between the supplier and the buyers (vertical competition), the competition between buyers (horizontal competition) is also incorporated into the capacity reservation model. For a given capacity reservation fee, the buyers from independent markets competitively reserve the supplier's capacity. The model is novel in that when there are multiple buyers a buyer with upturn in demand can end up utilizing (exercising) capacity (option) reserved by another buyer with downturn in demand. Interestingly, the authors observe that in contrast to single-buyer models, in a multi-buyer model the supplier may have incentives to create capacity beyond her total reservation amount anticipating that the buyers will need more than their reservations. Typically, this happens when buyers have low profit margins and cannot afford to reserve sufficiently. In order to achieve coordination, *uplifting contracts* are proposed. Uplifting contracts stipulate that the buyers share not only the downside risk with the supplier but also their upside gains. To get a share of a buyer's upside gain, the supplier charges the buyer extra to exercise (utilize) any option (capacity unit) that was not previously purchased (reserved) by her but was rather reserved by another buyer with demand downturn.

In a setting somewhat similar to Erkoc and Wu (2002), Cheng et al. (2003) study option contracts assuming exogenous wholesale prices. Their model examines capacity procurement contracts that include a combination of firm and partial commitments. While the exogenous wholesale price is applied to firm commitments, partial commitments are modeled by call options in which both the prices and execution fees are determined by the supplier. Before demand is realized the buyer determines the committed order quantity and the number of options to be purchased. Similar to Erkoc and Wu (2002), Cheng et al. (2003) observe that options will be appealing for the buyer only if the reservation (option) fee is below a certain threshold value. The authors also develop a put option model in which by purchasing a pull option the buyer acquires the right to return a surplus unit to the supplier after demand is realized. They show that call and put options can be reduced to one another and cannot coordinate the channel unless the salvage revenues are equal for both players. To ensure channel coordination they propose a profit-sharing contract.

Burnetas and Ritchken (2003) compare call and put options in a supply chain in which the end-customer demand curve is downward sloping in sale price set by the buyer. From put-call parity, they also show that for a put option there exists a call option that leads to an identical outcome for the trading parties. This result implies that if the supplier offers both call and put options, the supplier's wholesale prices and the buyer's selling prices will be unchanged from what they would have been if only call options were offered. A combination of call and put options is employed by Erkoc and Wu (2002) to coordinate the channel when the uncertainty on demand is only partially resolved when the buyer must exercise her capacity reservations. This situation arises when, like the capacity lead-time, the production lead-time is also substantial, even though the product lifecycle may be relatively short, creating tremendous pressure for the buyer (e.g., an OEM manufacturer) to place her order early. Thus, it is often the case that some level of uncertainty still remains when a firm order is placed. The authors show that coordination can be achieved if a call option (reservation) is combined with a put option (buy-back) in a compound way. In this arrangement, exercising the call option yields a put option for the buyer in that she has the right to return excess orders to the supplier after all demand uncertainty is resolved.

The majority of the literature on capacity reservation contracts assumes perfect information in that all parties in the supply chain have full information regarding cost structures and demand distributions. Like Cachon and Lariviere (2001), an exception is the model studied by Ozer and Wei (2003), in which the authors assume that in a supplier leading supply chain, the buyer has private information on end-customer demand. In contrast to Cachon and Lariviere (2001), they treat the private information as a continuous random variable. Ozer and Wei (2003) conclude that although truthful information revelation is achievable under the capacity reservation contracts, they do not guarantee channel coordination. They extend the results of Erkoc and Wu (2002) by illustrating that the capacity reservation contract becomes more favorable when the capacity cost is high under information asymmetry as well. They show that coordination can be achieved under information asymmetry by combining an advance purchase contract with an appropriate buy-back agreement. The advance purchase contract provides the buyer with the option of making firm orders in return for a discounted rate.

OPERATIONAL MODELS: OPTIMIZATION AND DECISION SUPPORT

We now consider operational models in high-tech manufacturing; the literature focuses on optimization and decision support models, which we categorize according to the level of detail they capture and the length of the planning horizon they consider. We will focus our attention on semiconductor manufacturing because most existing work in this area concentrates on that industry.

Benavides, Duley, and Johnson (1999) consider several fab installations of different sizes that can achieve the same through-put levels measured in wafer starts per month. They compare multiple fabs of smaller capacity, a single fab of larger capacity and one or more fabs of expandable capacity, all of which provide the same total wafer output level. The tradeoffs are that smaller and multiple plants do not achieve economies of scale as good as a single larger plant of the same total capacity; however, building small plants reduces the risk of underutilization. However, long installation times for an individual plant (up to 18 months) may cause lost business opportunities. The decisions are when and how much capacity to install under demand uncertainty, which is modeled as a continuous time Brownian motion. The objective is to identity the configuration that maximizes the total expected profits, which is found by applying a stochastic control approach. This study shows that, in general, it is more costly to install a fab before demand is realized than to install it late. In that regard, large-capacity expandable fabs provide a good compromise by delaying the capacity increments and also achieving economies of scale better than low-capacity fabs.

