

# **QRM and POLCA: A Winning Combination for Manufacturing Enterprises in the 21st Century**

Rajan Suri

Director

Center for Quick Response Manufacturing

University of Wisconsin-Madison

1513 University Avenue

Madison, WI 53706, USA

[www.qrmcenter.org](http://www.qrmcenter.org)

## Summary

Today's CAD/CAM technology gives us the ability to customize products for individual customers without incurring high additional costs. When combined with customer demands for personalized products and internet-based individualized ordering capabilities, this means that there will be increasing demand for customized products in the 21<sup>st</sup> century. In addition, after placing their orders both OEMs (original equipment manufacturers) and end consumers expect these products to be delivered quickly. Although Lean Manufacturing techniques can be powerful in certain situations, for companies making a large variety of products with variable demand or companies making highly engineered products, Lean Manufacturing has several drawbacks. Quick Response Manufacturing (QRM) can be a more effective competitive strategy for companies targeting such markets. In this article, we give an overview of the QRM strategy, which focuses on lead time reduction throughout the enterprise. We also explain why the Lean Manufacturing strategies of flow, *takt* time and pull don't work well for these markets, and why QRM has greater competitive potential. Next we describe POLCA, a material control system to be used as part of QRM. Again, we show why a kanban system (used in Lean Manufacturing for material control) is not appropriate for these markets. Instead, POLCA provides an effective method to support both manufacturing and material control for companies serving these markets. The combination of QRM and POLCA will provide companies with significant competitive advantage through their ability to deliver customized products with short lead times.

---

This paper is a revised and updated version of "Quick Response Manufacturing: A Competitive Strategy for the 21st Century," by the same author. This paper contains new results on the impact of QRM strategy, as well as updated references to recent case studies on the application of QRM at various companies.

## 21st Century Markets Are Here

In the past few years we have seen a rapid growth in the number of options provided by manufacturers to their customers. Even beyond providing pre-specified options though, is the fact that today's CAD/CAM technology has given companies the ability to custom-engineer and then manufacture products for individual clients without incurring the high additional costs that such customization would have required two decades ago. Along with this has come the power of the internet, which allows customers to easily view many different options and select from them, often even having the ability to specify additional choices that may require engineering. All of these developments mean that there will be increasing demand for customized products in the 21<sup>st</sup> century. In the rest of the article, we shall use the terms "21<sup>st</sup> century markets" to refer to markets that require a large variety of products and have variable demand for each type of product, and/or markets that require custom-engineered products. In addition to these characteristics of 21st century markets is the fact that customers today (both OEMs and end consumers) expect products to be delivered with a much shorter lead time than was acceptable in the past.

In their attempts to cope with this changing market scenario, executives in manufacturing corporations have been seeking new competitive strategies. While today's world abounds with new acronyms and business strategies, one strategy that has become popular recently is Lean Manufacturing. Actually, Lean is based on the Japanese just-in-time (JIT) manufacturing techniques, which have now been described and popularized under the name of "Lean Manufacturing" (Womack et al., 1990; Womack and Jones, 1996). Although Lean has produced impressive results in many companies and is certainly an effective strategy, we will show that to serve 21st century markets effectively, Lean would have several shortcomings. In fact, the JIT strategy on which Lean is based was designed for situations with relatively stable demand and largely for replacement products (Womack and Jones, 1996). We will see that key Lean principles of flow, *takt* time, level scheduling and pull (Kanban) all break down when attempting to serve 21st century markets.

This is where Quick Response Manufacturing (QRM) comes in. QRM provides an enterprise-wide strategy that is particularly effective at serving 21st century markets (Suri, 1998). Specifically, it enables companies to dramatically shorten their lead times to deliver products for these markets, while at the same time improving product quality and reducing cost. Thus QRM provides companies with the potential for creating significant competitive advantage: such companies can progress towards serving (or even creating!) 21st century markets in their business areas and generating a growing base of satisfied customers, while their competition struggles to find cost-effective ways of dealing with the myriad of options or customizations demanded by their customers.

