Quick Response Manufacturing: A Competitive Strategy for the 21st Century

Rajan Suri
Director
Center for Quick Response Manufacturing
University of Wisconsin-Madison
1513 University Avenue
Madison WI 53706
www.engr.wisc.edu/qrm

Abstract

JIT/Lean Manufacturing techniques are considered to be on the forefront of manufacturing strategy and are being implemented by companies around the world. However, JIT was perfected by Toyota over 30 years ago. To stay ahead of the competition, we need to find new ways to compete. While Lean Manufacturing can be a powerful strategy, we show that for certain markets, Lean Manufacturing has several drawbacks. Quick Response Manufacturing (QRM) can be a more effective competitive strategy for companies targeting such markets. Specifically, QRM is more effective for companies making a large variety of products with variable demand, as well as for companies making highly engineered products. It is our view that with the growing use of CAD/CAM as well as internet-based ordering, serving such markets will be increasingly important for competitiveness. In this article, we give an overview of the QRM strategy. Then we show why, in these market segments, Lean Manufacturing strategies are not appropriate, and QRM has greater competitive potential. Companies implementing Lean Manufacturing thus need to evaluate whether, in some of their markets, QRM would give them more competitive advantages. We conclude this article with the prerequisites for success in QRM implementation.

Introduction

In the last two decades, the success of just-in-time (JIT) manufacturing techniques has led to their adoption by corporations throughout the world. More recently, the strategy based on these Japanese methods has been described and popularized under the name of "Lean Manufacturing" (Womack et al., 1990; Womack and Jones, 1996). Based on the impressive results achieved by companies that have adopted JIT/Lean methods, most industrial corporations are turning to JIT/Lean. On the other hand, we must remember that JIT/Lean strategies were developed and perfected by Toyota over 30 years ago. How can we expect to stay ahead of the competition by adopting 30-year-old strategies? In fact, we will show that in some markets, JIT/Lean strategies have major shortcomings. Instead, recently a few companies have been leapfrogging their competition via a powerful strategy called Quick Response Manufacturing (QRM).

Today's manufacturing world abounds with new acronyms and strategies. In the last few decades, executives have had to consider implementing JIT/Lean, TQM, TPM, Kaizen, and Reengineering, to name only a few. So how does QRM differ from these strategies, and why would a company want to use QRM instead of the currently popular Lean Manufacturing? In this article, we give an overview of the QRM theories. We also show that while manufacturing companies are trying to reduce their lead times, most managers still support policies that *increase* their company's lead time. Then we show why, in certain market segments, Lean Manufacturing strategies are not appropriate, and QRM has greater potential for creating competitive advantage. Finally, we end with the prerequisites for success in QRM implementation.

What is QRM?

QRM is a companywide strategy that pursues the reduction of lead time in all aspects of a company's operations. To gain more insight, it is useful to define QRM in two contexts: (i) Externally, as perceived by customers, QRM means responding to those customers' needs by rapidly designing and manufacturing products customized to those needs. In so doing, we will show that QRM goes beyond the established goals and even the capabilities of Lean Manufacturing. (ii) Internally, in terms of a company's own operations, QRM focuses on reducing the lead times for all tasks in a company, resulting in improved quality, lower cost, and of course, quick response.

QRM is a practical strategy: it embodies the mindset of pursuing lead time reduction, along with detailed management principles, manufacturing methods, analysis techniques and tools, and a step-by-step methodology to achieve the desired reduction in lead times.

A typical company implementing QRM has realized the following benefits: reduction in lead times of 80-95%, reduction in product cost of 15-30%, on-time delivery performance improving from 60% to 99%, and reduction in scrap and rework by 80% or more (see Tubino and Suri, 2000).

What Differentiates QRM from Lean

Although we will provide detailed discussions later in this article, the reader may find it useful to have a quick summary of the points that differentiate QRM from JIT/Lean strategies.

- First and foremost is the QRM mindset: the driver for all the principles and strategies in QRM is reduction of lead time. This time-based mindset results in many operating policies that are different from traditional ones. In contrast the driver in JIT/Lean is waste reduction. Below we will see examples of where the QRM mindset may be more effective in some markets.
- Although the business press has been talking about the importance of lead time reduction, or "speed", for over a decade, we find that most companies still lack the knowledge and the tools to effectively reduce their lead times. Worse still, policies are in place that are lengthening, rather than shortening, lead times. QRM devotes a substantial amount of effort in educating management and workers on why these traditional policies result in long lead times, and in showing them the QRM principles that must be put in place instead.
- QRM is a companywide strategy. While the original implementation of JIT/Lean at Toyota may well have encompassed the whole company, most Western implementations of JIT/Lean have focused on manufacturing and materials management. In many cases, JIT/Lean has been interpreted even more narrowly as merely implementing a "pull" system with "kanban" cards. In contrast, QRM clarifies at the outset that it is a companywide strategy with implications far beyond the shop floor, and principles for other company areas, such as office operations, are clearly presented as part of the QRM philosophy.
- QRM provides rational principles and tools for lead time reduction. QRM uses an understanding of system dynamics, and exploits this understanding to define the best structures and policies that will reduce lead times. QRM begins by educating employees and giving them insight into these system dynamics. This then helps justify, to both management and workers, the need for changes in policies. State-of-the-art analysis tools such as the MPX software package (see the Bibliography) incorporate this analysis of system dynamics and help to derive the specific changes needed and to quantify the benefits that would be achieved.
- For companies making a large variety of products with variable demand, as well as for companies making highly engineered products, the JIT/Lean strategy of "pull" is either wasteful or breaks down altogether. For such companies, QRM provides an alternative strategy called POLCA which combines the best features of "push" and "pull" without their drawbacks. (POLCA is described later in this article.)
- While the JIT/Lean approach tries to eliminate variability, QRM recognizes that in certain markets responding to this variability may provide competitive advantage.

Instead of eliminating variability, QRM creates an effective organization structure to cope with it and serve the market. QRM does this by exploiting its understanding of system dynamics.

- A specific example of the difference between JIT/Lean and QRM is the issue of delivery of material or components. "On-time delivery" is a cornerstone of JIT/Lean implementation. And yet QRM's understanding of organizational dynamics shows that promoting on-time delivery results in dysfunctional dynamics with longer lead times and higher costs. QRM provides alternative metrics based on lead time reduction that promise greater improvement in the long run.
- The QRM approach extends to supply management as well, and is called time-based supply management (Ericksen, 2000). Companies such as John Deere are finding that, particularly for smaller suppliers, the time based mindset and QRM principles offer an effective approach to target improvements at the supplier's operation. In addition, rapid results can be achieved, with significant improvements in supplier deliveries and quality, and reduction in supplier cost and lead time (Golden, 1999; Ericksen, 2000; Nelson, 2000).

To summarize, QRM pursues the relentless reduction of lead time – all QRM principles stem from this singular driving concern. Instead of management announcing dozens of programs and acronyms, QRM enables management to present one unified message to the organization, and all policies follow from this one driving strategy.

Time-Based Competition and QRM

Coincidentally, QRM also finds its roots in a strategy used by Japanese enterprises. In the late 1980s this strategy was documented by several U.S. authors (Stalk, 1988; Schmenner, 1988) and became known as *time-based competition* or TBC (also see Blackburn, 1991). The basis of TBC is the use of *speed* to gain competitive advantage: a company that uses a TBC strategy delivers products or services faster than its competitors. While a TBC strategy can be applied to any business, including banking, insurance, and hospitals, the focus here is on its application in a manufacturing firm. We call this specific application of TBC strategy *Quick Response Manufacturing* or *QRM.* By focusing on manufacturing companies, QRM sharpens the principles of TBC and also adds a number of new dimensions.

