ELSEVIER

Int. J. Production Economics 70 (2001) 1–19

international journal of
production economics

www.elsevier.com/locate/dsw

Learning and process improvement during production ramp-up

Christian Terwiesch[a,*], Roger E. Bohn[b]

[a]Department of Operations and Information Management, The Wharton School, 1319 Steinberg Hall-Dietrich Hall, Philadelphia, PA 19104-6366, USA
[b]University of California, San Diego, CA, USA

Received 2 June 1999; accepted 18 October 1999

Abstract

Rapid product lifecycles and high development costs pressure manufacturing firms to cut not only their development times (time-to-market), but also the time to reach full capacity utilization (time-to-volume). The period between completion of development and full capacity utilization is known as production ramp-up. During that time, the new production process is ill understood, which causes low yields and low production rates. This paper analyzes the interactions among capacity utilization, yields, and process improvement (learning). We model learning in the form of deliberate experiments, which reduce capacity in the short run. This creates a trade-off between experiments and production. High selling prices during ramp-up raise the opportunity cost of experiments, yet early learning is more valuable than later learning. We formalize the resulting intertemporal trade-off between the short-term opportunity cost of capacity and the long term value of learning as a dynamic program. The paper also examines the tradeoff between production speed and yield/quality, where faster production rates lead to more defects. Finally, we show what happens if managers misunderstand the sources of learning.

Keywords: Yield; Ramp-up; Start-up; Learning curve; Experimentation

1. Introduction

Many high-tech industries are characterized by shrinking product lifecycles and increasingly expensive production equipment and up-front costs. The market window for selling many products has shrunk to less than a year in industries such as disk-drives and telecommunications. These forces pressure organizations to cut not only their development times (time-to-market), but also the time it takes to reach full production volume (time-to-volume) in order to meet their financial goals for the product (time-to-payback).

The period between the end of product development and full capacity production is known as *production ramp-up*. Two conflicting factors are characteristic of this period: low production capacity, and high demand. High demand arises because the product is still "relatively fresh" and might even be the first of its type. Thus, customers are ready to pay a premium price. Yet output is low due to low production rates and low yields. The production process is still poorly understood and, inevitably, much of what is made does not work properly the first time. Machines break down, setups are slow, special operations are needed to

* Corresponding author. Tel.: + 215-898-854.
E-mail address: terwiesch@wharton.upenn.edu (C. Terwiesch).

0925-5273/01/$ - see front matter
PII: S0925-5273(00)00045-1

correct product and process oversights, and other factors impede output. Over time, with learning about the production process and equipment, yields and capacity utilization go up (although in many industries they never reach 100%). Due to the conflicts between low effective capacity and high demand, the company finds itself pressured from two sides, an effect referred to as the "nutcracker" [1].

A recent example of the importance of ramp-up can be found in AMD's efforts to compete with Intel in the microprocessor market. AMD had several generations of product that were slow to ramp, leading to limited market acceptance and financial difficulties for AMD. More recently, Intel experienced problems ramping up the yield of its 0.18 micron version of the Pentium. Industry observers speculate that an effective ramp-up of AMD's K7 processor will allow AMD to compete in the high end segment of the PC market (Electronic Buyers' News, June 21, 1999).

In this article, we analyze the interactions among capacity utilization, yields, and yield improvement (learning) during ramp-up. Traditional learning-curve models implicitly assume that manufacturing performance increases with cumulative output from the plant, more or less independent of managerial decisions. This is clearly an oversimplification, and there is much that managers can do to affect the rate of learning [2]. We concentrate on deliberate learning through experiments such as engineering trials, which are controlled experiments using the production process as a laboratory. Such trials are essential for diagnosing problems and testing proposed solutions and process improvements. But they also use scarce production capacity. This creates a paradoxical trade-off between regular production for revenue and experimentation for learning. We formalize this intertemporal trade-off between short-term revenues and long term learning benefits in form of a dynamic program, and derive solutions for the cost, value, and level of experimentation.

The trade-off between short-term output and experiments, as well as more generally the phase of production ramp-up, is of substantial managerial importance. Launches of high-tech products are often either delayed or scaled back because of ramp-up problems. For example ramp-up problems in the production of video chips led to substantial losses during the launch of the Sega Dreamcast video console [3]. Similarly, pharmaceutical companies are struggling with ramping up the production of biotechnology-based drugs, leading to sales losses at the time when prices are at their premium [4]. This article models the complex dynamics of a new product's ramp-up, and assists decision making by providing concrete values for the cost and benefits of learning efforts. Specifically, we show that a misperception about the underlying drivers of learning can result in substantial financial losses over the lifecycle of a new product.

The remainder of this article is organized as follows. Section 2 provides more background on the assumptions of our model, and discusses several strands of related literature. Section 3 describes the type of production environments our analysis is appropriate for and presents a simple model that captures the interaction among capacity utilization, process knowledge, and yields. The analysis of this (static) model will be the basis for our dynamic model of learning and process improvement during production ramp-up, presented in Section 4. Our results are illustrated by several numerical examples in Section 5, where we show that different cost and demand situations call for different ramp-up strategies. Section 6 provides a summary, managerial implications, and future research directions.

2. Background

This paper draws on three strands of research, as it is about manufacturing *learning* during *ramp-up*, with *yields* the primary dependent variable. Each new product introduced into a factory must undergo a ramp-up. A new product's ramp-up can last a quarter of the product life cycle – several months for a hard disk drive, for example. During this period, yields and production rates gradually increase as learning takes place. Important types of learning typically include adjusting the process recipe, modifying tooling and equipment to reduce defects and downtime, and developing better and faster inspection methods.

Ramp-ups also occur when a new process or a new plant starts up. These are often more difficult and dramatic than new product ramp-ups, since many additional variables need to be learned about. Not all ramp-ups are successful, in either technical or business terms. Sometimes the plant is unable to raise yields to the breakeven level, or it takes so long that the product never earns enough revenue to repay its fixed costs.

Yields are an important state variable during ramp-up because they have a major effect on process economics and because low yields reflect gaps in process understanding and are closely linked to knowledge and learning [5]. The economic impact of low yields can be much larger than their impact on costs, since foregone revenue is usually a large opportunity cost during ramp-up [6]. Production speed and good output are also useful measures of progress during ramp-up. As we model in Section 3, the process manager often trades off yield and speed, for example when considering how hard to attempt to rework a bad unit before scrapping it.[1] Therefore, we will model the optimal trajectories of both yield and production rate over the course of a ramp-up.

There is little published research on ramp-ups, despite their ubiquity. Langowitz [7] conducts an exploratory study of ramp-up of four electronics products. Benfer [8] discusses the general problem of rapid ramp-up at Intel. Both emphasize the relationship between development and successful ramp-up. Clawson [9] discusses ramp-up in aerospace. Wasserman and Clark [10] document a problematic ramp-up of high-performance semiconductors, in which yields remained close to zero for months. Leachman [11] shows examples of semiconductor fabs taking years to raise their yields above 50 percent. As we model in Section 5, protracted or ultimately unprofitable ramp-ups can arise when managers assume that learning will occur automatically through experience, and therefore underinvest in deliberate learning through experimentation.

[1] In some assembly processes such as auto assembly it is economical to rework all defectives, and final yields are therefore very high. In this situation, first-pass yield or defect levels are a better measure of technological understanding and status during the ramp-up. For simplicity, this paper models situations where final yields are a good measure, such as semiconductors, disk drives, and parts fabrication processes.

2.1. Learning and experience curves

Although the importance of learning is widely recognized, many models of manufacturing and business strategy have focused on a single causal explanation, the cumulative volume of production. This is captured in the so-called *experience curve* model, surveyed critically in Dutton et al. [12]. This model postulates that per unit costs fall as the log of cumulative production [13]. Although this model may provide a good fit to costs ex post, that does not make it accurate or useful as a normative guide. "But in their current forms progress functions also have serious limitations. In offering cumulative volume as the only policy input variable, they fail to match the complex, underlying dynamics of firms' costs and imply that building cumulative volume is the only way to achieve progress. However, examination of progress-function studies reveals that sustained production often provides producers with opportunities to effect cost efficiencies that have little to do with cumulative volume [2]".

These criticisms are especially appropriate when looking at ramp-ups, where the central goal is to manage progress as rapidly as possible, and where a naive experience curve model would suggest that the rate and success of ramp-up are predetermined, completely predictable, and beyond managerial control.