In this article, we give an overview of the QRM principles and explain the POLCA system. (POLCA is an acronym explained later.) We 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, for 21st century markets, Lean Manufacturing

principles do not work well, and QRM combined with POLCA has greater potential for creating competitive advantage. We end with the prerequisites for success in QRM implementation.

### ***QRM Defined***

QRM is a company-wide strategy that pursues the reduction of lead time in all aspects of a company's operations, both internally and externally. Specifically, from a customer's point of view, QRM means responding to that customer's needs by rapidly designing and manufacturing products customized to those needs. This is the *external* aspect of QRM. Also, in terms of a company's own operations, QRM focuses on reducing the lead times for all tasks across the whole enterprise, resulting in improved quality, lower cost, and of course, quick response. This is the *internal* aspect of QRM.

In the last few years, dozens of companies have implemented QRM strategies with astounding results. Typical results include reduction in lead times of 80-95% (both in manufacturing and in office operations), reduction in product cost of 15-50%, on-time delivery performance improving from 40% to 98%, and reduction in scrap and rework by 80% or more (see the Tables later in this article, as well as the numerous case studies in three conference proceedings (Suri, 2000; Suri et al., 2001; Suri and Rath 2002).

QRM achieves these lead time reductions and other results through detailed management principles, manufacturing methods, analysis techniques and tools that use basic concepts of system dynamics, and a step-by-step methodology. In addition QRM puts a great deal of emphasis on creating the mindset of pursuing lead time reduction. We will provide an overview of all these points below.

### ***Lean, Time-Based Competition, and QRM***

Later in this article we will discuss in more detail the differences between Lean and QRM, but a quick summary is provided here. Also, the idea of lead time reduction, or more generally, using *speed* to gain competitive advantage, is not new. In fact, this concept was documented by several U.S. authors (Stalk, 1988; Schmenner, 1988; Blackburn, 1991; Charney, 1991; Stalk and Hout, 1992) and became known as *time-based competition* (TBC). So we will also summarize how QRM goes beyond the original TBC strategy.

We begin by contrasting QRM with Lean:

- The driver for all the principles and strategies in QRM is reduction of lead time. In contrast the driver in JIT/Lean is waste reduction. Below we will see examples of where the QRM mindset can be more effective in 21st century markets.
- 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 company-wide strategy with implications far beyond the shop floor – 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. It uses an understanding of system dynamics, and exploits this understanding to define the best structures and policies that will reduce lead times.
- For serving 21st century markets, the Lean strategy of “pull” (Kanban) is either wasteful or breaks down altogether. For such companies, QRM provides an alternative strategy called POLCA, which combines the best features of “push” (MRP) and “pull” without their drawbacks.
- The QRM approach extends to supply management as well, and is called time-based supply management. Companies such as John Deere are finding that this can produce dramatic reductions in both supplier lead time *and* cost (see the results in Ericksen and Suri, 2001).

Now we discuss the development of QRM beyond time-based competition (TBC). TBC strategies were introduced as being applicable to any business, including banking, insurance, and hospitals. In that sense they were rather general, and details for application to manufacturing enterprises were missing. By focusing on manufacturing companies, QRM has sharpened the TBC strategy, and has also added a number of new dimensions:

- QRM has produced specific and detailed principles for a manufacturing enterprise. Indeed, QRM provides a company-wide 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 (Suri, 1998).
- QRM pursues the relentless reduction of lead time – all its 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.
- QRM capitalizes on fundamental principles of manufacturing system dynamics to provide insight into how you can best reorganize an enterprise to achieve quick response.
- QRM provides a whole new material planning and control approach, called POLCA, described later in this article.
- The QRM approach also takes pains to clarify the misconceptions managers have about how to achieve lead time reduction, and gives an understanding of what it takes to implement QRM to ensure lasting success. These misconceptions had not been pointed out and tackled in articles prior to our work on QRM.

To illustrate the magnitude of misconceptions that exist in industry, let us present a simple fact. A few years ago, 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.