As previously stated, QRM pursues the reduction of lead time in all aspects of a company's operations. What is unique about QRM is that it espouses a *relentless emphasis on lead time reduction* that has a long term impact on every aspect of a firm. Though QRM uses this viewpoint first proposed in TBC philosophy, we now can capitalize on over a decade of observing manufacturing companies that have applied TBC to go beyond the original TBC strategy. QRM has refined TBC by:

- Focusing only on manufacturing, which has enabled us to sharpen the concepts and produce specific principles for manufacturing enterprises.
- Taking advantage of basic principles of manufacturing system dynamics to provide insight into how you can best reorganize an enterprise to achieve quick response.
- Clarifying the misconceptions managers have about how to apply time-based strategies.
- Providing specific QRM principles on how to rethink manufacturing process and equipment decisions.
- Developing a whole new material planning and control approach.
- Developing a novel performance measure.
- Providing a companywide strategy by furnishing principles that cover all major aspects of a manufacturing organization, from purchasing to sales, from engineering to accounting, from the shop floor to order processing.
- Understanding what it takes to implement QRM to ensure lasting success.

The reader may wonder what is new here. TBC has been around for over a decade. The reader may have personally read books on time-based competition and time-based manufacturing. Be that as it may, let us present a simple fact. During 1995-96, we interviewed over 400 U.S. executives and managers in dozens of industries, and even though all of them were from firms that were trying to cut their lead times, *over 70% of the policies in use by these managers and their companies were major obstacles to lead time reduction.* Worse yet, it was not as if these managers were working on changing the policies. In most cases they had no awareness that these policies were the source of the problem. If over two-thirds of the policies in use at an average U.S. firm are preventing it from cutting its lead times, what's the chance that companies you are associated with also suffer from this malady? We now present the quiz that was given to the managers.

What Managers Believe About Implementing QRM

Through the use of QRM techniques, companies have achieved substantial reductions in lead time, over 75% in new product introduction time and over 90% in time to fill orders for existing products. Furthermore, the beauty of QRM is that the very act of looking for ways to speed up existing procedures results in manifold benefits: successful QRM programs result in quality improvement and cost reduction as well. *Implementing QRM simultaneously achieves low cost, high quality, and rapid delivery.*

You would thus expect that a QRM strategy would be a priority for any manufacturing firm. But we have found that making QRM a priority and making it work are two entirely different matters. While there has been much publicity about competing on speed, and many books and articles have been written on the subject (Stalk, 1988; Schmenner, 1988; Charney, 1991; Blackburn, 1991; Stalk and Hout, 1992), our experience with hundreds of managers shows that there are many misconceptions about how to implement QRM. These misconceptions prevent successful results. Early experiences in implementing QRM led us to develop a simple quiz on implementing QRM, which we have used over the past several years to document the state of American management strategy.

Before we present the results, the reader may find it interesting to take this "QRM Quiz" (see box). If you are in industry, complete the quiz as follows. For each of the assertions in the quiz, ask yourself: "Do the key managers in my company consider this statement to be True or False?" If you are in academia, choose a company you know that is struggling with lead time reduction, and ask: "Do the key managers in that company consider this statement to be True or False?" Let's set some ground rules though, to make sure you are being completely *ruthless* in your evaluation. Take the first statement in the quiz as an example:

1. Everyone will have to work faster, harder, and longer hours, in order to get jobs done in less time.
   ❏ True ❏ False

As you look at this, you surely think, "We all *know* that to be False. We need to work smarter, not harder." But then, ask yourself, "Does the company frequently use overtime? Does it take a lot of expediting to get jobs out on time? Do people at the company often work on weekends?" If the answer to any of these is yes, then it is clear that key managers in the company believe item #1 is True! Use this same probing mindset as you approach each of the remaining items.

Mark your answers in the boxes, then read on to evaluate the results.

Quiz on Implementing QRM

For each statement below, ask yourself: Would the key managers in my company consider this statement to be True or False? Mark your responses in the boxes, then compare them with the answers given in the text.

1. Everyone will have to work faster, harder, and longer hours, in order to get jobs done in less time.
   ❑ True ❑ False

2. To get jobs out fast, we must keep our machines and people busy all the time.
   ❑ True ❑ False

3. In order to reduce our lead times, we have to improve our efficiencies.
   ❑ True ❑ False

4. We must place great importance on "on-time" delivery performance by each of our departments, and by our suppliers.
   ❑ True ❑ False

5. Installing a Material Requirements Planning (MRP) System will help in reducing lead times.
   ❑ True ❑ False

6. Since long lead time items need to be ordered in large quantities, we should negotiate quantity discounts with our suppliers.
   ❑ True ❑ False

7. We should encourage our customers to buy our products in large quantities by offering price breaks and quantity discounts.
   ❑ True ❑ False

8. We can implement QRM by forming teams in each department.
   ❑ True ❑ False

9. The reason for implementing QRM is so that we can charge our customers more for rush jobs.
   ❑ True ❑ False

10. Implementing QRM will require large investments in technology.
    ❑ True ❑ False

The author's experience with dozens of QRM projects has shown the following: for successful implementation of QRM it is necessary for a company's key decision-makers to believe that *every* assertion in the quiz is *False*! This may be obvious to the reader in some cases, such as item #1, where you know you have to find ways to work smarter. But what could be wrong with improving efficiencies (item #3)? And isn't on-time delivery (item #4) a cornerstone of every JIT program? And what about teams (item #8)? Aren't they all the rage these days, in everything from shop floor work to office operations? How could all those assertions possibly be False?

It is precisely these surprising points that we discuss briefly in this article, and cover in depth in Suri (1998). Many of these points are new not just to practitioners in industry but also to academics.

Now let us return to your own experience with the quiz: how well did *your* chosen firm score? Give your company a score of 0 for each True and 1 for each False. Count up the number of times you checked the False box, and that is your company's score. This score is on a scale of 0 to 10, where 0 denotes a company that will have to undergo a gargantuan change to succeed at QRM, while 10 denotes a company that is a "veteran" of QRM.

In reality, most companies will score somewhere in between. Do not be surprised if your company's score is low. We have given this Quiz to hundreds of employees at seminars around the U.S., and the typical score for a North American company is between 3 and 4. Interestingly, this average remains true across industry segment, from equipment manufacturers to parts suppliers and from electronics assembly firms to plastic injection molders. The score also seems to be independent of company size, with firms ranging in size from fifty employees to several thousand scoring in a similar range (Suri, 1998). In other words, 60-70% of the policies in use at North American companies are working against lead time reduction.

The peril of this situation is that not only are the wrong principles in operation, but *managers may not know* that these principles are wrong. More important than the correct response to each Quiz item, however, is an in-depth understanding of *why* it is the correct response, as well as the numerous issues that must be addressed to change from the current way of operation to the QRM way. Only when management clearly understands the basis for each QRM principle can it lead the organization along the QRM journey.

Next we give an overview of the reasoning behind the correct answers to the QRM Quiz. Then we will focus the remainder of the article on the differences between QRM and Lean Manufacturing.

Overview of QRM Principles

This section provides a summary of the 10 QRM principles that must replace the 10 traditional beliefs presented in the quiz. Space does not permit us to provide detailed case studies and examples to argue and illustrate all the points. Nevertheless, we hope that the discussion here will stimulate the reader to read the additional details in Suri (1998).

