



## Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Capacity Planning in the Semiconductor Industry: Dual-Mode Procurement with Options

Chen Peng, Feryal Erhun, Erik F. Hertzler, Karl G. Kempf,

To cite this article:

Chen Peng, Feryal Erhun, Erik F. Hertzler, Karl G. Kempf, (2012) Capacity Planning in the Semiconductor Industry: Dual-Mode Procurement with Options. *Manufacturing & Service Operations Management* 14(2):170-185. <http://dx.doi.org/10.1287/msom.1110.0361>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Capacity Planning in the Semiconductor Industry: Dual-Mode Procurement with Options

Chen Peng, Feryal Erhun

Management Science and Engineering, Stanford University, Stanford, California 94305  
{chenpeng@alumni.stanford.edu, ferhun@stanford.edu}

Erik F. Hertzler, Karl G. Kempf

Intel Corporation, Chandler, Arizona 85226 {erik.f.hertzler@intel.com, karl.g.kempf@intel.com}

To help a firm reduce inefficiencies associated with equipment capacity planning, we propose a *dual-mode equipment procurement* (DMEP) framework. DMEP combines dual-source (i.e., a less-expensive-but-slower base mode and a faster-but-more-expensive flexible mode) procurement with option contracts in three layers: a *contract negotiation* layer, where the firm chooses the best combination of lead time and price for each mode from the supply contract menu; a *capacity reservation* layer, where the firm reserves total equipment procurement quantities from the two supply modes before the planning horizon starts; and an *execution* layer, where the firm orders equipment from the two supply modes based on the updated demand information. We first investigate the execution layer as a dynamic dual-source capacity expansion problem with demand backlogging and demonstrate that the optimal policy lacks structure even under the simplest setting. Thus, we propose a heuristic solution for the execution-layer problem, which also serves as a building block for the other two layers. Through numerical analysis, we quantify the value of the added flexibility of DMEP for the firm. The DMEP framework has been implemented at Intel Corporation and has resulted in savings of tens of millions of dollars for one process technology.

**Key words:** capital-intensive industries; equipment procurement; capacity expansion; forecast revision; dual-mode procurement; dual sourcing; option contracts; heuristic; OM practice

**History:** Received: September 8, 2010; accepted: September 29, 2011. Published online in *Articles in Advance* February 2, 2012.

## 1. Introduction

Capacity planning is a complex balancing act. In the semiconductor industry, one of the most capital-intensive industries in the world, a single piece of semiconductor manufacturing equipment commonly costs tens of millions of dollars. Over the last five years, as a key player of this industry, Intel Corporation has spent on average \$5.3 billion annually on capital additions to its property, plant, and equipment (Intel Corporation 2009). Compounding the problem of high costs are long supplier lead times and a volatile consumer market. The order-to-production cycle for semiconductor manufacturing equipment can take up to 16 months, which exacerbates the difficulty of forecasting demand accurately. In this practice-based paper, we address the challenges of capacity planning in the semiconductor industry and describe Intel's efforts to address these challenges by continuously improving the set of rules for its engagement with its suppliers.

The marginal cost of unmet demand is significantly higher than the marginal cost of idle capacity in the semiconductor industry (Fleckenstein 2004). Thus, despite the astounding costs, semiconductor firms

often err on the side of having excess capacity and keep some equipment idle to ensure customer goodwill and loyalty. Traditionally, an easy way for firms with strong bargaining power (including Intel) to partially mitigate this capacity risk and avoid purchasing too much excess equipment has been through soft orders. To secure procurement contracts, equipment suppliers have allowed firms to overorder capacity and to cancel some of the excess orders without paying severe penalties as higher-confidence demand forecasts become available; this process is similar to the "phantom ordering" common in personal computer and electronics industry (Cohen et al. 2003). As a result, suppliers have carried a significant demand risk. However, such a relationship is no longer sustainable. To keep pace with the technology requirements of maintaining Moore's Law (Moore 1965), the cost of capital equipment in the semiconductor industry has been steadily rising, with no end in sight. Building a fabrication plant now costs in excess of \$5 billion, up from \$6 million in 1970 and around \$2 billion in 2001 (Kanellos 2003). Hence, soft orders are costly to suppliers. In addition, the recent trend of supply consolidation (Armbrust 2009) has further

increased suppliers' bargaining power. As a result, suppliers are more reluctant to carry demand risk through soft orders. Thus, semiconductor manufacturing firms are challenged to derive innovative ways to order the right amount of equipment at the right time and price, in advance of demand realization, to minimize unnecessary equipment purchases while maintaining a high level of service. Further, this goal should be achieved without pushing all the risk onto suppliers.

The industry leader Intel and its suppliers are always working to refine the set of rules for their engagement. At Intel, the continuous improvement of capital acquisition and installation processes is an ongoing corporate priority. Being able to respond rapidly to changing customer demand for products while minimizing unnecessary capital expenditures is vital to both the semiconductor market and the firm itself. From Intel's perspective, demand uncertainties caused by extremely long procurement lead times add too much idle capacity to the system and jeopardize the agility of the supply chain. Intel prefers to have tighter control of its capital supply chain by shortening equipment procurement lead times and improving the accuracy of demand forecasts used in capacity planning. In return for this flexibility, Intel is willing to take on some risk from its suppliers and is considering risk-sharing mechanisms.

We develop a framework that combines dual-source procurement with option contracts to effectively address the real-world dynamics and constraints of the equipment procurement problem discussed above. With this framework, which we call a *dual-mode equipment procurement* (DMEP) framework, Intel procures equipment from its supplier using two supply modes with complementary lead times and prices: a *base* mode (the regular procurement mode), which is less expensive but has a longer procurement lead time  $L_b$ , and a *flexible* mode, which is more expensive but has a shorter procurement lead time  $L_f$ . (We refer to the modes as "base" and "flexible" to be consistent with Intel's parlance. In the literature, these modes are often referred to as "slow" and "fast," respectively.) The flexible mode allows Intel to learn more about demand before committing to purchase capacity. In return, to share the risk with its suppliers, Intel makes an up-front payment to secure certain base and flexible capacity levels (i.e., it buys capacity options) ahead of time and exercises these options over the planning horizon. In addition, Intel faces prohibitive costs to cancel an order once it has been placed. Thus, the framework supports Intel's goal of continuously improving its equipment procurement strategies to satisfy customer demand without over-purchasing costly capacity and still committing to

fairness to Intel's suppliers through a risk-sharing mechanism.

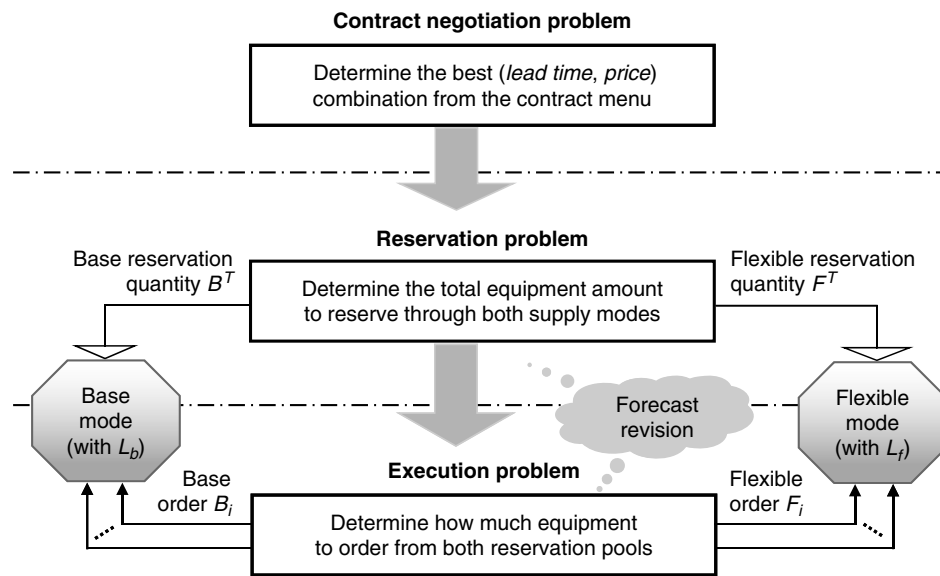
The rest of this paper is organized as follows: In §2, we introduce the DMEP framework. Although designed for Intel's continuous improvement efforts on equipment procurement, this framework is versatile enough to be adapted by firms in other capital-intensive industries such as electronic, automotive, and pharmaceutical industries. Section 3 provides a brief literature review for each layer of the framework. In §4, we study the execution problem (i.e., a dynamic dual-source capacity expansion problem with backlogs) and show that the optimal ordering policy lacks structure even in the simplest setting. Hence, we construct a heuristic approach for this problem in §5. The execution heuristic serves as a building block for the tactical and the strategic layers. In §6, we present a numerical analysis to quantify the value of the added flexibility that DMEP provides (§6.2) and to study the decisions of a risk-averse firm (§6.3). Section 7 concludes the paper. All proofs, as well as additional material on §4, are presented in the online supplement (available at <http://msom.journal.informs.org/>).

## 2. The Dual-Mode Equipment Procurement Framework

DMEP captures different phases of the relationship between Intel and its suppliers and is composed of three stages (Figure 1). During the *contract negotiation* stage, several years before the adoption of a new process and the procurement of the necessary equipment, Intel and its supplier agree on the parameters of the supply modes. That is, they negotiate the lead times and prices (a reservation price and an execution price) of the base and flexible modes. During the *reservation* stage, several quarters before the planning horizon starts, Intel reserves the total equipment procurement quantities  $B^T$  and  $F^T$  from the base and flexible supply modes at respective unit reservation prices. The reservation payment provides capital to the supplier to prepare its own production capacity and to help the supplier give Intel the guaranteed flexibility of ordering equipment when needed. That is, the reservation stage allows Intel to share risk with its supplier. During the *execution* stage, in each period  $n$ , Intel orders specific amounts of equipment from the two supply modes at respective unit execution prices, given the latest demand forecast as well as the realized demand information from the market.

DMEP is a multistage decision hierarchy that guides Intel during different phases of equipment procurement. It provides decision support in approaching questions such as: How can Intel quickly evaluate different flexible options during contract

Figure 1 Dual-Mode Equipment Procurement Framework



negotiations? Under what circumstances does the flexible mode create value for the firm? How much of the total capacity should be reserved through the flexible mode? When should this capacity be exercised? With its three stages, DMEP addresses such questions by evaluating various trade-offs inherent in this setting. First, the contract negotiation stage balances accurate demand information with sourcing from the expensive mode. Procuring from the flexible mode allows Intel to react to the market condition later with better demand information; however, the flexible mode is costlier than the base mode. Furthermore, Intel pays more for shorter procurement lead times, and choosing the right lead time–cost combination is crucial. Second, the reservation stage balances flexibility (by reserving adequate capacity) with high reservation costs. By having a large reservation pool up front, Intel can guarantee the punctual delivery of its future orders; however, the reservation payment implies a high opportunity cost if the equipment is not ordered. Finally, the execution stage balances equipment holding costs with backlogging costs. Early equipment procurement reduces the potential risk of having unsatisfied demand; however, it also leads to a higher equipment holding cost and a high opportunity cost if the equipment remains idle.