Various researchers have gone beyond the experience curve to investigate learning processes in manufacturing in more detail, in an attempt to "open the black box" and derive managerially useful lessons. Several have done detailed investigations into how factory problems are solved and learning occurs, emphasizing activities by the engineers. Lapré et al. [14] show that many directed improvement projects in fact have zero or negative effects. They find that both sound theory and empirical validation by experimentation are needed before process changes are justified. Von Hippel and Tyre [15] examine information flows and other aspects of problem solving for a variety of new process introductions in plants.

Zangwill and Kantor [16] point out that learning can be viewed as happening in cycles, where the result of one cycle is the starting point for the next cycle. Each cycle can be viewed as removing some "waste" from the manufacturing system, whether that waste is defects, causing yield loss, waste time, causing slower production, or something else such as excess inventory. This provides a very general framework in which many different forms of process improvement can be modeled. As we discuss later, this cyclic model of learning fits many aspects of ramp-up, such as the diminishing returns to experimentation in any one learning cycle. In their model, the effort needed for each halving of waste requires a roughly constant effort, giving processes which approach asymptotically to a zero-waste condition. Stata [17] provides extensive documentation of halving times for many kinds of waste reduction in a semiconductor company. In our models, zero waste equates to 100% yield and production at 100% of maximum theoretical speed. We show in Section 4.4 that a constant amount of experimentation is indeed needed for each halving of waste, and that it is not optimal to keep the experimentation level the same all the way through ramp-up.

Many others have looked at learning at a more aggregate level, to determine what factors drive performance improvement. These include [18–25]. Most of these articles emphasize empirical fits to data rather than conceptual models. Mody [26] provides a model, which explicitly examines engineering effort as a driver of learning. Dorroh et al. [20] have a related model of make to order production, with a production function that takes knowledge and other resources as inputs. Knowledge is produced by explicit investment in learning, independent of production. They examine the effects of discounting and other parameters on the decisions of how much and when to produce and learn.

2.2. *What drives learning*?

Although our model emphasizes deliberate learning through experiments, we also allow for learning to take place through cumulative experience. Management directly sets the rate of experimentation and the production rate, subject to the constraint of machine capacity. In the terminology of Dutton and Thomas, experimentation is a form of *induced* (deliberate) learning, while production experience is *autonomous* (automatic) learning. Fig. 1 shows the flow of causality.

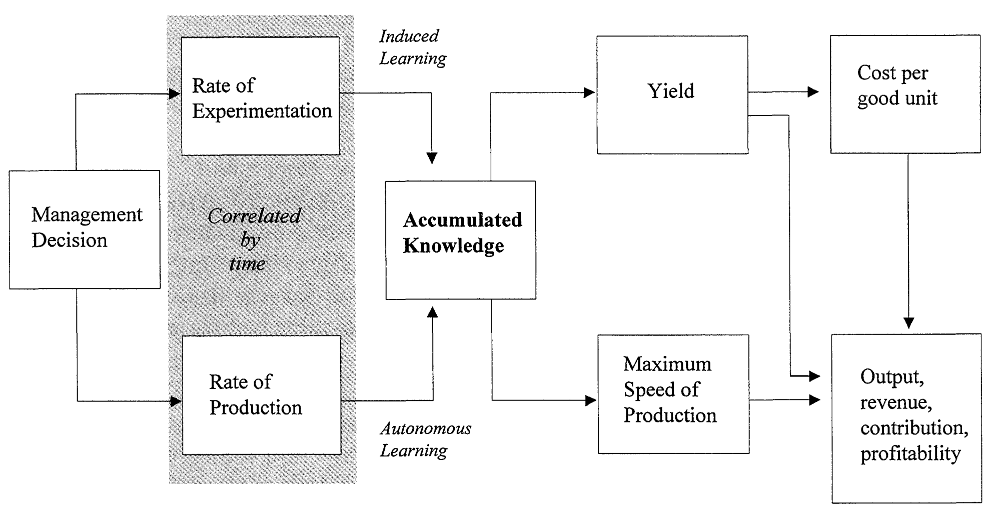

Fig. 1. Causes of learning and improvement.

Note that it may be difficult for an outside observer to know whether experimentation or experience is the principal driver of learning and thereby of improved performance. Both accumulated experiments and accumulated experience are correlated with time and therefore with each other. Hence, it is very difficult to use historical data to disentangle their effects, especially since experimentation is almost never carefully tracked. Therefore, we view most of the "experience curve" research, which purports to show that increased production leads to learning, as irrelevant to the question of what actually causes learning and how learning should be managed. In Section 5.4, we show what happens if the relative value of autonomous vs. induced learning is over-estimated.

If experimentation is a key driver of learning, what limits the rate of experimentation? Pisano [4] points to the scarcity of production capacity during ramp-up. Many yield ramp-ups involve strenuous debates about capacity allocation between process engineers, who are measured by yield, and production managers, who are measured by short-term throughput. For example semiconductor companies will restrict the number of "hot lots", i.e. expedited engineering trials, on the grounds that such lots cause a disproportionate reduction in production and increase in service times for normal production [27]. The rate of experimentation is not the only driver of learning, of course. For one thing, the *effectiveness* of experimentation varies dramatically depending on a variety of statistical and non-statistical issues [28].

The contributions of this article are as follows. First, we analyze the interaction between capacity utilization and yields, a trade-off of fundamental importance during production ramp-up. The model is far more detailed than any of the previous studies and thus provides a more micro-level analysis of ramp-up. Second, using dynamic programming techniques, we explicitly derive the cost and value of experimentation. These results support management in trading-off the short term opportunity cost of experimentation with the long-term value of increased processing capability. Finally, we explain a number of different ramp-up patterns that can be observed in various industries and suggest which ones are best under which circumstances.

3. Yield and output during ramp-up

Our model focuses on the production ramp-up of high-tech products, such as electronics. We define high-tech as meaning the company is on the cutting edge of what is currently understood in process engineering. High-tech products frequently experience high but rapidly falling prices, and the only opportunity to achieve higher than competitive prices is early in the product lifecycle. This forces management to bring the product to market long before the manufacturing process is fully understood. Production techniques start at low stages of knowledge and yield losses are still substantial.

During ramp-up, the goal is to raise both yield and production rate (starts per hour) as rapidly as possible. At each moment, there is a tradeoff between the two, as the likelihood of a defect is an increasing function of processing speed. There are many causes of such tradeoffs. Consider the operation of a robotic watch assembly line as described in [29]. Faster robot movement causes vibrations which decrease the precision of the assembly and thus increase the likelihood of a defect. Similar speed–precision–defect interactions occur in many automated placement and assembly operations. Similar issues apply for assembly or test operations performed by operators.

A second cause of speed vs. yield tradeoffs is rework. If there is a fixed capacity available for overall production, an increase in starts reduces the amount of capacity that can be used for rework. This reduces the number of rework loops that can be spent per defective item and thus, ultimately, final yields [6].

Third, because of process variability, allowing more work in progress (WIP) between operations increases total throughput. However, by Little's Law this raises average waiting time and thus the time between the occurrence of problems at upstream operations, and their detection at downstream test or inspection points. Once problems are detected, there is more bad WIP to be purged or reworked. Problem solving may also take longer, again lowering yields. Note that production variability is usually higher in ramp-ups, making this tradeoff especially pointed.

Fourth, consider the time spent for calibration, inspection and maintenance of equipment. These operations take time, which reduces production rates. However, badly calibrated or maintained machines will be more likely to produce defective parts.

Finally, many continuous and batch processes involve the application of power over time, such as baking, heat treating, and etching. The time–energy profile of such processes can be widely varied by adjusting temperatures, voltage, conveyor speed, and other parameters. A process can be optimized for raw speed by raising the power level and decreasing the time. However, this speed maximizing setting is usually not the quality/yield optimal setting, creating a tradeoff.

Each of these five explanations forces management to trade off an increased level of raw throughput against production yields. In the present article, we abstract from this detailed causality and develop a framework that is generalizable across various industrial settings. We will first develop a (static) model, analyzing the interaction between capacity utilization, processing capability, and yields. This model formalizes the starts vs. yields trade-off and is applicable beyond situations of production ramp-up. It will later allow us to show under what circumstances during production ramp-up management should focus on starts or on yields (Section 3.3). In Section 4, we use the same model as the starting point for exploring the trade-off between experiments and production.

3.1. Notation

Define time units such that it takes one unit of time to produce one unit of output, if the operation is executed at its maximum speed. In the presence of a speed versus yield trade-off, it might be beneficial to slow down the corresponding operation by a certain time x to $1 + x$ units of time per unit of output. Let Γ be the time available for production (e.g. machine hours) in a period. Given the definition of time units, this also corresponds to the maximum achievable output (theoretical capacity) of the process. For a fixed "level of care", x, chosen by management, the theoretical capacity is utilized at a percentage $u = 1/(1 + x)$.

