Journal of Operations Management 18 (1999) 41–59

JOURNAL OF OPERATIONS MANAGEMENT

www.elsevier.com/locate/dsw

The economics of yield-driven processes

Roger E. Bohn [a,1], Christian Terwiesch [b,*]

[a] University of California, San Diego, CA, USA

[b] The Wharton School, University of Pennsylvania, Steinberg Hall–Dietrich Hall, Philadelphia, PA 19104-6371, USA

Received 5 February 1998; accepted 6 April 1999

Abstract

The economic performance of many modern production processes is substantially influenced by process yields. Their first effect is on product cost — in some cases, low-yields can cause costs to double or worse. Yet measuring only costs can substantially underestimate the importance of yield improvement. We show that yields are especially important in periods of constrained capacity, such as new product ramp-up. Our analysis is illustrated with numerical examples taken from hard disk drive manufacturing. A three percentage point increase in yields can be worth about 6% of gross revenue and 17% of contribution. In fact, an eight percentage point improvement in process yields can outweigh a US$20/h increase in direct labor wages. Therefore, yields, in addition to or instead of labor costs, should be a focus of attention when making decisions such as new factory siting and type of automation. The paper also provides rules for when to rework, and shows that cost minimization logic can again give wrong answers.

Keywords: Production yields; Cost of quality; Product cost; Rework; Ramp-up; Location decisions; International operations

1. Introduction

Many modern production processes and services are driven by process yields. Not every unit of material that starts into the production process makes it to the end as a sellable, high quality product. Some "fall-out" along the way due to problems of various kinds. Often, some of the fall-out can be reworked, but always a fraction of it must be scrapped. This means that materials and effort go into making something that ultimately cannot be not sold.

The effect of yield losses on the economics of the product, factory, and business can be dramatic. The comprehensive Berkeley project on semiconductor manufacturing has documented many examples of integrated circuit factories with yields below 50% for years (Leachman, 1996). The impact of this is, crudely, that costs per good unit are multiplied by two compared with what they would be at 100% yield. The impact on profit is much greater.

* Corresponding author. E-mail: terwiesch@wharton.upenn.edu

[1] E-mail: rbohn@ucsd.edu.

0272-6963/99/$ - see front matter
PII: S0272-6963(99)00014-5

The main purpose of this paper is to analyze the economics of yield-driven production processes. Despite the widespread and important role of yields, their impact on economic performance is treated casually in management accounting systems, and has received little attention by operations management researchers. The result, we observe, is that some decisions are driven by analysis and intuition developed from inadequate models.

A secondary purpose of this paper is to compare the importance of yields with that of labor costs. Specifically, we show that under common conditions in ''high-tech'' industries, the impact of direct labor wage rates can be overshadowed by the effect of yields. Even eliminating direct labor entirely can have less effect on profit than modest changes in yield levels. Thus, yields matter when asking questions such as ''Where to site the next factory?'' and ''Should we automate a process?''

Our analysis is illustrated with examples from a high-tech industry, hard disk drives (HDDs). Disk drive production starts with the fabrication of key components (heads, media disks, and semiconductors). All of these fabrication processes are strongly yield-driven, i.e., much less than 100% of what goes ''in'' to the process comes ''out'' as good components. The components are then assembled in multi-step, labor- and testing-intensive processes. These assembly steps are also yield-driven. The industry is sensitive to yield issues, as illustrated by the following quotation. Nonetheless, it has not had good tools for quantifying their effects.

> It is how you can improve your yield that will get your productivity up. We are not in a business where you have a 99% yield. In many cases, there are initial yields on high-end products that are in the 50% range. So a 5% or 10% improvement in these yields is significant (Richard Downing, a senior VP of manufacturing at Seagate, quoted in Electronic Business Asia, Feb. 1997, p. 35).

Section 2 of this paper reviews the existing literature on yield-driven processes. Section 3 analyzes yields in multi-stage production process. Section 4 motivates our analysis by describing the yield-driven nature of the disk drive industry, and the yield-related decisions its managers must make. Section 5 examines the economics of rework and scrap in detail for a simple process. It concentrates on variable cost and output as the main effects of yield. Section 6 gives our conclusions and points at needed future research.

2. Prior research on yields

The subject of process yields has received considerable attention in various disciplines. We can group this research into four streams. First, engineering reports describe yield problems in specific industrial processes and provide technical solutions. Second, operations management and operations research models support production management of yield-driven processes. Typical concerns are inventories, inspection plans, order releases, scheduling and sequencing, and other issues related to production planning. Third, there is an organizational learning literature on how to improve yields and reduce ''waste''. Much of it is empirical- or case-based. Fourth, quality management research outlines a number of principles to reduce the ''cost of quality''. Yield losses correspond to internal quality problems, i.e., problems caught before goods leave the factory.

There are a number of engineering articles and technical reports describing methods of dealing with yield-driven production processes, especially in the semiconductor industry. For example, IEEE Transactions on Semiconductor Manufacturing has several articles per issue related to yields or ''defects''. The emphasis is on methods, concepts, and tools that will improve yields by detecting diagnosing and solving specific problems. Examples include methods of defect classification (Breaux and Kolar, 1996), yield-loss modeling (Stamenkovic et al., 1996), in-line product inspection (Wang et al., 1996), statistical software to analyze process control data (Burggraaf, 1996), and expert systems to provide estimates on quality of certain batches (Khera et al., 1994). This literature is vital to continued technological progress in these industries. As new products and processes push the state of the art, yields fall, and new cycles of yield improvement are needed.

The random nature of yield-driven processes and the resulting challenges for managing production have attracted a number of operations management researchers. Most of this literature takes the production technology, and thus, the yield problems, as given and provides models supporting standard production decisions such as how to manage work-in-process and congestion (e.g., Chen et al., 1988), inspection plans and quality improvement (e.g., Barad and Bennett, 1996), scheduling and sequencing (e.g., Ou and Wein, 1995), and other issues related to production planning (e.g., Denardo and Tang, 1997).

A smaller group within the operations management literature argue that the overall yields of a production process can be improved by effective management of the process. Proposed methods for yield improvement include inspection policies for quick feedback on the quality of the process (e.g., Tang, 1991), keeping the work in progress level low (e.g., Wein, 1992), and effectively combining items from different batches (e.g., Seshadri and Shanthikumar, 1997). In contrast to the engineering literature, these papers focus on improving various performance measures, including yields, without really changing the underlying production technology. This makes them more general across processes, but limits their potency.

The third stream of yield research is at the intersection between production management and organizational research, especially organizational learning, and has contributed some in-depth empirical studies on yield improvement. Mukherjee et al. (1995) categorized various quality projects undertaken at a major manufacturer of wire cord depending on the type of learning approach taken in the projects. A follow-up study (Lapré et al., 1996) links these quality projects to waste reduction (yield improvement). Bohn (1995a; b) looks at factors which influence the speed of yield improvement in semiconductor manufacturing. Kantor and Zangwill (1991) give a theoretical model of waste reduction learning. Like the engineering literature, the organizational literature has little to say on the economic value of yield improvement. For the most part, yield improvement is implicitly treated as a way to reduce costs without looking at other effects.

Finally, under the ''cost of quality'' paradigm as outlined in (Juran and Gryna, 1993), yield losses are viewed as part of internal failure costs and thus, as one of the main drivers of the costs of quality. Juran and Gryna emphasize the need to assign economic values to these quality costs, to make them easier to understand for top management decision-makers. The cost of quality approach is valuable in its recognition of hidden effects from quality problems, and its emphasis on quantifying them. For example, this approach would show that when first-pass yields get high enough, in-process inspection can be eliminated, which has various desirable effects. However, one of the main benefits of yield improvement is ignored, namely the improvement in effective capacity and output.