### 3. Literature Review

The contract negotiation stage of DMEP builds on the literature of supply contracts. We refer the reader to Cachon (2003) for a review of that literature. The contract that we study is closely related to the one analyzed in Yazlali and Erhun (2010). The authors investigate a dual-supply contract with minimum order quantity and maximum capacity restrictions.

This contract provides supply chain partners with an enhanced mechanism to share and manage demand uncertainty in the context of inventory management. We adopt a similar structure for capacity procurement. The negotiation of price–lead time combinations is also related to the literature on pricing and lead time quotation. This literature often assumes that the buyer's order quantity is a function of the price and lead time quoted by the supplier and thus focuses on the supplier's optimal price–lead time decision (Palaka et al. 1998, Liu et al. 2007). We, in contrast, take the buyer's perspective and give him a decision-support tool that he can use while negotiating the price and lead time with his supplier.

The reservation stage addresses a capacity expansion problem, which has been extensively studied under single-sourcing; see Van Mieghem (2003) for a review of this literature. Also see Wu et al. (2005) for a review of the literature on capacity planning in the high-tech industry. DMEP is also closely related to the literature on procurement and option contracts (e.g., Bassok and Anupindi 1997, Vaidyanathan et al. 2005). In particular, the reservation stage of DMEP builds on the paper by Vaidyanathan et al. (2005), which focuses on capacity contracts at Intel and discusses capacity options to better enable factory ramps. In this paper, we extend and merge these two streams by studying capacity expansion and option contracts in a *dual-mode* setting.

The execution stage of DMEP is closely related to the dual-sourcing problem that has been studied extensively in the context of inventory (e.g., Daniel 1963, Fukuda 1964, Feng et al. 2006, Yazlali and Erhun 2009). Dual-source inventory management has also been commonly adopted as an operational risk-hedging strategy by firms in different industries for



many years, e.g., Mattel (Johnson 2005) and Hewlett-Packard (Billington et al. 2002). Despite the existence of literature on the dual-sourcing inventory problem, research concerning the dual-sourcing *capacity procurement* problem is scarce, with the exception of the recent paper by Chao et al. (2009), in which the authors investigate a dual-source capacity procurement problem with lost sales. They establish that the optimal policy for the fast source is base stock; when the capacity obsolescence rate is deterministic, the optimal policy for the slow source is also base stock. In this paper, we study the dual-sourcing capacity procurement problem *with backlogging* and show that the optimal ordering policy lacks structure even in the simplest setting. Thus, we construct a heuristic approach for this problem.

Our execution heuristic relies on rolling-horizon decision making with information updating (e.g., Baker 1977, Yildirim et al. 2005, Lian et al. 2010). In particular, Yildirim et al. (2005) study a stochastic dual-source production problem where the sources have the same delivery lead time but different costs and capacity limits. The authors, similar to most of the existing literature in this area, consider an infinite-period problem with a finite-period rolling horizon. During each period, the information update only refers to the demand realization for the current period as well as the forecast for the additional period that has just entered the rolling window; the forecast information for the remaining periods is not updated. To incorporate a forecast updating mechanism (e.g., Graves et al. 1986, Fisher and Raman 1996, Eppen and Iyer 1997, Donohue 2000, Özer and Wei 2006), Lian et al. (2010) investigate a rolling-horizon inventory replenishment model where the buyer can update demand information and modify the previously committed order quantities. The authors limit their analytical discussion to a two-period model. In this paper, we use a rolling-horizon algorithm with a systematic range forecast updating mechanism in a finite-horizon setting. Unlike Lian et al. (2010), we consider dual sourcing. Unlike Yildirim et al. (2005), our sources have different, nonconsecutive lead times in addition to different costs and capacity limits.

#### 4. Analysis of the Execution Layer Problem: Dynamic Dual-Source Capacity Expansion Problem with Backlogs

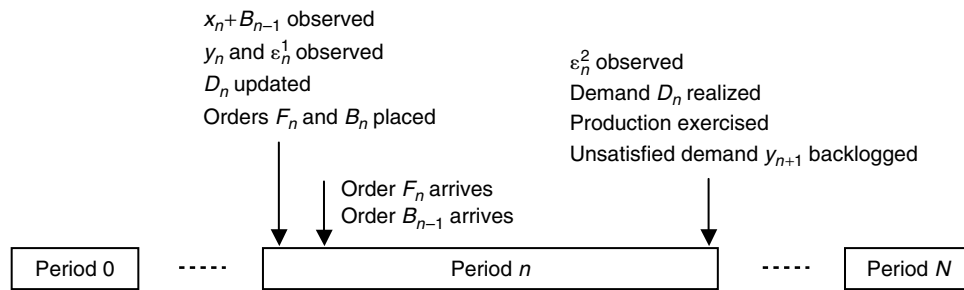
Intuitively, the equipment procurement decision that firms face during the execution stage fits in the scope of a dual-source capacity expansion problem and can be formulated as a dynamic programming model. In this section, we investigate such a formulation and demonstrate that the optimal ordering policy lacks

structure even in the simplest setting. This result will motivate our reliance on a heuristic solution in §5.

We consider a finite-horizon, periodic-review, dual-source capacity expansion model with demand backlogging and forecast updates. A firm needs an efficient equipment-procurement strategy to optimally match its production capacity with a growing but fluctuating demand over time. During each period  $n$ ,  $n = 0, 1, \dots, N$ , the firm orders capacity from two modes: an inexpensive-but-slow *base* mode with a one-period lead time and a unit cost of  $c_b$ , and a fast-but-expensive *flexible* mode with zero lead time and a unit cost of  $c_f$  ( $> c_b$ ). It does this to maximize its total expected profit over the entire planning horizon. Demand  $D_n$  in period  $n$  has three parts: a deterministic component  $\mu_n$ , the initial market information  $\varepsilon_n^1$ , and the final market information  $\varepsilon_n^2$ . The firm knows  $\mu_n$ , and observes  $\varepsilon_n^1$  at the beginning of period  $n$  and  $\varepsilon_n^2$  at the end of period  $n$ . Both  $\varepsilon_n^1$  and  $\varepsilon_n^2$  are random variables. Market information for different periods and different market information in the same period are assumed to be independent of each other. Given these assumptions,  $D_n$  can be expressed as  $D_n = g(\varepsilon_n^1, \varepsilon_n^2, \mu_n)$ , where  $g(\cdot)$  is any Borel-measurable function.

The firm determines the optimal amount of equipment to order from both modes at each period  $n$ . The sequence of events is as follows (Figure 2): (i) At the beginning of period  $n$ , the firm observes the current capacity position  $x_n + B_{n-1}$ , where  $x_n$  is the on-hand capacity and  $B_{n-1}$  is the on-order capacity from the base supplier in period  $n - 1$ . It also observes the initial market information  $\varepsilon_n^1$  and any backlogged demand  $y_n$  from the previous period. (ii) The firm places a flexible order  $F_n$  at unit price  $c_f$  and a base order  $B_n$  at unit price  $c_b$ . Note that in period 0 only base orders are placed because demand will not materialize until period 1; and in period  $N$  only flexible orders are placed because this is the last period in the selling horizon. (iii) Orders  $F_n$  and  $B_{n-1}$  arrive. (iv) The final market information  $\varepsilon_n^2$  is revealed and demand  $D_n$  is realized. (v) Given the on-hand equipment capacity  $x_n + B_{n-1} + F_n$ , production is carried out to satisfy demand at a unit profit margin of  $p_n$ . Without loss of generality, we assume each unit of capacity can be used to process only one product every period. (vi) Any unsatisfied customer demand  $y_{n+1}$  is backlogged. Note that when base orders are placed for period  $n$  ( $B_{n-1}$ ), the buyer still faces significant demand uncertainty ( $\varepsilon_{n-1}^2$ ,  $\varepsilon_n^1$ , and  $\varepsilon_n^2$ ). However, it has much improved demand information (both in terms of demand realization  $\varepsilon_{n-1}^2$  and an updated forecast  $\varepsilon_n^1$ ) with the flexible mode  $F_n$ . As such, our demand structure does not limit the value of the flexible mode to demand realizations but does capture

Figure 2 Sequence of Events in the Planning Horizon



forecast updates. That is, when a large demand forecast update  $\varepsilon_n^1$  is observed, the flexible mode can be used to dampen the risk of capacity underage.

Several cost and revenue parameters affect the firm's decisions. We assume that the unit profit margin  $p_n$  is decreasing in  $n$ ; hence, it is more profitable to satisfy demand earlier rather than later. Because we allow for backlogs, the decreasing margin also functions as the backlog penalty. On-hand capacity incurs unit holding/maintenance cost  $c_n$  per period. There is zero salvage value for on-hand capacity after the horizon ends. This can be justified by the fact that leading firms do not cheap-sell their idle capacity for fear of revealing crucial technology to competitors. Any unsatisfied demand after the terminal period  $N$  incurs an additional unit penalty  $c_u$ , which may be because an expensive alternative mode is used to satisfy this demand. We also assume that all random variables have finite mean and variance, and we impose a discount factor  $\delta$ ,  $0 < \delta < 1$ , per period. Note that the problem analyzed in this section closely mimics Intel's setting but is stylized to enable analytical tractability.

The sequential decision problem of choosing the optimal  $B_n$  and  $F_n$  for all  $n$  can be formulated as a dynamic programming model. We present the details of this model in the appendix. Proposition 1 demonstrates that an optimal structural policy exists for the flexible mode, but not for the base mode. Let  $\tilde{x}_n$  be the capacity position at the beginning of period  $n$  and  $\tilde{y}_n$  be the modified backlog level. Defining  $S_n^F(\tilde{y}_n)$  as the optimal base-stock level for the flexible mode and  $S_n^B(\tilde{y}_n) \geq S_n^F(\tilde{y}_n)$  as the optimal partial base-stock level for the base mode, we can prove the optimality of a partial base-stock policy for the base mode.