Christie and Wu (2002) consider capacity planning across multiple fabs. Microelectronics technologies are aggregated and distinguished by their capacity consumption rate. Individual fab capacity is modeled as a single resource whose level is a random variable and measured in wafer output per period. This approach takes into account yield variability at an aggregate level. Uncertainty in demand for technologies and uncertainty in capacity levels are reconciled in the form of discrete scenarios in a multi-period, multistage, stochastic programming model. The objective is to minimize the expected mismatch between planned and actual capacity allocation as defined in the scenarios. The authors illustrate various methodologies for preparing demand scenarios as input to their optimization model.

Karabuk and Wu (2003) build on the previous work by expanding their model to include capacity expansion decisions and aiming to minimize the sum of capacity expansion and expected fab reconfiguration costs. An efficient solution approach, inspired by the decentralized decision-making environment is developed. The proposed approach captures the tradeoff between the extent of the information shared between participating decision makers and the quality of capacity planning decisions. It also quantifies the value of information sharing.

Another line of research tries to determine whether there is enough tool capacity to satisfy the anticipated demand through medium term planning horizon and if so, when and in what numbers are additional tools required? These are very important decisions because tools cost up to several million dollars each and have long delivery lead times (6 to 12 months).

Further group research in these tactical planning problems under three categories differentiated by the level of detail the models capture and their respective solution approaches: mathematical programming, queuing networks and simulation and other heuristics. Mathematical programming approaches distinguish processes only by production rate which are otherwise treated as generic. Yield and cycle time variability is modeled as a constant yield factor and as such congestion effects are not modeled. The planning period is up to two years which reflects the delivery lead time for newly acquired tools. Some studies also incorporate the assignment of operations to tool groups that are capable of doing multiple operations at different rates into their optimization models. These are also referred to as routing decisions.

The batch processing nature of semiconductor manufacturing lends itself well to modeling with queuing networks. This approach can compute the cycle time and throughput without running computationally expensive simulation models. Although this approach has been applied extensively in performance evaluation of manufacturing systems (and semiconductor manufacturing as well), we restrict our scope to studies that use this approach within a scheme to make tool capacity planning decisions. In such studies, a queueing network model guides the search for a minimum cost tool configuration that achieves a desired performance level for a wafer fab. One restrictive assumption of this approach in the literature observed so far is that the manufacturing flow, the assignment of operations to tools and the routing of batches is assumed to be predetermined.

Detailed simulation models are also required to test the solutions obtained by optimization methods or by heuristics. However, the substantial computational requirements of these models often prohibit their use for evaluating more than a handful of alternative configurations. These and some ad hoc approaches for capacity planning at tactical levels (such as spreadsheet applications) fall into the last category that we consider.

Mathematical Programming Models

There are a few researchers who study the problem with a deterministic model. Bermon and Hood (1999) develop a linear programming (LP) based decision support system for tactical planning at an IBM wafer fab which they refer to as the CAPS system. The decision problem is to find the most profitable product mix for a given tool set and an expected demand profile. Production volumes are allowed to vary within a lower and an upper bound which are parameters at the planners disposal. The modeling challenges dealt with are (1) how to represent operation assignment preferences among parallel tool sets with different operation times and (2) how to represent capacity availability as a function of product mix with an LP model, without using integer variables. Although demand is very volatile and changes frequently, it is left to the planners to account for with the help of a software tool that implements the LP model. An important simplifying assumption is that there is no inventory or back-order between two consecutive planning periods. Catay, Erenguc, and Vakharia (2003) develop a deterministic integer model that relaxes this assumption. Their model minimizes the total cost of inventory, tool acquisition and tool operating costs such that the expected demand is satisfied. They develop a Lagrangian relaxation based heuristic which is empirically shown to produce solutions with a small duality gap.

However, demand volatility and long manufacturing lead times in the semiconductor industry make it important to integrate demand uncertainty into any decision model. Hence, most studies use a stochastic programming approach where the demand distribution is included as a set of discrete scenarios in the model. Swaminathan (2000) studies tool procurement problem in a single period under the simplifying assumption that operation-to-tool assignments are already known (rather than optimized by the model). Although the underlying model is simplified, the study illustrates the benefits of incorporating demand scenarios versus optimizing with respect to expected demand only. Later, Swaminathan (2002) generalizes his earlier work by extending it to a multi-period setting and including operation-to-tool assignments as decision variables. Heuristic solution methods are developed and their performance is investigated. The study reinforces the benefits of the demand scenario approach and applies the solution methods to an industry data set.

Barahona et al. (2003) build on their earlier work on the CAPS system and extend it to a two-stage mixed integer stochastic programming model. Their model finds the quantity and timing of tool acquisitions over a discrete time planning horizon and under a budget constraint. The objective is to minimize the expected unsatisfied demand. Even with a simplifying assumption that results in a two-stage stochastic programming model (as opposed to a multi-stage model over the planning horizon), the very large size of the resulting model when applied to an instance observed at IBM (2,500 integer variables, 230,000 continuous variables and 140,000 constraints) prohibits an exact solution approach. Cuts that strengthen the LP relaxation are developed and applied to several heuristic solution methods that are based on relaxing integrality constraints on a rolling horizon basis. Hood, Bermon, and Barahona (2003) highlight the previous reference from a business perspective. They elaborate on the business implications of and requirements for implementing the stochastic programming model. Several approaches for generating discrete demand scenarios are discussed within the scope of relevant business processes.