In the quality literature, yield loss is the extreme form of a defect — the product is unsalable. Therefore, much of the quality improvement literature is applicable to yield improvement. Probably most important are the tools and concepts of statistical process control to yield monitoring and improvement. Again, this is most active for semiconductors; see the survey/tutorial by Spanos (1992). Typical issues include how to detect a defective machine quickly, what inspection policies to set, and how to modify SPC tools such as control charts to cope with the huge amount of data produced by automated semiconductor manufacturing lines.

Although the literature reviewed above has substantially improved our understanding of yield issues in production processes, none of it has provided the basic economic analysis of how yields matter. We attempt to extend the literature in three directions:

- we assign concrete economic values to yield issues (Juran and Gryna, 1993)
- we do not take yields as given, rather, we concentrate on the value of improvement
- we look beyond the cost impacts of yield improvement.

This article can be viewed as an effort to evaluate the value of yield improvement.

3. Multi-stage yield-driven production processes

In this section, we discuss production processes consisting of a sequence of sub-processes, of which at least one has yield below 100%. Although defects can occur anywhere, they are detected mainly at test points. An important question in designing processes with yield losses is the positioning of tests or inspections. Tests are costly, and can sometimes reduce yields themselves. There are various formulations of where to put them. Common rules are to position them before expensive or irreversible operations, at the end of modules in modular subassembly, after low-yield operations (to avoid adding more value to bad units), or immediately after operations targeted for process improvement (to provide fast feedback for learning).

At each test point, items are classified into ''good items'' and various categories of ''defective items''. Whereas good items can continue processing at the next operation, defective items are removed from the line. They can then be either reworked or scrapped.

3.1. Yields and rework

Rework means that some operations prior to the defect detection point must be redone, or defects must be otherwise repaired. Thus, rework changes the capacity utilization profile of the process. In analyzing the influence of yields (and rework) on process capacity, we need to distinguish between bottleneck and non-bottleneck machines. If rework involves only non-bottleneck machines with a large amount of idle time, it has a negligible effect on the overall process capacity.

In many cases, however, rework is severe enough to make a machine a bottleneck (or, even worse, rework needs to be carried out on the bottleneck machine). As the capacity of the bottleneck equals the capacity of the overall process, all capacity invested in rework at the bottleneck is lost from the perspective of the overall process.

A second complication related to rework, which affects bottleneck and non-bottleneck machines, is related to the amount of variability in the process. A yield of 90% means not that every 10th item is bad, but that there is a 10% chance that a given item is bad. Thus, yield losses increase variability, which is the enemy of capacity. The best stochastic case is that yields are Bernoulli, i.e., that the process has no memory. Suppose that bad items at an operation are immediately reworked by repeating the operation. Even if the actual processing time of the operation is itself deterministic, the yield losses force items into multiple passes, and thus make the effective processing time for a *good* item a random variable. Hopp and Spearman (1996) (Section 12.3) show for this case that the variability (measured by the squared coefficient of variation) in the effective processing time increases linearly with $(1-y)$.

Capacity losses due to variability can be partially compensated by allowing WIP after each operation with yields below 100%. The larger these buffers, the more the capacity-reducing impact of variability can be reduced. However, additional WIP increases costs, lead times, and throughput times; it also can hurt problem detection and solution, thereby reducing yields.

3.2. Yields and scrap

Scrap occurs when bad items are discarded. Final output is correspondingly reduced. Rework is generally preferable, but sometimes, it is technically infeasible or uneconomic. An economic comparison of scrap and rework is given in Section 5.

Strictly speaking, scrap is a special form of rework, where the rework loop includes all operations between the defect generating machine and the beginning of the process. The impact of scrap losses on system capacity are even stronger than in the rework case, since additional capacity must be added at *all* stations upstream of yield test points, with the most capacity needed at the start of the process. It does not matter where the defective unit is actually created, only where it is detected. In order to get 100 good parts at the end of the process, more than $100/y$ must be started at the beginning, where y is the cumulative yield all the way through the process. Further, the stochastic variation in load is felt at *all* stages downstream of the yield loss, not just at the stages involved in the rework loop.

This points to the importance of capacity planning in yield-driven processes. If yields and resulting rework requirements are known at the time a line is laid out and remain roughly constant, then capacity planning and

Table 1
Summary of yield effects on cost

	Rework is done	Scrap
Material-related costs	Incremental material to replace bad components	All material up to failed test is lost
Labor-related costs	Rework labor	All labor up to failed test is lost
Capacity-related costs	More capacity needed in the rework loops of process	More capacity needed at all stages upstream of failed tests
Variability-related costs	WIP cost to buffer variability	WIP still needed but less effective; more capacity needed to counteract
	Lead time variability in make to order systems	Extra large lots needed in make-to-order systems
		Line never perfectly balanced; more capacity needed to counteract

line balancing is done by increasing the capacity at each station enough to handle its anticipated yield-caused extra load. With scrap, it takes the form of increasing the capacity at all upstream stations enough that they can keep up with demand at the end of the process. Usually, however, yields are neither known accurately in advance nor are they constant over time. Instead, the aggregate yield shows both a positive trend (learning) and a week-by-week variation which cannot be buffered out economically, even by finished goods inventory. Therefore, once a process starts up, the actual capacity at each stage usually will be ''sub-optimal'' by static criteria.

A related complication arises in make-to-order situations with scrap. To respond to a customer order of N units, we must start N/y at the beginning to compensate for the expected yield losses. This approach would work fine, if yields were deterministic. However, since they are not, the production scheduler has to trade off the costs of making too much against the cost of making too little. Mathematically this is a newsboy-type problem (Table 1).

3.3. Cost and value at different stages of the process

In addition to its effect on capacity, yields determine the value that a good unit of WIP has at various stages in the process. This information is, for example, important in deciding where to concentrate process improvement efforts. A two-point yield improvement has different value at different places in the process.

The value of a good unit of WIP also help to decide whether it is more economical to scrap a defective item or to rework it. For example, suppose that after a test a defective item can be reworked for a labor cost of US\$10, with a 90% chance of success and a 10% chance that the item must be scrapped. Is it better to pay for rework, or to scrap the item? Clearly if x is the value of a good item at that point, the decision rule is to rework if $10 < 0.9x$. However, determining x is not trivial.

At the beginning of the process, the value of a good item equals the cost of raw materials. At the end of the process, the value is given by the marginal revenue from a good item that can be sold. The value of a good item increases as it moves through the process, even if no additional material is being added. Let y_n be the yield at the nth stage. If there are no binding capacity constraints, the value leaving stage n is approximately $1/y_n$ times the sum of the value entering stage n and the variable costs at stage n.

This gives two different ways to calculate value: cost-based working forward, and price-based working backwards. The two will be equivalent if there is no binding capacity constraint, and differ if there is one. The discontinuity in value comes at the bottleneck operation(s). After the bottleneck, value is based on selling price; before the bottleneck, it is based on cost. An analogous effect will be formally discussed in Section 5. It can have surprising consequences when cheap raw materials are transformed into expensive products (e.g., semiconductors).