## Misconceptions About Implementing Quick Response

Given the manifold benefits of short lead times, most companies are attempting to improve their responsiveness. However, despite all the articles written on lead time reduction, there are many misconceptions about how to implement quick response. These misconceptions prevent successful results. Our early experiences in implementing QRM led us to develop a simple quiz which we have used 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. *You need to answer the quiz based on the policies in use at the company, not based on your own opinion of what is correct.* 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

Developed by Rajan Suri  
Center for Quick Response Manufacturing  
www.qrmcenter.org

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

Copyright © 1997 R. Suri.

Now we present the answers. Our 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.

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 managers and 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 POLCA strategy for material control.

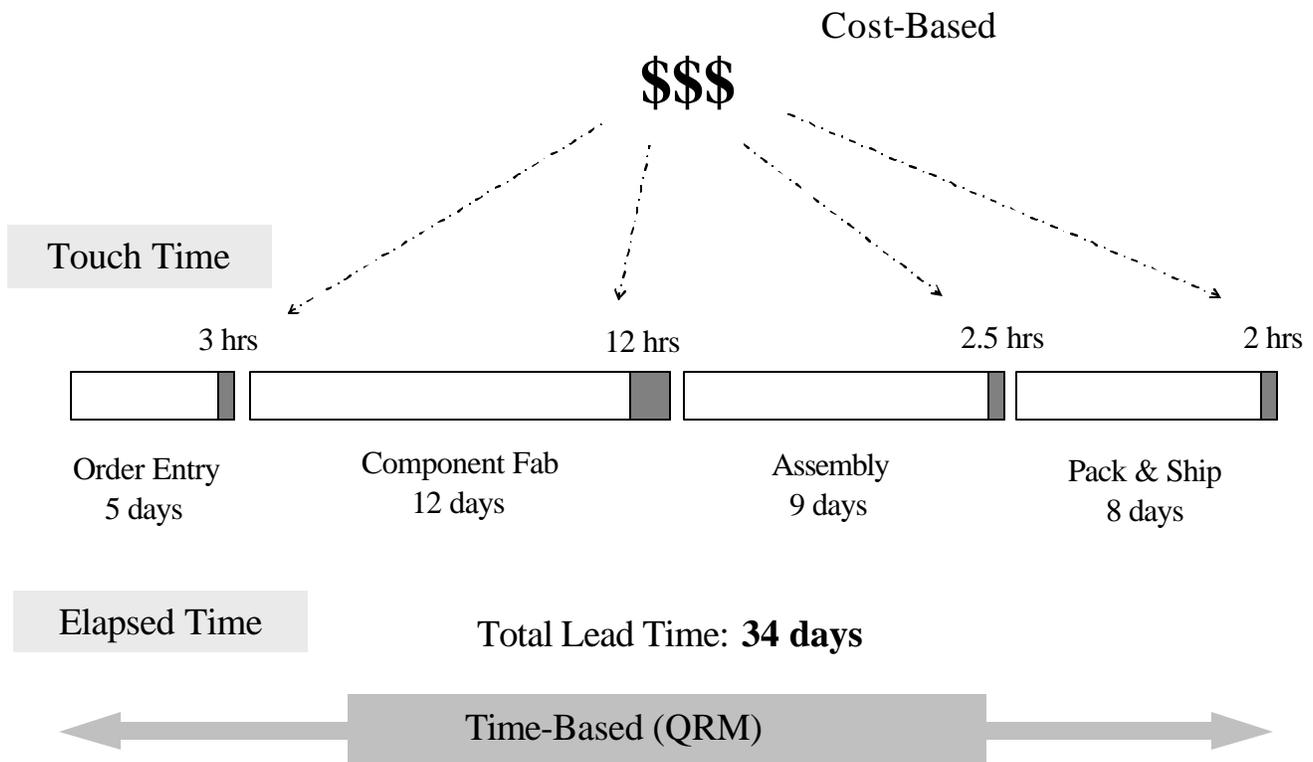
## Overview of QRM Principles

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 2.5 out of the 34 days. The rest of the time is the “white space” in the diagram, where nothing is happening to the job. Traditional approaches focus on reducing the touch time (gray space), while the QRM approach focuses on reducing the total elapsed time.