Next, we have to describe the relationship between the number of units started into the process, $\Gamma/(x + 1)$, and yields. An increase in the operation time x will reduce the likelihood of a defect. Define $y(x, \alpha)$ as the yield level as a function of x. This yield level is jointly determined by the operation time x and a parameter $\alpha > 0$ which measures processing capability. The higher the processing capability α, the more the process can be accelerated without major yield losses: $\partial y(x, \alpha)/\partial\alpha > 0$. We assume diminishing returns of the extra operation time x, so that $\partial y(x, \alpha)/\partial x > 0$ and $\partial^2 y(x, \alpha)/\partial^2 x < 0$. Output is then starts times yields or $y(x, \alpha)(\Gamma/(x + 1))$.

Throughout the article, we assume that capacity is a binding constraint. This is characteristic of production ramp-ups since the product is still relatively fresh and thus in strong demand, while output is restricted as we will see. All units produced can be sold at a selling price $\tilde{p}$, and the variable cost per start (e.g. raw material) is c. Before we turn to a dynamic version of the model, with learning (increase in processing capability α) or falling prices, we need to develop some simple insights about the static trade-off between starts and yields. Looking at one period in isolation, the operation time x is chosen to maximize the contribution (sales minus variable costs), which we can write as

$$\pi(\tilde{p}, \alpha, x, c) = \tilde{p}y(x, \alpha)\frac{\Gamma}{x + 1} - \frac{\Gamma}{x + 1}c.$$

3.2. The starts versus yields trade-off

Good output is not necessarily a monotonic increasing function in the number of starts. Starting too many units can disturb the production process so badly that not only yields fall, but even the net number of good units produced decreases. Contribution falls even more than good output since the contribution measure π also takes the costs of a start into account.

To simplify analysis, we now assume a specific functional form for the relationship among yields, processing capability, and operation time. Define yields as $y(x, \alpha) = y_0(1 - 1/\alpha x)$, which – consistent with our argument above – shows that the operation time x reduces the likelihood of a defect, but with diminishing returns. The parameter y_0 captures a base yield which is independent of the speed of the operation and cannot be improved, such as yield problems in operations that are downstream to the bottleneck production line. Without loss of generality, we discount the selling price by the base yields, i.e. define $p = \tilde{p}y_0$. Good output per period is now given by $y_0(1 - 1/\alpha x)(\Gamma/(x + 1))$ and the per period contribution is

$$\pi(p, \alpha, x, c) = p\left(1 - \frac{1}{\alpha x}\right)\frac{\Gamma}{x+1} - \frac{\Gamma}{x+1}c.$$

Let x^*_{out} be the operation time that maximizes good output. It is characterized by the balance between the marginal gains from higher quality of one particular unit (increased likelihood that an item started becomes good output) and the marginal losses resulting from a lower overall production rate. An additional unit started at a high level of utilization is not only likely to be defective itself, it also forces an increased processing speed on *all* other items, making them more likely to be defective as well. Thus, an increase in utilization is connected with a decrease in yields, and pushing utilization beyond $u^*_{\mathrm{out}} = 1/(x^*_{\mathrm{out}} + 1)$ actually decreases the overall output. At this point the effective capacity decreases.

In order to calculate the contribution optimal level of operation time x^*_{cont}, we need to take the costs per start into account, as well as the selling price. The general optimal solution is characterized by the balance between the marginal gains from higher quality of one particular unit (increased likelihood that an item started can be sold) and the marginal losses resulting from a slower overall production. In terms of Fig. 2, this yields a downward adjustment of utilization. Thus, the contribution optimal solution has higher yields and lower utilization than the output optimal solution. We assume $p/c > 1$, i.e. prices adjusted for downstream yield losses are high enough to cover the variable cost of production. Proposition 1 formalizes these ideas.

Proposition 1 (Static model). *The contribution optimal solution x^*_{cont} and the output optimal solution x^*_{out} have the following properties*:

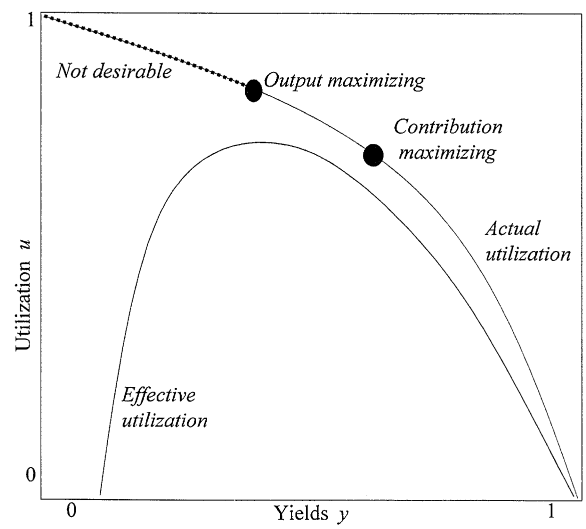

Fig. 2. Uitization versus yields ($y_0 = 1$).

- *Both the output maximizing operation time and the contribution maximizing operation time are strictly positive, i.e. $x^*_{\mathrm{out}} > 0$ and $x^*_{\mathrm{cont}} > 0$. As a result of this $u^*_{\mathrm{out}}, u^*_{\mathrm{cont}} < 1$, which corresponds to a deliberate under-utilization of the capacity.*
- *The output maximizing operation time x^*_{out} and the corresponding contribution level Π^*_{out} are given by*

$$x^*_{\mathrm{out}} = \frac{1 + \sqrt{1+\alpha}}{\alpha},$$

$$\Pi^*_{\mathrm{out}} = \pi(p,\alpha, x^*_{\mathrm{out}},0) = \Gamma p \frac{1}{1 + (2/\alpha)(1 + \sqrt{1+\alpha})}. \quad (3.1)$$

- *The contribution maximizing operation time x^*_{cont}, the corresponding yield level $y(x^*_{\mathrm{cont}}, \alpha)$, and the resulting contribution level Π^*_{cont} are given by*

$$x^*_{\mathrm{cont}} = \frac{1 + \sqrt{1 + \alpha - c\alpha/p}}{\alpha(1 - c/p)},$$

$$y^*_{\mathrm{cont}} = y(x^*_{\mathrm{cont}}, \alpha) = \frac{\sqrt{1 + \alpha - c\alpha/p} + c/p}{\sqrt{1 + \alpha - c\alpha/p} + 1}, \quad (3.2)$$

$$\Pi^*_{\text{cont}} = \pi(p, \alpha, x^*_{\text{cont}}, c)$$
$$= \Gamma\alpha\frac{(p-c)^2}{p}\frac{1}{2\sqrt{1+\alpha-c\alpha/p}+2+\alpha-c\alpha/p}.$$

- *The contribution maximizing operation time x^*_{cont} decreases with the selling price p being increased relative to cost c (p/c increases). For large values of p/c the contribution optimal solution approaches the output optimal solution: $x^*_{\text{cont}} \to x^*_{\text{out}}$.*

Proof. For easier readability, all the proofs are given in the appendix.

The first part of Proposition 1 shows the difference between utilization and effective utilization: it is both contribution and output optimal not to operate the production line at its maximum speed. Pushing utilization above u^*_{out} is not beneficial, as the yield losses more than offset the gains from starting more units. At this point, the effective utilization of the plant is maximized.

Proposition 1 also shows how the optimal operation times x^*_{cont} and x^*_{out} depend on the various parameters, especially the processing capability α. Yields and contribution can also be written as functions of α. The last point in Proposition 1 states that a decrease in selling price p will – everything else equal – reduce the number of starts and increase the resulting yields. Thus, with falling prices, the production line needs to put an even higher emphasis on quality.

Finally, it is interesting to observe the difference between (3.1) and (3.2). For the special case where the cost per start $c = 0$, the two are identical. For $c > 0$, utilization is adjusted downwards in favor of yields. This confirms the intuition generated by Fig. 2. Thus, a simple corollary of Proposition 1 is that $y^*_{\text{out}} \leqslant y^*_{\text{cont}}$ and, for the corresponding utilization levels, $y^*_{\text{out}} \leqslant y^*_{\text{cont}}$.

3.3. Yield emphasis versus volume emphasis

Consider a sequence of periods similar to the one described above. The only difference between each period is the processing capability α: over time, the organization learns more about its production process, which corresponds to an increase in α. Note that an increase in α allows for higher yields at the same level of starts, or more starts at the same level of yields. In this section, we are not explicit about *how* the learning occurs. It might be driven by volume, by an organizational learning effort, or by time alone.

