

UC San Diego

High-Technology Manufacturing

Title

Learning and Process Improvement during Production Ramp-Up

Permalink

<https://escholarship.org/uc/item/5zf5q453>

Authors

Christian Terwiesch
Roger E. Bohn

Publication Date

1998-08-06

Learning and Process Improvement during Production Ramp-Up

Christian Terwiesch Roger E. Bohn
The Wharton School University of California San Diego

September 18, 1998

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 it takes them 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 still 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 experimentation. Experiments such as engineering trials lead to process improvements, but they also 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. Depending on the relationship between selling price and variable cost, it may be optimal to follow a "yield first" or a "speed first" policy. Several numerical examples illustrate our results. We show that misunderstanding the causes of learning, in particular whether it is experience driven or experiment driven, leads to suboptimal behavior and outcomes.

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-drives¹ 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 production 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 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 capacity and high demand, the company finds itself pressured from two sides, an effect referred to as the “nutcracker” (McIvor *et al.* 1997).

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, however, and there is much that managers can do to affect the rate of learning (Dutton and Thomas 1984).

Learning in manufacturing takes place through many mechanisms. 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 and experimentation. On one hand, high market prices create an incentive to maximize output and minimize engineering trials. On the other hand, it is during the ramp-up that the knowledge level of the plant is at its lowest, so the need for learning

¹This research was funded by the Sloan Center for the Information Storage Industry at the University of California, San Diego.

is especially high. 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 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. Ramp-up is the transition a manufacturing process makes from zero to full production. Each new product introduced into a factory must undergo a ramp-up. Ramp-ups are also needed for new lines and new factories; these are sometimes called “start-ups”. A new product ramp-up can take 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 more effective and faster inspection methods at all stages of the process. Thus production rates and yields go up, while defects, maintenance time, and rework go down.

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.

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 (Bohn 1994). 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 (Terwiesch *et al.* 1997). 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 it². 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 (1988) conducts an exploratory study of ramp-up of four electronics products. Benfer (1993) discusses the general problem of rapid ramp-up at Intel. Both emphasize the relationship between development and successful ramp-up. Clawson (1985) discusses ramp-up in aerospace. Wasserman and Clark (1986) document a problematic ramp-up of high performance semiconductors, in which yields remained close to zero for months³. 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. Leachman (1996) show 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.

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.* 1984). This model postulates that per unit costs fall as the log of cumulative production (Argote and Epple 1990). 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

²In 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.

³Other ramp-up case studies include (Bohn and Jaikumar 1986; Freeze and Clark 1986; Freeze and others 1984; Langowitz and Wheelwright 1986).

provides producers with opportunities to effect cost efficiencies that have little to do with cumulative volume (Dutton and Thomas 1984).

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. Such a simplistic view of the learning process offers no managerial guidance for improving the ramp-up of a new product.

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, responsible for systematic problem solving and learning. Lapré *et al.* (1996) show that many directed improvement projects in fact have zero or negative effects. Both sound theory and empirical validation by experimentation are needed before process changes are justified. Tyre (1990) and von Hippel (1993) examine information flows and other aspects of problem solving for a variety of new process introductions in plants.

Many others have looked at learning at a more aggregate level, to determine what factors drive performance improvement. These include Adler (1990), Adler and Clark (1991), Dorroh *et al.* (1994), Epple *et al.* (1996), Gruber (1994), Mishina (1998), Lapré and Wassenhove (1998) and Mukherjee *et al.* (1998). Most of these articles emphasize empirical fits to data rather than conceptual models. Three conceptual papers are related to our approach. Mody provides a model, which explicitly examines engineering effort as a driver of learning (Mody 1989). Dorroh (1994) has 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.

Zangwill and Kantor (1998) propose that learning be modeled as “waste removal” rather than directly as cost reduction. Rather than yield or production rate as the dependent variable, they suggest using “percent defective” and “waste time” respectively. They propose that each halving of waste requires a roughly constant amount of effort. Earlier empirical work in Stata (1989) provides extensive documentation of halving times for various kinds of waste reduction efforts. Our model in Section 4 is related to these ideas with respect to production rate, although it models yields directly.

2.2. What drives learning?

Although our model emphasizes deliberate learning through experiments, we also allow for learning to take place through experience. Management directly sets the rate of experimentation and the production rate, subject to the constraint of machine capacity. Both experimentation and production (experience) can cause learning, which leads to higher knowledge about how to produce. In the terminology of Dutton and Thomas, experimentation is a form of *induced* (deliberate) learning, while production experience is *autonomous* (automatic) learning. Both can potentially operate, and we will investigate the consequences of learning by experimentation compared with by experience. Figure 1 shows the flow of causality.