4. Yield-related problems in disk drive production

In this section, we discuss the managerial importance of production yields based on a particular industry, HDDs. We describe the production of HDDs as well as the managerial questions related to yields. The answers to these questions will be provided by the economic analysis in Section 5.

4.1. Product and process technology

A HDD is a magnetic data storage device that reads, writes and stores digital data. The main components are the head disk assembly (HDA) and the printed circuit board assembly (PCBA). The HDA includes the head, media (disks), head positioning mechanism (actuator) and the spindle motor. The HDA is sealed in an enclosure that shields the HDD from dust and other particles, keeping a contamination-free environment over the life of the product. The PCBA includes custom-integrated circuits, an interface connector to the main computer and a power connector.

The manufacturing of HDDs is a complex process. The sub-micron flying heights of the head over the media make the HDD vulnerable to contamination by small particles, requiring a clean room environment for many steps in production. A second challenge in the assembly of a HDD is given by the high degree of miniaturization of the components (especially the head) and the extremely small tolerances in putting the parts together. Third, magnetic tolerances are very tight. These challenges make the production of HDDs a yield-driven process.

Assembly of HDDs starts with the assembly of the actuator mechanism, head sub-assembly, disks, and spindle motor in a housing to form the HDA. Although this process can be partially automated, it typically is a largely manual operation. After the HDA is assembled, an operation known as servo writing is putting a basic logical format on the disks. This is followed by several optical and functional tests, which are typically highly automated. Finally, the PCBA is added to the HDA and the completed unit is formatted and tested prior to packaging and shipment.

Table 2 includes information about typical component prices and other production data, including yields. As we will discuss below, it is typically beneficial to conduct rework on HDDs. The information for the rework process is also given by Table 2.

4.2. Competitive pressures on HDD production

The HDD industry is a typical ''high technology'' industry, meaning that to survive, companies must be on the cutting edge, with rapid product innovations. Most product generations last less than 1 year. Furthermore, because competitive pressure forces products to be brought to market before they or their manufacturing

Table 2
Typical HDD data

	Initial production		Rework
Material cost	135 (US$/drive)		27 (US$/drive)
Direct labor	0.9 (h/drive)		1.35 (h/drive)
Yield rate	60 (%)		70 (%)
Testing time	1 (h/drive)		2 (h/drive)
Set of heads	1 (unit/drive)		0.25 (unit/drive)
Selling price		300 (US$/drive)	
Demand		150,000 (drives/month)	
Wage rate		US$6/h	

processes are fully understood, production techniques are at low stages of knowledge. A low stage production process is one that is not well-understood and may behave unpredictably (Bohn, 1994).

This situation usually has two key consequences. First, production yields are well below 100%. Because the production process is poorly understood, inevitably much of what is made does not work properly. Over time, as more is learned and process problems are identified and solved, yields increase, but they never reach 100% and often never get close to it.

The second consequence of being on the cutting edge is that the product is in short supply. Initial production volumes are usually low, because of a variety of problems at the manufacturer or its key suppliers. If the product is successful, demand exceeds supply. Over a period of months, the manufacturing plant strives to increase output through a process known as ''ramp-up,'' the gradual acceleration of manufacturing output from zero to full capacity. Although other forces also come into play, again the key driving force behind ramp-up is usually learning of various kinds. Machine downtime decreases as causes are identified and fixed. Bottlenecks are detected and circumvented. More workers are trained for the labor-intensive production steps.

Notice that low-yields exacerbate the problem of short supply. After all, the other production problems are dealt with and units are produced, not all of them work properly. Thus, one way in which output increases is by increasing yields.

4.3. Managerial questions

The central research question of this article is *''What is the economic value of an x% yield improvement?''* The following three managerial decisions are driven by this economic value.

First, the economic value of yield improvement is a crucial input in making *process improvement decisions*. Most process improvement decisions in the HDD industry are geared towards yield increases. Several consulting companies (including a company called Yields-Up) promise to improve production yields. Similarly, purchasing decisions of new equipment or formation of Kaizen teams can lead to higher yields. Whereas the economic cost of such projects can be computed easily, understanding their economic pay-back requires an answer to our value-of-yield-improvement question.

Second, when companies make decisions about new plants and processes, they often have to *choose among a range of geographic locations*, technologies, and workforces. The disk drive industry is characterized by a strong separation into two geographic clusters: most product development is done in the US, whereas 65% of the assembly is done in Southeast Asia. Further, there is a trend towards moving some manufacturing to countries with even cheaper labor, such as the Philippines and mainland China (for a detailed description of the global patterns of this industry, see Gourevitch et al., 1997). In many cases, moving into a new country has the potential to affect yields, particularly during ramp-up of advanced products. Workers and engineers in the new factory will not be as fast at debugging problems, or as able to communicate with developers for joint problem solving. In addition, infrastructure differences among countries may affect ramp-up and yields.

To what extent is there actually a trade-off between wage rates and yields in HDDs? Evidence on this is sketchy and anecdotal, in part, because of the general confidentiality of yield information, and in part, because it is a lot easier to measure wage effects of a workforce than to measure yield effects. One disk media company, HMT, says publicly that it manufactures in California because it is easier to ramp-up new products to high-yield quickly there. However, many of HMT's competitors are building their capacity additions near their customers' assembly plants in SE Asia. In HDD-assembly, there is general agreement that Singapore today has assembly capability and yield as good or better than anywhere in the world. One executive who played a key role in the early SE Asian migration of the industry stated that in 1988, yields in Thailand could not compete with those in Singapore, due to superior worker ability to ''tweak'' production processes (Interview with S.G. Tien, cited in Doner, 1998). Therefore, the effects of yields need to be evaluated at the same time as other consequences of factory location, and are likely to have a similar magnitude of impact on the bottom line.

Third, *automation generally improves yields*, especially as components get smaller and smaller. At the same time, automation reduces the need for labor. Again, an informed choice concerning an automation equipment purchase decision requires a detailed understanding of the value of yield improvement.

The last two questions (location and automation) broaden the scope of our earlier discussion to include different wage rates. Thus, when comparing different locations or technologies, we not only need to consider the economic effect of yields, but also look at the effect of changes in wage rates. In order to support managerial decisions, our analysis therefore must be extended to: *''What effect do wages have on the contribution? What is the effect of yields? Which of the two effects dominates under what conditions?''*

A final question we aim to answer in our analysis is related to the above discussion of the value of a good unit. As yield losses in HDD processes occur at various stages in the process, including upfront operations like component fabrication, as well as back-end operations such as final assembly, the value of a good unit changes drastically throughout the value chain. This leads to the question *''When is it beneficial to rework, and when to scrap, a defective item?''*

4.4. Current practice

In the past, HDD producers answered these questions using standard cost accounting techniques. However, accounting systems are quite poor at dealing with yield issues, both prospectively and retroactively. Scrap costs are often treated as a separate cost pool, which is not carefully allocated back to individual points in the process. Even more basic, accounting systems only look at the cost-based numbers, not the price-based values. Sensitivity analysis on the effects of alternative production methods with different yields is very difficult with conventional cost accounting. Because of these problems with accounting numbers, experienced managers in yield-driven industries often rely on intuition for relevant decisions, while inexperienced managers make mistakes. Even the decision on what to rework and what to scrap, seemingly a technical decision, turns out to be an economic choice, and one not captured in a cost-based accounting system.