Traditional Belief #1*: Everyone will have to work faster, harder and longer hours, in order to get jobs done in less time.*

QRM Principle #1*: Find whole new ways of completing a job, with the focus on lead time minimization.*

To see the importance of this focus, refer to Figure 1 which shows the typical progress of an order through a company. The figure shows the "touch time" when someone is actually working on the job, compared with the elapsed time. We can see that touch time accounts for just 2 out of the 34 days. Traditional approaches focus on reducing the touch time, while the QRM approach focuses on reducing the total elapsed time.

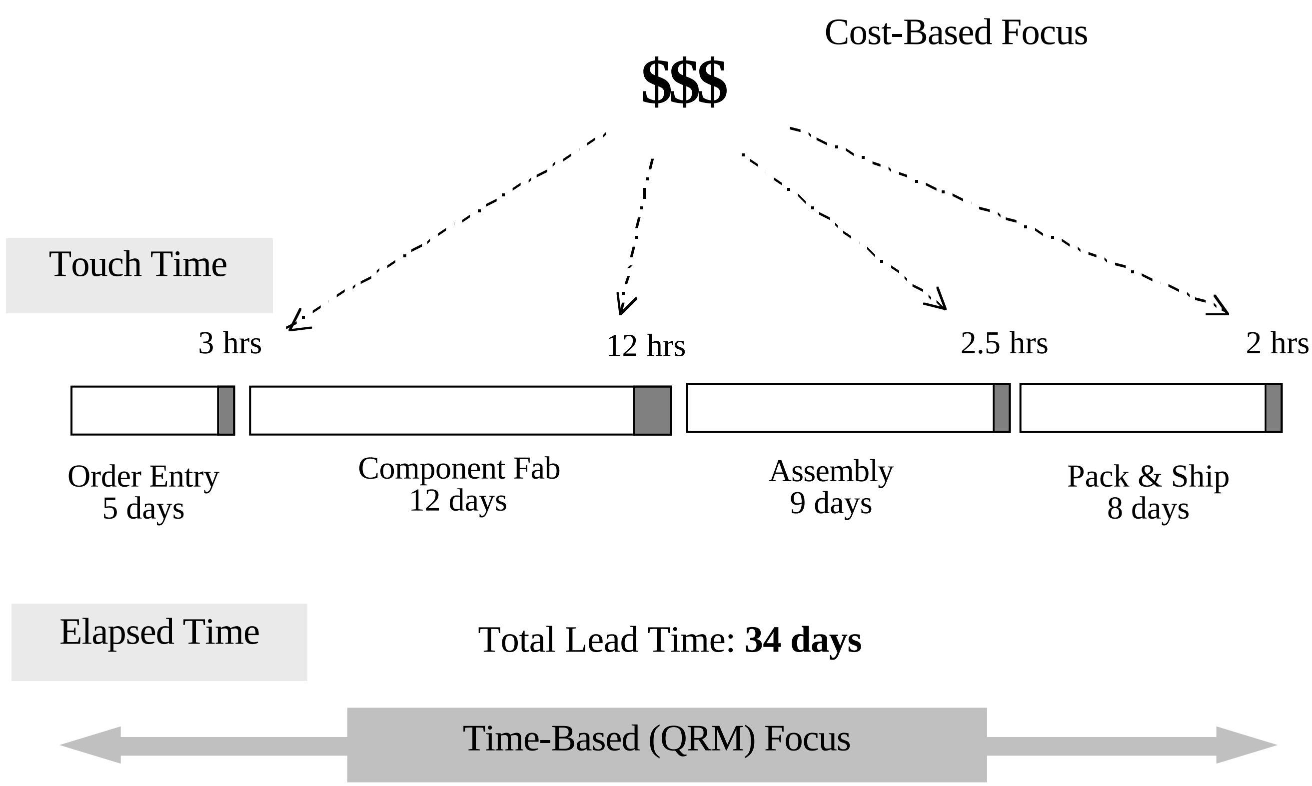

Figure 1. Comparison of Cost-Based and Time-Based (QRM) Approaches

However, our organizations are not designed to manage this total elapsed time. *Organizational structures, accounting systems, and reward systems are based on* *managing scale and cost*. A legacy of scale/cost based management systems, and the greatest enemy of QRM efforts, is the functional organization with specialized departments. Another legacy is the *Response Time Spiral*: an increasing spiral of lead times which results from scale/cost-based management systems. Taking time out of the system requires completely rethinking how you organize production, materials supply, and white collar work. The result is a cellular organization, both in the office and the shop floor, with each cell aimed at a focused target market segment, along with a new approach to materials planning and control, and new supplier strategies.

Although the cellular organization has been discussed for over two decades, our experience shows that companies still struggle with the proper implementation of cells. QRM theory provides a fresh perspective on implementing cells by combining engineering and management principles. These lead to creative rethinking for cellular manufacturing, overcoming many of the traditional obstacles. Also, new operating methods such as time-slicing are described, to help cells share non-cell resources. Finally, we thoroughly discuss many management and employee concerns with cell implementation and how to overcome them.

Traditional Belief #2. *To get jobs out fast we must, keep our machines and people busy all the time.*

QRM Principle #2*: Plan to operate at 80% or even 70% capacity on critical resources.*

Most managers' reaction to this principle is: "We can't afford to do that. We will be wasting our resources and our costs will go up!" However, QRM will eliminate the complex series of dysfunctional interactions that result from the present 100% utilization policy, such as growing queues, jobs spending a lot of time waiting for resources, and resulting long lead times. The QRM approach is to show managers that these interactions result in system-wide costs that exceed the cost of the idle capacity. Thus, QRM theory shows how *idle capacity actually serves as a strategic investment* that will pay for itself many times over in increased sales, higher quality, and lower costs.

Traditional Belief #3*: In order to reduce our lead times, we have to improve our efficiencies.*

QRM Principle #3*: Measure the reduction of lead times and make this the main performance measure. Eliminate traditional measures of utilization and efficiency.*

The problem with this traditional belief is not the *concept* of efficiency, but that most *measures* of efficiency work counter to lead time reduction. For example, one measure of efficiency on the shop floor results in an incentive to run large batches. For managers however, the QRM principle seems too bold a step: "If I eliminate any measures of utilization and efficiency, how will I know that my operating costs are not going out of control?" Yet a case study in Suri (1998) shows how lead times for a line of spare parts dropped from 36 days to 6 days using reduction of lead time as the main performance measure. To accomplish this though, it is important for everyone in a manufacturing firm, and especially for senior managers, to understand the dynamics of factory operations. They need to study the interactions between capacity utilization, efficiency measures, and lot sizing policies, and their effects on lead time. Using simple mathematics understandable to managers, QRM theory shows that lot sizes appropriate for quick response bear little relation to the values calculated by the Economic Order Quantity (EOQ) formula, which fails to consider many costs of large lots, and ignores the value of responsiveness. Nor can good lot sizes for QRM be predicted by the MRP system, since it assumes fixed queue times regardless of workload. This is why companies like John Deere are using the MPX modeling tool, based on QRM theory, to help their suppliers determine lot sizes which minimize lead times (Golden, 1999; Network Dynamics, 2000).