Zhang et al. (2004) focus on the large number of demand scenarios that are required to capture uncertainty accurately, since this aspect is one of the main reasons for the large size of the resulting model. Their model is developed under the simplifying assumptions that there is no inventory and there are no backorders between periods and that the allocation of products to tools follow simple rules rather than explicitly being optimized by the model. The multi-variate nature of the demand distribution (for all products combined) complicates the model considerably. The authors approach the problem using a novel representation of the joint demand distribution with a collection of demand rays, rather than discrete demand points. The resulting formulation is reduced to a minimum cut network model which is solved relatively easily. Their approach, although it requires simplifying assumptions, finds solutions to problems of realistic size.

Although demand volatility is the main source of uncertainty in semiconductor tool capacity planning, yield variability, especially for new technologies, is an important consideration too. Karabuk and Wu (2002) incorporate both demand and yield uncertainty into a stochastic programming model via discrete scenarios. A scenario is defined as a combination of a vector of yield realization factor applied to expected wafer starts and a demand vector through the planning horizon. The resulting large-scale model is decomposed with respect to manufacturing and a marketing problem and a price-based coordinated solution is facilitated by mathematical decomposition. Two coordination schemes based on augmented Lagrangian are developed and their properties are studied.

Queuing Networks and Stochastic Processes

Chen et al. (1988) are among the first to use a queuing network model to study the performance of a wafer fab as an alternative to a detailed simulation study. The advantage of a queueing model is that it provides a closed form solution that relates control policies and fab configuration to system performance (e.g., manufacturing cycle time, throughput, work in progress.) Such an analytical model makes it easy to explore the performance consequences of different facility configurations and operating policies without heavy computational requirements. However, since the approach relies on certain assumptions and approximations, the resulting solutions must be considered in conjunction with a simulation study for verification before being put to use. The authors verify their model by comparing the results to actual observations and to detailed simulation output. Their results show that the accuracy of their model is within 10–30% with respect to cycle time and 5–10% with respect to throughput and that it only requires the mean processing time and variance information for each machine operation. This study can be considered as a proof of the concept on the application of queuing networks to semiconductor wafer manufacturing. Connors, Feigin, and Yao (1996) extend the previous work by incorporating rework and scrapping of wafer lots and by extending the queuing network model by incorporating tool characteristics: single wafer tools and batch processing tools whose efficiency depends on job size (hence importance of rework and scrap). They use this model to plan for tool capacities. Kumar and Kumar (2001) provide a tutorial overview of queuing networks in semiconductor wafer fab performance analysis.

Bard, Srinivasan, and Tirupati (1999) make use of the closed-form solution of a queuing model to develop a set of nonlinear equations that represent the cycle time as a function of the total number of tools in the system. A nonlinear mixed integer optimization model is developed, which aims to minimize cycle time with respect to a given demand profile and a given budget available for tool purchases. The authors develop several heuristic solution approaches and validate them with a set of industry data.

Kao and Chou (2000) develop a tool portfolio planning methodology that iterates between an aggregate capacity model and a queuing model and considers cycle time, tool investment costs and throughput. A tool portfolio refers to the quantity and type of tools in a fab. Later, Chou, et al. (2001) extend the previous work with the use of a simulation model in a hierarchical fashion. The methodology is then used to find a minimum cost tool portfolio subject to throughput and cycle time constraints. Chou and Everton (1996) develop a multi-criteria decision making framework that considers throughput, cycle time, tool capacity and investment cost simultaneously. A utility function is developed and used for evaluating alternative solutions with respect to the above criteria. The resulting decision model is applied to describe and quantify the advantages of capacity sharing between partner plants.

Hopp et al. (2002) also combine a queuing network model with simulation. The study attempts to find a minimum cost capacity configuration while the cycle times are within acceptable limits and throughput is as desired. The queuing model is used to guide the search of the optimization model to achieve this feat Batch processes, machine changeovers and re-entrant flows are incorporated. Iwata, Taji, and Tamura (2003) also use a queuing network model to derive cycle time and production costs as functions of the number of tools in tool groups and of the throughput for products. The model is used to determine the number and size of tool groups with the criteria being the tradeoff between manufacturing costs and the product throughput. They use the model to analyze the impact of fab scale on cycle time and production costs. They show that the degree of versatility of the tool groups determines the required fab capacity to achieve a given performance: as the versatility increases the required fab scale decreases. Agile mini-fabs could therefore be achieved by installing versatile tools.

Heuristics and Simulation Based Approaches
There are a large number of studies that develop rough-cut heuristic approaches for finding easy-to-understand practical capacity planning solutions. These are often reported by practitioners and published in conference proceedings. Some of these approaches are based on simulation. However, development complexity and high runtime, requirements preclude simulation from being a mainstream application tool. These kinds of studies are numerous and we therefore provide only a limited sample. Optimization does not take place but rather closed-form formulations are developed drawn from common best planning practices. As such, most of these models lend themselves to easy implementation via spreadsheet applications.