More recently, high-tech companies are using cost-of-ownership (COO) analysis to address the above questions. Consider, for example, choosing between an automated and non-automated machine. As discussed above, the automated machine is likely to have higher production yield and lower labor cost, but will also require a larger upfront investment. To support the purchasing decision, the company performs a so-called COO analysis of the two machines (typically implemented in form of a spreadsheet). Yields and capacity utilization are important inputs for such COO models. The COO analysis computes the production cost of a good HDD from the automated machine and compares it with the cost from a non-automated process. The total economic value of the yield and wage differences is then computed using an estimated total volume of drives produced over the life of the equipment. If this lifecycle cost of owning the non-automated machine (with lower yields and higher labor cost) exceeds the difference in purchasing cost, the automated machine is acquired.

COO models are better than a pure accounting approach, but are still inadequate, as we will now show.

5. Economic analysis

To address the questions raised above, we have, based on research at several companies in the information storage industry, developed a simple mathematical model. The model is targeted towards a managerial audience and is based on a strong simplification of the complex HDD production process. It allows us to demonstrate how the current managerial practice of analyzing yield-driven processes *dramatically underestimates* the value of yield improvements. The real value of yield improvement is even larger than what is captured in our model, since we abstract from the capacity reducing effects of variability discussed in Section 3.

5.1. A two-stage model

Our model is based on the abstracted production process depicted in Fig. 1. The first stage corresponds to the "normal" production process while the second stage is a special rework process where defective items are repaired to eventually meet the quality specifications. We will index our variables with *in* for initial production, and *rew* for rework process. The amount of direct labor required in stage in is denoted by L_{in} [h/unit], with a wage rate w [US$/h]. M_{in} denotes the material costs per incoming item at stage in [US$/unit]. Initial production ends with a quality inspection of every single item. We define the first-pass yield (y_{in}) as the proportion of items that pass this test and can thus be put on the market. For the moment, we will assume that all good items are sold, at a price p.

If the item does not pass the test at the end of the initial stage, it enters a rework process where some of its components are replaced, adjusted, or repaired. Rework can be a rather complicated production process, including multiple workstations, a substantial amount of testing and diagnosing devices, and multiple passes. For our economic analysis, we do not need the microstructure of this rework process. All we need are data describing the aggregated behavior of the rework system, including:

y_{rew}: rework yield (proportion of items that are successfully reworked so that they can pass the quality test)

L_{rew}, M_{rew}: average direct labor requirements/average material costs, per part entering rework.

If the rework is successful, the item can be sold at price p; thus, the full functionality of the product can be reached in the rework. Otherwise, the item is scrapped.

Let K be the number of items started at stage in in a given time period, e.g., a month. Of these K items, Ky_{in} can be put on the market without any rework. $K(1-y_{\mathrm{in}})$ enter the rework process, of which $K(1-y_{\mathrm{in}})y_{\mathrm{rew}}$ can be reworked to become a sellable output. This creates an overall (composite) yield of:

$$y_{\mathrm{c}} = y_{\mathrm{in}} + (1-y_{\mathrm{in}})\,y_{\mathrm{rew}}. \tag{5.1}$$

Thus, rework raises effective yields from y_{in} to y_{c}, an improvement of $(1-y_{\mathrm{in}})y_{\mathrm{rew}}$.

In most high-tech industries, specifically in HDDs, rework is more difficult than initial production, and therefore, y_{rew} is often less than y_{in}. Reasons rework is harder include the need to disassemble the defective items, which may cause damage; faulty initial diagnosis so that the real problem is not what gets repaired; and some problems simply cannot be fixed but rather the whole item must be discarded. On the other hand, the rework process can be repeated several times to improve its yield ("rework of the rework"). As a result, y_{rew} can be higher than y_{in}. Typical values in assembly of a high-end disk drive are $y_{\mathrm{in}} = 60\%$, $y_{\mathrm{rew}} = 70\%$ and therefore, $y_{\mathrm{c}} = 88\%$ (see Table 2).

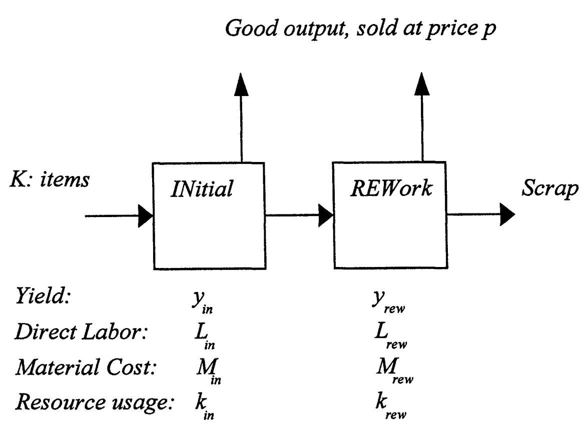

Fig. 1. Production process with rework.

How much good output will come out of the process in Fig. 1, if K units are started? There are two cases, capacity-constrained and capacity-unconstrained. The easier case is the one where production has sufficient capacity to keep up with demand. The factory can make as much as it can sell.

Let D denote the demand per period. To have final output D, the factory must start $K = D/y_c$ items into the initial stage. We assume that D is known and that demand is fulfilled at a market price p, which is — as a result of many other suppliers on the market — out of control of the individual company. Similarly, we do not consider any price discounts or other aspects of marketing or competitive interaction.

The capacity constraint is related to the production process, and happens especially during ramp-up periods. The factory cannot produce as much volume as it would like to. Such constraints can be a result of component shortages, limited production equipment, limited trained workers, or other factors.

Let K_{max} describe the number of units available of a scarce resource. In disk drives, K_{max} could be the number of heads available from a supplier or the overall testing capacity in the period. The K_{max} units of the scarce resource are consumed at a rate of k_{in} units per incoming item at initial production and at a rate k_{rew} per reworked item. Then for K items started, Kk_{in} units of resources will be used at stage in, and on average $K(1-y_{in})$ items will need rework, consuming a further $K(1-y_{in})k_{rew}$ units of scarce capacity. This must be kept at or below available capacity K_{max}, giving us:

$$K \le \frac{K_{max}}{k_{in} + (1-y_{in})k_{rew}} \tag{5.2}$$

Combining the two cases, production unconstrained and production capacity-constrained, we have:

$$K = M_{in}\left\{ \frac{D}{y_{in} + (1-y_{in})y_{rew}}, \frac{K_{max}}{k_{in} + (1-y_{in})k_{rew}} \right\} \tag{5.3}$$

This is the number of starts into the factory. The effective output level is given by Ky_c.

5.2. The effect of yield on contribution

Our analysis focuses primarily on one performance measure: contribution per period, which we define as revenues minus variable costs (materials and labor). Contribution roughly equals profit before tax plus fixed costs, so that a US$1 change in contribution gives a US$1 change in pre-tax profits. The advantage of working with contribution per period is that it includes both cost and revenue aspects. Traditional analysis using Cost of Goods Sold (COGS) or other cost-based measures neglects the positive effects of yields on sales and revenues. We will discuss the relationship between these measures further below.Define π as the per period contribution. Then:

$$\pi = \underbrace{Ky_c p}_{revenues} - \underbrace{K\left[\overbrace{wL_{in} + M_{in}}^{initial} + \overbrace{(1-y_{in})(wL_{rew} + M_{rew})}^{rework}\right]}_{\cos ts} \tag{5.4}$$

The objective of our analysis is to develop a better understanding of how changes in the wage rate w, or changes in process performance such as yields y_{in} and y_{rew}, affect the overall economics. We will do this in two steps. First, by deriving the partial derivatives of π with respect to yield rates and wage rates, we develop a number of qualitative insights about how these variables influence economic performance. Second, we link the analysis to the managerial questions raised in Section 4 and estimate concrete economic values for changes in yields and wage rates.