The result of these changes is a sequence of models similar to the static model described above. This constitutes the first step toward a "dynamic ramp-up problem". A natural question to ask in this model is: What should the plant do with its increased processing capabilities, produce more or further increase yields (at the cost of output)? To illustrate this trade-off, we extend Fig. 2 by showing various levels of α. Each of the points on the additional lines corresponds to a pair of yield and utilization level, i.e. a set of (y^*_t, u^*_t), that is computed using (3.2) for a changing level of processing capability α. We define this path of utilization/yield combination as the ramp-map. This is summarized in Fig. 3. Let a *yield-emphasizing ramp* be a ramp with high initial yields (relative to utilization, u^*/y^* is small) where the learning is used to increase utilization. Let a *utilization-emphasizing ramp* be a ramp with low initial yields (relative to utilization, so u^*/y^* is large) with high initial utilization. With increasing processing capability α, yields are increased.

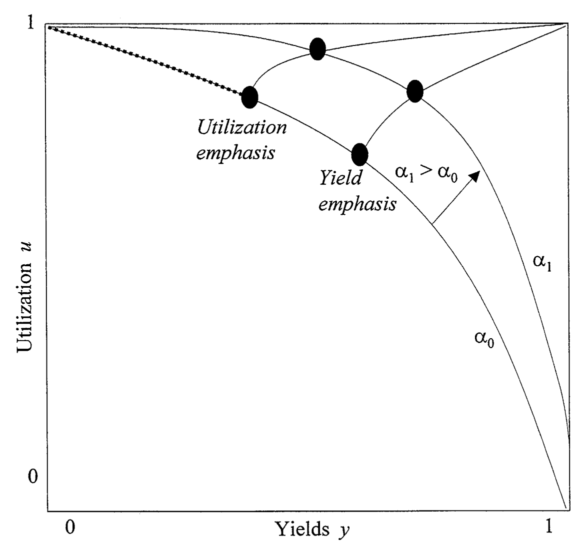

Fig. 3. Increasing processing capability: the ramp-map ($y_0 = 1$).

Proposition 2 (Ramp-map). *If learning occurs exogenously to the model, i.e. the operation times x_t have no impact on any future α_i, then the ramp-map has the following properties*:

- *Long-run behavior*: *for* $\alpha \to \infty$, $y^* \to y_0$ *and* $u^* \to 1$ *and as a result* $u^*/y^* \to 1/y_0$.
- For small α, u^*/y^* *is an increasing function in* p/c. *Large values of* p/c *favor a utilization emphasizing ramp.*
- *For large values of* p/c, $u^*/y^* \to u^*_{\text{out}}/y^*_{\text{out}} = \alpha/(1+\alpha)y_0$, *which characterizes the maximum possible utilization emphasis. The ramp-map above this path is empty.*

Proposition 2 is interesting in several ways. First, we see that regardless of cost per start c and selling price p the long-run behavior for increasing levels of processing capability α is always perfect yields ($y^* \to y_0$) and 100% utilization ($u^* \to 1$). This provides the end-point of the ramp-map. Second, the ratio u^*/y^* helps us to further specify the location of the start-point. For large ratios of price to cost p/c, following Proposition 1, the only focus is on output, thus u^*/y^* is maximized. At this point, the ratio between utilization and yields is characterized by $\alpha/(1+\alpha)$. For smaller values of p/c, there is an extra focus on yields, which means the path through the ramp-map shifts to the lower right.

To illustrate the implications for different industrial processes, compare disk drive assembly and semiconductor wafer production. For drives, the costs of raw material are very close to the market price of the finished good, which makes scrapping a drive extremely expensive. Even if first-pass yields are low, rework is used intensively to reach high final yields. Rework corresponds to an extended operations time, x_t, in our model. Proposition 2 predicts a strong yield emphasis in the ramp-up, which is consistent with empirical research in this industry [6].

For semiconductors, the main cost driver is the equipment, rather than the raw material. Thus, the value of the finished wafer is many times its cost per start and scrapping a defective wafer loses little in terms of direct cost. Following Proposition 2, the main focus is on output and the ramp-up is characterized by extremely low initial yields. Various studies in the semiconductor industry show that production can sometimes continue at low yields for a prolonged period, if competition is low and prices high (e.g. [11]).

4. Learning in ramp-up

In this section, we extend our analysis presented in Section 3 by explicitly modeling the sources of learning. We do this by adding a second managerial decision variable, learning efforts, which will come in the form of controlled experiments. The results of Proposition 1 allow us to replace the decision variable "operations time", x_t, with its contribution optimal level x^*_{cont} and thereby to focus our discussion purely on learning efforts.

The benefit of learning efforts lies in an increased knowledge about the production process, which is captured in the processing capability parameter α in the model. However, learning also has drawbacks. First, experiments consume capacity which could otherwise be used for regular production (e.g. setups of experiments, experimental output might not be salable, disruption from expediting experimental "hot lots"). Second, experiments are a deviation from what is currently believed to be the optimal process control. This lowers yields (e.g. in case of trying out a new recipe).

This creates a dual role of the production process: it not only produces salable output, it also provides the environment for conducting experiments [28]. Looking at the per period contribution Π^*_{cont} in (3.2), we see that the overall contribution is proportional to the available capacity Γ. So, spending more machine time for experimentation creates an opportunity cost of lost regular production.

The difficulty of choosing between these conflicting goals has also been observed by Pisano [4]. Based on his study of process development in the biotechnology industry, Pisano reports: "Setting aside capacity and time for development runs, training operators in experimental design so that they will be better able to work with R&D Scientists, and implementing appropriate controls and documentation to enable experimentation without threatening the quality of commercial production are some of the prerequisites for making plants

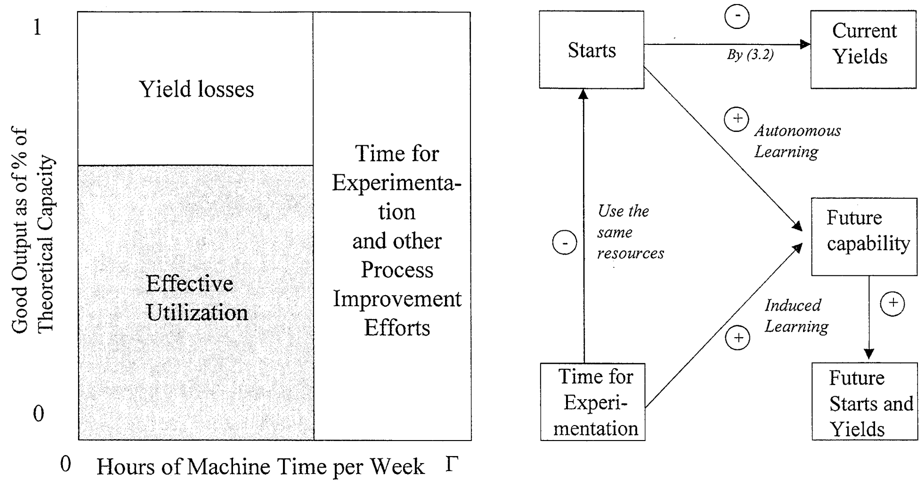

Fig. 4. Effects of experimentation.

producers of knowledge as well as of products". These conflicting goals and their interaction are illustrated in Fig. 4.

4.1. A mathematical model

In order to formalize the tradeoff between current production and learning, we define z as the fraction of the overall processing capacity Γ that is used for experiments. Together with the operations time (level of care) x, the experimentation time represents our second managerial decision variable. The overall output of the production process is then

$$(1-z)y(x,\alpha)\frac{\Gamma}{x+1},$$

where $0 \leqslant z \leqslant 1$.

Process improvement corresponds to an increased value of processing capability α. Consider the case where we improve the process by increasing the processing capability from α_{old} to a higher level $\alpha_{\text{new}} := \lambda\alpha_{\text{old}} > \alpha_{\text{old}}$. The magnitude of the improvement λ from one period to the next will depend on the two learning mechanisms, learning by doing and learning by experimentation. Similar to a production function, we assume

$$\lambda = \beta_1^z \beta_2,$$

where β_1 captures the relative importance of learning by experimentation to learning by doing and β_2 the learning rate of learning by doing in itself.[2] In the absence of experimentation ($z = 0$), process improvement is solely a result of learning by doing ($\lambda = \beta_2$). Learning by doing is thus driven by the cumulative time (e.g. machine hours) the production line has been processing the new product. β_1 measures how much additional progress would occur if the line were dedicated to experiments for the *whole* period.