Traditional Belief #4*: We must place great importance on "on-time" delivery performance by each of our departments and our suppliers.*

QRM Principle #4*: Stick to measuring and rewarding reduction of lead times.*

Almost every book on modern manufacturing discusses on-time delivery and says that it is a cornerstone of JIT. What we have observed though, is that while on-time performance is desirable as an outcome, it is the emphasis on it as a performance measure that is dysfunctional. Instead of trying to reduce lead times, internal departments and external suppliers alike tend to pad their quoted lead times so that their on-time deliveries look good. As a result, the Response Time Spiral takes over the organization. With QRM, organizational changes promote shorter lead times, supported by a novel performance measure we call the QRM Number (which measures lead time *reduction*). These shorter lead times, in turn, eliminate the Response Time Spiral, and delivery problems disappear, resulting in on-time performance.

Traditional Belief #5*: Installing a material requirements planning (MRP) system will help in reducing lead times.*

QRM Principle #5. *Use MRP for high level planning and coordination of materials. Restructure the manufacturing organization into simpler product-oriented cells. Complement this with a new material control method that combines the best of push and pull strategies.*

MRP systems serve an important function of assisting with materials supply but they cannot solve lead time problems because the underlying model in MRP is flawed (Hopp and Spearman, 1996). In fact, the fixed lead time assumption in MRP promotes growth of the Response Time Spiral. In the redesigned QRM organization, MRP is used for higher level planning but not for micromanaging of work centers. Teams should run their own cells, and they should be provided with simple tools to manage their capacity and continually improve their responsiveness. A novel material control strategy, called POLCA, combines the best of push and pull methods to limit congestion while at the same time providing a high degree of flexibility, enabling even custom engineered products to be made. (POLCA is discussed further in a later section.)

Traditional Belief #6. *Since long lead time items need to be ordered in large quantities, we should negotiate quantity discounts with suppliers.*

QRM Principle #6*: Motivate suppliers to implement QRM, resulting in small lots at lower cost, better quality, and short lead times.*

The more a company purchases items in large batches, the longer its suppliers take to make them, motivating the company to put in orders for even larger batches. This creates another dysfunctional Response Time Spiral. This spiral is exacerbated by traditional purchasing policies and incentives, which also motivate the use of large batches. Instead, case studies from John Deere show the effectiveness of the QRM approach: by working with its suppliers to implement QRM, John Deere has obtained cost reductions while also getting better quality and shorter lead times for smaller order quantities (Golden, 1999; Ericksen, 2000; Nelson, 2000).

Traditional Belief #7*: We should encourage our customers to buy our products in large quantities by offering price breaks and quantity discounts.*

QRM Principle #7*: Educate customers on your QRM program, and negotiate a schedule of moving to smaller lot sizes at reasonable prices.*

This is the reverse of #6. A company's salesforce is motivated to offer quantity discounts. The customer's behavior of ordering larger batches then degrades the company's delivery performance, which further encourages the customer to order ahead with large quantities. With QRM companies form strategic partnerships with their customers and show how QRM will allow them to receive smaller batches at lower cost and with shorter lead time.

Traditional Belief #8*: We can implement QRM by forming teams in each department.*

QRM principle #8*: Cut through functional boundaries by forming a Quick Response Office Cell (Q-ROC), which is a closed-loop, collocated, multifunctional, cross-trained team responsible for a family of products aimed at a focused target market segment, and empower the Q-ROC to make necessary decisions.*

Some of the team implementations that follow the above traditional belief are the result of the quality (TQM) movement. True, a team with all its members in one functional department may result in local quality improvements. For the purpose of QRM however, such a team will do little to cut overall lead time for office operations. Instead, the team for QRM must be the Q-ROC with characteristics as above. (Cells and other QRM changes are not restricted to the shop floor.) Such Q-ROCs result in significant reduction of lead times for jobs such as cost estimating, quoting, and order processing: Ingersoll Cutting Tool Company, in Rockford, Ill. reduced its engineering and order processing time for customized cutters from 10 days to half a day. *Closed-loop* means that all the required steps can be done within the team which means,*you will have to cut across functional boundaries and change reporting structures*. This is not just an application of Reengineering, however. By using principles of system dynamics in the design of Q-ROCs, providing specific engineering and management principles for manufacturing organizations, plus by changing management principles and performance measures and adopting a companywide approach, QRM goes much deeper than Reengineering.

Traditional Belief #9*: The reason for implementing QRM is so that we can charge our customers more for rush jobs.*

QRM Principle #9*: The reason for embarking on the QRM journey is that it leads to a truly lean and mean company with a more secure future.*

While customers may pay more for speedy delivery, and this may be a good short term result of better response, it should not be the main reason for engaging in QRM. Searching for ways of squeezing time out uncovers quality problems and wasted efforts. Fixing these results in higher quality, lower WIP, less waste, lower operating costs, and greater sales. While Lean Manufacturing methods have put a lot of emphasis on elimination of waste, certain types of waste caused by long lead times are ignored in those approaches. With its broader definition of waste, QRM can create an even leaner enterprise that will remain a formidable competitor for years to come.

Traditional Belief #10*: Implementing QRM will require large investments in technology.*

QRM Principle #10*: The biggest obstacle to QRM is not technology, but "mindset." Management must recognize this and combat it through training. Next, companies should engage in "low-cost" or "no-cost" lead time reductions, leaving expenisve technological solutions for a later stage.*

New technologies, such as rapid prototyping and CAD/CAM, offer great opportunities for time reduction. While these are important, there are several steps that precede them, like education. Education must be a company's first step, or else other efforts will fail. In particular, the mindset of all employees, from the shop floor to the boardroom, from desk workers to senior managers, must be realigned to QRM principles. Also, in order to bring about the mindset change, organizations will need to thoroughly rethink existing performance measures. Performance measurement is intimately tied in with the cost accounting system, which is an obstacle to implementing an effective QRM program. QRM does not rely on complex changes such as activity-based costing (ABC) to address this issue. Rather, two simple fixes to the accounting system, involving strategic pools created by management, can go a long way towards making the accounting system support QRM.

QRM Compared With Lean Manufacturing

QRM is best applied by two types of companies: (i) those that make highly engineered products in small batches, even one-of-a-kind, and (ii) companies that do not need to engineer each product, but have a very large number of different products with highly variable demand for each. With this context in mind, and to understand the differences between QRM and JIT (or Lean Manufacturing) we will briefly review three key concepts of Lean Manufacturing from Womack and Jones (1996), and contrast them with the corresponding principles in QRM. The three concepts of Lean Manufacturing are:

1. Elimination of *muda* (Japanese for waste)
2. Implementing flow
3. Implementing pull

We now discuss these three concepts. A basic emphasis of JIT, and repeated in Womack and Jones (1996), is the systematic elimination of *muda* through eliminating non-value-added waste, resulting in improved quality, reduced costs and reduced lead times. In contrast, QRM emphasizes the relentless reduction in lead time, resulting in the elimination of non-value-added waste, improved quality, and reduced costs. While this distinction may not seem substantial, once you adopt the QRM approach many many additional forms of waste are uncovered that do not immediately surface when applying JIT.

As an example of this, in 1997 the author was part of a team working on implementing QRM at a large factory of an American metal fabricator. The factory had already worked with some consultants using a JIT approach, and they had identified what they thought were the key forms of waste that needed to be eliminated. They presented this list throughout the plant to motivate employees. Soon after, when our QRM team conducted a workshop with a group of 30 managers and employees to identify waste due to long lead times, we came up with over a dozen important items that the JIT group had not identified. We were told later that this list was an eye-opener for many of the managers.