Two cases have to be considered, depending whether the capacity constraint is binding. If it is not, we assume that input and production capacity are unlimited and the only constraint is provided by the market. In this case, the number of starts required to meet demand can be computed as $K = (D/y_c)$. The second case is the opposite. The factory can sell whatever it can make, but there is a limited supply of components and capacity. The two cases are now analyzed in greater detail.

5.2.1. Analysis of case 1: market limit only

If the factory is not capacity-constrained, it will make and sell D units. The number of starts is provided by $K=(D/y_c)$, which, substituted into Eq. (5.4), gives an overall per period contribution which we denote by π_{market}:

$$\pi_{market} = \underbrace{Dp}_{revenues} - \overbrace{\frac{D}{y_{in} + (1 - y_{in})y_{rew}}}^{starts-to-meet-demand} \overbrace{[wL_{in} + M_{in} + (1 - y_{in})(wL_{rew} + M_{rew})]}^{cost-per-start} \quad (5.5)$$

The first term in Eq. (5.5) describes the revenues and the second term the direct costs. Note that the second term in Eq. (5.5) can be interpreted as D times the cost per good unit, which is cost-per-start divided by composite yields.

There are several interesting questions about the production process which this model allows us to answer. Specifically, what is the effect of increases in the yield at each stage? What effect do wages have on the contribution? We can approach these questions by looking at the derivatives of contribution with respect to each of these variables. First, consider wages.

The partial derivative of contribution with respect to wage rate can be computed as:

$$\frac{\partial \pi_{market}}{\partial w} = -D\frac{L_{in} + (1 - y_{in})L_{rew}}{y_{in} + (1 - y_{in})y_{rew}} \quad (5.6)$$

From Eq. (5.6) we see that an increase in wage rate reduces the contribution proportionally to $(L_{in} + (1 - y_{in})L_{rew})$, a factor that describes the expected amount of labor that is spent per start. Second, consider the partial derivative with respect to first-pass yield:

$$\frac{\partial \pi_{market}}{\partial y_{in}} = \frac{D}{[y_{in} + (1 - y_{in})y_{rew}]^2}[wL_{in}(1 - y_{rew}) + M_{in}(1 - y_{rew}) + wL_{rew} + M_{rew}]. \quad (5.7)$$

As yield enters the contribution expression only through costs, the partial derivative does not include any changes in revenues. Thus, in the market-limited case, the yield increase pays off purely in the form of cost reduction. This cost reduction effect can be decomposed into:

- savings in the rework process at rate $M_{rew} + wL_{rew}$ as fewer items have to be reworked
- savings in the production process, as, with increased yields, fewer items have to be started in order to fulfill the same amount of demand D.

Finally, the partial derivative with respect to rework yields is similar to Eq. (5.7) and is:

$$\frac{\partial \pi_{market}}{\partial y_{rew}} = \frac{D(1 - y_{in})}{[y_{in} + (1 - y_{in})y_{rew}]^2}[wL_{in} + M_{in} + wL_{rew}(1 - y_{in}) + M_{rew}(1 - y_{in})] \quad (5.8)$$

the impact of a rework yield increase on the production cost has to be scaled by $(1 - y_{in})$, because only that fraction of starts enter rework. Otherwise, the effect is qualitatively similar to improving first-pass yield.

5.2.2. Analysis of case 2: capacity constraint

Binding constraints on both raw material and equipment are typically observed during the ramp-up of a new process. For a new generation disk, the key components are redesigned, pushing the component supplier to the frontier of what currently is producible. This translates into low-yields and low production volumes at the supplier, and ultimately to a shortage of components for the disk manufacturer. An example of this is the shortages of platters (media). If a drive contains four platters, and on average one platter must be replaced per reworked drive, we have $k_{in} = 4$ and $k_{rew} = 1$.

Another scarce resource during the ramp-up is the capacity of testing equipment. Given the substantial cost of testing equipment, testing is frequently the bottleneck activity in the overall process. Rework is especially testing intense, and one drive can need four h of testing ($k_{rew} = 4$). For the normal production process, the numbers are lower, but still reach 1 h per start ($k_{in} = 1$). In fact, k_{in} and k_{rew} tend to fall during ramp-up. Initially, drives must be ''extra-tested'' to ensure all problems are caught. Later, engineers learn how to test more narrowly for specific problems, allowing faster testing.

If the factory is constrained to K_{max} units of the scarce resource per period (e.g., 100,000 available heads per month, 10,000 h available testing time), the number of starts is restricted to $K = ((K_{max})/(k_{in} + (1 - y_{in})k_{rew}))$, as stated in Eq. (5.2). Substituting this into Eq. (5.4) gives an overall per period contribution of:

$$\pi_{capacity} = \overbrace{\frac{K_{max}}{k_{in} + (1 - y_{in})k_{rew}}[y_{in} + (1 - y_{in})y_{rew}]p}^{\text{revenues}} \underbrace{- \frac{K_{max}}{k_{in} + (1 - y_{in})k_{rew}}[wL_{in} + M_{in} + (1 - y_{in})(wL_{rew} + M_{rew})]}_{\text{production costs}} \quad (5.9)$$

where $\pi_{capacity}$ is defined as the per period contribution in the capacity-constrained case. As in Eq. (5.9), the first term is revenue, and the second is production cost.As before, the effect of a change in wage rate is straightforward:

$$\frac{\partial \pi_{capacity}}{\partial w} = -\frac{K_{max}}{k_{in} + (1 - y_{in})k_{rew}}[L_{in} + (1 - y_{in})L_{rew}] \quad (5.10)$$

which again is linear in the expected amount of labor per start.The effect of first-pass yield is more complicated, with a partial derivative of:

$$\frac{\partial \pi_{capacity}}{\partial y_{in}} = \frac{K_{max}\, p}{[k_{in} + (1 - y_{in})k_{rew}]^2}[k_{rew} + (1 - y_{rew})k_{in}] + \frac{K_{max}}{[k_{in} + (1 - y_{in})k_{rew}]^2}[k_{in}(wL_{rew} + M_{rew}) - k_{rew}(wL_{in} + M_{in})]. \quad (5.11)$$

Note that unlike the market-constrained case, a change in yield now affects the revenue as well as the costs. Thus, during the ramp-up, an improved yield not only reduces unit costs, but also creates an increase in net capacity. This is captured in Eq. (5.11) where the first expression describes the increased revenue resulting from a better usage of the K_{max} units of the scarce resource. The revenue effect comes from two sources. One is the direct effect of yield improvement on output. The second is an indirect effect of reducing capacity consumption during rework, which allows an increase in starts, and therefore in output and revenues.Finally, consider the partial derivative of contribution with respect to y_{rew}:

$$\frac{\partial \pi_{capacity}}{\partial y_{rew}} = \frac{K_{max}\, p}{k_{in} + (1 - y_{in})k_{rew}}(1 - y_{in}). \quad (5.12)$$

Similar to Eq. (5.11), we see that for the capacity-constrained case, a yield improvement creates an increase in net capacity (and thus in revenues). There is no impact on total costs, as once an item has entered the rework process, the investment for labor and material is determined. The same holds for capacity consumption, as once the item has entered rework, it uses up k_{rew} units of capacity. Thus, the only effect of y_{rew} is on output. Of course, the average cost per good unit falls as a result of the increased output.