**PROPOSITION 1.** *For the flexible mode, a state-dependent base-stock policy is optimal. For the base mode, there exists a state-dependent partial base-stock policy with parameter  $\underline{S}_n^B(\tilde{y}_n)$  satisfying  $\underline{S}_n^B(\tilde{y}_n) \geq S_n^F(\tilde{y}_n)$  such that*

- (i) *if  $\tilde{x}_n \leq S_n^F(\tilde{y}_n)$ , it is optimal to expand the capacity position to  $S_n^B(\tilde{y}_n)$ ;*
- (ii) *if  $S_n^F(\tilde{y}_n) < \tilde{x}_n \leq \underline{S}_n^B(\tilde{y}_n)$ , the optimal expand-to capacity position depends on  $\tilde{x}_n$  and  $\tilde{y}_n$ ; and*

- (iii) *if  $\tilde{x}_n > \underline{S}_n^B(\tilde{y}_n)$ , it is optimal not to order from the base mode.*

The failure of the optimality of a state-dependent base-stock policy for the base mode is a stark departure from the related inventory literature and is counterintuitive at first. In the capacity expansion setting, the cumulative capacity automatically serves as an additional mode. The existence of this mode is inconsequential when there are no backlogs, as in Chao et al. (2009). However, if there are backlogs, then this mode acts as a fictitious delivery source for which the firm does not even have to place an order; i.e., it can be thought of as a delivery mode with a lead time of  $-1$ . Therefore, even a dual-source problem with consecutive zero-one lead times acts as a multisource problem. With this interpretation, our result is consistent with Feng et al. (2006), where the authors show that only the fastest two modes (in our case, the fictitious mode and the flexible mode) have a base-stock policy. Thus, the dynamic dual-source capacity expansion problem with backlogs is inherently different and more complex than its inventory counterpart as well as the dynamic dual-source capacity expansion problem with lost sales.

The above analysis suggests that modeling the DMEP problem as a dynamic program may not be the best approach for two reasons. First, this is a complex problem and the dynamic programming model is rather difficult to solve. Even for the simplest setting, where the total order quantities are uncapacitated and lead times are consecutive zero-one, only the flexible orders follow a state-dependent base-stock policy. The base orders, however, follow only a partial base-stock policy; the expand-to capacity position for the base mode could be decreasing in the initial capacity level because of the effect of backlogging. (Figure O1 in the online supplement provides the details.) The general case with nonconsecutive lead times is even more complex. Second, although the dynamic programming model may help us identify the optimal equipment procurement strategy, it comes with strict assumptions, such as fixed and known distributions of all uncertain factors and consecutive lead times, which can hardly be justified in practice. Therefore,

to avoid the above restrictions, in §5 we propose an open-loop simulation model with a rolling horizon as a heuristic approach. The goal of this approach is to provide Intel with a fast and accurate decision support tool with what-if capabilities that can guide the firm in answering the questions we pose in §1.

## 5. The Dual-Mode Equipment Procurement Heuristic

To map the framework presented in Figure 1, the DMEP heuristic consists of the same three layers: the outermost is the contract negotiation layer that identifies indifference curves of lead time and price combinations so that Intel can pick the best alternative from the contract menu; the middle is the reservation layer that calculates the equipment reservation quantities  $B^T$  and  $F^T$  for the two supply modes; and the innermost is the execution layer that determines the equipment order quantities from the two supply modes in each period. Before elaborating on each of these three layers in more detail, we first consider the demand forecast revision process, which is one of the main drivers of the problem.

### 5.1. The Forecast Revision Mechanism

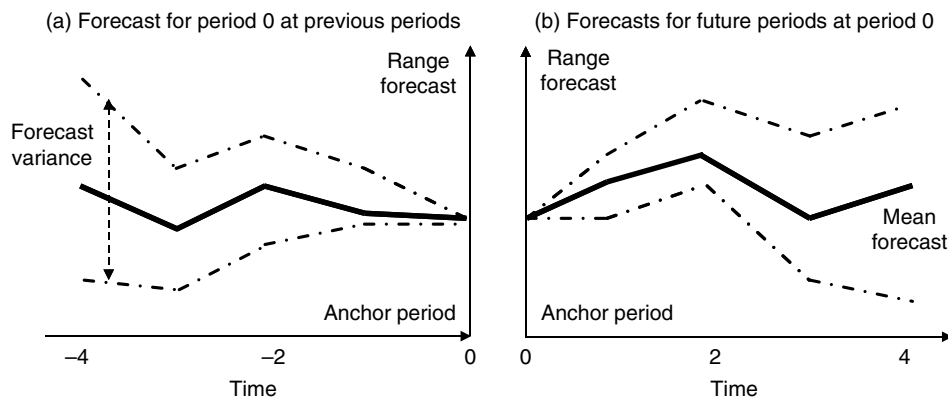
Demand volatility is a major concern for Intel during capacity planning. To guide Intel in its reservation and execution decisions in this highly volatile environment, DMEP should demonstrate a sound underpinning of the process by which the firm continuously adjusts its anticipation of the future demand distribution based on the available information. Such a model should be detailed enough to incorporate the main dynamics, but simple enough to transfer to practice seamlessly.

Using historical data, Figure 3(a) illustrates the nature of demand forecasts at Intel. The figure displays how forecasts made in earlier periods for period 0 demand evolve over time (period 0 is chosen as an anchor period and does not necessarily

correspond to the starting period of the demand ramp). The solid line represents the mean forecast evolution path, and the interval between the dashed lines denotes the variance range, which shrinks as the forecasting lead time decreases. Furthermore, this improvement is almost linear with respect to the forecasting lead time. Based on these observations, we characterize the forecasts made in a period for all future periods' demand as illustrated in Figure 3(b): the solid line represents the mean forecast and the dashed interval denotes the variance range, which diverges as the forecasting lead time increases. That is, the forecasting accuracy decreases as one forecasts further into the future.

Following the above guidelines and the characterization illustrated in Figure 3, we modify the classical *martingale model of forecast evolution* (MMFE; Hausman 1969, Graves et al. 1986) by decomposing it into a mean evolution process and a variance evolution process. This decomposition is driven by the empirical observation of independent mean and variance evolutions based on 10 years of historical data (equivalent to approximately 5 manufacturing processes) at Intel. In particular, we define  $\mathcal{D}_n^m$  as the demand forecast for period  $n$  made in period  $m$  ( $m = 1 - L_b, \dots, N - L_f$ ;  $n = 1, \dots, N$ ;  $m < n$ ). We use  $\mu_n^m$  and  $cv_n^m$ , respectively, to represent the mean and the coefficient of variation (c.v.) associated with  $\mathcal{D}_n^m$  that follows a normal distribution.  $\mathcal{D}_n^m$  then refers to the realized demand of period  $n$ . We assume that the mean forecast evolution for a certain period  $n$ 's demand follows a Markovian process with either an additive form:  $\mu_n^m = \mu_n^{m-1} + \varepsilon_n^m$  ( $m = 2 - L_b, \dots, n - 1$ ), where all  $\varepsilon_n^m$  are independent random variables with mean zero; or a multiplicative form:  $\mu_n^m = \mu_n^{m-1} \hat{\varepsilon}_n^m$  ( $m = 2 - L_b, \dots, n - 1$ ), where all  $\hat{\varepsilon}_n^m$  are independent positive random variables with mean value  $E\hat{\varepsilon} = 1$ . For both cases, the initial mean forecast profile at period  $1 - L_b$  is given by  $\bar{\mu}^{1-L_b} = (\mu_1^{1-L_b}, \dots, \mu_N^{1-L_b})$ . The choice between these two forms depends on the specific business environment. Despite its analytical

Figure 3 Illustration of the Demand Forecast Updating Process at Intel



complexity, the multiplicative form may be preferred for two reasons (Hurley et al. 2007): First, the additive form may lead to negative demand values; second, industry forecasts are usually updated in a relative sense rather than an absolute sense. We further assume that the forecast variance evolution is predictable and the coefficient of variation of forecasts is uniquely determined by the forecasting lead time:  $cv_n^m = f(n - m)$ , where  $f(\cdot)$  is an increasing function. Different functional forms (e.g., linear, quadratic, logarithmic, etc.) can be used for  $f(\cdot)$  to reflect how forecasting accuracy changes with the forecasting lead time.

Our modifications may be considered a notational complication to the standard MMFE method. However, these modifications capture Intel's forecast updating process in an easy-to-implement format. Especially when compared to the original multiplicative MMFE process under which  $\mathcal{D}_n = \mu_n^{1-L_b} \prod_{m=2-L_b}^n \hat{\varepsilon}_n^m$ , our approach avoids calculating cumulative distributions that involves the multiplication of several random variables. Thus, our approach achieves a trade-off between analytical rigor and managerial applicability. Note that the overall DMEP framework is independent of this forecast update mechanism (which we devise specifically for Intel) and can be combined with other forecast update mechanisms that may be a better fit for another firm's practice. Also note that prior to our analysis, Intel only had a formal process that combined range forecasting (i.e., the variance of forecasts as we discussed above) with scenario analysis to update the next forecast. The company did not model the evolution of forecasts for any periods beyond that. Therefore, our forecast revision process extends Intel's forecast update approach to a formal process for updating all future forecasts.

## 5.2. The Execution Problem

The execution module is the core of the DMEP heuristic. Given the reservation quantities  $B^T$  and  $F^T$ , the execution module characterizes how Intel should place the base and the flexible orders in each period. At the beginning of each period, Intel obtains the latest demand realization and forecast updates for all future periods. Based on this information—all previously placed base and flexible orders, and the backlog quantity from the preceding period—the open-loop execution algorithm calculates the myopic optimal base and flexible order quantities for the remaining periods.