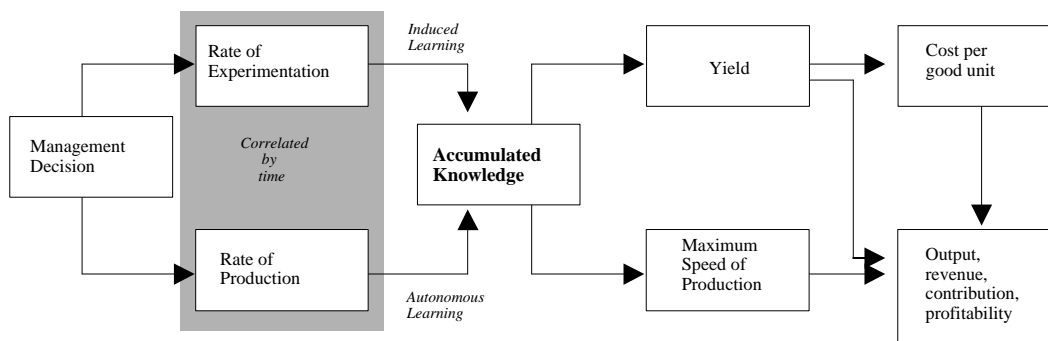


Figure 1: Causes of learning and improvement

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.

If experimentation is a key driver of learning, what limits the rate of experimentation? In section 4 we emphasize the production capacity constraint, i.e. that experiments decrease output. Many yield ramp-ups involve strenuous debates 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 (Ehteshami *et al.* 1992).

There are at least three other limits that could be considered. First, engineers and other learning workers are in short supply, and they have more problems to work on than time. Second, some kinds of experiments incur out of pocket costs e.g. for special laboratory analysis. Third, the effectiveness of experimentation varies dramatically depending on a variety of statistical and non-statistical design issues (Bohn 1987).

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 suited 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. Further, high-tech environments 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 are at low stages of knowledge and yield losses are still substantial.

During ramp-up, the goal is to raise both yield and production rate 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 the processing speed. There are many causes of such tradeoffs. First, consider the operation of a robotic watch assembly line as described in Jaikumar and Bohn 1992. 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. (Terwiesch *et al* 1997)

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 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 output or on yields (Subsection 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

Consider an operation which takes γ units of time [hours/unit], if executed at its maximum speed. In presence of a speed versus yield trade-off, it might be beneficial to slow down the corresponding operation by a certain operation time x to $x + \gamma$ units of time [hours/unit]. Assume a total capacity of Γ machine hours per period [hours/period]. Then the number of units started in the process is $s = \frac{\Gamma}{x+\gamma}$ at a utilization of $u = \frac{\gamma}{x+\gamma}$ of the theoretical capacity. x represents a deliberate “level of care” built into the operation and is selectable by management.

Next, we have to describe the relationship between starts s and yields. Spending more time on an operation 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 x and a parameter $\alpha > 0$ which measures processing capability. The higher α , the more the process can be accelerated without major yield losses: $\frac{\partial y(x, \alpha)}{\partial \alpha} > 0$. We assume diminishing returns of the extra operation time x , so that $\frac{\partial y(x, \alpha)}{\partial x} > 0$ and $\frac{\partial^2 y(x, \alpha)}{\partial x^2} < 0$. Output is then starts times yields or $y(x, \alpha) \frac{\Gamma}{x + \gamma}$.

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 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, x is chosen to maximize the contribution (sales minus variable costs), which we can write as:

$$\pi(\tilde{p}, \alpha, x, c) = \tilde{p}y(x, \alpha) \frac{\Gamma}{x + \gamma} - \frac{\Gamma}{x + \gamma} c \quad (3.1)$$

3.2. The starts versus yields trade-off

Good output is not necessarily a monotonic increasing function in the number of starts or in the level of utilization. Starting too many units can disturb the production process that badly that not only yields fall, but even the overall 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. Let

$$y(x, \alpha) = y_0 \left(1 - \frac{1}{\alpha x} \right) \quad (3.2)$$

which - consistent with our argument above - shows that 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 standardize the minimum production time $\gamma = 1$ and discount the selling price by the base yields, i.e. define $p = \tilde{p}y_0$. Good output per period is now given by $y_0 \left(1 - \frac{1}{\alpha x} \right) \frac{\Gamma}{x + 1}$ and the per period contribution is:

$$\pi(p, \alpha, x, c) = p \left(1 - \frac{1}{\alpha x}\right) \frac{\Gamma}{x+1} - \frac{\Gamma}{x+1} c \quad (3.3)$$

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 $u_{out}^* = \frac{1}{x_{out}^* + 1}$ actually decreases the overall output. At this point the effective capacity decreases.