As another example, the JIT system requires inventory in many intermediate stages of the materials replenishment system (see the discussion of pull below). *But in the QRM approach this kind of inventory is truly "waste," because these are products for which there is, as yet, no end demand. And yet JIT does not recognize this as waste. Quite the contrary, it is institutionalized in the JIT pull system.*

A second key concept in lean manufacturing is *flow:* the goal is to make the value-creating steps flow. First this involves tackling the organization of functional departments with a batch-and-queue mode of operation. This is accomplished by focusing on a given product and laying down all the resources for it so that an order can proceed continuously without any backflows or stoppages. In doing this, one also rethinks specifics of work practices and machines used. Thus far, this sounds similar to the QRM approach, with one exception: in QRM it is not necessary for cells to have unidirectional flow (see the examples in Suri, 1998). However, three additional aspects of flow distinguish it even more from QRM. These aspects are *takt* time, *heijunka,* and *flex fences*.

Takt time is the time between completion of each piece, if the shipping rate to customers is to be maintained. Once *takt* time has been defined, the goal is "to determine how to adjust every [operation step] so that it takes exactly [the *takt* time]. This can often be done through careful development of *standard work,* in which every aspect of the task is carefully analyzed, optimized, and then performed in exactly the same way each time..." (Womack and Jones, 1996). On the other hand, QRM is designed for companies whose products are customized with highly differing specifications, or where there are a large number of products with very variable demand for each one. In this situation, the organizational structure as a whole needs to be more flexible, allowing more general organization of work within a cell, as well as a more flexible organization across cells, which we discuss below. In addition, the variability in processing needs and demand for products makes the use of *takt* time impractical. One could define an average time based on a given month's orders, but the daily demands on a given machine could be so different that the *takt* time concept would not work.

Books on JIT do mention that if there is a change in demand then you need to redefine the *takt* time and reoptimize tasks using the approach above, or you may even need to add or subtract machines. However, the detailed nature of the task optimization approach described above, or the expense of adding and subtracting machines, makes it clear that this is not an activity that one would undertake on a daily basis. When you have relatively stable demand that shifts little from week to week or month to month, the flow approach makes sense. But in the QRM context, the variabilities described above can, from one day to the next, lead to huge swings in work content for a given operation. The *takt* time approach is just too simplistic or unrealistic for QRM then. Instead, the QRM approach tackles this variability in requirements while still managing short lead times, using a number of principles. Key among them are:

- Organizational flexibility (resulting from rethinking all three: product design, process design, and organization structure).
- Understanding and exploiting system dynamics that result from this type of interaction and variability, and implementation of novel constructs such as time-slicing.
- Use of queueing models to plan and manage capacity and lot sizing, in order to plan for the variability.

Implementation of flow also requires level scheduling (*heijunka* in JIT language). A part of level scheduling in JIT is finding ways to reduce setup times and run smaller batch sizes. Here QRM and JIT agree. But another part is that in order to level the schedule across multiple upstream and downstream steps one also needs to freeze it within some time horizon. Thus *both a frozen schedule and level scheduling are needed.* A frozen schedule though, is the antithesis of responsiveness to customers. In contrast, the QRM approach recognizes that variabilities can be ingrained in the nature of a company's business. Indeed, the QRM approach is attractive to these kinds of businesses for this very reason. Hence the QRM organization is designed to cope with these variabilities.

Another requirement when implementing flow is the use of *flex fences* to deal with suppliers with long lead times. If demand increases, even though *takt* time may be shortened inside the factory, these components may not be available. The flow manufacturing approaches attempt to finesse this issue by setting "flex fences," which are ranges of demand increases that a supplier should be able to provide at short notice. Since the suppliers have inherently long lead times, they usually accomplish this by maintaining a buffer sufficient to handle the flex fence. Again, this is *muda,* yet it is institutionalized in the lean system! In a company with a wide range of products, this implies that someone along the supply chain needs to maintain high buffer stocks. More extreme, in a company with custom designed products, the company often will not know it needs a component until after an order is received and engineered. In either case, the flex fence approach is not practical. As described in detail in Suri (1998), the QRM approach instead changes both the operation of the suppliers as well as the very structure of the interaction between a company and its suppliers.

The third key concept in Lean Manufacturing is that of *pull.* We can illustrate this by summarizing the example of a car bumper that is described by Womack and Jones (1996). When a customer arrives at a Toyota dealer and needs a new bumper, the dealer will "pull" this bumper from its inventory. The sale of this part to the customer triggers a pull signal to the Toyota Parts Distribution Center (PDC). When the PDC ships the bumper to the dealer, that triggers a pull signal to the preceding link in the supply chain, which is the Toyota Parts Redistribution Center (PRC). As this center ships a bumper to the

PDC, it sends a pull signal to Bumper Works, the factory that makes the bumpers. Bumper Works, in turn, as it uses up raw material to make the bumpers, sends a pull signal to its supplier of sheet steel.

In addition to the pull signals across organizations, there are pull signals operating within each organization. Taking Bumper Works as an example, a shipment to the PRC does not immediately trigger a pull signal to the steel supplier. Instead the shipment from finished goods triggers a pull signal to final assembly to replenish the parts. As each operation at Bumper Works uses up material to produce a replenishment for the next stage of production, it sends a pull signal to the previous stage, and so on, upstream all the way to the point where the pull signal leaves Bumper Works and goes to its sheet steel supplier.