Table 3 summarizes the effects of wage rate and yield changes on contribution per period. First, we see that the effect of wage rate is the same for both cases, and is linear in the expected amount of labor time per item started. Second, depending on which of the two constraints on starts is binding, an improvement in first-pass yield creates cost reduction and an increase in revenue (capacity is binding) or a cost reduction effect only (demand is binding).

Typically, in industries with long product lifecycles and few introductions of new products, the cost benefit is dominant. This explains why previous OM literature has largely emphasized cost aspects of yield management. In industries with short lifecycles and a high rate of product replacements, however, the importance of the revenue effect is higher and one major effect of yield improvement is an increase in net capacity. As Table 3 shows, this effect is even stronger than linear.

5.2.3. Incentive problems caused by yield accounting

The ambiguous effect of yield improvement on production costs can lead to incentive problems. The second expression in Eq. (5.11) defines the change in the overall labor and material costs. Interestingly, total variable costs can rise or fall. Suppose:

$$\frac{(wL_{\text{rew}}+M_{\text{rew}})}{k_{\text{rew}}} > \frac{(wL_{\text{in}}+M_{\text{in}})}{k_{\text{in}}}.$$

That is, rework cost per unit of capacity consumed in rework is larger than the production cost per unit of capacity consumed in production. Then, the overall costs actually increase rather than decrease if yields improve. An example of such situation would be the case of testing devices being the capacity constraint. An improved first-pass yield frees up testing capacity in the rework process. This capacity can be reallocated to the normal production process, enabling the factory to increase its starts, which ultimately leads to an increase in total costs. This effect, of course, is more than compensated by the increase of revenue, so that in any case, there is an economic gain from the yield improvement.

This situation where both costs and profits go up can have perverse incentive effects on production managers. In one company we studied, overhead costs of the factory got allocated to production lines based on their direct production costs. The manager of a production line who improved her yield had higher total cost, and got punished through a larger allocation of overhead costs. The revenue and profit enhancing benefits of the yield

Table 3
Effects of wage rate and yield changes on contribution per period

	Limited market	**Capacity-constrained**
Effect of wage rate reduction	+ Linear increase of contribution per period (proportional to the expected labor time per started item)	
Effect of increase in first-pass yield	+ reduction in stage 1 costs + reduction in stage 2 costs ⇒ lower total and unit costs	+ more output (direct effect) + better use of scarce resource, this allows more starts (indirect effect) ⇒ overproportionally more revenue + reduction in stage 2 costs − increase in stage 1 costs ⇒ lower unit costs, total costs unclear
Effect of increase in rework yield	+ reduction in stage 1 costs + reduction in stage 2 costs lower total and unit costs	+ more output (direct effect) ⇒ linearly more revenue, lower unit costs

improvement were not included in the accounting system. Note that such perverse incentive effects are created by cost-driven measures. Thus, any measure that does not take into account the revenue aspects of changes in the production process is misleading. A further complication is that initial production and rework costs may be charged to different cost centers. If yields go up, rework costs come down but initial production costs always rise, possibly further embarrassing the manager whose yields improved.

5.2.4. Financial data from the HDD industry

To explore the effects of yield and wage changes on contribution, we now reproduce Table 3 with actual cost data (Table 4). For the case where the factory can produce all the demand (first column), we assume that 150,000 units can be sold per month. For the next column, we are working with 150,000 read–write heads available, and for the last column the constraint is given by 200,000 h of available testing time. Table 4 shows the impacts of different process changes. As a result of the different constraints, the output varies across columns: 150,000 in the capacity-unconstrained case, 120,000 for the read–write heads, and 98,000 for the case when the testing equipment provides the binding constraint. All numbers are thousands of dollars per month of contribution, unless stated otherwise.

The effect of a reduction in wage rate is straightforward. As seen in Eqs. (5.6) and (5.10), a change in wage rate has a linear effect on contribution. The contribution gains are, depending on the availability of capacity, between US$160,000 and 245,000 per month.

The effect of a 5% improvement (five percentage points, from 60% to 65%) in first-pass yields is more complicated. Consider the market-constrained case first. The yield improvement results in a cost improvement at both regular production and rework. This translates into a reduction of US$4.90 per good unit, or, in other words, a US$735,000 per month improvement in contribution and a 2.7% cost reduction.

For the read–write head constrained case, there is no improvement in production costs. Some savings can be achieved in the rework process, but the costs in the regular production go up rather than down. This is a result of having 1.15% more starts, enabled by a better usage of the read–write heads. More important, the increase in starts provides a 2.87% increase in output (and thus in revenue), corresponding to a US$1 million increase in monthly contribution. These effects are even stronger in the testing constrained case. Although production costs increase by over US$0.8 million per month, this is more than offset by a US$2 million increase in revenues, with a net-benefit of US$1.4 million per month.

Table 4
Effects of wage rate and yield changes in the disk drive case (All numbers are thousands of dollars of contribution per month, unless otherwise noted)

	Limited market	**Capacity-constrained (read–write heads)**	**Capacity-constrained (testing equipment)**
Output per month	150,000 units	120,000 units	98,000 units
Revenue per month	45,000	36,000	29,333
Contribution per month	18,675	14,940	12,173
Effect of US$1/h wage reduction	+245	+196	+160
Effect of 5% increase in the first-	+401 (at stage 1)	2.87% more output	7.69% more output
pass yield (60% to 65%)	+334 (at rework)	1.15% more starts	5.88% more starts
	⇒ 735 (overall)	⇒ 1034 (in revenues)	⇒ 2254 (in revenues)
	(unit cost reduced by	−220 (at stage 1)	−917 (at stage 1)
	US$4.90/disk)	+220 (at rework)	+114 (at rework)
		⇒ 0 (in costs)	⇒ −803 (in costs)
		⇒ 1034 (overall)	⇒ 1451 (overall)
	+532 (at stage 1)	818 (overall)	666 (overall)
	+53 (at rework)		
	⇒ 585 (overall)		

Note that depending on what measure is used, the evaluation of changes differs substantially. To properly capture the overall effects, any metric used for evaluating changes in yields must include revenue aspects.

- Looking at total production cost alone is obviously misleading, since in the last column production, costs actually increase when yields improve.
- Working with unit costs is also misleading. In all three cases, unit costs went down by US$4.90 per unit. Looking at the economic performance, however, we can see that the contribution effect in column three is actually twice as strong as in the first column: 1451 vs. 735.

The effect of an improvement in rework yield is similar. Again, the improvements are stronger for the capacity-constrained cases than for the market-limited case.

5.3. Automation and location decisions

We now have derived the value of yield improvements as well as the benefits of wage changes. However, as discussed previously, specifically the automation and location decisions require a relative comparison between the yield and the wage effect.

The following equations show the ratio between the change in contribution from yield improvement and the change in contribution from wage rate reduction. These apply to the disk drive example introduced in Table 2. As before, we have to distinguish between the market-limited case and the capacity-constrained case:

$$\frac{\partial \pi_{\text{market}}}{\partial y_1} \Big/ \frac{\partial \pi_{\text{market}}}{\partial w} = \frac{149}{-245} = -0.6$$

$$\frac{\partial \pi_{\text{capacity}}}{\partial y_1} \Big/ \frac{\partial \pi_{\text{capacity}}}{\partial w} = \frac{274}{-160} = -1.71$$

In other words, a 1% improvement in initial yield has the same value as a US$0.60/h reduction in wage rate, in the market-limited case. From the numbers, we see that the yield effect is relatively dominating, especially in the capacity-constrained case.