The selling season of the product is  $N$  periods and starts in period 1. To prepare for the demand ramp, Intel starts placing orders  $L_b$  periods before the first demand realization and stops placing orders  $L_f$  periods before the end of the selling season. Therefore, the

length of the planning horizon is  $N + L_b - L_f$  periods. The firm orders only from the base mode in periods  $1 - L_b, \dots, -L_f$ ; it orders from both modes in periods  $1 - L_f, \dots, N - L_b$ ; and it orders only from the flexible mode in periods  $N - L_b + 1, \dots, N - L_f$ . In periods  $N - L_f + 1$  to  $N$ , Intel simply satisfies demand with the existing capacity. Specifically, in period  $m$  of the planning horizon ( $m = 1 - L_b, 2 - L_b, \dots, N - L_f$ ), the execution module solves the following stochastic optimization problem:

$$\begin{aligned} \max_{\bar{B}_{1:N}; \bar{F}_{1:N}} \quad & \mathbb{E}_{d_{1 \vee (m+1)}, \dots, d_N} \left[ \sum_{i=1}^N \delta^i \{p_i s_i - c_b B_i - c_f F_i - c_h k_i\} \right. \\ & \left. - \delta^{N+1} c_u d_{N+1}^{\text{rem}} \right] \quad (1) \end{aligned}$$

subject to:

For each realized sample path  $d_{1 \vee (m+1)}, \dots, d_N$ :

$$s_i = \min\{k_i, d_i + d_i^{\text{rem}}\} \quad \text{for } i = 1, \dots, N; \quad (2)$$

$$k_i \geq \psi d_i + d_i^{\text{rem}} \quad \text{for } i = 1 \vee (m + L_f), \dots, N; \quad (3)$$

$$k_i = k_{i-1} + B_i + F_i, \quad \text{for } i = 1, \dots, N \text{ with } k_0 = 0; \quad (4)$$

$$d_i^{\text{rem}} = (d_{i-1} + d_{i-1}^{\text{rem}} - k_{i-1})^+, \quad \text{for } i = 1, \dots, N \text{ with } d_1^{\text{rem}} = 0; \quad (5)$$

$$\sum_{i=1}^N B_i \leq B^T; \quad \sum_{i=1}^N F_i \leq F^T; \quad (6)$$

$$\bar{B}_{1:N} \geq 0; \quad \bar{F}_{1:N} \geq 0; \quad (7)$$

$$\bar{B}_{1:m-1+L_b} = \bar{B}_{1:m-1+L_b}^*; \quad \bar{F}_{1:m-1+L_f} = \bar{F}_{1:m-1+L_f}^*. \quad (8)$$

We define  $x \vee y := \max(x, y)$  and use the notation  $\vec{x}_{1:j}$  to represent the vector  $[x_1, x_2, \dots, x_j]^T$ . Additionally,  $p_i$  is the profit margin for period  $i$ ;  $c_b$  the base mode execution price;  $c_f$  the flexible mode execution price;  $c_h$  the unit equipment holding cost;  $c_u$  the unit penalty cost for unmet demand after the planning horizon ends;  $d_i$  the incoming demand in period  $i$ ;  $d_i^{\text{rem}}$  the unsatisfied demand remaining from period  $i - 1$ ;  $s_i$  the actual sales quantity during period  $i$ ;  $k_i$  the cumulative capacity position at period  $i$ ;  $\psi$  the service level target; and  $\delta$  the discount factor.  $B_i^*$  and  $F_i^*$  represent the optimal decisions that were already executed in previous periods. We note that the subscripts used for order quantities denote the period when the ordered equipment arrives.

The objective function maximizes the expected profit by considering the profit margin, equipment procurement costs, and inventory-related costs. The expectation is taken over all future demand, and the execution cost is calculated when the orders arrive. We note that as our goal is to understand the strategic- and tactical-level capacity decisions,



we suppress the operational-level product inventory problem and assume that production in a period will not exceed demand. Constraint (2) guarantees that sales in a given period cannot be larger than the demand or the supply; constraint (3) forces Intel to satisfy at least  $\psi$  percent of the incoming demand during each period after fulfilling the backlogs. Constraints (4) and (5) enable state transitions for the capacity and backlogs, respectively. Constraints (6) and (7) guarantee that the base and flexible orders are within their reservation limits and are nonnegative. Finally, constraint (8) freezes the already executed orders before finding the optimal order quantities. One crucial feature of this rolling-horizon algorithm is that, in period  $m$ , only the orders  $B_{m+L_b}^*$  and  $F_{m+L_f}^*$  (if  $m + L_f \geq 1$ ) will actually be placed, and the rest of the decisions will be postponed to future periods.

To accurately model Intel's business practice, we make several assumptions: First, we analyze a setting where the unmet demand during a certain period is backlogged (instead of lost, which is assumed in most of the capacity planning literature). We penalize the backlogged demand in two ways: (i) At the end of the planning horizon, any remaining unsatisfied demand incurs a terminal penalty; and (ii) the profit margin of the product is decreasing over time, which implies that there is a loss of revenue associated with backlogging. Therefore, satisfying the demand earlier is always preferable if it is possible. Second, the on-hand equipment capacity incurs a unit holding cost, which may include the opportunity cost of investment or costs such as a utility fee, maintenance expenditure, or cost of floor space. Third, investment in the equipment capacity is irreversible; i.e., capacity contraction is not allowed. Finally, the raw material inventory is always sufficient for the production process, and we only concentrate on the equipment procurement decisions. In addition, we also assume that Intel is risk-neutral (we relax this assumption in §6.3) and the supplier has sufficient capacity to always satisfy the firm's orders whether they are base or flexible as long as they are within the reservation limits.

Note that our formulation does not include the option of holding product inventory. This modeling choice is based on two things. First, at the top level of the framework we are solving a strategic problem possibly years in advance of actual demand realization. Given this long lead time, there is huge uncertainty regarding not only the demand but also the supply process (e.g., yields). Therefore, we simplified the execution-level problem and concentrated on the most critical decisions (i.e., the capacity levels to be executed) to avoid including additional noise. Second, the decision to not include product inventory is also consistent with Intel's strategy of "erring on the side

of ordering too much equipment" and was considered to be a better initial approach than tackling too many frontiers at the same time. If we were to include product inventories, we would see reductions in the total capacity level, as capacity and inventory are substitutes. We would like to note that the execution-level algorithm can be modified to include the option of holding product inventory easily. One must define a decision variable for product inventory for each period and parameters for inventory holding cost and salvage value. As a result, the execution-level problem would be slightly more complicated as, in addition to the base and flexible capacity execution levels, the optimization will also need to calculate the optimal product inventory levels.

To illustrate the decision-making process more clearly, assume that there are six periods in the selling season, the base mode lead time is four periods, and the flexible mode lead time is two periods. At the beginning of the planning horizon, period  $-3$ , the latest demand forecast information ( $\mu$ s and  $c$ v)s for period 1 to period 6 is revealed. The algorithm runs the aforementioned stochastic optimization to maximize the expected total profit across the entire planning horizon under the current demand information, taking the six base orders ( $B_1$  to  $B_6$ ) and six flexible orders ( $F_1$  to  $F_6$ ) as the decision variables. Once the optimal order quantities are obtained, the firm only needs to commit to  $B_1$  at period  $-3$ ; the rest of the decisions ( $B_2$  to  $B_6$  and  $F_1$  to  $F_6$ ) are postponed until later. At the beginning of period  $-2$ , the updated forecast information (new  $\mu$ s and  $c$ v)s is obtained; the firm then solves a new stochastic optimization to maximize the horizonwide expected profit under the newly obtained demand information, fixing  $B_1$  and taking  $B_2$  to  $B_6$  and  $F_1$  to  $F_6$  as the decision variables. Similarly, at period  $-1$  only the decision  $B_2$  needs to be executed, and the rest of the decisions are left for later periods. Following this logic,  $B_3$  and  $F_1$  will be executed in period  $-1$ , etc. This rolling-horizon decision-making process continues until period 4, when the final order  $F_6$  is committed. It is important to emphasize that starting from period 1, the actual realized demand is treated as part of the updated demand information and should be taken into consideration when calculating the total expected profit.

We solve the above stochastic program using the standard sample average approximation method based on Monte Carlo simulation (Entriken and Infanger 1990, Shapiro 2008). As we show in the next proposition, the problem is easy to execute; therefore, we are able to consider a large number of samples, which improves the accuracy of the solution. Hence, we do not have to rely on refinements of Monte Carlo simulation, such as importance sampling, which aim

to reduce the variance of estimation and improve its power.

**PROPOSITION 2.** *Assuming that demand  $d_n$  is discrete and takes finitely many values, the above stochastic programming model can be converted into a linear programming model.*

The execution algorithm provides a handy roadmap that indicates when and how much to order from the two supply modes. It captures the evolution of demand information and enables timely decision making. Unlike rule-of-thumb approaches Intel previously relied on (such as executing flexible orders only during the peak period), the necessity of the flexible mode is evaluated in each period of the planning horizon. On the flip side, the execution algorithm is myopic in the sense that it finds the “optimal” ordering scheme based on the current information without considering the opportunity to make contingent decisions based on the actual realized demand at each stage. Fortunately, this disadvantage is mitigated by the rolling-horizon nature of the algorithm.

### 5.3. The Reservation Problem

As equipment suppliers gain more power because of the trend of supply consolidation, simply letting the supplier bear most of the procurement risk is no longer viable. The reservation procedure of the DMEP heuristic, therefore, functions as a mechanism for risk-sharing between Intel and its supplier. Intel, by paying an up-front reservation fee, shares the risk of capacity building and installation with the supplier and enjoys the guaranteed delivery of equipment in return.

Determining how much capacity to reserve from the two supply modes, especially the flexible mode, is based on a trade-off between the reservation cost and the potential benefits from the guaranteed flexibility. Specifically, the optimal reservation quantities  $B^T$  and  $F^T$  are determined according to a scenario analysis of the future demand profiles. Intuitively, if the future demand scenario involves no uncertainty, then flexibility has no value; it is never optimal to order from the expensive flexible mode. In contrast, if the future demand scenario is highly uncertain and the demand mean forecast is very likely to be modified during the updating process, then flexibility has a high value and we should expect a higher flexible reservation level. In general, the reservation quantities maximize the expected horizonwide profit over all possible demand scenarios.

More precisely, the reservation algorithm determines the optimal  $B^T$  and  $F^T$  based on a Monte Carlo simulation performed on the mean forecast evolution trajectories, which are generated according to the forecast revision mechanism introduced in §5.1. Assuming that the mean forecast for the demand in different

periods evolves according to a Markovian process, we then have

$$P(\mu_1^m, \dots, \mu_N^m | \mu_1^{m-1}, \mu_1^{m-2}, \dots, \mu_N^{m-1}, \mu_N^{m-2}, \dots) \\ = P(\mu_1^m, \dots, \mu_N^m | \mu_1^{m-1}, \dots, \mu_N^{m-1}), \quad \text{for } m < 1, \quad (9)$$

where  $P(\cdot)$  is the probability mass (density) function if  $\mu$  takes discrete (continuous) values. Therefore, at the beginning of the planning horizon, given the initial demand forecast profile  $\bar{\mu}^{1-L_b}$  for the entire horizon and  $P(\cdot)$ , the algorithm enumerates a large number of possible mean forecast evolution paths. For each path it calls the execution module to calculate the specific order quantities as well as the expected horizonwide profit. The algorithm then chooses the reservation quantities  $B^T$  and  $F^T$  that maximize the average total profit across the entire planning horizon.