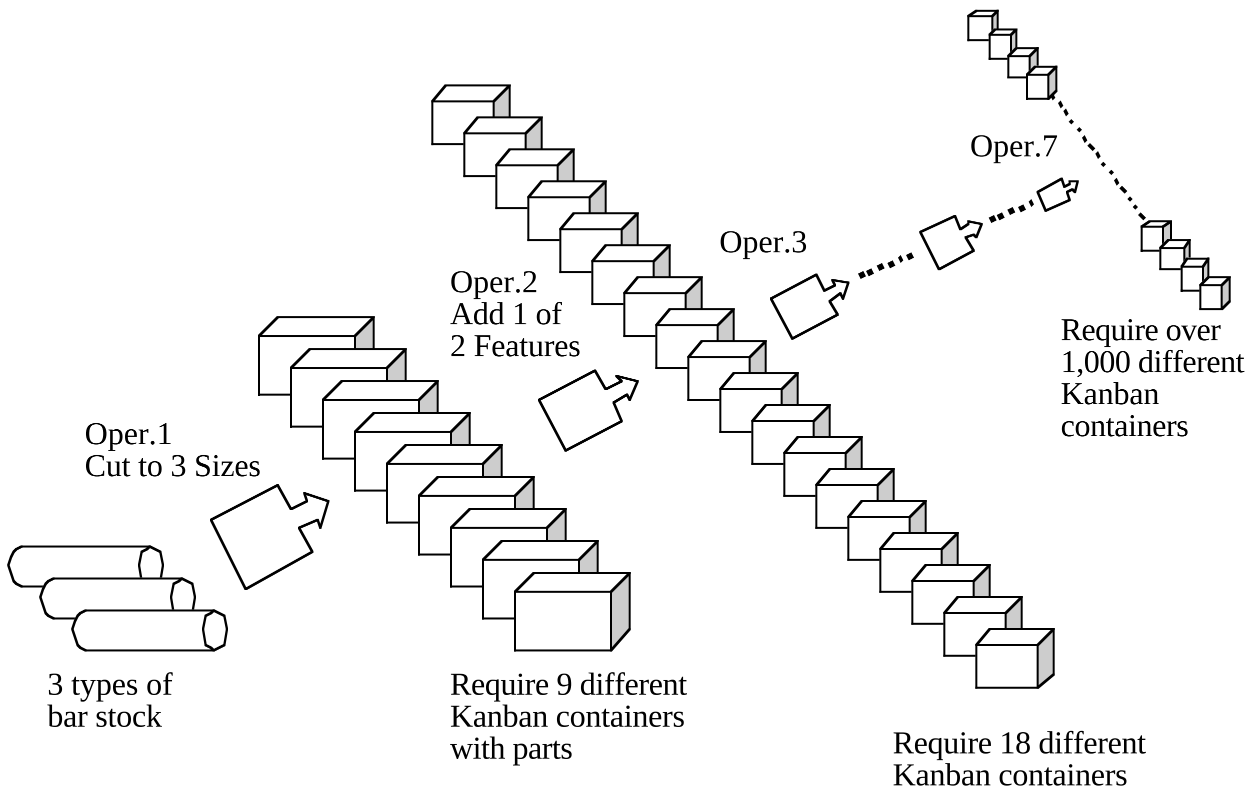

Figure 2. Potential for Proliferation of WIP in a Pull System

While pull systems have been widely touted, their disadvantages are less widely publicized. Consider the implications of the above system for a QRM company. Womack and Jones (1996) state that the philosophy behind pull is “ship one; make one.” But what this means is that the company has “one” in finished goods, all made and ready to ship. For the QRM company that makes thousands of different items, this implies that there is inventory of each of these items at *each* stage of the supply chain. Not only does this mean that the company has inventory at the end point of each of the organizations in the chain, such as Bumper Works and the PRC, but also it implies inventory of each item's partly manufactured stock between each operation within each manufacturing organization (Figure 2). This is a vast amount of inventory: is this not *muda*? There is a worse situation for a company that custom designs and fabricates each product: here the pull system fails at the very first step above. There is no product in finished goods, since the parameters of the product are not known till the order is received. Similarly, the intermediate stages cannot have the required inventory to pull from either, since stages whose operations depend on the parameters of the final product cannot start production until the actual order is engineered.

The preceding discussion makes it clear that there are two prerequisites for success of a Lean Manufacturing strategy. One is a limited degree customization. Specifically, lean systems are designed to make products that involve minor customization of a main product, or products where the customization involves choosing from a set of predefined options, as opposed to totally custom engineered products. The second prerequisite for lean manufacturing methods to be applicable is a marketplace with relatively stable demand. A fundamental basis for lean thinking is the premise that "end-use demand of customers is inherently quite stable and largely for replacement" (Womack and Jones, 1996). While this may be true in large segments of the market, it is our hypothesis (verified by many companies that are adopting QRM) that there is a great deal of opportunity in pursuing other markets, such as: (i) emerging market segments where the pattern of demand is unpredictable and product requirements are changing fast, and (ii) markets where companies need to tailor their products in detail to individual customers. Suri (1998) contains a detailed case study of how Ingersoll Cutting Tool Company hypothesized the existence of such a market segment, went after it, and having found it was able to increase its market share by over 500%. Indeed, the QRM strategy helps companies find such market niches wherever they may exist, and once found, they can expand those niches into broad markets, while at the same time being the only company that can provide products for those now large markets.

Looking to the future, we believe that in this era of computer-aided design and manufacturing, and internet-based ordering and supply management, the marketplace is headed for increasingly more custom engineered and individually tailored products. This means that in the future increasingly more market segments will be suited to the QRM strategy.

POLCA: The Material Control System for QRM

As just discussed, pull systems (which usually involve some sort of material control mechanism such as kanban) are not appropriate for a QRM company. At the same time, push systems using MRP have their own drawbacks in terms of exacerbating the Response Time Spiral and promoting ever-longer lead times. For QRM there is thus a need to develop a whole new material control method. We have devised such a method called POLCA (an acronym that is explained below).

To understand the basis for POLCA, let us review where the QRM strategy is most effective. It is best applied: (i) at companies make custom designed products in small batches (or even one-of-a-kind), and (ii) at companies that don't custom design each product, but still have such a wide variety of options and combinations of specifications that they cannot afford to store inventory for all these options at various stages of their manufacturing system. QRM strategy organizes these companies as follows: first the company creates cells focusing on subsets of the production process for similar parts, and then it processes a given customer order through differing cells depending on the needs of that order.

As an example, consider a company called CFP Corp., that makes customized faceplates, such as rating plates used on products ranging from small electrical appliances to large earthmoving equipment, and faceplates on calculators and instruments. The plates are made from different materials, range in size from under an inch square up to over eighteen inches in either dimension, have information printed on them, along with features such as holes, notches, and fasteners to assist in mounting them. CFP's strategy is not to compete with the high volume manufacturers for large markets, but to go after companies that need small batches of plates for specialized markets.

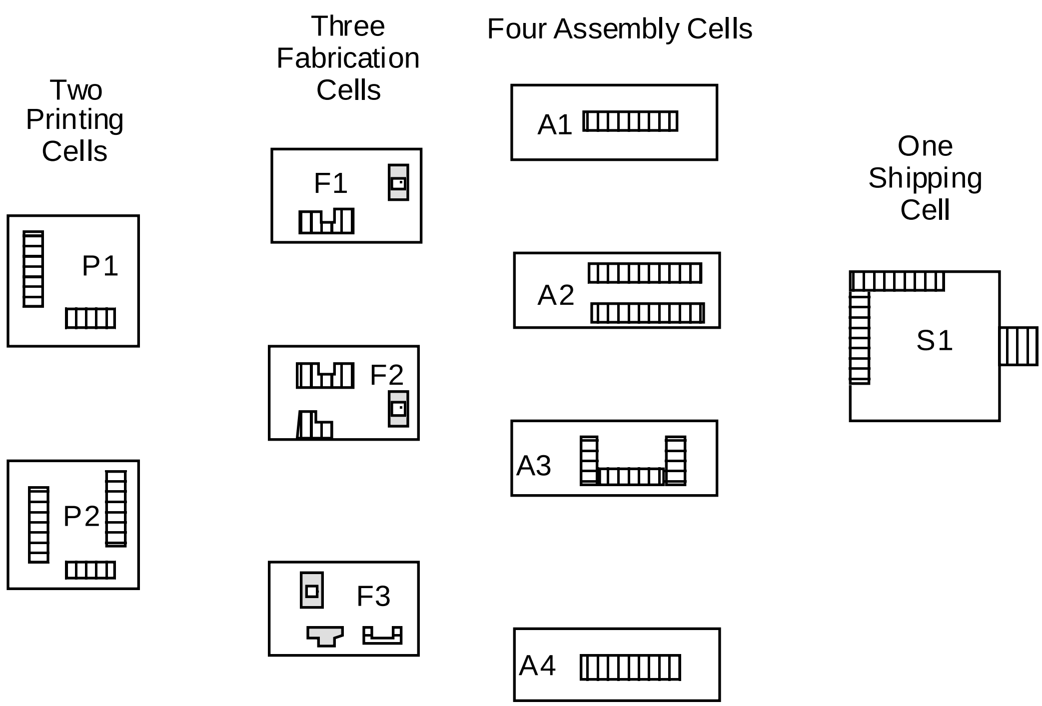

Figure 3. Organization of Cells at CFP Corporation

CFP has created several cells to serve its highly varied markets via a QRM strategy (Figure 3). First there are two Printing Cells: P1 focuses on screen printing, P2 on lithographic printing. Next are three Fabrication Cells, F1, F2 and F3, which convert the printed sheets into individual plates with the desired features. Operations include punching holes and notches, cutting the sheets, and bending. Cell F1 focuses on plastic plates, F2 on light gauge aluminum, and F3 on heavy gauge aluminum. After the fabrication operations, plates go to one of four Assembly Cells, A1 through A4. Here finishing operations such as deburring, attaching fastners, and packaging are carried out. The four cells differ in terms of the size of products handled, the types of fasteners to be attached, and the form of packaging to be used. Finally, all orders go to the Shipping Cell S1, where the packaged plates are placed in shipping containers and then loaded onto trucks.