To further explore wages vs. yields, we plot the monthly contribution over a range of possible yields for different wage levels in Fig. 2. We assume for this graph that rework and first-pass yields move in concert. The lower curve corresponds to a wage rate of US$20/h, a level representative of the US or Europe. The middle

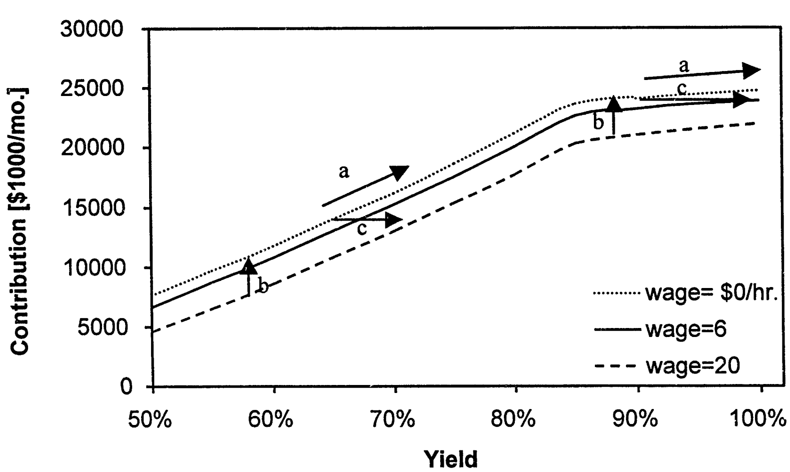

Fig. 2. Impact of yield changes (both types) and wages on contribution.

curve is based on a wage rate of US$6/h, which is about a typical number in Singapore. The upper curve is the extreme case where labor is free, thus, the wage rate is equals to zero. Note the distinct change in slope at about 85% yield. This is where production capacity catches up to market demand.

There are three directions to look at the graphs in Fig. 2. Direction (a) is similar to Table 4. It shows how a yield improvement influences contribution. The left of the two (a)-arrows is at a low-yield. This typically corresponds to a new product. With an increase in yields, the contribution goes up quickly, until at some point, market demand is satisfied. Hereafter, any further improvement in yield has a much lower effect because it does not increase revenue.

Direction (b) corresponds to decreasing the wage rate from US$20/h to US$6/h and ultimately down to US$0/h. The effect is constant in the sense that regardless of the yield level (or the stage in the lifecycle if we take a more dynamic perspective), it gives the same improvement in contribution.

Finally, direction (c) shows the relative comparison between yields and wage rates. Consider the left arrow labeled (c) first. By moving from the beginning to the end of the arrow, we see that an 8% improvement in yields corresponds to moving from a high wage country into a country with zero (!) wages, and still getting the same contribution. Improve the process by eight percentage points and get all labor for free! If we look at the right arrow labeled (c), this picture changes dramatically. At high-yields, a much larger yield improvement is needed to compensate for any wage hike. The reason is that capacity is no longer constrained, so that yield affects costs but not revenue. This confirms our earlier analysis that yield effects are large relative to wage effects, especially when capacity is constrained. The reader may notice that Fig. 2 has steeper slopes than some of the earlier calculations. The reason is that the figure is based on both initial and rework yields improving simultaneously. This is commonly what happens — new knowledge benefits both.

Fig. 2 was generated by assuming that the labor hours per drive remained constant, and the wages per hour changed. Since only the product of these two factors determines labor cost, it can be reinterpreted as showing what happens when the process is partially or fully automated. The wage = 0 line corresponds to a fully automated process. One partnership in the industry, MKE/Quantum, is noted for running ''lights-out'' factories. Seagate tends to use highly labor intensive methods, while others are in-between. A more sophisticated model of automation would look at differential automation levels for production and rework. Our equations can be used in this way, but Fig. 2 implicitly assumes the same degree of labor displacement for both stages.

For automation, we interpret Fig. 2 as showing not how automation should be *traded-off* against yields, but as how automation should be *evaluated* with respect to both yields and labor costs. For example, in a US$20/h wage factory where automation will reduce labor requirements by half (i.e., halfway between the US$20 and US$0 lines in Fig. 2), a yield improvement of four percentage points will double these benefits during the ramp-up period. Once capacity catches up with demand, the labor-saving benefits continue while the yield-improving benefits get smaller.

5.4. When to rework?

So far, we have assumed that rework is always desirable. Sometimes, rework is technically infeasible — rework yield is zero. All our equations hold for this situation by setting:

$$0 = y_{\text{rew}} = k_{\text{rew}} = L_{\text{rew}} = M_{\text{rew}}. \tag{5.13}$$

Not surprisingly, the effects of first-pass yield become larger when rework is infeasible.

More interesting is the case where rework is technically feasible, but uneconomic. It turns out that the criteria on whether to rework depend on all the cost and profit parameters, including whether the factory is market- or capacity-limited. We look at the market limit case first. In this situation, the decision rule on when to rework is to compare contributions with and without rework. More precisely, choose the larger of π_{market} in Eq. (5.5) compared with π_{market} when Eq. (5.13) is substituted into Eq. (5.5). After manipulation, this leads to:Optimal rework rule based on cost: Do rework if

$$\frac{wL_{in} + M_{in}}{y_{in}} > \frac{wL_{rew} + M_{rew}}{y_{rew}}. \tag{5.14}$$

In other words, compare the *cost* per good unit from new production with the *cost* per good unit from rework. Since, in most situations, $M_{rew} \ll M_{in}$ and labor costs are small compared with material costs, this means that rework is almost always a good idea, unless it has very low-yield compared with initial production.

A generalization of Eq. (5.14) covers the case of multiple defect types or symptoms. Rework each defect type j for which (5.14) holds with appropriate values of y_j, L_j, M_j on the right-hand side. Sometimes, a few defect types are so unlikely or expensive to repair that units with that defect type should be scrapped.

The logic for analyzing the capacity-constrained case is similar, with Eq. (5.9) the relevant contribution equation, but the result is very different. The decision rule becomes:Optimal rework rule based on contribution/capacity: Do rework if

$$\frac{py_{in} - (wL_{in} + M_{in})}{k_{in}} < \frac{py_{rew} - (wL_{rew} + M_{rew})}{k_{rew}} \tag{5.15}$$

The numerators of Eq. (5.15) are the contribution per unit started into new production (if no rework is done) and into rework. So Eq. (5.15) amounts to comparing the *contribution per unit of scarce capacity*. Thus, the decision of when to rework is based on very different criteria in the market-limited and capacity-constrained cases.

It is possible that the decision of whether to rework will change during ramp-up. For example, at the start, components may be the scarcest resource, with rework needing many fewer components than new build ($k_{rew} \ll k_{in}$) so that rework is good. Later, as the vendor ramps up component production, test capacity may be the scarcest resource, with $k_{rew}/k_{in} = 4$, and selling prices high so that the expected contribution is about the same for rework as for new build. In this situation, rework reduces profit. Finally, when ramp-up ends, capacity is no longer a constraint, cost becomes the deciding criterion, and rework is again desirable.