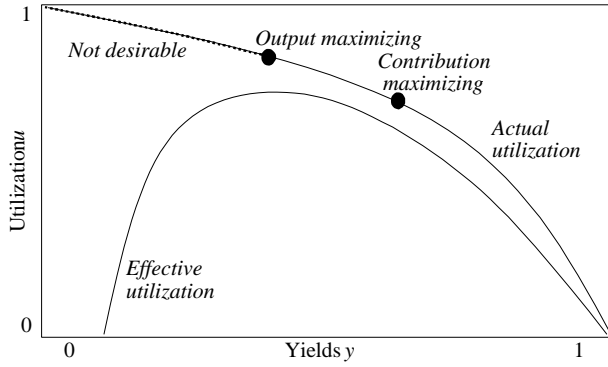


Figure 2: Utilization versus Yields

In order to calculate the contribution optimal level 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 Figure 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 $\frac{p}{c} > 1$, 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:

- Both the output maximizing operation time and the contribution maximizing operation time are strictly positive, i.e. $x_{out}^* > 0$ and $x_{cont}^* > 0$. As a result of this $u_{out}^*, u_{cont}^* < 1$, 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:

$$x_{out}^* = \frac{1 + \sqrt{1 + \alpha}}{\alpha}, \quad \Pi_{out}^* = \pi(p, \alpha, x_{out}^*, 0) = \Gamma p \frac{1}{1 + \frac{2}{\alpha}(1 + \sqrt{1 + \alpha})} \quad (3.4)$$

- The contribution maximizing operation time x_{cont}^* , the corresponding yield level $y(x_{cont}^*, \alpha)$, and the resulting contribution level Π_{cont}^* are given by:

$$x_{cont}^* = \frac{1 + \sqrt{1 + \alpha - \frac{c\alpha}{p}}}{\alpha \left(1 - \frac{c}{p}\right)}; \quad y_{cont}^* = y(x_{cont}^*, \alpha) = \frac{\sqrt{1 + \alpha - \frac{c\alpha}{p}} + \frac{c}{p}}{\sqrt{1 + \alpha - \frac{c\alpha}{p}} + 1} \quad (3.5)$$

$$\Pi_{cont}^* = \pi(p, \alpha, x_{cont}^*, c) = \Gamma \alpha \frac{(p - c)^2}{p} \frac{1}{2\sqrt{1 + \alpha - \frac{c\alpha}{p}} + 2 + \alpha - \frac{c\alpha}{p}}$$

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

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 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 (3.4) and (3.5). For the special case $c = 0$, the two are identical. For $c > 0$, utilization is adjusted downwards in favor of yields. This confirms the intuition generated by Figure 2. Thus, a simple corollary of Proposition 1 is that $y_{out}^* \leq y_{cont}^*$ and, for the corresponding utilization levels, $y_{out}^* \leq y_{cont}^*$.

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 from (3.2), 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 Figure 2 by showing various levels of α . Each of the points on the additional lines corresponds to a set of (y_t^*, u_t^*) that is computed using (3.5) for a changing level of α . We define this path of utilization / yield combination as the Ramp-Map. This is summarized in Figure 3. Let a *yield-emphasizing ramp* be a ramp with high initial yields (relative to utilization, $\frac{u^*}{y^*}$ 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 $\frac{u^*}{y^*}$ is large) with high initial utilization. With increasing α , yields are increased.

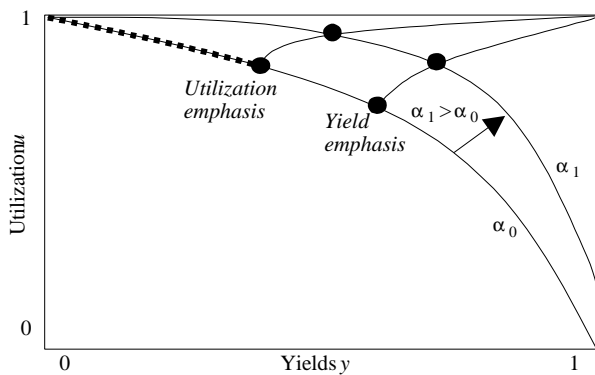


Figure 3: Increasing processing capability: the Ramp-Map

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

- Long run behavior: for $\alpha \rightarrow \infty$, $y^* \rightarrow y_0$ and $u^* \rightarrow 1$ and as a result $\frac{u^*}{y^*} \rightarrow 1/y_0$
- For small α , $\frac{u^*}{y^*}$ is an increasing function in $\frac{p}{c}$. Large values of $\frac{p}{c}$ favor a utilization emphasizing ramp.

- For large values of $\frac{p}{c}$, $\frac{u^*}{y^*} \rightarrow \frac{u_{out}^*}{y_{out}^*} = \frac{\alpha}{(1+\alpha)y_0}$, 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 α is always $y^* \rightarrow y_0$ and $u^* \rightarrow 1$, where production is perfect. This provides the end-point of the Ramp-map. Second, the ratio $\frac{u^*}{y^*}$ helps us to further specify the location of the start-point. For large values of $\frac{p}{c}$, following Proposition 1, the only focus is on output, thus $\frac{u^*}{y^*}$ is maximized. At this point, the ratio between utilization and yields is characterized by $\frac{\alpha}{1+\alpha}$. For smaller values of $\frac{p}{c}$, 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 wafer production. For the former, the costs of raw material are very close to the market price of the finished good. Proposition 2 predicts a strong yield emphasis in the ramp-up, which is consistent with empirical research in this industry (Terwiesch *et al.* 1997). Even if first-pass yields are low, rework is used intensively to reach high final yields. For semiconductors, the value of the finished wafer is many times its cost per start. Thus, 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. Leachman 1996).