Equation (10) formulates the reservation problem mathematically. We first denote the optimal value function of the period  $m$  execution problem (1)–(8) as  $J^m(B^T, F^T, \bar{\mu}^m, \bar{c}v^m)$  by decomposing the demand information  $d$  into its two components  $\mu$  and  $cv$ . We choose  $B^T$  and  $F^T$  to maximize the expected horizonwide profit  $J^{N-L_f}$  because period  $N - L_f$  is the last period during which a decision can be made. The stochastic optimization is given as

$$\max_{B^T \geq 0; F^T \geq 0} \mathbb{E}_M [J^{N-L_f}(B^T, F^T, M, \Sigma | \bar{\mu}^{1-L_b})] \\ - r_b B^T - r_f F^T, \quad (10)$$

where the expectation is taken with respect to the mean forecast evolution space  $M$ :

$$M = \{\bar{\mu}^m: \bar{\mu}^m = (\mu_{1 \vee (m+1)}^m, \dots, \mu_N^m), \\ m = 1 - L_b, \dots, N - L_f\}.$$

As expressed in (9), the mean forecast profile  $\bar{\mu}^m$  follows a Markov process with initial state  $\bar{\mu}^{1-L_b}$  and transition probability space  $P(\cdot)$  such that a particular realization of  $M$  occurs with probability  $P(M | \bar{\mu}^{1-L_b}) = P(\bar{\mu}^{2-L_b} | \bar{\mu}^{1-L_b})P(\bar{\mu}^{3-L_b} | \bar{\mu}^{2-L_b}) \dots P(\bar{\mu}^{N-L_f} | \bar{\mu}^{N-1-L_f})$ . The variance forecast evolution space is  $\Sigma = \{\bar{c}v^m: \bar{c}v^m = (cv_{1 \vee (m+1)}^m, \dots, cv_N^m), m = 1 - L_b, \dots, N - L_f\}$ . Based on our assumption that the forecasting variance is uniquely determined by the forecasting lead time,  $\Sigma$  is deterministic. Proposition 3 establishes the concavity of the reservation problem. In addition, the objective function goes to negative infinity as  $B^T$  or  $F^T$  tends to infinity, given that demand is finite. Thus, the reservation problem has finite optimal reservation quantities for the supply modes and can be solved using either an optimization software or a search algorithm.

**PROPOSITION 3.** *The objective function in (10) is concave in  $(B^T, F^T)$  and, given that demand is finite, goes to negative infinity as  $B^T$  or  $F^T$  tends to infinity.*

The reservation algorithm helps Intel determine the optimal amount of equipment to reserve for both supply modes. It builds on the philosophy of scenario analysis and chooses the reservation quantities that guarantee the maximum expected return to the firm. By adjusting the mean forecast transition probability  $P(\cdot)$ , we can easily create different demand scenarios; hence the reservation algorithm is applicable to a wide range of business settings.

#### 5.4. Contract Negotiation Problem

During the contract negotiation stage, Intel and the equipment supplier determine the lead times ( $L_b, L_f$ ) of the two delivery modes as well as the unit reservation ( $r_b, r_f$ ) and execution ( $c_b, c_f$ ) prices associated with these times. Our algorithm here involves a sensitivity analysis: for different (lead time, price) combinations, we run the reservation and execution heuristic and obtain the corresponding expected horizonwide profits. The decision maker can then choose the (lead time, price) pair from the contract menu that leads to the highest expected return. The strength of this part of the heuristic is that it helps Intel make strategic-level decisions by considering potential tactical- and operational-level contingencies. Thus, the three stages of DMEP constitute a stable decision-support pyramid, where the decisions are made in a top-down sequence while the underlying algorithm follows an embedded bottom-up order.

### 6. DMEP as a Decision-Support Tool

We revisit the questions that we asked in §2 and illustrate our approach to them with numerical examples. In §6.1, we provide the parameter values that we use. We then explore the value of DMEP as a decision-support tool with an emphasis on these questions (§6.2). Finally, we study the decisions of a risk-averse firm (§6.3). We execute the DMEP heuristic using the convex optimization tool CVX (<http://cvxr.com/cvx/>) implemented in Matlab.

#### 6.1. Parameter Values for the Numerical Examples

The parameter values we use are based on the business environment of Intel and its suppliers. However, the values presented here are either publicly available or have been disguised to protect the firms. For all numerical examples, the selling season is  $N = 6$  quarters. Each piece of equipment can process 12,000 wafers per quarter, and each wafer can be further sawed into 1,425 chips. The profit margin (including the equipment cost) of a single chip is  $p_0 = \$33.75$  before the first quarter of the life cycle starts; the profit margin  $p_t$  in quarter  $t$  satisfies an exponential decreasing formula:  $p_t = p_0 e^{-\alpha t}$  (Leachman 2007), where the coefficient  $\alpha$  is equal to 0.23, roughly implying that the margin decreases by 50% per year. The

unit penalty cost for unmet demand at the end of the selling season is  $C_u = \$50$ . The service level is 95%. The discount factor per quarter is  $\delta = 0.96$ .

In this setting, the lead time and price for the base mode are usually fixed. However, suppliers offer a menu of contracts for the flexible mode that commonly includes the available flexible lead time options as well as the price associated with each lead time. The shorter the flexible lead time, the higher the flexible price. Using this fact, we set the base lead time  $L_b$  to 4 quarters and the total equipment price associated with the base mode  $p_e^b$  to \$25 million. We impose a minimum base reservation price of  $r_b = \$0.1$  million to eliminate the trivial case of an infinite base reservation quantity. The base execution price is thus given as  $c_b = p_e^b - r_b$ . In our examples, we consider a piece of equipment with a long lead time, as these tools are the bottleneck in capacity planning and are usually the ones which are very expensive. Hence they are difficult to manage, and so the target of DMEP. The price associated with the flexible mode is determined by two parameters: the equipment price increase ratio  $\theta$  and the flexible reservation price ratio  $\lambda$ . That is, the total equipment price of the flexible mode is  $p_e^f = \theta p_e^b$ , where  $\theta \geq 1$  because a faster mode implies a shorter preparation lead time for the supplier and hence a higher supply cost. The reservation price of the flexible mode is  $r_f = \lambda p_e^f$ . The remaining  $(1 - \lambda)$  portion, together with a fast shipment premium of \$50,000, is paid as the flexible execution price; i.e.,  $c_f = (1 - \lambda)p_e^f + \$50,000$ .

For the forecast revision process, we make additional assumptions. First, we use a multiplicative form ( $\mu_n^m = \mu_n^{m-1} \varepsilon_n^m$ ) to model the mean forecast evolution and introduce a mean-adjustment factor  $\beta$  to capture the value of  $\varepsilon_n^m$ : i.e.,  $\varepsilon_n^m = \{1 + \beta, 1, 1 - \beta\}$  with equal probability 1/3. (The mean forecasts for demand in different periods evolve independently.) Second, we assume that the coefficient of variation  $cv_n^m$  is linearly increasing in the forecast lead time;  $cv_n^m = \gamma \times 0.01(n - m)$ . We note that  $\beta$  controls the solid path in Figure 3 and  $\gamma$  controls the width of the dashed variance interval. Finally, the demand in each period follows a *truncated* normal distribution.

Table 1 displays the initial demand forecast provided by Intel. Using this initial forecast, which is symmetric with periods 3 and 4 being peaks, we study three different forecast scenarios. In the first scenario, the initial forecast is not adjusted and the realized demand in each period matches the corresponding mean forecast. The only randomness in this system is the variance coefficient  $\gamma$ . Because the mean forecast evolution process is degenerate and hence  $\beta = 0$  for all periods in this case, we call this scenario the *stationary demand* scenario. The second scenario investigates demand *forecast shocks*: during



**Table 1** Nonstationary Demand Forecast Scenarios (Unit: Wafer-Start per Week)

Scenarios		Period 1	Period 2	Period 3	Period 4	Period 5	Period 6
Initial forecast	$\bar{\mu}^{-3}$	4,788	9,577	14,365	14,365	9,577	4,788
Stationary demand	$\bar{\mu}^{-3}$	4,788	9,577	14,365	14,365	9,577	4,788
Forecast shock	$\bar{\mu}^{-1}$	<b>7,000</b>	9,577	14,365	14,365	9,577	4,788
	$\bar{\mu}^1$	6,850	9,577	<b>19,320</b>	14,365	9,577	4,788
Realization shock	$\bar{\mu}^1$	<b>14,000</b>	9,577	14,365	14,365	9,577	4,788

Note. Bold figures denote either a forecast shock or realization shock to the original forecast.

period  $-1$  (1), the forecast for the mean demand in period 1 (3) is adjusted upward. The third scenario corresponds to a situation with demand *realization shock*: the initial mean forecast profile  $\bar{\mu}^{-3}$  is not updated in periods  $-2$ ,  $-1$ , and 0; in period 1, however, the actual realized demand is much higher than the previous mean forecast.

## 6.2. The Value of DMEP as a Decision-Support Tool

**6.2.1. Under What Circumstances Does the Flexible Mode Create Value for the Firm?** To capture the settings under which the flexible mode creates value for Intel, we first investigate the optimal procurement decisions given the reservation quantities  $B^T$  and  $F^T$ . The optimal procurement decisions can be characterized in terms of the ratio between the flexible execution price  $c_f$  and the base execution price  $c_b$  and the size of the forecast shocks and/or demand realization shocks. As expected, if  $c_f/c_b \leq 1$ , then Intel procures equipment only from the flexible mode, because it is not only faster but also cheaper. If  $c_f/c_b > 1$  and the forecast and realization shocks are small enough, then it is optimal to procure only from the base mode. That is, in the stationary demand scenario or when shocks are small, cost is the only critical parameter: Intel prefers to single source using the less expensive mode as long as there is available reservation quantity. When there are upward forecast shocks or large realization shocks, however, this threshold policy ceases to apply and both modes are used. Now the flexible mode adds value with its shorter lead time, even when it is more expensive. Intel can wait until the last minute to learn more about demand before placing orders and thus maintains an agile environment.

At the tactical capacity reservation level, Figure 4 shows that as the procurement risk increases (i.e., either the mean evolution jump size  $\beta$  or demand variance coefficient  $\gamma$  increases), Intel relies on the flexible mode more heavily. The flexible mode is especially valuable in dealing with demand realization shocks because the firm's contingencies are very limited in that case. Figure 5(a) shows that as we gradually increase the service level constraint from 87.5% to 99%, the value of the flexible mode increases. Note that the expected profit associated with a 99% service

level is 1.6% less than that associated with a 95% service level. This is by no means a small compromise, and this trade-off should be considered (against other factors such as reputation, customer satisfaction and loyalty, competitive advantage, etc.) when making service-level decisions. Finally, Figure 5(b) shows that even when the flexible mode is considerably more expensive than the base mode, with a very conservative estimate, there would be at least a 0.4% increase in profit with dual-mode procurement, corresponding to tens of millions of dollars at Intel.