More realistically, instead of on/off all-or-nothing policy shifts for rework, the decisions about which kinds of defects and how to rework them will change in increments over time. In the beginning, rework almost everything, at least enough to salvage all scarce components. In the middle, rework only a few defect types. At the end, rework most defects, but perhaps, still scrap units with symptoms that are hard to diagnose, so that probability of success is low. A further complication is that static profit maximization is not the only reason to rework. Ramp-ups should be managed for rapid learning and yield improvement (Jaikumar and Bohn, 1992). Under this condition, rework can also give useful information about how to change the product or process to further improve yields. This means working on failure types which, at present, are not profitable to rework according to the above formulas.

6. Discussion and conclusion

6.1. Concluding implications

This paper has provided an economic model of yield-driven processes. By applying the model to the specific example of one process, HDD assembly, we compared the economic values of wage reduction, yield improvement, and automation. We find that the effect of yield improvement in increasing contribution and profit can be very strong. Especially during ramp-up periods of scarce resources or capacity, it is critical to focus attention on yields. During ramp-up, it is generally not the cost per unit from the product that most affects the company's bottom line, but its total contribution margin. Both traditional accounting approaches and COO models fail to deal with this properly.

We have applied the analysis to the location decision of a disk drive manufacturer. At least, during the initial phases of the product lifecycle, there is no wage rate low enough to compensate for even modest yield losses. A yield drop of 8% has a bigger effect on contribution than going from US$20/h wages down to free labor. Thus, the quality of work done in a specific location by a specific labor force is more important than their wage. Once the product is mature, wages become more important relative to yields, and in some situations, a cheaper labor force could be justified even if it reduced yields. However, the calculations have to be done explicitly. Even a ''productivity-adjusted wage rate'' will not properly adjust for the effect of a cheaper labor force on yields, since yields affect material costs and revenues, and not just labor costs.

Of course, wages and yields are not the only things affected by siting decisions. We have investigated overseas factory siting by the HDD industry, and find that a number of other cost and non-cost factors, such as tax incentives, appear to be important (Gourevitch et al., 1997). Thus, when a country and a firm consider tax incentives, they should investigate yield issues in as much detail as labor cost issues, and weigh their effects against other criteria.

An analogous situation exists for the choice of technology, such as the extent and nature of automation. Many automated technologies affect yields, often for the better. The yield effects of a technology can easily be more important than its effects on labor costs.

6.2. Further research

Our model is static, yet change is a key element of production ramp-up. We are developing a dynamic model of the phenomena, including learning, price reduction, and changes in demand (Terwiesch and Bohn, 1998). As a process successfully ramps up yields and capacity, the value of yield improvement falls, either gradually or abruptly. Over the life cycle of a product, how much net present value is gained by higher initial yields, faster yield learning, or faster capacity ramp-up? Where should managerial attention be focused?

Next, we need to explore other industries and processes to see the relative importance of the factors in our model. For example, component fabrication segments of the HDD have rather different economics than assembly.

Finally, our focus in this article has been on production, simplifying the competitive environment of the producing firm. If firms improve yields more rapidly, they have some choice about whether to exploit this as higher margins or for lower prices.

In addition to the managerial lessons, we believe this paper has political implications. Our findings suggest that US wage rates are not very relevant to ''bringing manufacturing jobs back to America.'' The keys in the high-tech manufacturing game are yields, and speed of bringing products into volume production. These are results of the organization's understanding of the production process. Therefore, training and education of all levels of the current and future workforce, as well as direct development of new technological capabilities at the organization and national levels, are crucial.

Acknowledgements

Funding for this research was provided by Alfred P. Sloan Foundation through grants to the UCSD Information Storage Industry Center. We thank the following for useful comments: Neil Brumberger, Scot Burton, Scott Hampton, Michael Lapré, and two anonymous referees. Rick Dehmel provided invaluable mentoring on yields in the semiconductor industry.

References

Barad, M., Bennett, G., 1996. Optimal yield improvement in multi-stage manufacturing systems. European Journal of Operational Research 95, 549–655.

Bohn, R.E., 1994. Measuring and managing technological knowledge. Sloan Management Review 36, 61–73.

Bohn, R.E., 1995a. The impact of process noise on VLSI process improvement. IEEE Transactions on Semiconductor Manufacturing 8, 228–238.

Bohn, R.E., 1995b. Noise and learning in semiconductor manufacturing. Management Science 41, 31–42.

Breaux, L., Kolar, D., 1996. Automatic defect classification for effective yield management. Solid State Technology 39, 89–96.

Burggraaf, P., 1996. Yield analysis software solutions [IC manufacture]. Semiconductor International 19, 79–85.

Chen, J.M.H., Mandelbaum, A., Van Ackere, A., Wein, L.H., 1988. Empirical evaluation of a queuing network model for semiconductor wafer fabrication. Operations Research 36, 202–215.

Denardo, E.V., Tang, C.S., 1997. Control of a stochastic production system with estimated parameters. Management Science 43, 1296–1307.

Doner, R.F., 1998. Thailand's hard disk drive industry. Information Storage Industry Center, Report 98-02, University of California San Diego.

Gourevitch, P., Bohn, R.E., McKendrick, D., 1997. Who is US? The Nationality of Production in the Hard Disk Drive Industry. Data Storage Industry Globalization Project, Report 97-01, University of California San Diego.

Hopp, W.J., Spearman, M.L., 1996. Factory Physics: Foundations of Manufacturing Management, Irwin.

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

Juran, J.M., Gryna, F.M., 1993. Quality Planning and Analysis, 4th edn., McGraw-Hill.

Kantor, P.B., Zangwill, W.I., 1991. Theoretical foundation for a learning rate budget. Management Science 37, 315–330.

Khera, D., Cresswell, M.W., Linholm, L.W., Ramanathan, G., Buzzeo, J., Nagarajan, A., 1994. Increasing profitability and improving semiconductor manufacturing throughput using expert systems. IEEE Transactions on Engineering Management 41, 143–151.

Lapré, M.A., Mukherjee, A.S., Van Wassenhove, L.N., 1996. Behind the Learning Curve: Linking Learning Activities to Waste Reduction. Working Paper 96/24/TM, INSEAD, 77305 Fontainebleau, France.

Leachman, R.C., 1996. Competitive semiconductor manufacturing survey: third report on the results of the main phase. Berkeley report CSM-31, University of California Berkley.

Mukherjee, A.S., Lapré, M.A., Van Wassenhove, L.N., 1995. Knowledge Driven Quality Improvement. Working Paper 95/48/TM, INSEAD, 77305 Fontainebleau, France.

Ou, J., Wein, L.M., 1995. Dynamic scheduling of a production/inventory system with by-products and random yield. Management Science 41, 1000–1017.

Seshadri, S., Shanthikumar, J.G., 1997. Allocation of chips to wafers in a production problem of semiconductor kits. Operations Research 45, 315–321.

Stamenkovic, Z., Stojadinovic, N., Dimitrijev, S., 1996. Modeling of integrated circuit yield loss mechanisms. IEEE Transactions on Semiconductor Manufacturing 9, 270–272.

Tang, C.S., 1991. Designing an optimal production system with inspection. European Journal of Operational Research 52, 45–54.

Terwiesch, C., Bohn, R.E., 1998. Learning and Process Improvement During Production Ramp-up. Working Paper, The Wharton School/UC San Diego, Philadelphia/San Diego.

Wang, P., Lee, F., Chan, K.M., Goodner, R., Ceton, R., 1996. Yield enhancement in a high-volume 8-inch wafer fab.: II. Yield enhancement programs. Semiconductor International 19, 217–222.

Wein, L.M., 1992. Random yield, rework and scrap in a multistage batch manufacturing environment. Operations Research 40, 551–563.