4. Learning in Ramp-up

Proposition 2 is based on the assumption that learning is exogenous to the model. In other words, it assumes that learning efforts are in no way affected by production. In this section, we alter the previous model by explicitly including the sources of learning into our analysis. We do this by adding a second managerial decision variable, learning effort. We assume that learning effort takes the form of controlled experiments.

The benefit of learning is 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. set-ups 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 (Bohn 1987). Looking at Π_{cont}^* in (3.5), we see that the overall contribution is proportional to Γ . So, spending more machine time for experimentation creates an opportunity cost of lost regular production. This situation is illustrated in Figure 4. The overall capacity is reduced because of yield losses and down time for experimentation. Although experiments reduce contribution in the short-run, they increase the future processing capability, which allows for more starts and higher yields in the following periods. The remainder of this section formalizes the model of Figure 4 focussing on the capacity reducing effect of experimentation.

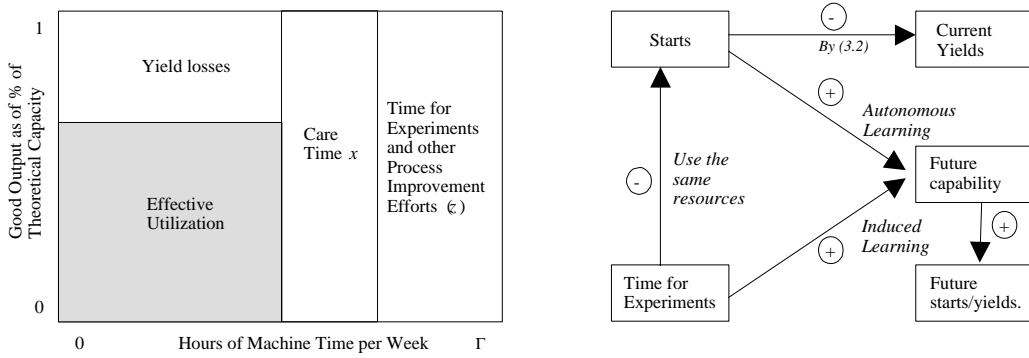


Figure 4: Effects of experimentation

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 processing speed (level of care x), the experimentation time represents our second managerial decision variable. The overall output of the production process is then:

$$(1 - z)y(x, \alpha) \frac{\Gamma}{x + 1} \quad 0 \leq z \leq 1$$

We model the dynamic influence of z on the processing capability α as $\alpha_{new} := \alpha_{old} \beta_1^z \beta_2$. Thus learning occurs both by experimentation z and by experience. 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 itself⁴. If $z = 0$, the next period's processing capability α_{new} is β_2 times larger than the current capability α_{old} . Learning by experience is thus driven by

⁴Note, that this formulation of learning is equivalent to assuming $\alpha_{new} := \alpha_{old} \beta_1^z \beta_2^{1-z}$, where β_1 is defined as the absolute importance of learning by experimentation and β_2 as the absolute importance of learning by doing.

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 knowledge, care level, 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 earnings 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 x_t) and the fraction of capacity used for experimentation z_t . This choice is basically a balance between the cost of experiments in form of lost production and the value of experiments in form of an increased future capability (see Figure 4).

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 α_t than spending 10% in the current period and 10% in the subsequent period. There are several reasons for such a *sub-additivity*. First, experimentation is normally done in cycles. It is more effective to wait for the results of one experiment before formulating ideas which become the basis of the next cycle. Second, although capacity is a key input, there are other inputs, especially engineering time. For a fixed number of engineers, there will be decreasing marginal returns to z . Third, conducting too many experiments at the same time increases the noise in the process, which makes it harder to learn.

We model this sub-additivity by discounting the learning rate β_1 by a factor $\Theta \leq 1$, so that $\beta_{eff}(z) = \beta_1 \Theta^z$. The new processing capability is then given by

$$\alpha^{new} := \alpha^{old} \beta_2 [\beta_{eff}(z)]^z = \alpha^{old} \beta_1^z \Theta^{z^2} \beta_2 \quad (4.1)$$

For small values of experimentation time z , the discount 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 . For larger values of z , an increase in z has diminishing returns. 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 overall discount 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 presence of changing

cost and value, this approach is unlikely to be optimal and the real optimal solution has to be chosen based on the overall optimization problem.