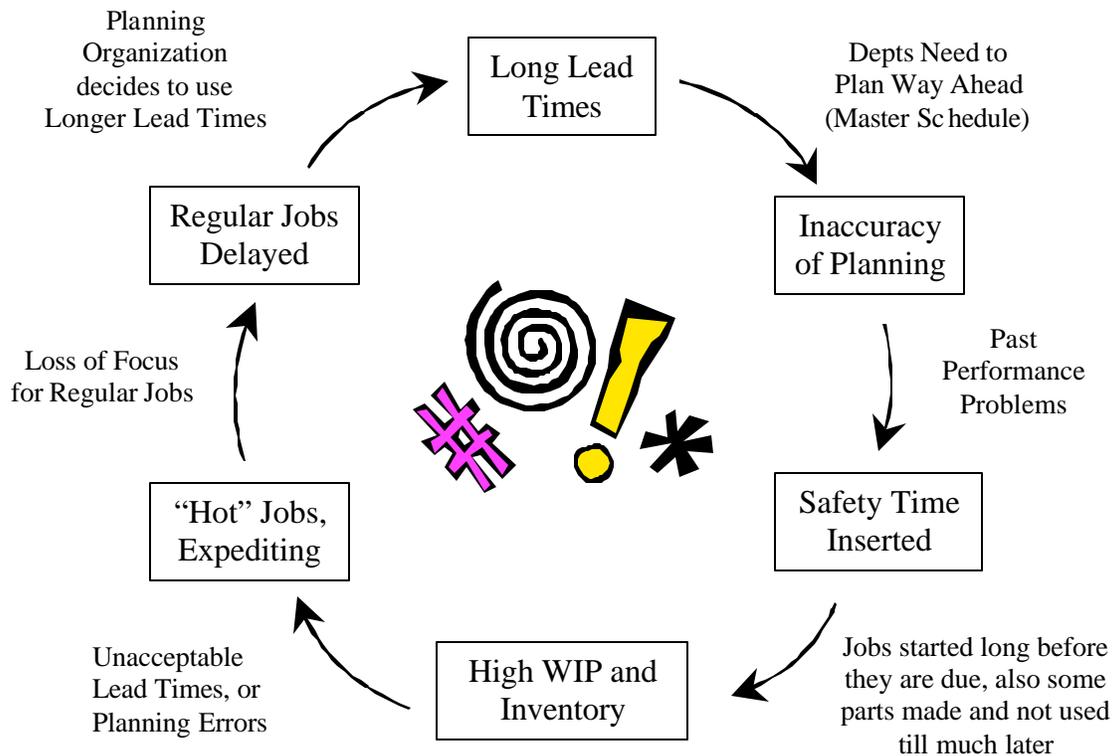
**Figure 1. Comparison of Cost-Based and Time-Based (QRM) Approaches**

However, our organizations are not designed to manage this total elapsed time. *Organizational structures, accounting systems, and reward systems are based on managing large scale operations and minimizing local cost.* For example, in many situations, in order to reduce the white space, we may need to *increase* the gray space. An example of this is running smaller batches, which reduce the queuing and waiting, but require more setups. However, traditional measurement systems will indicate that increasing the gray space is bad (e.g. they predict an increase in standard cost or a reduction in efficiency for that department). The problem is that traditional measurement systems do not recognize the “cost of the white space”. Long lead times result in many additional activities and costs at various levels of an organization, resulting in a substantial addition to the overhead costs. Conversely, significant reductions in lead time can shrink or even eliminate many overhead activities and costs (see Table 1.) Put in concrete terms, a dollar more in labor or machine costs might save tens of dollars in overhead costs. Unfortunately, our accounting systems will not predict this, since there is no direct tie between lead time reduction and overhead reduction in any accounting system. So, as part of its education for management, QRM theory explains that it may be justifiable to increase the gray space if it leads to a substantial reduction in the white space.