We now extend our static model of Section 3 to a T-period dynamic model. The previously introduced variables α, x and p for processing capability, operations time, and price are now indexed by time, i.e. α_t, x_t and p_t. Over the periods, prices fall at a rate of δ_p per period, and future cash flows (revenues and cost) are discounted with a factor δ_d. As this paper focuses on the dynamics *inside* the plant, we view this price fall as exogenous to our model.

In every period, management needs to decide on both speed of the line (in form of the operations time x_t) and the fraction of capacity used for experi-

[2] This formulation of learning is equivalent to assuming $\lambda = \gamma_1^z \gamma_2^{1-z}$, where γ_1 is defined as the absolute importance of learning by experimentation and γ_2 as the absolute importance of learning by doing.

mentation z_t. Whereas the costs of an experiment are additive over time (the first 10% of production capacity has the same opportunity cost as the last 10%), the value of an experiment typically is not. For example, spending 20% of the capacity for experimentation in one period will yield a smaller increase in processing capability α_t than spending 10% in the current period and 10% in the subsequent period.

There are several reasons for diminishing returns to experimentation within one period. First, experimentation is normally done in cycles, rather than in one single batch [16]. It is more effective to wait for the results of one experiment before formulating ideas which become the basis of the next cycle [30]. Second, although capacity is a key input for experimentation, there are others, specifically engineering time. Third, conducting too many experiments at the same time increases the noise in the process, which makes it harder to learn. Thus, process improvement returns will be reduced, if management decides to "jam" all experimentation efforts in one or few periods.

To capture these diminishing returns to experimentation, we adjust the learning rate of experimentation β_1 by multiplying it with a factor Θ^z, which is decreasing in the amount of experimentation, z, carried out in that period ($\Theta < 1$). This reduces the effective learning rate of experimentation to $\beta_1 \Theta^z$, and therefore results in an overall process improvement of

$$\lambda := \beta_2 [\beta_1 \Theta^z]^z = \beta_1^z \Theta^{z^2} \beta_2. \tag{4.1}$$

For small values of experimentation time z, the factor Θ^z is close to one, i.e. for the first units of experimentation, the marginal gains are close to the ideal learning rate β_1. In the extreme case of $z = 1$, only Θ percent of the ideal learning rate is achieved.

For example, we can write $\alpha_2 = \alpha_1 \beta_1^{z_1} \beta_2 \Theta^{z_1^2}$ and $\alpha_3 = \alpha_2 \beta_1^{z_2} \beta_2 \Theta^{z_2^2} = \alpha_1 \beta_1^{z_1 + z_2} \beta_2 \Theta^{z_1^2 + z_2^2}$. We can see immediately that the effect of Θ is minimized if the experiments are evenly spread over the periods. Therefore, if cost and value of the experiments were constant over time, it would be optimal to have $z_1 = z_2$. However, in the presence of changing cost and value, the optimal solution has to be chosen based on the overall optimization problem.

4.2. Dynamic programming formulation

As the operation time x_t has no effect on any future processing capabilities α_{t+i}, $i = 1, \ldots, T-t$, we can *decompose* the overall optimization problem. For every period, we choose the optimal operation time x^*_{cont} given current capability α_t, which is given by Proposition 1. Anticipating that we will choose this optimal operation time, we then are left with finding the optimal learning effort z_t for every period. This requires the analysis of a dynamic program with the processing capability α_t as the state and the experimentation level z_t as the decision variables,

$$\Pi^{\text{total}}(z_1, \ldots, z_T) = \max_{z_1, \ldots, z_T} \sum \delta_{\text{d}}^{t-1} (1 - z_t) \pi_t(\alpha_t), \tag{4.2}$$

where α_t is connected to α_{t-1}, z_{t-1} and x_{t-1} by the improvement rate given in (4.1) and the immediate pay-offs per period are defined as $\pi_t(\alpha_t) = \pi(\delta_{\text{p}}^{t-1} p,\ \alpha_t, x_t^*, c)$. We define

$$F_t(\alpha_t) = \max_{z_t} \{(1 - z_t)\pi_t(\alpha_t) + \delta_{\text{d}} F_{t+1}(\alpha_t \beta_1^{z_t} \Theta^{z_t^2} \beta_2)\}, \tag{4.3}$$

$$F_T(\alpha_T) = \pi_T(\alpha_T) + \text{Terminal value}(\alpha).$$

For the last period T, there is no direct value of experimentation in our model. However, higher processing capability beyond period T typically has some value, e.g. in lower unit costs for the residual product lifecycle or in increased knowledge for future product generations. For a general period t, we can see from (4.3) that the first part of $F_t(\alpha_t)$ is decreasing linearly in the experimentation time z_t. As we will show more formally below, the returns to experimentations are marginally decreasing, which makes (4.3) a sum of two concave functions. Thus, the optimal solutions $z_1^*, \ldots, z_T^*$ are uniquely identified and can be computed by backward induction.

4.3. Costs of an experiment

In order to understand how much of its scarce production capacity the organization should invest in experimentation, we need to understand the cost and benefit of one unit of experimentation time. At first sight, the analysis looks quite simple: costs of experimentation are given by the opportunity costs of not producing and the benefits of experimentation are given by the increased process knowledge that we have already formalized in (3.2).

Although this intuition is correct, the actual analysis is more complicated, as both opportunity cost and value of increased process knowledge are functions of time and current processing capability. Time is important as it relates to selling price and thus to the opportunity cost of not producing. The current knowledge is important as it influences how much is still to be learned from an experiment as well as the opportunity cost. Having the line not produce is cheaper at a low level of knowledge than at a high level of knowledge.

Let $k(t,\ \alpha_t, z_t) = \mathrm{Max}\{z_t \pi_t(\alpha_t),\ 0\}$ denote the cost of experimentation if a fraction z_t of capacity is used for experimentation at time t and state α_t. We can then prove the following proposition.

Proposition 3a (Cost of an experiment). *The cost $k(t,\ \alpha_t, z_t)$ of doing z_t units of experimentation in time t and state α_t is an increasing function of the processing capability α_t, and a decreasing function of time t.*

Proposition 3a means that the cost of experimentation can, over the periods $1, \dots, T$, go either up or down. Increasing levels of processing capability α_t bring the opportunity cost up, as at a high α the production line can produce more and at higher yields. However, falling prices, which also drive the opportunity cost, are pushing the opportunity cost down. As over time the organization increases its processing capability, α_t and t move together, allowing the cost of experimentation to go either up or down.

Before we turn to the value of an experiment, we compute the costs of increasing the processing capability from α_t to $\lambda\alpha_{t+1}$. This extends Proposition 3a which derived the costs of experimentation per unit of experimentation time.

Proposition 3b (Doubling α). *Increasing the processing capability by a factor λ carries the following costs:*

- *the amount of experimentation required as a function of λ is given by*

$$z(\lambda) = -\frac{1 \log \beta_1}{2 \log \theta} - \sqrt{\frac{\left[\frac{\log \beta_1}{\log \theta}\right]^2}{4} - \frac{\log \frac{\lambda}{\beta_2}}{\log \theta}}, \tag{4.4}$$

- *the corresponding cost is given by $k(t,\ \alpha_t, z(\lambda))$.*

We can see that although the required amount of experimentation to increase the processing capability from α to $\lambda\alpha$ is independent of time, the associated costs $k(t,\alpha_t,\ z(\lambda))$ are not. This is a result of Proposition 3a. Eq. (4.4) shows the relationship between experimentation time and the learning parameters β_1, β_2, and θ. These parameters provide an upper bound of how much improvement can be achieved within one period. We can see that a doubling of the processing capability α becomes more and more expensive as α increases, and has to be justified by large benefits. These benefits are now analyzed in greater detail.

4.4. Value of an experiment

The value of an experiment depends on three factors: how far the product has advanced in the lifecycle (the period t), the current processing capability α, and the level of experimentation z. Similar to the cost of an experiment, both time t and processing capability α (the two state variables of the DP) have an influence on the value. The value of a unit of experimentation depends on how much additional experimentation is conducted in that particular period. We define $v(t,\ \alpha_t, z_t)$ as the value of doing z_t units of experimentation in time t and state α_t.

We can express $v(t,\ \alpha_t, z_t)$ using the recursive definition of $F_t(\alpha_t)$ in (4.3):

$$v(t,\ \alpha_t, z_t) = \delta_{\mathrm{d}}[F_{t+1}(\alpha_t \beta_1^{z_t} \Theta^{z_t^2} \beta_2) - F_{t+1}(\alpha_t \beta_2)]. \tag{4.5}$$

Doing z_t units of experimentation will bring the processing capability at period $t+1$ from α_t to $\alpha_t \beta_1^{z_t} \Theta^{z_t^2} \beta_2$. The net present value of this is given by $F_{t+1}(\alpha_t \beta_1^{z_t} \Theta^{z_t^2} \beta_2)$. If we decide to not invest into process improvement, the new state will be $\alpha_t \beta_2$ with the associated net present value of $F_{t+1}(\alpha_t \beta_2)$. Therefore, we define the value of an experiment as the net present value difference between two scenarios, corresponding to two different α-trajectories, starting at period $t+1$. The first scenario is based on optimal experimentation in period t, the second scenario forces $z_t = 0$. Note that the two scenarios are likely to have different experimentation policies beyond period $t+1$.