Customer orders to CFP are served by using the appropriate combination of cells needed to print, fabricate and assemble each order. Orders can have very different demands within the cells too: the punching requirements for an order for large plates with lots of holes may use a lot of time on a CNC Turret Punch in F3 and not much time on the Shear, while another order for small plates may have very little punching time but take a lot of time for shearing. In addition, the routing of products *within* each cell can differ from order to order. For all these reasons, the lean concepts of flow, takt time and level scheduling are not applicable.

In order to serve its market niche for customized plates, CFP Corp. thus has three key requirements for its materials management system: (1) the ability to route products through different combinations of cells, as needed by a given order; (2) within a cell, the ability for products to use machines in different sequences; and (3) a good deal of flexibility in terms of capacity requirements for each operation in a cell. We should note that the original intent of MRP systems was to enable a company such as CFP to achieve these requirements, but for reasons that are by now well-dcumented in the literature (e.g. see Hopp and Spearman, 1996; Suri, 1998) that intent was not fulfilled. Also, as we just discussed, a pull system will not work for this organization.

Instead, the material control system that we have devised for use in a QRM company is called *Paired-cell Overlapping Loops of Cards with Authorization* (POLCA). This is a material control system that operates in the context of a high level material requirements planning system (HL/MRP) and a cellular organization, in other words, these are prerequisites for a POLCA system. We briefly describe the key features on which POLCA is based.

First, for each order, *Release Authorization times* are created via HL/MRP. Similar to an MRP system, the POLCA system generates times when each cell may begin work on a particular order. However, unlike in a standard push system where a workcenter should start work at that time, POLCA simply authorizes the beginning of the work, but the cell cannot start without other conditions being satisfied, as we discuss next.

Production control cards, which we call *POLCA cards*, are used to communicate and control the material movement between cells. While this may seem similar to kanban, there are some important differences. First, the cards are only used to control movement *between* cells, not within cells. (For material control between workstations *within* a cell, cells have the freedom to use various other procedures.) Second, the POLCA cards, instead of being specific to the product, as in a pull system, are assigned to *pairs of cells*. Figure 4 shows the POLCA card flows for a particular order at CFP Corp. This order's routing takes it from P1 to F2, then to A4 for assembly, and finally to S1 to be shipped. This order will therefore proceed through the POLCA card loops with the pairs P1/F2, F2/A4 and A4/S1, as shown in the figure.

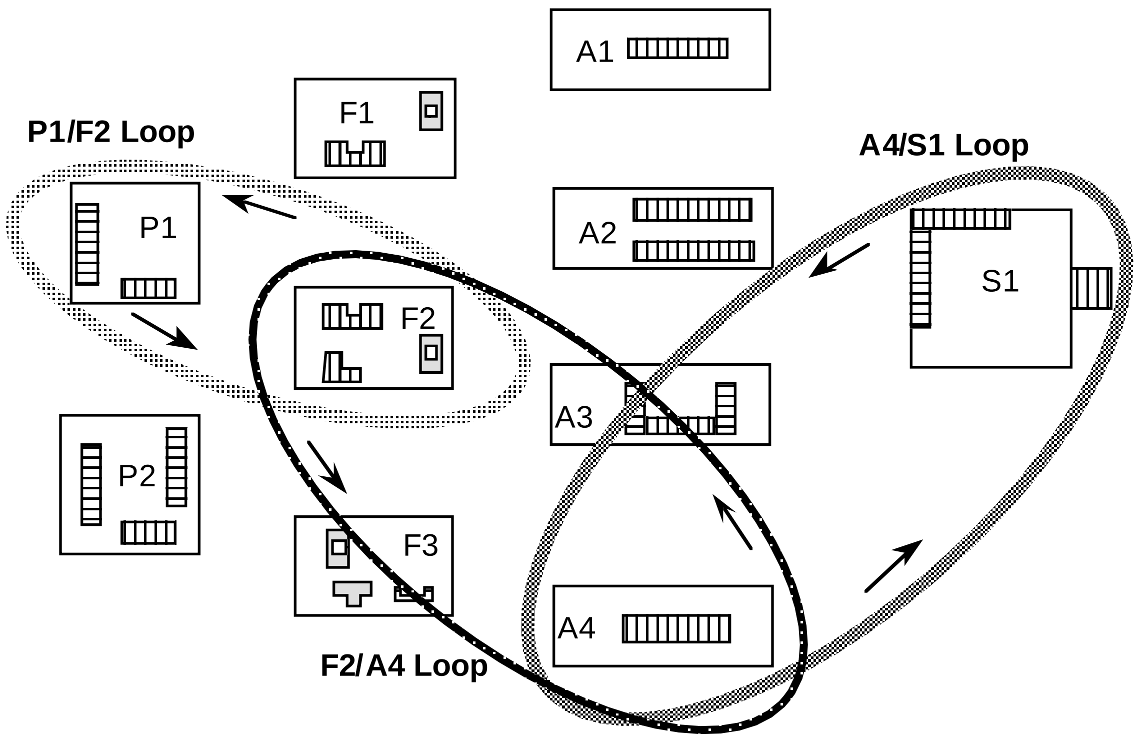

Figure 4. POLCA Card Flows for a Particular Order at CFP Corp.

The third difference from kanban is that the POLCA cards for each pair of cells stay with a job during its journey through *both* cells in the pair before they loop back to the first cell in the pair. For example, a P1/F2 card above would be attached to a job as the job entered cell P1, it would stay with this job through cell P1 and as it goes to cell F2, continue to stay with the job until cell F2 has completed it, and while the job moves on to its next cell (A4), this P1/ F2 card would be returned to cell P1. Since most cells will belong to more than one pair of cells, there will be multiple loops of cards that *overlap* in each cell, as seen in Figure 4.

The detailed procedures used with POLCA cards include rules for sequencing jobs at each cell, movement of jobs between cells, and return of POLCA cards to the originating cells, and can be found in Suri (1998).

Advantages of POLCA Over Both MRP and Pull Systems

Now we discuss how in the context of QRM the POLCA system overcomes the drawbacks of both MRP and pull systems. First, the use of POLCA cards assures that each cell only works on jobs that are destined for downstream cells that will *also* be able to work on these jobs in the near future. In other words, if a POLCA card from a downstream cell is not available, it means that cell is backlogged with work (or cells downstream from it are backlogged). Working on a job destined for that cell will only increase inventory in the system since somewhere downstream there is a lack of capacity to work on this job. It is more expedient to hold off putting organizational resources into such a job; those resources would be better used in other ways. This is similar to the logic used in a typical pull system.