**Table 1: Examples of Organizational Waste Due to Long Lead Times**

<p>These are examples of activities and costs that are incurred today, but would shrink or be eliminated if lead times were reduced substantially:</p> <ul style="list-style-type: none"> <li>• Expediting of hot jobs or late orders:             <ul style="list-style-type: none"> <li>– Requires Systems, Air Freight, People, even Top Management time</li> </ul> </li> <li>• Production Meetings required to change and update priorities</li> <li>• Overtime costs for trying to speed up late jobs</li> <li>• Time spent by Sales, Planning, and other Departments to develop and update forecasts</li> <li>• WIP and Finished Goods holding costs, including space</li> <li>• Obsolescence of parts made to forecast but not used</li> <li>• Quality problems not detected till much later; lots of rework or scrap</li> <li>• Opportunity for:             <ul style="list-style-type: none"> <li>– Order changes or even cancellations</li> <li>– Feature and scope creep</li> <li>– Loss of sales to competition</li> </ul> </li> <li>• Sales time devoted to expediting and explaining delays to customer</li> <li>• Complex systems required to manage the dynamic environment</li> </ul>
---

Another legacy of scale/cost based management systems, and the greatest enemy of QRM efforts, is the functional organization with specialized departments. Along with this comes the *Response Time Spiral*. This is an increasing spiral of lead times that results from scale/cost-based management systems (see Figure 2). In essence, the functional department structure requires organizations to plan for sufficient lead time for jobs to make their way through the company. As this lead time is often quite long, there are always “hot jobs” required in less than the normal lead time. These jobs push aside the regular jobs, which get delayed, so the planning organization decides to use even longer lead times in its planning. And so the spiral grows.



**Figure 2. The Response Time Spiral for a Make-to-Order Company**

To understand the spiral, start at the top box (“Long Lead Times”) and follow the arrows and descriptions. Each time around the spiral, lead times get longer and the other problems get exacerbated. Similar Response Time Spirals occur in Make-to-Stock and Engineer-to-order companies, see Suri (1998).

What QRM teaches instead, is that you need to take time out of the system, not allow more time in the system. But that requires substantial reorganization. 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. The POLCA material control system helps coordinate production across multiple cells. 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: Strategically plan for spare capacity – plan to operate at 80% or even 70% capacity on critical resources.*

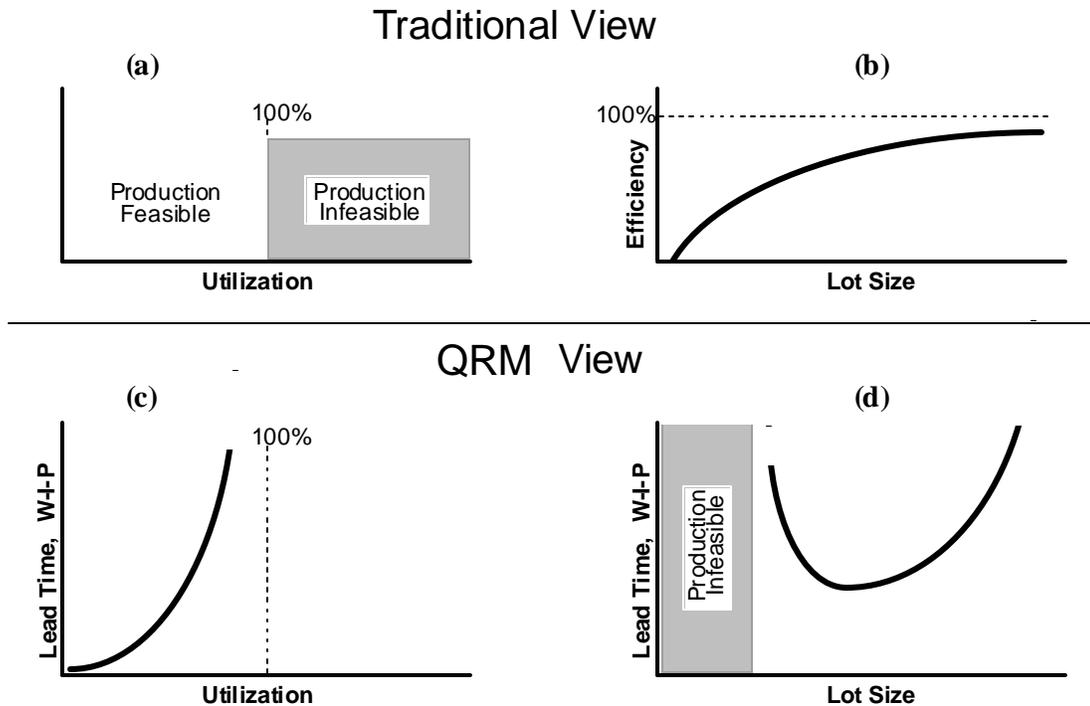
Most managers' reaction to this principle is: "We can't afford to do that. We will have excessive resources and our costs will go up!" However, QRM will eliminate the complex series of dysfunctional interactions that result from the present 100% utilization policy, such as growing queues, jobs spending a lot of time waiting for resources, and resulting long lead times with expediting and other organizational costs, as in Table 1. The QRM approach is to show managers that these dysfunctional interactions result in system-wide costs that exceed the cost of the spare capacity. Thus, QRM theory shows how *spare capacity actually serves as a strategic investment* that will pay for itself many times over in increased sales, higher quality, and lower total 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 the 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 #3 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 Becker (2001) shows how lead time for a line of spare parts for the oil drilling industry dropped from 40 days to 5 days using reduction of lead time as the main performance measure in a manufacturing cell. 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 understand the effects of capacity utilization, efficiency measures, and lot sizing policies on lead time. Using simple mathematics understandable to all levels of the organization, QRM theory demonstrates all these effects. It 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. Managers need to have a basic knowledge of manufacturing system dynamics to understand the impact of their policies on lead times (see Figure 3).