Proposition 4 (Value of experiment). *The value of increasing the processing capability from α to $\lambda\alpha$ goes down in α (diminishing physical returns) as well as in t.*

Proposition 4 shows the value of an experiment falls over time. There are three reasons for this. First, the residual lifetime of the product, to which the new knowledge might be applied, is shrinking. Second, the value of an experiment falls as prices fall. This makes early knowledge more valuable than late knowledge.

Third, in addition to those two effects, that are purely driven by calendar time, the value of an experiment falls as α increases. To illustrate this, compare two situations. In the first situation, the processing capability α is small and yield losses are still high. Increasing α at this point has substantial leverage, as there are still plenty of opportunities for improvement. In the second situation, the process is close to being perfect. Both utilization, u, and yields, y, are close to one, so an improvement in α, even if of substantial size, will not have much impact on the bottom line.

This is similar to the argument of Zangwill and Kantor [16]. Instead of looking at process yields, they make waste (defined as 1-yields) their key variable. The authors argue that the effort required for a proportional waste reduction is constant. For example, getting yields from 50 to 75% and getting them from 75 to 87% both correspond to a halving of waste, and require the same effort, but the first improvement is more valuable then the second. This is consistent with our model, if we define waste as $1/\alpha x$, and Proposition 3b, which requires a constant effort for each proportional change in α. Thus, there exists a constant α-improvement (in form of a multiplier) for each halving of waste.

We now turn to a series of numerical examples, which illustrate how qualitatively different optimal behavior can arise from different market, technological, and learning parameters.

5. Numerical illustrations

We solve a number of numerical examples in this section. They shed light on the structure of the optimal solution to the general profit maximizing problem as stated in (4.2). Consider an example of a low price to cost ratio process, such as disk drive assembly. The initial price is $p = \$3$/unit, prices fall at $\delta_p = 0.95$ per period (month), the discount factor is $\delta_d = 0.98$. We assume cost per start to be $c = \$1$/unit and consider only the final assembly, so there are no substantial yield losses further downstream ($y_0 = 1$). Let the initial processing capability be $\alpha_1 = 1$ and the learning rates be $\beta_1 = 2.80$ and $\beta_2 = 1.01$ for learning by experimentation and learning by doing respectively. The overall capacity available for production *and* experimentation is $\Gamma = 1000$ units per month, and the lifecycle is $T = 12$ months.

5.1. High experimentation capability

To begin with, consider the case where the experimentation capability is high, i.e. $\theta = 1$. Engineers can conduct a large number of experiments and still get the maximum learning out of each of them. We can compute the cost of an experiment in the first period using Proposition 3a. With no experimentation, the optimal first period profit is $\pi_1 = 25.4$. Thus, each percent of experimentation time creates an opportunity cost of $k(1, \alpha_1, 0.01) = 0.254$. The value of experimentation is driven by the future periods' increased capability.

Fig. 5 plots the cost and value of experimentation for this specific case. Over the first four periods, the value of complete experimentation exceeds its cost indicating that the full period should be spent on experimentation. This changes from period 5 onwards. As the processing capability α_5 is substantially higher than earlier, the opportunity cost goes up to $k(5, \alpha_5, 0.01) = 1.04$ per percentage experimentation time. At the same time, the value of the experiment has decreased for the reasons discussed

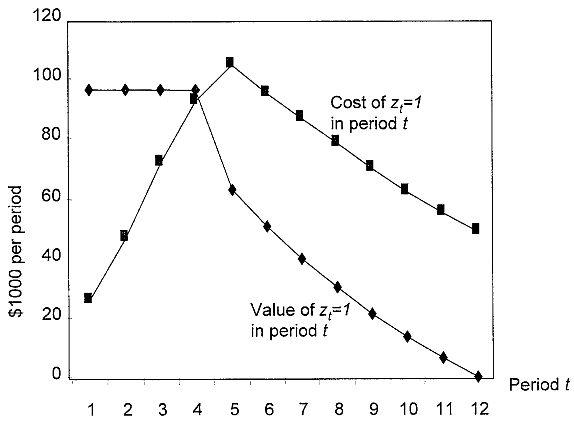

Fig. 5. Cost and value (in dollars) of experimentation.

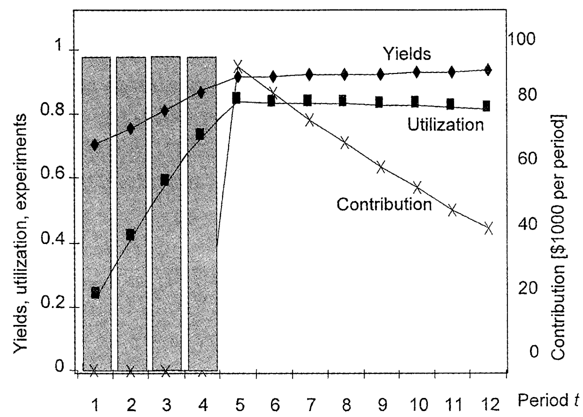

Fig. 6. Optimal solution for $\theta = 1$; the bars indicate the optimal z_t.

in connection with Proposition 4. First, there are fewer periods left to which the additional knowledge can be applied. Second, because of the physical diminishing returns, a further increase in processing capability has less value.

As a result, no time is spent on experimentation in the fourth period and beyond. Fig. 5 also shows that experiments are inexpensive in the beginning and in the end of the product lifecycle, but most costly in the middle.

Fig. 6 summarizes the optimal solution. The four bars indicate the optimal experimentation policy: full experimentation at the start, then none. Fig. 6 also shows yields, utilization, and per period contribution. We see that the initial focus of the plant is on yields (start at 72%) rather than utilization (start at 26%).[3] This yield emphasizing ramp is a result of the relatively small price to cost ratio (initially 3 : 1).

5.2. Low experimentation capability

Next, consider the case of lower experimentation capability, e.g. $\theta = 0.5$. The cost per unit of experimentation is the same in the first period, but its value goes down drastically. Spending $z_1 = 1$ units on experimentation now only results in a second period capability of $\alpha_2 = 1.46$. This is driven by the sub-additivity argument. The lower α_2 also translates into a lower second period opportunity cost of the coming periods ($k(2, \alpha_2, 0.01) = 0.31$, $k(3, \alpha_3, 0.01) = 0.36$, $k(4, \alpha_4, 0.01) = 0.40$). The decrease in $k(2, \alpha_2, 0.01)$ together with the reduced first period value of experimentation creates an incentive to move some experiments from period one to period two.

Fig. 7 shows the optimal solution for the case $\theta = 0.5$. Again, the emphasis of the ramp is on yields, rather than utilization. As opposed to the previous example (and Fig. 6), production starts in the first period, so the plant is actually producing at the initial yields of 72%. Fig. 7 demonstrates the harsh economic reality that most companies face during ramp-up. Given its low learning capability captured in $\theta = \frac{1}{2}$, it is not until period 9 (75% into the lifecycle) that the plant reaches its maximum contribution. However, rapidly falling prices quickly erode even the remaining 25% of the lifecycle, so that the time that can be used to pay back development expenses is extremely short.

These first two examples have illustrated the importance of the sub-additivity parameter θ. The first example is similar to a production ramp-up on a pilot line. The market introduction of the product is delayed (despite falling prices) and all the capacity is used for process engineering. In the second example, the product is introduced to the market earlier and process engineering is spread out over

[3] Note that these values are "not realized", as the complete first periods are dedicated to experimentation.

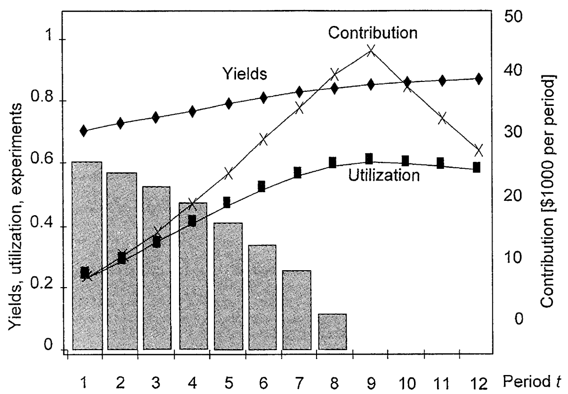

Fig. 7. Optimal solution for $\theta = \frac{1}{2}$; the bars indicate the optimal z_t.

several periods. Although this approach allows for some early profits and high prices, it also forces management to run the production line at low yields and utilization. As expected, the overall profit in the first example ($\pi = 522.5$) exceeds profits in the second one ($\pi = 296.3$).