As the speed of the line x_t has no effect on any future $\alpha_{t+i}, i = 1..T - t$, we can decompose the overall optimization into T optimization problems with x_t as a decision variable and one optimization problem that chooses z_1, \dots, z_T assuming an optimal speed in each period. The optimal x_t can be computed based on (3.5). The optimal z_t can be computed by solving a simple dynamic program with α_t as the state and z_t as the decision variable.

$$\Pi^{total}(z_1, \dots, z_T) = \underset{z_1, \dots, z_T}{Max} \sum \delta_d^{t-1} (1 - z_t) \pi_t(\alpha_t) \quad (4.2)$$

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

$$\begin{aligned} F_t(\alpha_t) &= \underset{z_t}{Max} \left\{ (1 - z_t) \pi_t(\alpha_t) + \delta_d F_{t+1}(\alpha_t \beta_1^{z_t} \Theta^{z_t^2} \beta_2) \right\} \\ F_T(\alpha_T) &= \pi_T(\alpha_T) + \text{Terminal value}(\alpha) \end{aligned} \quad (4.3)$$

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 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^*, \dots, z_T^* are uniquely identified and can be computed by backward induction.

4.2. Costs of an Experiment

In order to understand how much of their 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 (3.5).

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 $k(t, \alpha_t, z_t) = \text{Max}\{z_t \pi_t(\alpha_t), 0\}$ 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 $1..T$ - 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 (i.e. $\alpha_{new} = \lambda\alpha_{old}$) carries the following costs:

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

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

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

We can see that although the required amount of experimentation to increase 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. (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. The sub-additivity parameter θ discounts the rate of learning by experimentation following (4.1). We can see that a doubling of 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.

4.3. Value of an Experiment

The value of an experiment depends on three factors, t , α , and 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 an experiment is - as opposed to the cost - not linear in z . Thus, 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):

$$v(t, \alpha_t, z_t) = \delta_d \left[F_{t+1}(\alpha_t \beta_1^{z_t} \Theta^{z_t^2} \beta_2) - F_{t+1}(\alpha_t \beta_2) \right] \quad (4.5)$$

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 $F_{t+1}(\alpha_t \beta_2)$. In other words, 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 the 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 to be falling over time. There are three reasons for this. First, the residual life-time of the product to which the new knowledge might be applied, is shrinking. An experiment in the next period will only help a later period, but will reduce the current regular production. 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 is also falling over increasing α . To illustrate this, compare two situations. In the first situation, 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 u and 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 (1998). 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 50-75% and getting them from 75-87% both correspond to a halving of waste, and require the same effort, but the first improvement is more valuable than the second. This is consistent with (3.2) in our model, where we define waste as $\frac{1}{\alpha x}$, 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, such as disk drive assembly. The initial price is $p = \$3/\text{unit}$, prices fall at $\delta_p = 0.95$ per period (month), the discount factor is $\delta_d = 0.98$. We assume $c = \$1/\text{unit}$ 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 $\beta_1 = 2.80$ and $\beta_2 = 1.01$. The overall capacity available for production *and* experimentation is $\Gamma = 1000$, 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.

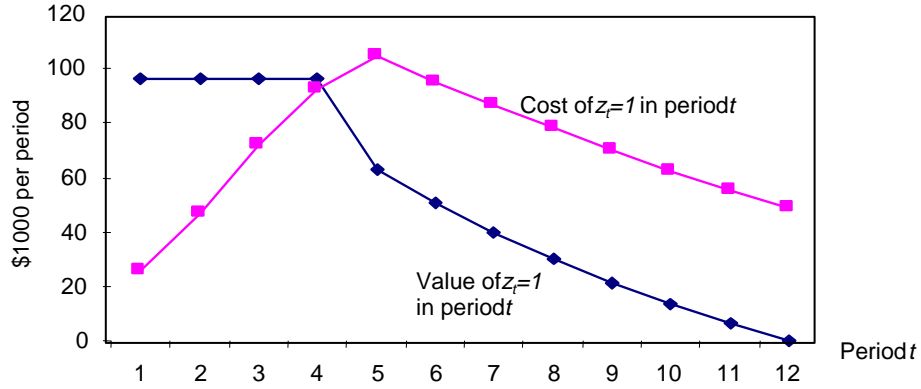


Figure 5: Cost and value (in dollar) of experimentation

Figure 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 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. Figure 5 also shows that experiments are inexpensive in the beginning and in the end of the product lifecycle, but most costly in the middle.