Second, the use of HL/MRP allows a make-to-order environment through flexible routings that use cells as needed, plus the use of authorization times prevents build-up of unnecessary inventory. As we showed in earlier examples, pull systems have the disadvantage of filling intermediate stages with inventory for the typical QRM context. By coupling the routing and authorization procedure using HL/MRP with the POLCA scheme we ensure that the company does not make products just because they have a pull signal; it makes products only when there is explicit demand for them.

Third, unlike a kanban system where workstations are tightly coupled via kanban cards, the POLCA cards flow in longer loops. There is coupling of cells, but it is more flexible. Remember that a kanban system is highly tuned to produce at a given rate. In fact, in designing a pull system, a good deal of effort is spent in determining the corresponding *takt* time. Indeed, the purpose of the tight coupling in a pull system is to find and eliminate the obstacles to achieving the consistent *takt* times. On the other hand, for QRM one needs to satisfy varying demand for multiple (even one-of-a-kind) products. Recall the example of CFP Corp. where some products need more shearing and others need more punching. Hence a company can set some average capacities – for example by using the MPX planning tool (Network Dynamics, 2000) – but the actual rates and bottlenecks will vary from day to day. This is the main reason for having the overlapping loops in POLCA. By making the card loops longer the additional jobs in the loop act as a buffer to absorb variations in demand and product mix. This allows each cell to balance its capacity as best as it can for the current mix, which cannot be done with a pull system because of the tight coupling through the kanban cards balanced carefully with the takt time calculations.

Implementing QRM: The Prerequisites for Success

As we have mentioned, although the idea of competing on speed has been with us for a decade there are still many misconceptions about how to implement this strategy. Also, while much has been written about competing on speed, not enough information is provided on many supporting topics critical to its successful implementation. Based on our numerous experiences with lead time reduction projects, the QRM strategy tackles these topics by laying down the following prerequisites for a successful implementation:

- *There must be a company-wide understanding of the basics of QRM*, *what it means, why it is necessary, how it works*. Such an understanding must be provided to everyone in the firm, not just to manufacturing workers and managers. To implement QRM a company needs *active* involvement from senior executives, staff and workers in *all* functional areas.
- *Workers and managers need to understand some basic dynamics of manufacturing systems*. Specifically they need to know how capacity planning, resource utilization, and lot sizing policies *interact* with each other and how they *impact lead time.* Without this, we have found that there will be no buy-in to the key techniques and policies for QRM.
- *The QRM program has to be implemented in both shop floor and office operations*. We have found that office operations constitute a significant portion of the total lead time for products, yet they are often overlooked as an opportunity for lead time reduction.
- *Firms must incorporate QRM policies in all areas*. This involves rethinking how the company operates in every area, not just obvious areas such as manufacturing and supply management, but also areas such as shipping, equipment purchase, employee hiring, accounting and performance appraisal. All these policies need to be made consistent with the QRM ideal.
- *Shop floor and office employees, as well as managers, need to thoroughly understand the concept of work cells*. While the concept of cellular manufacturing has been with us for over two decades, cells are *not* passé: we continue to find lack of organizational will to implement cells, as well as incorrect implementations or outright failures. Most of these problems can be attributed to a misunderstanding of a few basic principles. Several case studies from our projects have shown that educating all employees and managers on the principles of work cells has turned failures around to resounding successes.
- *Obstacles to implementation should be anticipated as much as possible, so everyone is prepared to combat them*. This involves tackling the traditional beliefs (listed earlier in the article) early in the cycle of implementation.

- *Even though you should create QRM education and awareness company-wide, top management should not attempt to reorganize the whole company for QRM right away*. Instead, QRM implementation should begin by focusing on a market segment where there is an opportunity via a quick response strategy, and a small part of the company should be reorganized using QRM principles, to serve this market. In this way, by trying QRM in one or two areas, management can minimize its risk and investment while it proves to itself and the rest of the company that this approach really works. After absorbing the lessons from this experience, additional parts of the company can be reorganized for QRM.
- *Concrete steps for implementing QRM should be identified at the start of the initiative*. By building on lessons learned from implementing QRM at dozens of companies, we are able to provide a roadmap for successful implementation. It is important for management to review the entire map early on in the intiative, so that they buy in to the whole plan. While some other approaches to manufacturing also offer roadmaps, the implementation plan for QRM has some unique aspects such as the use of three different types of teams; we have found these unique ideas maximize the potential for success.

As we stated earlier, a key aspect of the QRM approach is that all the principles stem from a single theme: *reduce lead times*. Some popular manufacturing management approaches appear as a collection of disjoint ideas; managers and employees have to remember a list of assertions such as the "five S's." In contrast, the entire set of principles in QRM strategy are derived from one theme, yet these principles are powerful enough to span the entire organization, from the shop floor to the office, from order entry to accounting, from purchasing to sales. Such an approach is more palatable to managers than a disparate collection of ideas, because it enables them to stick with a consistent message to the organization.

Another lesson we have learned from all our QRM projects is that *lead time reduction cannot be done as a tactic; QRM has to be an organizational strategy led by top management*. To significantly impact lead times firms must change the traditional ways of operating and redesign organizational structures. Such changes cannot be made without total commitment from top management. Hence educating senior managers on QRM strategy, and getting them to buy in to the roadmap for implementation, must be the first step in a QRM program.

Bibliography

Blackburn, J.D. (Ed.). 1991. *Time-Based Competition: The Next Battle Ground in American Manufacturing,* (Business One Irwin, Homewood, IL).

Charney, C. 1991. *Time to Market: Reducing Product Lead Time,* (Society of Manufacturing Engineers, Dearborn, MI).

Ericksen, P. 2000. "Time-Based Supply Management," in R. Suri (Ed.), *Proceedings of the Quick Response Manufacturing 2000 Conference* (Society of Manufacturing Engineers, Dearborn, MI).

Golden, P. 1999. "Quick Response Manufacturing Drives Supplier Development at John Deere," *IIE Solutions,* July.

Hopp W.J. and M.L. Spearman. 1996. *Factory Physics: Foundations of Manufacturing Management,* (Richard D. Irwin, Burr Ridge, IL).

Nelson, D. 2000. "Supporting Supplier Improvement Using Quick Response Manufacturing," in R. Suri (Ed.), *Proceedings of the Quick Response Manufacturing 2000 Conference* (Society of Manufacturing Engineers, Dearborn, MI).

Network Dynamics, Inc. 2000. MPX Rapid Modeling Software. Framingham, Mass. See `www.networkdyn.com` for more information.

Schmenner, R. 1988. "The Merit of Making Things Fast," *Sloan Management Review,* Fall, pages 11-17.

Stalk, G. Jr. 1988. "Time—The Next Source of Competitive Advantage," *Harvard Business Review*, July-August, pages 41-51.

Stalk, G. Jr. and T.M. Hout. 1992. *Competing Against Time,* (The Free Press, New York, NY).

Suri, R. 1998. *Quick Response Manufacturing: A Companywide Approach to Reducing Lead Times,* (Productivity Press, Portland, OR).

Tubino, F. and R. Suri. 2000. "What Kind Of Numbers Can A Company Expect After Implementing QRM?" in R. Suri (Ed.), *Proceedings of the Quick Response Manufacturing 2000 Conference* (Society of Manufacturing Engineers, Dearborn, MI).

Womack, J.P. and D.T. Jones. 1996. *Lean Thinking* (Simon and Schuster, New York, NY).

Womack, J.P., D.T. Jones and D. Roos. 1990. *The Machine that Changed the World* (HarperPerennial, New York, NY).