5.3. Rapidly falling prices

Next, consider a situation of rapidly falling prices. Suppose R&D has come up with a radical new product that is the first of its kind. For the first period, we can charge a monopoly price of $p = 5$. Afterwards, competitors enter the market and prices drop sharply to $p = 3$, and from then onwards fall at $\delta_p = 0.95$. All other parameters are identical to the second example above. This example is interesting as it demonstrates that the capacity dedicated to experimentation should not necessarily decrease over time. The cost of experimentation in the optimal solution are given by $k(1, \alpha_1, 0.01) = 0.58$ and $k(2, \alpha_2, 0.01) = 0.28$. Compared to the second period, prices are high in the first period, which drives up the opportunity cost of not producing. The initial processing capability $\alpha_1 = 1$ is sufficiently high to create profits. This picture changes in the second period. As little time was spent on experimentation in the first period, $\alpha_2 = 1.39$, which is not a substantial increase. However, given the drastic fall in prices, management now faces a situation where the selling price is substantially lower than before. This price level together with the current processing capability makes experimentation now less expensive, yielding a second period experimentation of $z_2 = 0.57$. From then onwards, the cost of experimentation follows the path we have seen in the first two cases: an initial increase because of the increased capability and a long-term decrease because of falling prices. The optimal level of z_2 to z_{12} is decreasing in time.

Taking the above examples together with Propositions 3 and 4, we can postulate three types of solution:

1. *Virtual pilot line*: The introduction of the product is delayed and all available capacity is used for experimentation ($z_t^* = 1$ for $t = 1, \ldots, n < T$). This approach is optimal if (a) prices are falling at a modest rate (δ_p is low), (b) high experimentation capability (weak sub-additivity: θ is close to 1).
2. *Mix of experiments and production*: Experiments are spread out over several periods, but more and more of the production capacity is used for regular production ($z_t > z_{t+1} > 0$). This approach is optimal for low experimentation capability.
3. *Delayed experimentation*: The time spent for experimentation is larger in the second period than in the first period ($z_2 > z_1 > 0$). This approach is optimal if prices fall faster in the beginning than they do later.

5.4. Knowing the sources of learning

The above examples assume that management is fully aware of the true sources of learning. This includes both the relative magnitude between β_1 and β_2 and their absolute magnitude. Following our discussion in Section 2, this is frequently not the case. The importance of "learning by experience" is frequently overestimated compared to the importance of controlled experiments. For example, Lapré et al. [14] study learning projects in a mature wire drawing plant, and found that all non-experimentation based improvement projects, which accounted for the majority of projects, had no effect or even a negative effect on waste. Although the environment was different from a high-tech ramp-up, this shows that management is frequently unaware of the "true" sources of learning and specifically underestimates the importance of learning by experimentation.

Consider a situation similar to the example of Section 5.1. The initial processing capability is $\alpha_1 = 1$ and the learning rates are $\beta_1 = 2.8$ and $\beta_2 = 1.01$. However, these underlying parameters are not known by managers, who have estimates of the learning rates in form of $\hat{\beta}_i$, $i = 1, 2$. Let $\hat{\beta}_1 = 2$ and $\hat{\beta}_2 = 1.5$, which corresponds to an overestimation of experience versus experimentation. Based on this assumption, management chooses the experimentation times to maximize lifecycle contribution according to (4.2). This yields lower than optimal z_t and a total contribution of $\pi = 235$. If management had followed the true optimal policy the discounted contribution would have been $\pi = 296$. In other words, 20% of the potential contribution is lost because of an incorrect estimate of the learning rate.

Next, consider the reverse case where the real learning rate is $\beta_1 = 2.8$, however engineering over-estimates the importance of experimentation, with $\hat{\beta}_1 = 5$. As a result of this, more time is allocated to experimentation than optimal. Instead of getting a contribution of $\pi = 296$, the product now only reaches contributions of $\pi = 267$.

Two remarks clarify these examples. First, in both of them management was aware that experimentation yields higher improvement rates than pure volume, but the magnitude of these rates were misestimated. If management thinks the only source of learning is regular production ("learning by doing"), no time will be spent on experimentation ($z_i = 0$). The resulting loss of contribution is even larger than in the two examples.

Second, a deviation of 10–20% in π seems to be relatively small, especially if compared to how far the $\hat{\beta}_i$ estimates were from the true values. However, this is looking at contribution, rather than profits. In presence of large fixed costs, a 10–20% contribution change will make the difference between bottom line profits and losses.

6. Conclusion, implications and future research

We have presented an analytical model of production ramp-up, which combines a static trade-off between yields and utilization with a dynamic trade-off of learning and process improvement. In today's rapidly changing environments, cutting development times (time-to-market) in itself is not sufficient. Another key to achieving high profit is a rapid production ramp-up of a new product. This includes quickly achieving both high yields and a high level of utilization.

Our findings have a number of managerial implications. Most basic is the need for managers to accept and deal with the inherent paradox of learning during production ramp-up. At the beginning of a ramp when prices are at their highest, and yields and output at their lowest, it is nonetheless still the moment to further reduce output in order to run engineering trials and work on yield and speed improvements. This paradox often creates, in our experience, strong pressures to take shortcuts in learning, such as experiments with overly small sample sizes relative to the process noise level, or not running validation trials before implementing process changes. While this keeps up-time higher in the short run, it often leads to problems which reduce performance for the rest of the ramp-up period and beyond. We deal with this paradox by explicitly calculating the cost and value of experimentation as functions of time and processing capability. Figs. 6 and 7 show the patterns that can result from optimal behavior.

Second, we show the importance of understanding the sources of learning. It is incorrect to treat learning as an exogenous process beyond managerial control. Rather, there are three key high-level inputs which should be explicitly allocated and managed. These are normal production experience, capacity withdrawn from production for experiments of many kinds, and engineering time. Only the first of these happens automatically. Only engineering time (which we modeled as being lumped together with experimentation) appears explicitly in a cost accounting system. But the dollar costs of experimentation time, although not captured in accounting systems, can be a large investment as well, and are integral to success.

Our analysis in Section 3 provides a first look at the important trade-off between yields and production speed. With different product economics, and at different times in ramp-up, the optimal levels of care and rework shift. It also serves as a strong reminder that in yield driven industries there is a large difference between utilization and effective utilization. Finally, this research illustrates the importance of time-to-volume compared with the still dominant paradigm of time-to-market. We show how different situations require different decisions during the ramp-up period.

We have kept the model as simple as possible in order to focus on structural results. This approach clearly has limitations. Our assumptions that the processing capability can be represented as a single number, as well as the assumptions concerning the functional forms of learning rates and sub-additivity are strong simplification of real ramp-up situations. Refinements of the model provide interesting avenues for future research. First, some of the assumptions could be relaxed. For example, prices and competitive behavior could be explicitly modeled. Spence [31] provided an influential analysis of the effect of learning on strategic competition. He modeled a firm investing in learning early, in order to deter entry by potential competitors. In his model, learning was an inherent by-product of production experience, so that the form of "investment" was to produce *more*. The firm uses low prices both to encourage demand, and to serve as a signal to competitors that it has made an investment. This leads to the prescription to "price ahead of the learning curve". In our model, firms can also invest in learning, but in the form of deliberate learning through more time for experiments (and *less* production). A combination of both models seems promising.

Second, we see a strong need for more empirical research on this topic. Detailed case studies on the ramp-up period will help to reveal additional variables [32]. Such case studies could try to develop a managerial check-list of items that need to be addressed before or during the ramp-up. Another empirical research opportunity lies in a detailed econometric analysis of yield and utilization curves over time, which tries to identify the most effective variables that help increase the effective capacity.

Finally, the issues of production ramp-up should be linked to the existing fields of product development and learning in manufacturing. Although in the present paper we try to explicitly include findings from the manufacturing learning literature, we do not sufficiently include aspects of product development. What happens during product development will have a strong impact on the initial processing capability as well as on the speed of ramp-up. Thus, linking the quality of the ramp-up to events during the product development process provides a third interesting avenue for future research.