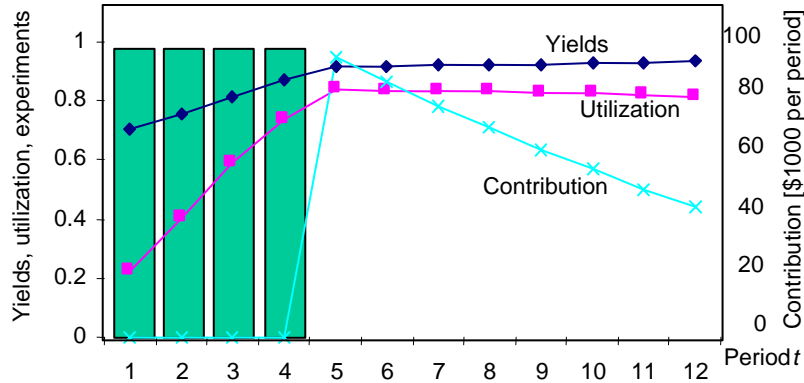


Figure 6: Optimal solution for $\theta = 1$; the bars indicate the optimal z_t

Figure 6 summarizes the optimal solution. The four bars indicate the optimal experimentation policy: full experimentation at the start, then none. Figure 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%)⁵. This yield emphasizing ramp is a result of the relatively small price to cost ratio (of initially 3:1).

5.2. Low Experimentation Capability

Next, consider the case of lower experimentation capability, e.g. $\theta = 0.5$. Whereas the cost per unit of experimentation remain unchanged in the first period, 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.

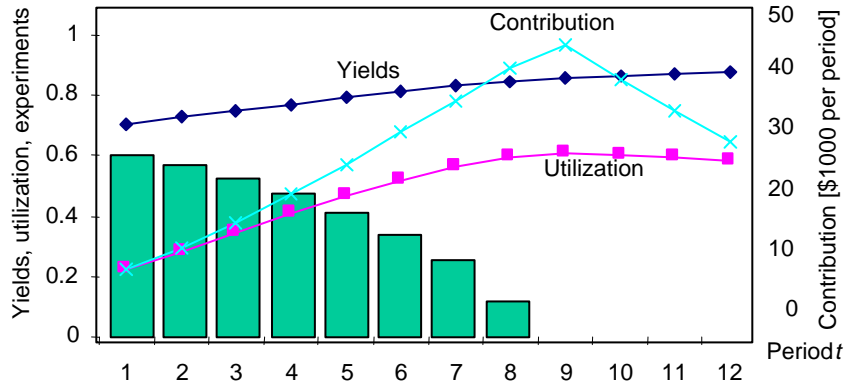


Figure 7: Optimal solution for $\theta = \frac{1}{2}$; the bars indicate the optimal z_t

Figure 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 Figure 6), production starts in the first period, so the plant is actually producing at the initial yields of 72%. Figure 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

⁵Note that these values are "not realized", as the complete first periods are dedicated to experimentation.

process engineering. In the second example, the product is introduced to the market earlier and process engineering is spread out over 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 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 $t = 1..n < T$). 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 all 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. Especially the importance of “learning by experience” is frequently overestimated compared to the importance of controlled experiments. In terms of our model, this corresponds to an underestimation of β_1 .

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 are 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, yielding $\hat{\beta}_1 = 5$. 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’s qualitative understanding concerning the sources of learning was correct. Management was aware that experimentation yields higher improvement rates than pure volume, but the magnitude of these rates were misestimated. If management thinks of the only source of learning as being regular production (“learning by doing”), no time will be spent for experimentation ($z_i = 0$). The resulting loss of contribution is even larger than in the other examples.

Second, a deviation of 20-40% 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 any reasonably large fixed costs, a 20-40% contribution change will make the difference between bottom line profits and losses.

6. Conclusion, Implication 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. The other key to achieving high contributions is a rapid production ramp-up of a new product. This includes quickly achieving both high yields and a high level of utilization.

6.1. Summary

We model the trade-off between yields and utilization by making yields a decreasing function of the number of items started in the process. Such start-vs-yield trade-offs are common from causes including variances increasing with processing speed, human failure, rework, increased WIP, or reduced maintenance. In this situation, it is both output and contribution beneficial to deliberately start fewer items than allowed by the available capacity. A larger ratio between price and cost allows a pure focus on output rather than sales or cost (Proposition 1).

In Proposition 2, we introduced the concept of the Ramp-map. Given an increasing processing capability, the production line is able to increase both yields and utilization. In the extreme case of perfect knowledge about the production process, yields and utilization reach 100%. Before this point, management can choose between different paths, emphasizing either high yields or high utilization. We show that high prices create an incentive to emphasize utilization more than yields. The model allows us to explain different empirical patterns of ramp-up.

Whereas in Proposition 2 the sources of learning are exogenous to the model, Proposition 3 explicitly includes a decision variable for learning efforts. Learning effort is modeled as controlled experiments, which create an opportunity cost of not satisfying the high demand. Over time, the costs of an experiment can go up or down. Costs tend to go up because of an increasing processing capability: the higher yields and utilization, the larger are the opportunity costs for shutting down the line for experimentation. On the other hand, prices are highest at the day of market introduction and fall thereafter. This can reduce the cost of experimentation.

The value of an experiment is falling, both in time and processing capability. The fall over time is a result of a shorter residual life-time as well as of falling prices. The fall over processing capability is based on physically diminishing returns: the same amount

of learning effort will create higher physical benefits (e.g. in terms of yields or output) and higher economic benefits for a low processing capability than for a high processing capability.