Chou and Everton (1996) apply simulation for capacity planning in a development wafer fab and look at the work in process (WIP), cycle time, bottleneck operations and the impact of randomness on the performance. Their study helps Fujitsu determine staffing levels and equipment acquisitions to support projected production levels. Witte (1996) describes a capacity planning tool developed at Harris Semiconductor. The model is based on company wide tool availability, process requirement and demand data which are used in simple equations to turn a first cut aggregate results into capacity requirements. A software tool which subsequently implemented the model provides a unifying framework across facilities to start their capacity analysis. Detailed analysis at individual facility level may follow. Mercier (1998) describes a spreadsheet model for identifying tool groups that are capacity bottlenecks.

Chen et al. (1999) develop a materials requirement planning model tailored to a twin wafer fab system. The model keeps track of WIP amounts, current capacity loads and also decides on wafer releases assuming unlimited capacity. Its main goal is to smooth loading across fabs and across planning periods. Kotcher and Chance (1999) describe an easy-to-apply methodology for assessing the sensitivity of tool capacity with respect to product mix. The methodology could be used in connection with a simulation model to identify bottleneck tool groups and make tool acquisition decisions. Occhino (2000) describes a spreadsheet application that implements equations that relate product mix to capacity consumption at a high volume DRAM manufacturer. Iwata and Wood (2002) develop simple equations that relate fab capacity to wafer cost. The resulting model decomposes facility costs into capacity-dependent and -independent components and takes into account processes variety, tool sizes and machine changeovers. The model is used to establish a relationship between tool and process characteristics and fab size. The conclusions are that increased process diversity and decreased tool versatility leads to larger scale fabs to achieve given delivery and cost performance.

Cakanyildirim and Roundy (2002a) evaluate tool and factory floor (shell space) capacity planning and expansion methodologies that are commonly practiced in industry and highlight their relative performance under different conditions. They make use of an optimal-seeking algorithm developed previously by the authors that finds the optimal tool purchase time and quantities such that the sum of expected lost demand and tool purchase costs are minimized. The results of this algorithm serves as a common comparison basis for the heuristic methods that are practiced. They show that the optimum-seeking approach reduces costs by 5–10%. In another experiment they study the effects of restricting capacity expansion decisions to the beginning of a planning period. The experiment concludes that the quality of decisions deteriorate only if the periods are longer than six months. In a related experiment they look into the impact of the frequency of planning and forecasting on the solution quality. The experiment reveals that the practitioners should revise forecasts and plan at least once every quarter.

Operational Planning and Scheduling

Although detailed planing and scheduling decisions usually take place after aggregate capacity planning decisions are made, they have a direct impact on efficient capacity utilization, and as such they indirectly influence capacity planning decisions. Due to the complexity of the wafer manufacturing process operational problems are equally challenging to solve. Uzsoy, Lee, and Martin-Vega (1992) list some characteristics of the wafer manufacturing process that complicate operational planning: the large number of operations a wafer has to go through, the re-entrant flow of wafer batches, the sequence-dependent machine setups, wide variety of machine characteristics—some process one wafer at a time and others do batch processing and time dependency of some operations. The fact that conflicting planning objectives such as reducing cycle time, increasing throughput and decreasing work in progress inventories are heavily interconnected makes operational planning even more challenging.

We will refer to major review papers published in this area as a starting point for the interested reader. Uzsoy, Lee, and Martin-Vega (1992, 1994) present an excellent review in two parts that covers existing literature and discusses research directions. In a more recent review, Fowler, Cochran, and Horng (1999) develop a classification scheme with respect to several dimensions including problem type, modeling technique and publication characteristics. They set up a public Internet site where researchers can search a bibliography database and obtain a listing of references and they pledge to keep it up to date. As of the publication date the authors report an availability of 559 articles.

CONCLUSIONS

While the literature on high-tech capacity planning and management is limited, related literature that addresses various aspects of this topic is vast. It is beyond the scope of this article to cover all relevant literature in this area; however, we believe that the article describes the general research landscape and points to several exciting new directions for capacity research.

REFERENCES

Abel, A.B., A.K. Dixit, J.C. Eberly, and R.S. Pindyck, "Options, the value of capital, and investment," *The Quarterly Journal of Economics*, 1996, Vol. 111(3), pp. 753–777.

Allprantis, C.D. and S.K. Ckakrabarti, *Games and Decision Making*. Oxford University Press, New York, 2000.

Amram, M. and N. Kulatilaka, *Real options: Managing strategic investment in an uncertain world*. Harvard Business School Press, Boston, 1999.

Angelus, A. and E.L. Porteus, "Simultaneous capacity and production management of short-life-cycle, produce-to-stock goods under stochastic demand," *Management Science*, Vol. 48, No. 3, 2002, pp. 399–413.

Angelus, A., E.L. Porteus, and S.C. Wood, "Optimal sizing and timing of modular capacity expansion," Working paper, Stanford University, Palo Alto, CA, 2000.

Armony, M. and E. Plambeck, "The impact of duplicate orders on demand estimation and capacity investment," Working paper, Stanford University, Palo Alto, CA, 2003.