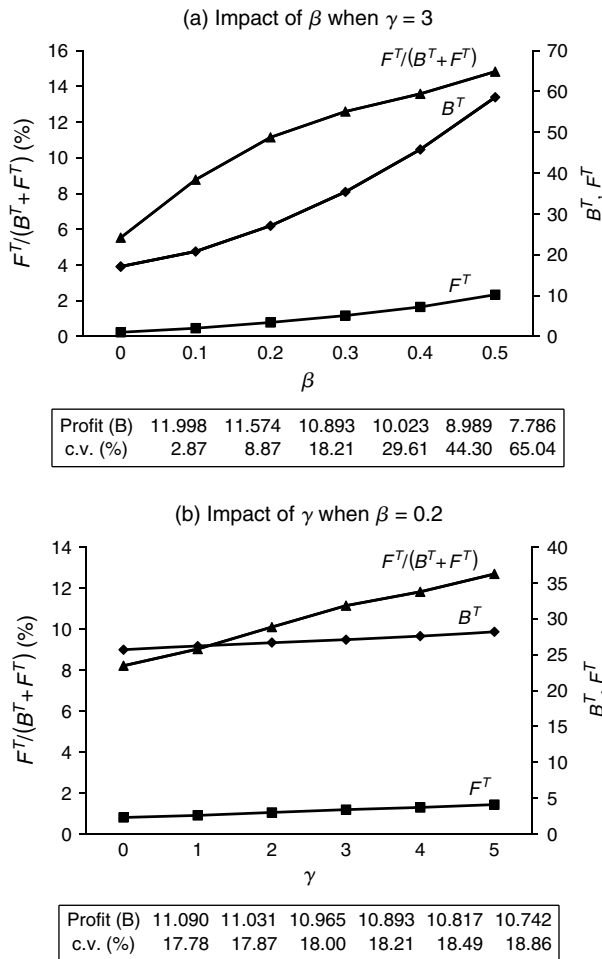
**6.2.2. How Much of the Total Capacity Should Be Reserved Through the Flexible Mode? When Should This Capacity Be Exercised?** We note that the conditions for dual-mode procurement correspond to the business environment where Intel operates; i.e., the forecast and realization shocks, the forecast uncertainty, and service levels are all high. Thus, dual-mode procurement should be seriously considered in this industry, which has traditionally used single-mode procurement. The notion that flexibility is only necessary for peak periods is also a misconception. Actually, the flexible mode may be optimally deployed whenever there is a forecast shock or a realization shock, both of which occur frequently during the ramp-up stage of the product life cycle, instead of just at the peak. Although the value of the flexible mode decreases as its total price (i.e.,  $\theta$ ) and/or reservation price (i.e.,  $\lambda$ ) increases, as long as the flexible lead time ( $L_f = 2$ ) is significantly shorter than the base, Intel should continue to reserve more than 8% of its total capacity through the flexible mode even when the flexible mode is 60% more expensive (Figure 6(a)) or when the firm has to pay 25% up front (Figure 6(b)). When the flexible lead time increases, however, the value of flexibility dramatically decreases and Intel tends to depend more on the base mode (Figure 5(b)).

## 6.2.3. How Can Intel Quickly Evaluate Different Flexible Options During Contract Negotiations?

One efficient way for Intel to select offers from the contract menu is to compare the position of different lead time and price combinations for the flexible mode on an iso-profit graph, where the flexible price is adjusted by two parameters: the price increase



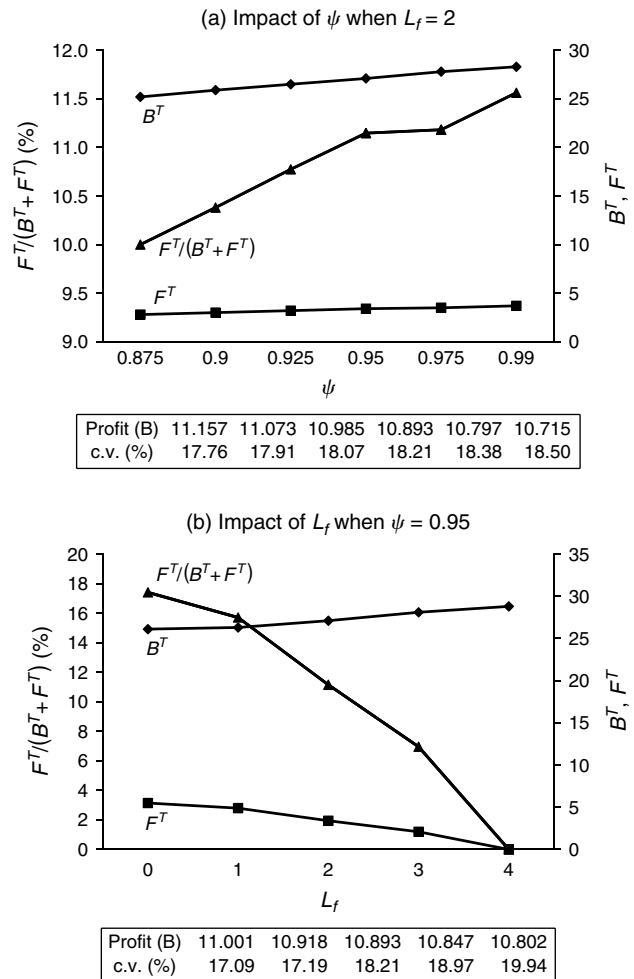
**Figure 4** Impact of Mean Evolution Jump Size  $\beta$  and Demand Variance  $\gamma$  ( $\theta = 1.3, \lambda = 0.15, \psi = 0.95, L_f = 2$ )



Note. (B) denotes billion.

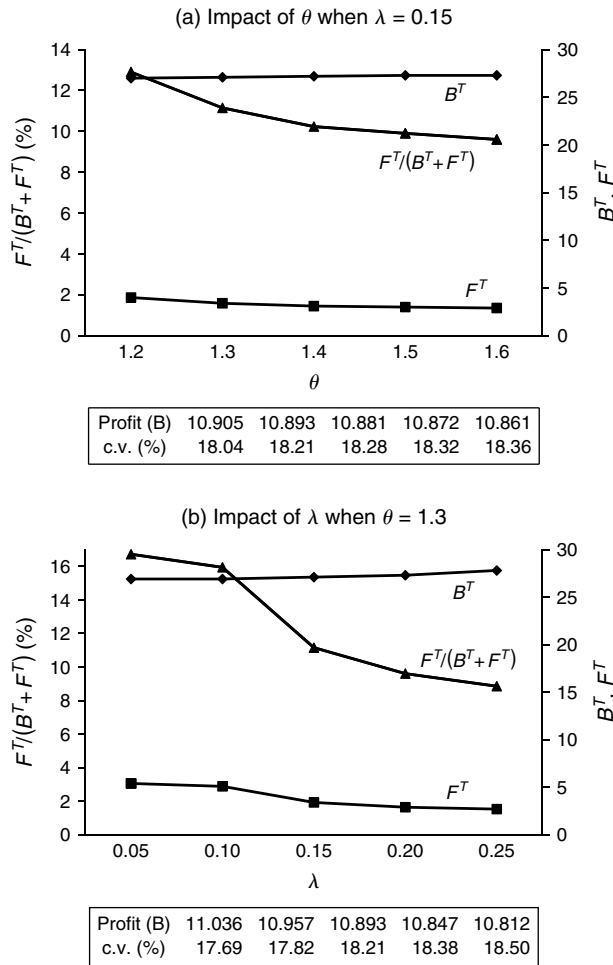
ratio  $\theta$  (Figure 7(a)) and the reservation price ratio  $\lambda$  (Figure 7(b)). The lead time–price pairs on each of the solid lines lead to the same expected total profit under the optimal reservation decision, while lines toward the lower-left corner correspond to higher profits than those toward the upper-right corner. Intel can utilize these curves in two ways. The curves demonstrate the dominance of different contract options: e.g., in Figure 7(a), contract A with  $L_f = 1$ ,  $\lambda = 14\%$ , and an expected profit of \$10.950 billion should be preferred to contract B with  $L_f = 3$ ,  $\lambda = 12.8\%$ , and an expected profit of \$10.893 billion. Alternatively, each curve quantifies the maximum reservation price Intel should be willing to pay for added flexibility; e.g., Intel can pay up to 20% of the total price up front and decrease the flexible lead time to 0, while still keeping its profits at the same level as in contract B. With the assistance of such iso-profit graphs, Intel will know the bottom-line impact of different alternatives while negotiating with its supplier and can make informed trade-offs between flexibility and cost.

**Figure 5** Impact of Service Level Target  $\psi$  and Flexible Mode Lead Time  $L_f$  ( $\theta = 1.3, \lambda = 0.15, \beta = 0.2, \gamma = 3$ )



Note. (B) denotes billion.

Note that for all our analyses (Figures 4–7), we also record the corresponding profit variability in terms of the coefficient of variation. For Figures 4–6, as demonstrated in the table below each figure, when cost, uncertainty, service level, or flexible lead time increase, the expected total profit decreases and the profit variability in terms of the c.v. increases. This information can be useful for Intel's decision-making process. In Figure 7, for each of the iso-profit curves, the profit variability first decreases then increases as we move from the left end of the curve to the right. For instance, on the \$10.950 billion curve in Figure 7(a), contract A with  $L_f = 1$  leads to the smallest profit c.v. of 17.20%, compared with the  $L_f = 0$  case with a c.v. of 17.53% and the  $L_f = 3$  case with a c.v. of 18.91%. This implies that Intel may prefer a contract term in the middle of an iso-profit curve rather than the contracts on the two ends, especially when risk preferences are also considered. That is the problem we investigate in the next section.

**Figure 6** Impact of Flexible Price Increase Ratio  $\theta$  and Reservation Price Ratio  $\lambda$  ( $\beta = 0.2, \gamma = 3, \psi = 0.95, L_f = 2$ )

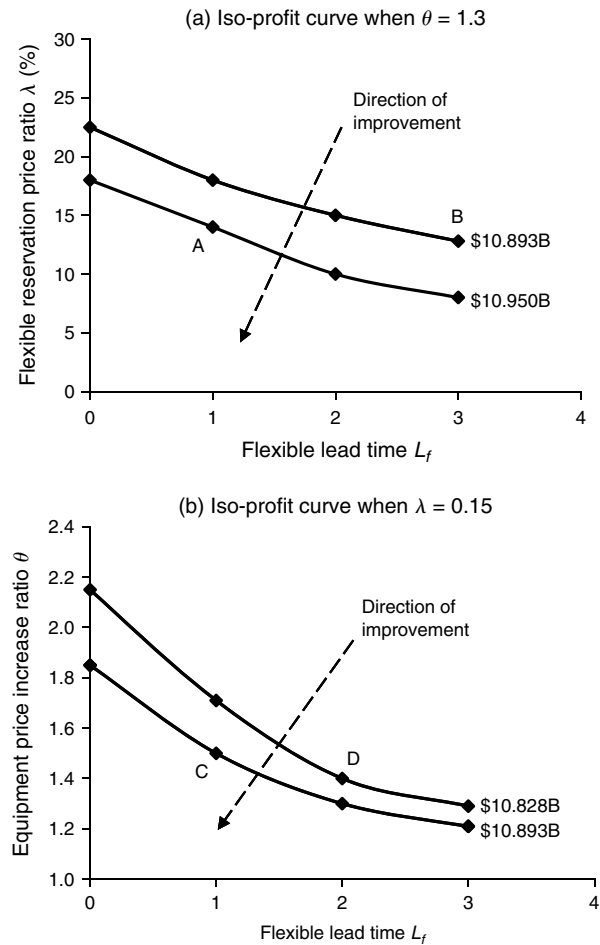
Note. (B) denotes billion.