Finally, we illustrate the structure of the optimal ramp-up policy based on a series of numerical examples. For slowly falling prices, low initial processing capability, and high experimentation capability, it is optimal to do all the experiments up front. This means the production process should be managed like a pilot line and regular production should take over only once yields and utilization have been increased. In presence of lower experimentation capability, it is beneficial to spread the experiments out over time. In this case, the proportion of capacity dedicated to experimentation typically decreases from period to period. However, in the extreme case of rapidly falling prices, it can be beneficial to delay the experiments to a latter period.

6.2. Managerial Implications

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. Figures 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.

On a related note, we provide a more explicit way to think about the effectiveness of learning, through our β_1 and θ parameters. Different plants, people, and learning methods have different abilities to extract useful process knowledge from a given amount of experimentation. Indeed whole books have been written on this, and there is much more research still to be done on it. There are also differential abilities to learn from plain production experience, modeled as our β_2 coefficient. Knowing the sources of learning (Figure 1) and then working to improve their effectiveness is a key form of managerial learning from one ramp-up to the next.

At a tactical level, 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.

6.3. Future Research

We have kept the model as simple as possible in order to focus on structural results rather than on concrete solutions to be taken as a quantitative guide. 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 (1981) 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 *lower* output). We hypothesize that, as in Spence-style models, firms will still use lowered prices as a signal of their increased production capability. Thus the within-factory prescriptions of the two models are very

different (maximize output in one case, reduce output and maximize learning in the other), yet they are similar at a strategic level.

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. 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.

A related refinement would be to make capacity and other fixed cost decisions endogenous. We treated capacity as a fixed constraint which was binding throughout the ramp-up period. But it is usually possible to buy extra capacity at the start, moving up the time when supply catches demand. Thus in some ways, investing in capacity can substitute for investing in learning.

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 manuscript 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.

7. References

- Adler, Paul S., "Shared Learning," *Management Science*, 36:8 [1990], pp. 938-957.
- Adler, Paul S. and Kim B. Clark, "Behind the Learning Curve: a Sketch of the Learning Process," *Management Science*, 37:3 [1991], pp. 267-281.
- Argote, Linda and Dennis Epple, "Learning Curves in Manufacturing," *Science*, 247:23 [1990], pp. 920-924.
- Benfer, Richard H., *Learning during Ramping: Policy Choices for Semiconductor Manufacturing Firms*, M.S. thesis, MIT, 1993.
- Bohn, Roger and R. Jaikumar, *Priam Corporation - Hard Disk Drive Manufacturing*, 1986, HBS case 685-096.
- Bohn, Roger E., *Learning by Experimentation in Manufacturing*, Harvard Business School, 1987, working paper 88-001.

- Bohn, Roger E., "Measuring and Managing Technological Knowledge," *Sloan Management Review*, 36:1 [1994], pp. 61-73.
- Clawson, Richard T. "Controlling the Manufacturing Start-Up." *Harvard Business Review*, May-June 1985, pp 6-20.
- Dorroh, James R., Thomas R. Gulledge, and Norman K. Womer, "Investment in knowledge: A generalization of learning by experience," *Management Science*, 40:8 [1994], pp. 947-958.
- Dutton, John and Annie Thomas, "Treating Progress Functions as a Managerial Opportunity," *Academy of Management Review*, 9:2 [1984], pp. 235-247.
- Dutton, John M., Anne Thomas, and John E. Butler, "The History of Progress Functions as a Managerial Technology," *Business History Review*, 58:No. 2 [1984], pp. pp. 204-233.
- Ehteshami, Babak, Raja G. Petrakian, and Phyllis M. Shabe, "Trade-Offs in Cycle Time Management: Hot Lots," 5:2 [1992], pp. 101-106.
- Epple, Dennis, Linda Argote, and Kenneth Murphy. "An empirical investigation of the microstructure of knowledge acquisition and transfer through learning by doing." *Operations Research* 44 (1996): 77-86.
- Freeze, Karen and Kim B. Clark, *Sun Microsystems, Inc. (B)*, Harvard Business School, 1986, HBS Case 686-134.
- Freeze, Karen J., Neil Wasserman, and Kim B. Clark, *Seeq Technology*, 1984, Harvard Business School, 1984, HBS Case 9-685-081.
- Gruber, Harald. "The yield factor and the learning curve in semiconductor production." *Applied Economics* 26 (1994): 837-843.
- Jaikumar, R. and R. Bohn, "A Dynamic Approach to Operations Management: An Alternative to Static Optimization." *International Journal of Production Economics*, 1992.
- Langowitz, Nan and Steven C. Wheelwright, *Sun Microsystems, Inc. (A)*, Harvard Business School, 1986, Case 686-133.
- Langowitz, Nan S., "An Exploration of Production Problems in the Initial Commercial Manufacture of Products," *Research Policy*, 17:[1988], pp. 43-54.
- Lapr e, M.A., A.S. Mukherjee, and L.N. Van Wassenhove, *Behind the Learning Curve: Linking Learning Activities to Waste Reduction*, INSEAD, 1996 Working paper 96/24/TM.
- Lapr e, M.A. and L.N. Van Wassenhove, *Productivity Improvement in a Network of Learning Factories: A Learning Curve Analysis*, BU School of Management, 1998, Working paper 98-02.