Atamturk, A. and D.S. Hochbaum, "Capacity acquisition, subcontracting, and lot sizing," *Management Science*, Vol. 47, No. 8, 2001, pp. 1081–1100.

Barahona, F., S. Bermon, O. Gunluk, and S. Hood, *Robust capacity planning in semiconductor manufacturing*, Tech. Report RC22196, IBM, T.J. Watson Research Center, Yorktown Heights, NY, 2003.

Bard, J.F., K. Srinivasan, and D. Tirupati, "An optimization approach to capacity expansion in semiconductor manufacturing facilities," *International Journal of Production Research*, Vol. 37, No. 15, 1999, pp. 3359–3382.

Barnes-Schuster, D., Y. Bassok, and R. Anupindi, "Coordination and flexibility in supply contracts with options," *M&SOM*, Vol. 4, No. 3, 2002, pp. 171–207.

Benavides, D.L., J.R. Duley, and B.E. Johnson, "As good as it gets: Optimal fab design and deployment," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 3, 1999, pp. 281–287.

Bermon, S. and S.J. Hood, "Capacity optimization planning system (caps)," *Interfaces*, Vol. 29, No. 5, 1999, pp. 31–51.

Bernstein, F. and G.A. DeCroix, "Decentralized pricing and capacity decisions in a multitier system with modular assembly," *Management Science*, Vol. 50, No. 9, 2004, pp. 1293–1308.

Birge, J.R., "Option methods for incorporating risk into linear capacity planning models," *Manufacturing & Service Operations Management*, Vol. 2, No. 1, 2000, pp. 19–31.

Bish, E. and Q. Wang, "Optimal investment strategies for flexible resources, considering pricing and correlated demands," *Operations Research*, Vol. 52, No. 6, 2004, pp. 954–964.

Black, F. and M. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, Vol. 81, No. 3, 1999, pp. 637–654.

Bollen, N.P., "Real options and product life cycles," *Management Science*, Vol. 45, No. 5, 1999, p. 670.

Bonser, J. and S.D. Wu, "Procurement planning to maintain both short-term adaptiveness and long-term perspective," *Management Science*, Vol. 47, 2001, pp. 769–786.

Bowman, E.H. and G.T. Moskowitz, "Real options analysis and strategic decision making," *Organisation Science*, Vol. 12, No. 6, 2001, pp. 772–777.

Bradley, J.R. and B.C. Arntzen, "The simultaneous planning of production, capacity and inventory in seasonal demand environments," *Operations Research*, Vol. 47, No. 6, 1997, pp. 795–806.

Broadle, M. and J.B. Detemple, "Option pricing: Valuation models and applications," *Management Science*, Vol. 50, No. 9, 2004, 1145.

Brown, A. and H. Lee, Optimal pay-to-delay capacity reservation with application to the semiconductor industry, Working paper. Vanderbilt University, Nashville, TN, 1998.

Burnetas, A. and P. Ritchken, Option pricing with downward sloping demand curves: The case of supply chain options, To appear in *Management Science*, Vol. 51, No. 4, 2005.

Cachon, G.P., Supply chain coordination with contracts. In Kok, A.G. and S.C. Graves (eds.), *Supply Chain Management: Design, Coordination and Operation, Handbooks in a chain, Operations Research and Management Science*, Elsevier, Amsterdam, The Netherlands, Vol. 11, 2003, 229.

Cachon, G.P. and M.A. Lariviere, "Contracting to assure supply: How to share demand forecasts in a supply chain," *Management Science*, Vol. 47, 2001, pp. 629–646.

Cakanyildirim, M. and R.O. Roundy, "Optimal capaciti, expansion and contraction under demand uncertain," Working paper, University of Texas, Dallas, TX, 2002.

Cakanyildirim, M. and R.O. Roundy, "Evaluation of capacity planning practices for the semiconductor industry," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 15, No. 3, 2002, pp. 331–340.

Catay, B., S.S. Erenguc, and A.J. Vakharia, "Tool capacity planning in semiconductor manufacturing," *Computers and Operations Research*, Vol. 30, No. 9, 2003, pp. 1349–1366.

Chen, H., J.M. Harrison, A. Mandelbaum, A. Van Ackere, and L.M. Vein, *Emprical evaluation of a queuing network model for semiconductor wafer fabrication*, *Operations Research*, Vol. 36, No. 2, 1988, pp. 202–215.

Chen, J.C., Y.C. Fan, J.Y. Wang, T.K. Lin, S.H. Leea, S.C. Wu, and Y.J. Lan, "Capacity planning for a twin fab, IEEE International Symposium on Semiconductor Manufacturing Conference Proceedings, 11–13 October 1999, Santa Clara, CA, pp. 317–320.

Cheng, F., M. Ettl, G.Y. Lin, M. Schwarz, and D.D. Yao, "Flexible supply contracts via options," Working paper, IBM T.J. Watson Research Center, Yorktown Heights, NY, 2003.

Chou, W. and J. Everton, Capacity planning for development wafer fab expansion, Advanced Semiconductor Manufacturing Conference and Workshop, ASMC 96 Proceedings, IEEE/SEMI, Cambridge, MA, 1996, pp. 17–22.