Appendix. Proofs of Propositions 1–5

Proof of Proposition 1. We need to show that there exists both a unique output maximizing care time x^*_{out} and a unique contribution maximizing care time x^*_{cont}. To do this, we use $s = \Gamma/(1+x)$ and show that there exists a uniquely defined optimal levels of starts. Given the 1:1 transformation between s and x, this also uniquely characterizes the optimal care levels x^*_{out} and x^*_{cont}. Output as a function of starts is given by $q(s) = s(1-(1/\alpha)(1/\Gamma/(s-1))$. As $\partial^2 q(s)/\partial^2 s = -2 - 4(s/\alpha(\Gamma - s)) - 2s^2/(\alpha(\Gamma - s)^2) < 0$, $q(s)$ is concave in s. The contribution maximizing problem can be restated as $pq(s) - cs \to \text{Max}$, which provides a linear combination of concave functions, and thus itself is concave.

Thus, both x^*_{out} and x^*_{cont} can be obtained from first-order conditions and the corresponding yield and contribution levels result from substituting x^*_{out} and x^*_{cont} into the corresponding definitions. Eq. (3.2) converges to (3.1) for large p $(c/p \to 0)$. □

Proof of Proposition 2. As $u^* = 1/(1+x)$, with $\alpha \to \infty$, we can see from (3.2) that $x \to 0$ and thus $u^* \to 1$. The same holds for yields y^*.

The ratio u^*/y^* indicates to what extent the process focuses on output or yields. From Proposition 1, we can determine

$$\frac{u^*}{y^*} = \alpha\,\frac{(p-c)}{p}$$

$$\frac{1+\sqrt{1+\alpha-c\alpha/p}}{[\sqrt{1+\alpha-c\alpha/p}+1+\alpha-c\alpha/p][\sqrt{1+\alpha-c\alpha/p}+c/p]}.$$

All of the proposed statements can be derived from the ratio u^*/y^*. □

Proof of Proposition 3. *Proof* (3a): We can see from Proposition 1 that $\pi_t(\alpha_t)$ is increasing with α_t, as the operation time (3.2) decreases with α (thus a high α allows for more starts) and the corresponding yield level also increases with α.

To show that costs are decreasing with p_t, define $m = \sqrt{p^2 + p^2\alpha - p\alpha} > 0$ and consider

$$\frac{\partial^2 \Pi^*_{\text{cont}}}{\partial^2 p} = \alpha(p-1)$$

$$\frac{pm\alpha + 2pm - \alpha m + 2p^2 + 2p^2\alpha - p\alpha + 2m + 2p - \alpha}{(p\alpha + 2p - \alpha + 2m)^2 m},$$

where, without loss of generality, $c = 1$. As $p > c = 1$, the second derivative is positive (i.e. costs are falling). This is, holding α_t constant, the only parameter changing with time.

Proof (3b): Eq. (4.4) can be derived by solving $\lambda\alpha = \alpha\beta_1^z \Theta^{z^2}\beta_2$ for z. After taking logarithms, we obtain a quadratic expression in z, yielding two solutions. As λ has to be smaller than $\beta_1\beta_2\Theta$, which is the maximum achievable improvement ($z = 1$), the optimal solution is given by (4.4). □

Proof of Proposition 4. To establish the diminishing returns we define m as above and compute

$$\frac{\partial^2 \Pi^*_{\text{cont}}}{\partial^2 \alpha} = -\frac{1}{2}(p-1)^2$$

$$p\,\frac{4p^2 + 3p^2\alpha - 6p\alpha - 4p + 3\alpha + 4pm - 4m}{(p\alpha + 2p - \alpha + 2m)^2(p + p\alpha - \alpha)m},$$

where m is defined as above. As $p > 1$, the second derivative is < 0 which shows that Π^*_{cont} grows slower for large values of α.

First, prices are falling over time, and second, the residual lifetime to which the new knowledge can be applied is shrinking. Each of the two arguments is sufficient in itself to make the value of an improvement go down in t. □

References

[1] R.J. McIvor, D. Martin, H. Matsuo, J. Ng, Profiting from process improvement in the new semiconductor manufacturing environment, Technology and Operations Review 1 (2) (1997) 1–15.

[2] J. Dutton, A. Thomas, Treating progress functions as a managerial opportunity, Academy of Management Review 9 (2) (1984) 235–247.

[3] S. Thomke, Project dreamcast: Serious play at Sega Enterprises Ltd, case study N9-600-028, Harvard Business School, Boston, 1999.

[4] G. Pisano, The Development Factory, Harvard Business School Press, Boston, 1995.

[5] R.E. Bohn, Measuring and managing technological knowledge, Sloan Management Review 36 (1) (1994) 61–73.

[6] R.E. Bohn, C. Terwiesch, The economics of yield-driven processes, Journal of Operations Management 18 (1999) 41–59.

[7] N.S. Langowitz, An exploration of production problems in the initial commercial manufacture of products, Research Policy 17 (1988) 43–54.

[8] R.H. Benfer, Learning during ramping: Policy choices for semiconductor manufacturing firms, M.S. Thesis, MIT, Cambridge, MA 1993.

[9] R.T. Clawson, Controlling the manufacturing start-up, Harvard Business Review 63(3) (1985) 6–20.

[10] N.H. Wasserman, K.B. Clark, Everest Computer, case study 685-085, Harvard Business School, Boston, 1986.

[11] R.C. Leachman, Competitive semiconductor manufacturing survey: third report on the results of the main phase, Berkeley Report CSM-31, 1996 (Chapter 2).

[12] J.M. Dutton, A. Thomas, J.E. Butler, The history of progress functions as a managerial technology, Business History Review 58 (2) (1984) 204–233.

[13] L. Argote, D. Epple, Learning curves in manufacturing, Science 247 (23) (1990) 920–924.

[14] M.A. Lapré, A.S. Mukherjee, L.N. Van Wassenhove, Behind the learning curve: linking learning activities to waste reduction, Management Science, forthcoming.

[15] E. von Hippel, M.J. Tyre, How learning is done: problem identification in novel process equipment, Research Policy 24 (1) (1995) 1–12.

[16] W.I. Zangwill, P.B. Kantor, Toward a theory of continuous improvement, Management Science 44 (7) (1998) 910–920.

[17] R. Stata, Organizational learning – the key to management innovation, Sloan Management Review 30 (1989) 63–74.

[18] P.S. Adler, Shared learning, Management Science 36 (8) (1990) 938–957.

[19] P.S. Adler, K.B. Clark, Behind the learning curve: A sketch of the learning process, Management Science 37 (3) (1991) 267–281.

[20] J.R. Dorroh, T.R. Gulledge, N.K. Womer, Investment in Knowledge: A generalization of learning by experience, Management Science 40 (8) (1994) 947–958.

[21] D. Epple, L. Argote, K. Murphy, An empirical investigation of the microstructure of knowledge acquisition and transfer through learning by doing, Operations Research 44 (1996) 77–86.

[22] H. Gruber, The yield factor and the learning curve in semiconductor production, Applied Economics 26 (1994) 837–843.

[23] K. Mishina, Learning by new experiences: Revisiting the flying fortress learning curve, in: N. Lamoreaux, D. Raff, P. Temin (Eds.), Learning by Doing in Markets, Firms, and Countries, University of Chicago Press, Chicago, 1998.

[24] M.A. Lapré, L.N. Van Wassenhove, Productivity improvement in a network of learning factories: A learning curve analysis, Working paper 98-02, BU School of Management, 1998.

[25] A.S. Mukherjee, M.A. Lapré, L.N. Van Wassenhove, Knowledge driven quality improvement, Management Science 44 (11) (1998) 35–44.

[26] A. Mody, Firm strategies for costly engineering learning, Management Science 35 (4) (1989) 496–512.

[27] B. Ehteshami, R.G. Petrakian, P.M. Shabe, Trade-offs in cycle time management: Hot lots, IEEE Semiconductor Manufacturing 5 (2) (1992) 101–106.

[28] R.E. Bohn, Learning by experimentation in manufacturing, Harvard Business School, Working paper 88-001, 1987.

[29] R. Jaikumar, R. Bohn, A dynamic approach to operations management: an alternative to static optimization, International Journal of Production Economics 27 (3) (1992) 265–282.

[30] C.H. Loch, C. Terwiesch, S. Thomke, Parallel and sequential testing of design alternatives, Working Paper at The Wharton School, Harvard Business School, INSEAD, 1999.

[31] M. Spence, The learning curve and competition, The Bell Journal of Economics 12 (1981) 49–70.

[32] C. Terwiesch, R.E. Bohn, K.C. Chea, An exploratory study of product transfer and production ramp-up in the data storage industry, Working Paper, University of California, San Diego, Sloan Center of the Information Storage Industry.