### 6.3. The Impact of Risk Attitude

Because a firm's capacity planning involves a substantial up-front investment with uncertain future revenues, one natural extension of our model is to consider the impact of risk aversion. Van Mieghem (2003) reviews several methods to model a firm's risk aversion and hedging behavior during capacity investment. One predominant approach is to use a concave Bernoulli utility function and assume that the firm operates to maximize its expected utility instead of profit. We adopt this approach and inspect how the firm's reservation decisions  $B^T$  and  $F^T$ , as well as the profit and its variability, would change if a concave increasing utility function  $\mathcal{G}(\cdot)$  is applied to the reservation stage:

$$\max_{B^T \geq 0; F^T \geq 0} \mathbb{E}_M \mathcal{G}(\mathcal{F}(B^T, F^T, M | \tilde{\mu}^{1-L_b})),$$

where  $\mathcal{F}(B^T, F^T, M | \tilde{\mu}^{1-L_b}) = J^{N-L_f}(B^T, F^T, M, \Sigma | \tilde{\mu}^{1-L_b}) - r_b B^T - r_f F^T$  and  $J^{N-L_f}(\cdot)$  is the value function of the last-stage execution-level optimization problem.

**Figure 7** The Iso-Profit Curves for Fixed  $\theta$  and  $\lambda$  ( $\beta = 0.2, \gamma = 3$ )

We investigate the case where  $\mathcal{G}(\cdot)$  is a power function:  $\mathcal{G}(z) = z^\rho$ , where  $0 < \rho \leq 1$  (see Liu and van Ryzin 2011 for a similar treatment). Note that  $\mathcal{G}(z)$  is concave increasing, and the smaller the  $\rho$ , the more risk averse the firm tends to be (i.e.,  $\rho = 1$  corresponds to the risk neutrality). Also note that  $A(z) = -\mathcal{G}''(z)/\mathcal{G}'(z) = (1 - \rho)/z$  is decreasing in  $z$  and  $R(z) = zA(z) = 1 - \rho$  is constant in  $z$ . Hence,  $\mathcal{G}(z)$  has decreasing absolute risk aversion and constant relative risk aversion—two properties that are consistent with experimental and empirical findings about the risk-averse behavior of individuals and corporations (Friend and Blume 1975).

In a numerical analysis (with parameters  $\theta, \lambda, \beta$ , and  $\gamma$  taking values from a wide range set), we observe that as the firm becomes more risk averse, it reserves more capacity from both the base and flexible modes. Furthermore,  $F^T/(B^T + F^T)$  increases, which implies that the flexible mode becomes a more attractive option. The firm's expected profit decreases because of risk aversion. However, the profit variability (represented by both the standard deviation and c.v.) also decreases. Table 2 presents a representative

**Table 2** Risk Aversion Case with Power Utility Function ( $\theta = 1.3$ ,  $\lambda = 0.15$ ,  $\beta = 0.2$ ,  $\gamma = 3$ ,  $\psi = 0.95$ )

$\rho$	Lead time							
	$L_b = 0$				$L_b = 4, L_f = 2$			
	1	7/8	5/8	3/8	1	7/8	5/8	3/8
$B^T$	25.7	25.8	25.9	27.3	27.1	27.3	27.3	27.6
$F^T$	0	0	0	0	3.4	3.5	3.7	3.9
$F^T / (B^T + F^T)$ (%)	0	0	0	0	11.15	11.36	11.94	12.38
Profit (B)	11.238	11.230	11.228	11.222	10.893	10.891	10.888	10.886
Profit std. dev.	1.961	1.957	1.954	1.950	1.984	1.975	1.965	1.959
Profit c.v. (%)	17.45	17.43	17.40	17.38	18.21	18.13	18.04	17.98

Note.  $\rho = 1$  corresponds to the situation where the firm is risk neutral.

scenario where all the key parameters take their standard values. As a benchmark, we also consider a setting where the base mode has zero lead time; i.e., the firm enjoys the maximum level of flexibility and cost efficiency. Even under this ideal setting, the coefficient of variation of the total profit is around 17%. This level is inevitable because of the demand forecast evolution process and the tactical nature of the problem. That is, risk aversion removes a limited portion of the profit variability arising from operational demand-supply mismatches.

## 7. Conclusion

Capital equipment purchasing is a crucial yet difficult task for many semiconductor, electronic, automotive, and pharmaceutical firms. In this paper, we have proposed a DMEP model to guide firms through this complex task. DMEP serves three purposes. On the strategic level, it provides decision support to contract negotiation by comparing alternatives with different levels of flexibility and costs. On the tactical level, it guides capacity reservation decisions by characterizing the amount of capacity that should be reserved from different procurement modes. On the operational level, it quantifies procurement amounts by considering the latest demand information as well as the installed capacity. By incorporating interactions among these three levels of decision making, DMEP enables a flexible supply chain that adapts effectively to changing demand conditions. It helps firms better manage their equipment procurement process, eliminate excess capacity, and thus lower their costs. It benefits suppliers by enabling a risk-sharing mechanism through up-front capacity reservation and the elimination of soft orders.

This paper details a successful academic-practitioner interaction that resulted in the development and implementation of a decision-support framework at Intel Corporation. DMEP formalizes and extends the approach that Intel has used in the past to price

and exercise capacity options with reduced lead times for a few types of equipment (Vaidyanathan et al. 2005). Intel has been incrementally improving this approach with each new manufacturing technology. The DMEP framework provides Intel with a much improved method with sound theoretical underpinnings for determining the number of options needed, the valuation of those options, and the appropriate timing for exercising those options (or not) as forecasts evolve. Our sensitivity analysis (e.g., Figures 4–7) provides a simple yet powerful decision-support tool for Intel executives. DMEP also complements recent improvements in demand forecasting methodologies at Intel (Wu et al. 2010). Thus, the primary contribution of this research is the comprehensive DMEP framework that structures the multistage decision hierarchy and captures the real-world dynamics and constraints for the equipment procurement problem. In the process of developing the framework, we also demonstrate that the dynamic dual-source capacity expansion problem with backlogs is categorically different and more complex than not only its inventory counterpart but also the dual-source capacity expansion problem with lost sales. Hence, we provide an efficient heuristic for the rather complex dual-source capacity expansion problem with general lead times and demand backlogging.

Our analyses demonstrate that, by reserving 8%–12% of its capacity through the flexible mode, Intel can achieve its goal of improving its equipment procurement strategies to satisfy customer demand without overpurchasing costly capacity through a risk-sharing mechanism with its suppliers. In fact, the value of the DMEP framework has been tested in a real context as it and its associated decision support tools have been implemented at Intel, influenced procurement decisions for the previous process technology at the level of tens of millions of dollars in savings, and are now in use for the next process technology (a recently announced multi-billion-dollar capacity expansion).

DMEP can be modified easily to include additional factors that may be relevant for other firms. First, DMEP does not include the option of holding product inventory. However, the execution-level algorithm can be modified to include that option easily as we discussed in §5.2. Second, the heuristic can be modified to include order bounds on either the total capacity that can be reserved from a mode or the capacity that can be exercised at a given period from a mode.

Despite its many advantages, DMEP also has some limitations. First, we investigate a situation where the firm procures only one type of equipment from its supplier. This simplification enables us to demonstrate the dynamics of the algorithm without introducing complexity. DMEP can be generalized to a multi-equipment scenario where (i) firms consider ordering from the base and flexible modes for all types of equipment from different suppliers with different lead times in each period and (ii) the available capacity in each period is constrained by the lowest capacity among all the types of equipment. As expected, as the number of types of equipment increases, the interactions and the complexity of the problem also increase. Having said that, firms should consider DMEP only for equipment on the critical path of capacity planning, as the rest will not impose additional constraints on the system (in Intel's case, the types that have shorter lead times and are cheaper). Yet alternative formulations for multi-equipment procurement (similar to the ones provided in Huang 2008 for a setting with a single mode for each tool) would be valuable. These alternative formulations should consider the multitool problem as a portfolio of tools and suppliers, and potentially utilize approaches other than DMEP. Second, we formulate the procurement problem as a linear program. As such, DMEP is better suited to providing the fraction of orders from each mode rather than the actual procurement quantities. However, if the intention is to use DMEP to obtain the specific order quantities, generalizing it to an integer program would be preferable. Finally, in this paper, we model the problem of factory ramp; we assume that investment in the equipment capacity is irreversible and capacity contraction is not allowed. Although not common for Intel, capacity contractions may be relevant for other firms. Thus, studying a setting with capacity contraction would be fruitful.

This paper takes an initial, yet important, step toward formulating the complex DMEP problem in a practical and insightful way. In capital-intensive industries, capital expenditures often constitute up to one quarter of the total revenue and up to two thirds of the manufacturing costs. Given the millions of dollars that are at stake, we believe that equipment procurement problems will attract more attention from academia.

## Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://msom.journal.informs.org/>.

## Acknowledgments

The authors thank Steve Graves as well as the anonymous review team whose valuable suggestions improved the content and exposition of the paper significantly. The authors express their deepest gratitude to Intel Corporation's Technology and Manufacturing Group for their continuous support. This work benefitted greatly from the valuable discussions with Robert Bruck (vice president of the Technology and Manufacturing Group), Blake Sacha (capital planning director, Intel Corporation), and Timothy Rosenberg (senior finance analyst, Intel Corporation). The authors also thank Yanchong Zheng, Yichuan Ding, and seminar participants at Stanford University, Arizona State University, and Intel Corporation. They gratefully acknowledge Intel Corporation's financial support. Chen Peng and Feryal Erhun were partially supported by the National Science Foundation [Career Award NSF/CAREER-0547021].