Chou, Y.C., C.S. Wu, C.E. Kao, and S.H. Hsieh, "Integration of capacity planning techniques for tool portfolio planning in semiconductor manufacturing," *International Journal of Industrial Engineering—Theory Applications and Practice*, Vol. 8, No. 4, 2001, pp. 279–288.

Christie, R.M.E. and S.D. Wu, "Semiconductor capacity planning: Stochastic modeling and computational studies," *IIE Transactions*, Vol. 34, No. 2, 2002, pp. 131–144.

Connors, D.P., G.E. Feigin, and D.D. Yao, "A queueing network model for semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 9, No. 3, 1996, pp. 412–427.

Dangl, T. "Irreversible Investment and capacity choice under uncertain demand," *European Journal of Operational Research*, Vol. 117, 1999, pp. 415–428.

Dixit, A. "Investment and employment dynamics in the short run and the long run," *Oxford Economic Papers*, Vol. 49, 1997, pp. 1–20.

Dixit, A. and R.S. Pindyck, *Investment under Uncertainty*. Princeton University Press, Princeton, NJ, 1994.

Dixit, A.K., "Investment and scale economies," *Journal of Economic, Dynamics and Control*, Vol. 19, Nos. 1–2, 1995, pp. 327–350.

Eberly, J.C. and J.A. van. Mieghem, "Multi-factor dynamic investment under uncertainty," *Journal of Economic Theory*, Vol. 75, No. 2, pp. 345–387.

Erkoc, M. and S.D. Wu, *Managing high-tech capacity expansion via reservation contracts*, To appear in Production and Operations Management, 2002.

Erkoc, M. and S.D. Wu, "Capacity reservation across multiple buyers," Working paper, Lehigh University, Bethlehem, PA, 2004.

Feinstein, S.P. "A better understanding of why npv undervalues managerial flexibility," *The Engineering Economist*, Vol. 47, No. 4, 2002, pp. 418.

Ferguson, M.E., G.A. DeCroix, and P.H. Zipkin, "When to commit in a multi-echelon supply chain with pertial information updating," Working paper, Georgia Tech, Atlanta, GA, 2002.

Fowler, J.W., U.K. Cochran, and S.M. Horng, "A group-technology-coded literature review of semiconductor manufacturing operations publications: The masmlab bibliography web site," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 2, No. 2, 1999, pp. 259–263.

Fundenberg, D. and J. Tirole, *Game Theory*. The MIT Press, Cambridge, MA, 1991.

Harrison, J.M. and J.A. van Mieghem, "Multi-resource investment strategies: Operational hedging under demand uncertainty," *European Journal of Operational Research*, Vol. 113, No. 1, 1999, pp. 17–29.

Herath, H.S.B. and C.S. Park, "Multi-stace captital investment opportunities as compound real options," *The Engineering Economist*, Vol. 47, No. 1, 2002, pp. 1–27.

Hicks, D.A., "Evolving complexity and cost dynamics in the semiconductor industry," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 9, No. 3, 1996, pp. 294–302.

Ho, T.H., S. Savin, and C. Terwiesh, "Managing demand and sales dynamics in new product diffusion under supply constraint," *Management Science*, Vol. 48, No. 2, 2002, pp. 187–206.

Hood, S.J., S. Bermon, and F. Barahona, "Capacity planning under demand uncertainty for semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 16, No. 2, 2003, pp. 273–280.

Hopp, W.J., M.L. Spearman, S. Chayet, K.L. Donohue, and E.S. Gel, "Using an optimized queueing network model to support wafer fab design," *IIE Transactions*, Vol. 34, No. 2, 2002, pp. 119–131.

Huh, W.T. and R.O. Roundy, "A continuous-time strategic capacity planning model," Working paper, Cornell University, Ithaca, NY, 2002.

Iwata, Y., K. Taji, and H. Tamura, "Multi-objective capacity planning for semiconductor manufacturing," *Production Planning & Control*, Vol. 14, No. 3, 2003, pp. 244–254.

Iwata, Y. and S.C. Wood, "Simple cost models of high-process-mix wafer fabs at different capacities," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 15, No. 2, 2002, pp. 267–273.

Jin, M. and S.D. Wu, "Modeling capacity reservation in high-tech manufacturing," Working paper, Lehigh University, Bethlehem, PA, 2001.

Johnson, B. and C. Billington, Creating and leveraging options in the high technology supply chain. In Harrison, T. P., Lee, H. L., and Neale, J. J. (eds.), *The Practice of Supply Chain Management: Where Theory and Application Converge, International Series in Operations Research & Management Science*, Kluwer Academic Publishers, 2003.

Kao, C.E. and Y.C. Chou, A tool portfolio planning methodology for semiconductor wafer fabs. Semiconductor Manufacturing Technology Workshop, 14–15 June 2000, Hsinchu, Taiwan.

Karabuk, S. and S.D. Wu, Coordinating strategic capacity planning in the semiconductor industry, *Operations Research*, Vol. 51, No. 6, 2003, pp. 839–849.