Leachman, Robert C. “Competitive semiconductor manufacturing survey: Third report on the results of the main phase,” (1996) Berkeley report CSM-31, Chapter 2.

McIvor, Robert, John D. Martin, Hirofumi Matsuo, Joseph Ng, “Profiting from Process Improvement in the New Semiconductor Manufacturing Environment”, *Technology & Operations Review*, 1 (1997), Number 2.

Mishina, Kazuhiro. “Learning by New Experiences: Revisiting the Flying Fortress Learning Curve,” in *Learning by Doing in Markets, Firms, and Countries*, ed. Naomi Lamoreaux, Daniel Raff, and Peter Temin, University of Chicago Press, 1998.

Mody, Ashoka, “Firm Strategies for Costly Engineering Learning,” *Management Science*, 35:4 (1989), pp. 496-512.

Mukherjee, A.S., M.A. Lapré, and L.N. Van Wassenhove, “Knowledge Driven Quality Improvement,” *Management Science*, forthcoming:[1998]

Spence, Michael. “The Learning Curve and Competition.” *The Bell Journal of Economics* 12 (1981): 49-70.

Stata, Ray, “Organizational Learning - the Key to Management Innovation”, *Sloan Management Review* 30 (1989), pp. 63-74.

Terwiesch, Christian, Roger E. Bohn, and Scott M. Hampton, *The Economics of Yield-Driven Processes*, University of California San Diego, Data Storage Industry Globalization Project, 1997, Working paper 97-04.

von Hippel, Eric and Marcie J. Tyre. “How learning is done: Problem identification in novel process equipment.” *Research Policy* 24 (1 1995): 1-12.

Wasserman, Neil H. and Kim B. Clark, *Everest Computer*, Harvard Business School, 1986, case study 685-085.

Zangwill, Willard I. and Paul B. Kantor. “Toward a Theory of Continuous Improvement,” *Management Science* (1998).

8. Appendix: Sketches of Proofs of Propositions 1 to 5

8.1. 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 $s = \frac{\Gamma}{1+x}$ 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 $q(s) = s(1 - \frac{1}{\alpha} \frac{1}{\Gamma/s-1})$. As $\frac{\partial^2 q(s)}{\partial^2 s} = -2 - 4 \frac{s}{\alpha(\Gamma-s)} - 2 \frac{s^2}{\alpha(\Gamma-s)^2} < 0$, $q(s)$ is concave in s . The contribution maximizing problem can be restated as $pq(s) - cs \rightarrow Max$, 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.

(3.5) converges to (3.4) for large p ($\frac{c}{p} \rightarrow 0$). \square

8.2. Proof of Proposition 2

As $u^* = \frac{1}{1+x}$, with $\alpha \rightarrow \infty$, we can see from (3.5) that $x \rightarrow 0$ and thus $u^* \rightarrow 1$. The same holds for yields y^* .

The ratio $\frac{u^*}{y^*}$ indicates to what extent the process focuses on output or yields. From Proposition 1, we can determine: $\frac{u^*}{y^*} = \alpha \frac{(p-c)}{p} \frac{1 + \sqrt{1 + \alpha - c\alpha/p}}{\left[\sqrt{1 + \alpha - c\alpha/p} + 1 + \alpha - c\alpha/p \right] \left[\sqrt{1 + \alpha - c\alpha/p} + c/p \right]}$

All of the proposed statements can be derived from the ratio $\frac{u^*}{y^*}$. \square

8.3. 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 (3.5) 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 $\frac{\partial^2 \Pi_{cont}^*}{\partial^2 p} = \alpha (p-1) \frac{pm\alpha + 2pm - \alpha m + 2p^2 + 2p^2\alpha - p\alpha + 2m + 2p - \alpha}{(p\alpha + 2p - \alpha + 2m)^2 m}$

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. \square

PROOF (3b): (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). \square

8.4. Proof of Proposition 4

PROOF: To establish the diminishing returns we define m as above and compute $\frac{\partial^2 \Pi_{cont}^*}{\partial^2 \alpha} = -\frac{1}{2} (p-1)^2 p^{\frac{4p^2+3p^2\alpha-6p\alpha-4p+3\alpha+4pm-4m}{(p\alpha+2p-\alpha+2m)^2(p+p\alpha-\alpha)m}}$ where m is defined as above. As $p > 1$, the second derivative is negative which shows that Π_{cont}^* grows slower for large values of α .

(2) 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 . \square