## Appendix. The Execution Layer Problem: Details of the Model

The sequential decision problem of choosing the optimal  $B_n$  and  $F_n$  for all  $n$  can be formulated as a dynamic programming model. To simplify the notation, we introduce a new state variable  $\tilde{x}_n$  as the capacity position at the beginning of period  $n$ , which is defined as  $\tilde{x}_n = x_n + B_{n-1}$ . The other state variables are  $y_n$ , the unmet demand from last period, and  $\epsilon_n^1$ , the observed initial market information at period  $n$  ( $\epsilon_n^1 = \epsilon_n^1$ ). We construct the  $(N+1)$ -stage dynamic programming model as follows ( $n = 0, 1, \dots, N$ ):

$$\begin{aligned} J_n(\tilde{x}_n, y_n, \epsilon_n^1) \\ = \max_{B_n, F_n} \mathbb{E}[p_n \min\{g_n(\epsilon_n^1, \epsilon_n^2, \mu_n) + y_n, \tilde{x}_n + F_n\} - c_b B_n - c_f F_n \\ - c_h(\tilde{x}_n + F_n) + \delta J_{n+1}(\tilde{x}_{n+1}, y_{n+1}, \epsilon_{n+1}^1)] \end{aligned}$$

subject to

$$B_n \geq 0, \quad F_n \geq 0, \quad \text{and} \quad J_{N+1}(\cdot) = -c_u y_{N+1}.$$

In the objective function above, the first term denotes the sales profit; the second and third terms are the base and flexible capacity ordering costs, respectively; the fourth term captures the holding cost of the on-hand capacity; and the last term is the discounted profit-to-go. In the final period  $N$ , the firm incurs a penalty cost for any unsatisfied demand at the end of the planning horizon. The states are updated according to the following equations: capacity position  $\tilde{x}_{n+1} = \tilde{x}_n + B_n + F_n$ , where  $x_0 = 0$ ; unmet demand  $y_{n+1} = (y_n + g_n(\epsilon_n^1, \epsilon_n^2, \mu_n) - (\tilde{x}_n + F_n))^+$ , where  $y_1 = 0$ ; and  $(x)^+ = \max\{x, 0\}$ .

Furthermore, we assume that demand  $D_n$  is in additive form:  $D_n = g_n(\epsilon_n^1, \epsilon_n^2, \mu_n) = \epsilon_n^1 + \epsilon_n^2 + \mu_n$ . The additive form is commonly used in the literature (e.g., Graves et al. 1986). With this form, we can simplify the model by combining the



unmet demand  $y_n$  with the initial market information  $\varepsilon_n^1$ , both of which are revealed before the ordering decision is made. We denote this new term  $\tilde{y}_n = y_n + \varepsilon_n^1$  as the *modified* backlog level. We also update the decision variables and work with expand-to capacity positions,  $\tilde{x}'_n = \tilde{x}_n + F_n$  and  $\tilde{x}''_n = \tilde{x}'_n + B_n$ . As a result, the original model can be rewritten as

$$J_n(\tilde{x}_n, \tilde{y}_n) = \max_{\tilde{x}_n \leq \tilde{x}'_n \leq \tilde{x}''_n} V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}'_n, \tilde{x}''_n), \quad \text{where}$$

$$V_n(\tilde{x}_n, \tilde{y}_n, \tilde{x}'_n, \tilde{x}''_n) = \mathbb{E}[p_n \min\{\tilde{y}_n + \varepsilon_n^2 + \mu_n, \tilde{x}'_n\}$$

$$- c_b(\tilde{x}''_n - \tilde{x}'_n) - c_f(\tilde{x}'_n - \tilde{x}_n) - c_h \tilde{x}'_n$$

$$+ \delta J_{n+1}(\tilde{x}''_n, y_{n+1} + \varepsilon_{n+1}^1)]$$

for  $n = 0, 1, \dots, N$ ;  $y_{n+1} = (\tilde{y}_n + \varepsilon_n^2 + \mu_n - \tilde{x}'_n)^+$ ; and  $J_{N+1}(\cdot) = -c_u y_{N+1}$ .

To determine the optimal capacity expansion policy for the above model, we first need to establish some structural properties of the objective function  $V_n$ , which enable us to derive our results in Proposition 1:

**PROPOSITION A1.**  $J_n$  is concave in  $(\tilde{x}_n, \tilde{y}_n)$  and  $V_n$  is concave in  $(\tilde{x}'_n, \tilde{x}''_n)$  for any given  $\tilde{x}_n$  and  $\tilde{y}_n$ .

## References

- Armbrust, D. 2009. Now's the time. SIA Board Meeting Presentation, November 5, SEMATECH, Albany, NY. Retrieved August 18, 2010, [http://www.sematech.org/corporate/messages/SIA\\_Oct27.pdf](http://www.sematech.org/corporate/messages/SIA_Oct27.pdf).
- Baker, K. 1977. An experimental study of the effectiveness of rolling schedules in production planning. *Decision Sci.* 8(1) 19–27.
- Bassok, Y., R. Anupindi. 1997. Analysis of supply contracts with total minimum commitment. *IIE Transactions* 29(5) 373–381.
- Billington, C., B. Johnson, A. Triantis. 2002. A real options perspective on supply management in high technology. *J. Appl. Corporate Finance* 15(2) 32–43.
- Cachon, G. P. 2003. Supply chain coordination with contracts. A. G. de Kok, S. C. Graves, eds. *Handbooks in Operations Research and Management Science—Supply Chain Management: Design, Coordination and Operation*, Chap. 6. Elsevier, Amsterdam, 229–340.
- Chao, X., H. Chen, S. Zheng. 2009. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty. *Oper. Res.* 57(1) 82–93.
- Cohen, M., T. H. Ho, Z. J. Ren, C. Terwiesch. 2003. Measuring imputed cost in the semiconductor equipment supply chain. *Management Sci.* 49(12) 1653–1670.
- Daniel, K. H. 1963. A delivery-lag inventory model with emergency. H. E. Scarf, D. M. Gilford, M. W. Shelly, eds. *Multistage Inventory Models and Techniques*. Stanford University Press, Stanford, CA, 32–46.
- Donohue, K. L. 2000. Efficient supply contracts for fashion goods with forecast updating and two production modes. *Management Sci.* 46(11) 1397–1411.
- Entriken, R., G. Infanger. 1990. Decomposition and importance sampling for stochastic linear models. *Energy* 15(7–8) 645–659.
- Eppen, G. D., A. V. Iyer. 1997. Improved fashion buying with Bayesian updates. *Oper. Res.* 45(6) 805–819.
- Feng, Q., S. P. Sethi, H. Yan, H. Zhang. 2006. Are base-stock policies optimal in inventory problems with multiple delivery modes? *Oper. Res.* 54(4) 801–807.
- Fisher, M., A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* 44(4) 87–99.
- Fleckenstein, B. 2004. Intel's twin woes: Excess capacity, slack demand. MSN.com. (July 19). Retrieved August 18, 2010, <http://moneycentral.msn.com/content/P85429.asp>.
- Friend, I., M. Blume. 1975. The demand for risky assets. *Amer. Econom. Rev.* 65(5) 900–922.
- Fukuda, Y. 1964. Optimal policies for the inventory problem with negotiable lead time. *Management Sci.* 10(4) 690–708.
- Graves, S. C., H. C. Meal, S. Dasu, Y. Qui. 1986. Two-stage production planning in a dynamic environment. S. Axsater, C. Schneeweiss, E. Silver, eds. *Lecture Notes in Economics and Mathematical Systems, Multi-Stage Production Planning and Inventory Control*, Vol. 266. Springer-Verlag, Berlin, 9–43.
- Hausman, W. H. 1969. Sequential decision problems: A model to exploit existing forecast. *Management Sci.* 16(2) 93–111.
- Huang, X. 2008. Capacity planning in a general supply chain with multiple contract types. Unpublished doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge.
- Hurley, G., P. Jackson, R. Levi, R. O. Roundy, D. B. Shmoys. 2007. New policies for stochastic inventory control models—Theoretical and computational results. Working paper, Cornell University, Ithaca, NY.
- Intel Corporation. 2009. Annual report. Retrieved July 28, 2010, <http://www.intel.com/annuals.cfm>.
- Johnson, M. E. 2005. Dual sourcing strategies: Operational hedging and outsourcing to reducing risk in low-cost countries. C.-Y. Lee, H. L. Lee, eds. *Supply Chain Excellence in Emerging Economies*, Chap. 5. Springer Science+Business Media, New York, 113–133.
- Kanellos, M. 2003. Soaring costs of chipmaking recast industry. CNET News.com. (January 22). Retrieved August 18, 2010, <http://news.cnet.com/2009-1001-981418.html>.
- Leachman, R. C. 2007. The economics of speed. *Fab Engrg. Oper.* (November 20), [http://www.feomag.com/documents.asp?grID=12&d\\_ID=12](http://www.feomag.com/documents.asp?grID=12&d_ID=12).
- Lian, Z., L. Liu, S. X. Zhu. 2010. Rolling-horizon replenishment: Policies and performance analysis. *Naval Res. Logist.* 57(6) 489–502.
- Liu, L., M. Parlar, S. X. Zhu. 2007. Pricing and lead time decisions in decentralized supply chains. *Management Sci.* 53(5) 713–725.
- Liu, Q., G. van Ryzin. 2011. Strategic capacity rationing when customers learn. *Manufacturing Service Oper. Management* 13(1) 89–107.
- Moore, G. E. 1965. Cramming more components onto integrated circuits. *Electronics Magazine* (April 19), 114–117.
- Özer, Ö., W. Wei. 2006. Strategic commitments for an optimal capacity decision under asymmetric forecast information. *Management Sci.* 52(8) 1238–1257.
- Palaka, K., S. Erlebacher, D. H. Kro. 1998. Lead time setting, capacity utilization, and pricing decisions under lead time dependent demand. *IIE Trans.* 30(2) 151–163.
- Shapiro, A. 2008. Stochastic programming approach to optimization under uncertainty. *Math. Programming* 112(1) 183–220.
- Vaidyanathan, V., D. Metcalf, D. Martin. 2005. Using capacity options to better enable our factory ramps. *Intel Tech. J.* 9(3) 185–191.
- Van Mieghem, J. A. 2003. Capacity management, investment, and hedging: Review and recent developments. *Manufacturing Service Oper. Management* 5(4) 269–302.
- Wu, S. D., M. Erkoc, S. Karabuk. 2005. Managing capacity in the high-tech industry: A review of literature. *Engrg. Economist* 50(2) 125–158.
- Wu, S. D., K. G. Kempf, M. O. Atan, B. Aytac, S. A. Shirodkar, A. Mishra. 2010. Improving new-product forecasting at Intel Corporation. *Interfaces* 40(5) 385–396.
- Yazlali, Ö., F. Erhun. 2009. Dual-supply inventory problem with capacity limits on order sizes and unrestricted ordering costs. *IIE Trans.* 41(8) 716–729.
- Yazlali, Ö., F. Erhun. 2010. Managing demand uncertainty with dual supply contracts on capacity and inventory: A heuristic approach. Working paper, Stanford University, Stanford, CA.
- Yildirim, I., B. Tan, F. Karaesmen. 2005. A multiperiod stochastic production planning and sourcing problem with service level constraints. *OR Spectrum* 27(2–3) 471–489.