Karabuk, S. and S.D. Wu, "Decentralizing semiconductor capacity planning via internal market coordination," Vol. 34, No. 9, 2002, pp. 743–759.

Karabuk, S. and S.D. Wu, Incentive schemes for semiconductor capapcity allocation: A game theoretic analysis, To appear in *Production and Operations Management*, 2003.

Kleindorfer, P.R. and D.J. Wu, "Integrating long and short term contracting via business-to-business exchanges for capital-intensive industries," *Management Science*, Vol. 49, No. 11, 2003, pp. 1597–1615.

Kotcher, R. and F. Chance, "Capacity planning in the face of product-mix uncertainty," *IEEE International Symposium on Semiconductor Manufacturing Conference Proceedings*, 11–13 October 1999, Santa Clara, CA, 1999, pp. 73–76.

Kouvelis, P. and J.M. Milner, "Supply chain capacity and outsourcing decisions: The dynamic interplay of demand and supply uncertainty," *IIE Transactions*, Vol. 34, No. 8, 2002, pp. 717–728.

Kulatilaka, N. and E.C. Perotti, "Strategic growth options," *Management Science*, Vol. 44, No. 8, 1998, pp. 1021–1031.

Kumar, S. and P.R. Kumar, "Queueing network models in the design and analysis of semiconductor wafer fabs," *IEEE Transactions on Robotics and Automation*, Vol. 17, No. 5, 2001, pp. 548–561.

Majd, S. and S. Myers, "Abandonment value and project life," *Advances in Futures and Options Research*, Vol. 4, 1990, pp. 1–21.

Mallik, S. and P.T. Harker, "Coordinating supply chains with competition: Capacity allocation in semiconductor manufacturing," *European Journal of Operational Research*, Vol. 159, 2004, pp. 330–347.

Manne, A.S., "Capacity expansion and probabilistic growth," *Econometrica*, Vol. 29, No. 4, 1961, pp. 632–649.

Mason, S.J., M.H. Cole, B.T. Ulrey, and L. Yan, "Improving electronics manufacturing supply chain aqility through outsourcing," *J. of Physical Dist. and Log. Man.*, Vol. 32, No. 7, 2002, pp. 610–620.

McDonald, R. and D. Siegel, "Investment and the valuation of firms when there is an option to shut down," *International Econom. Review*, Vol. 26, 1985, pp. 331–349.

Mercier, J.R., Dunamic capacity modeling, IEEE/SEMI Advanced Semiconductor Manufacturing Conference and Workshop, 23–25 September Boston, MA, 1998, pp. 148–150.

Miller, L.T. and C.S. Park, "Decision making under uncertainty—Real options to the rescue?" *The Engineering Economist*, Vol. 47, No. 2, 2002, pp. 105–150.

Murto, P., E. Nasakkala, and J. Keppo, "Timing of investments in oligopoly under uncertainty: A framework for numerical analysis," *European Journal of Operational Research*, Vol. 157, No. 2, 2004, pp. 486–500.

Narongwanich, W., I. Duenyas, and J.R. Birge, "Optimal protfolio of reconfigurable and dedicated capacity under uncertainty," Working paper, University of Michigan, Ann Arbor, MI, 2002.

Occhino, T.J., "Capacity planning model: The important inputs formulas and benefits," *Advanced Semiconductor Manufacturing Conference and Workshop*, IEEE/SEMI, 2000, pp. 455–458.

Ozer, O. and W. Wei, "Strategic commitment for optimal capacity decision under asymmetric forecast information," Working paper, Stanford University, Palo Alto, CA, 2003.

Pindyck, R.S., "Irreversible investment, capacity choice, and the value of the firm," *The American Economic Review*, Vol. 78, No. 5, 1988, pp. 969–985.

Pindyck, R.S., "A note on competitive investment under uncertainty," *The American Economic Review*, Vol. 83, No. 1, 1993, pp. 273–277.

Ridsdale, D., "FSA reports good wafer demand for tables companies," www.semiconductorfabtech.com

Ryan, S.M., "Capacity expansion for random exponential demand growth with lead times," *Management Science*, Vol. 50, No. 6, 2004, pp. 740–748.

Serel, D.A., M. Dada, and H. Moskowitz, "Sourcing decisions with capacity reservation contracts," *European Journal of Operactional Research*, 131, 2001, pp. 635–648.

Spialer, S. *Capacity Reservation for Capital-Intensive Technologies: An Options Approach*. Springer-Verlag, Berlin, 2003.

Spinler, S., A. Huchzermeier, and P.R. Kleindorfer, "The valuation of options on capacity," Working paper, the Wharton School, University of Pennsylvania, Philadelphia, PA, 2002.

Swaminathan, J.M., "Tool capacity planning for semiconductor fabrication facilities under demand uncertainty," *European Journal of Operational Research*, Vol. 120, No. 3, 2000, pp. 545–558.

Swaminathan, J.M., "Tool procurement planning for wafer fabrication facilities: A scenario-based approach," *IIE Transactions*, Vol. 34, No. 2, 2002, pp. 145–155.

Tan, B., "Subcontracting with availability auarantees: Production control and capacity decisions," *IIE Transactions*, Vol. 36, No. 8, 2004, pp. 711–724.

Taylor, T.A. and Plambeck, E.L. "Supply chain relationships and contracts: The impact of repeated interactions on capacity investment and procurement," Working paper, Stanford University, Stanford, CA, 2003.

Tomlin, B., "Capacity investments in supply chains: Sharing the gain rather than sharing the pain," *M&SOM*, Vol. 5, No. 4, 2003, pp. 317–333.

Trigeorgis, L., *Real Options: Managerial Flexibility and Strategies in Resource Allocation*, MIT Press, Cambridge, 1996.

Uzsoy, R., C.Y. Lee, and L.A. Martin-Vega, "A review of production planning, and scheduling models in the semiconductor industry part i: System characteristics, performance evaluation and production planning," *IIE Transactions*, Vol. 24, No. 4. 1992, pp. 47–60.

Uzsoy, R., C.-Y. Lee, and L.A. Martin-Vega, "A review of production planning and scheduling models in the semiconductor industry part ii: Shop-floor control," *IIE Transactions*, Vol. 26, No. 5, 1994, pp. 44–55.

van Mieghem, J.A., "Investment strategies for flexible resources," *Management Science*, Vol. 44, No. 8, 1998, pp. 1071–1078.

van Mieghem, J.A., "Coordinating investment, production, and subcontracting," *Management Science*, Vol. 45, No. 7, 1999, pp. 954–971.

van Mieghem, J.A., "Capacity management, investment, and hedging: Review and recent developments," *Manufacturing and Service Operations Management*, Vol. 5, No. 4, 2003, pp. 269–302.

van Mieghem, J.A. and M. Dada, "Price versus production postponement: Capacity and competition," *Management Science*, Vol. 45, No. 12, 1999, pp. 1631–1649.

van Mieghem, J.A. and N. Rudi, Newsvendor networks: Inventory management and capacity investment with discretionary activities, *M&SOM*, Vol. 4, No. 4, 2002, pp. 313–335.

Witte, J.D., Using static capacity modeling techniques in semiconductor inanufacturing, Advanced Semiconductor Manufacturing Conference and Workshop, 1996. ASMC 96 Proceedings, IEEE/SEMI 1996, Cambridge, MA, pp. 31–35.

Wu, D.J. and P.R. Kleindorfer, Competitive options, supply contracting and b2b exchanges, To appear in *Management Science*, Vol. 51, No. 3, 2005.

Wu, D.J., P.R. Kleindorfer, and J.E. Zhang, "Optimal bidding and contracting strategies for capital intensive goods," *European Journal of Operational Research*, Vol. 137, 2002, pp. 657–676.

Zhang, F., R. Roundy, M. Cakanyildirim, and W.T. Huh, "Optimal capacity expansion for multi-product, multi-machine manufacturing systems with stochastic demand," *IIE Transactions*, Vol. 36, No. 1, 2004, pp. 23–37.

BIOGRAPHICAL SKETCHES

S. David Wu is currently Iacocca professor and Dean of the P.C. Rossin College of Engineering at Lehigh University. Professor Wu's research interests are in the area of production and logistics systems, more specifically in game theoretic decision-making, supply chain coordination and discrete optimization. He has published widely in these areas, including journals such as *Operations Research, Management Science, IEEE Transactions*, *IIE Transactions* and others. Professor Wu is co-founder of the Center for Value Chain Research, a research center focusing on planning and development strategies in high-value manufacturing and pharmaceutical industries. His research has been supported by NSF, DOD and various industrial sources such as Lucent Technologies, Agere Systems, GM and HP. Dr. Wu received his M.S. and Ph.D. degrees from the Pennsylvania State University. In 1995 to 1996, he was a visiting professor at University of Pennsylvania. He currently serves as department editor of *IIE Transactions*, and is on the editorial boards of *M&SOM* and *JMS*.

Murat Erkoc is an assistant professor of industrial engineering at University of Miami. He has B.S. degree from Istanbul Technical University, Turkey, and an M.S. degree from Bogazici University, Istanbul, Turkey, both in industrial engineering. He received his Ph.D. in industrial engineering from Lehigh University in January 2003. His current research interests include game theory; distributed decision making; and stochastic processes with applications to capacity investment, logistics, inventory control and supply chain management problems. He is a member of IIE, INFORMS, and POMS.

Suleyman Karabuk is an assistant professor of Industrial Engineering at the University of Oklahoma. Suleyman received his Ph.D. from Lehigh University in 2001 and he worked for Milliken & Company as an Operations Research Analyst until he joined the University of Oklahoma in 2003. He works on supply chain logistics problems and he applies stochastic programming, simulation and game theoretic approaches. His work is published in IIE transactions, Operations Research and Production and Operations Management Journals. In industry, Suleyman worked on logistics problems at various levels of the textile supply chain. At Lehigh University, he participated in industry linked research projects related to planning in semiconductor industry. He worked as a software developer in an international company for two years prior to pursuing a Ph.D. degree and involved in development and implementation of MRP-II systems. He received his M.S. degree in 1994 from Bilkent University, where he did research on scheduling and flexible manufacturing systems.