**Figure 3. Traditional Versus QRM Views of Capacity and Lot Sizing**

Traditional performance measures of *utilization* and *efficiency* encourage managers to maximize resource utilization, and only think about their capacity limit as a boundary between feasible and infeasible production targets as shown in (a), and to run large lot sizes, as in (b). With QRM's focus on reducing lead time, it is important to understand the impact of utilization on lead time, see (c), as well as the effect of lot size on lead time (d). QRM theory includes fundamental principles of manufacturing system dynamics which provide insights such as these about the impact of management policies on the enterprise's lead time.

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, the emphasis on it as a performance measure 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, lead times get longer, and on-time deliveries actually get worse! 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 improved on-time performance – see the results in Table 2.

**Table 2. Impact of Lead Time Reduction on On-Time Performance\***

<b>Company (Product Type)</b>	<b>% Reduction in Lead Time</b>	<b>% On-Time Performance (Before → After)</b>
Hydraulic Motors	57	40 → 97
Seat Assemblies	80	40 → 95
Hydraulic Valves	93	40 → 98
Wiring Harnesses	94	43 → 99

\*Data taken from Ericksen and Suri (2001).

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 POLCA, 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. QRM begins by restructuring the organization into cells. Then, in the

redesigned QRM organization, MRP is used to provide high level planning and coordination of materials from external suppliers and across internal cells. This High Level MRP (HL/MRP) is used to set overall lead time for each cell, but not to micromanage each of the workcenters in the cells. Instead, 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, helps to manage and coordinate the material flow across cells. 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 described 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 lot deliveries 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 ordering 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 (see the results in Ericksen and Suri, 2001).

*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 deliveries at reasonable prices.*

This is the reverse of #6. A company's sales force 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 demonstrate how QRM will allow them to receive smaller batches with shorter lead times *and* at lower prices.

*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 (Suri, 1998). *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. 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 company-wide 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 productive company with a more secure future. Also, lower cost/price, higher quality and shorter lead times result in highly satisfied customers.*

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. (Also it is risky, since charging more might motivate your customers to look for alternative sources.) The driving reason for implementing QRM is that searching for ways of squeezing time out uncovers quality problems and wasted efforts. Fixing these results in higher quality, lower WIP, less overhead, 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.

Table 3 shows examples of results from a few QRM projects. Not only were dramatic lead time reductions achieved (over 90% in some cases) but there was also a significant impact on product cost and quality. Additional results can be found in Tubino and Suri (2000).

**Table 3. Impact of Lead Time Reduction on Cost and Quality\***

Company (Product Type)	% Reduction in Lead Time	% Reduction in Product Cost	Quality in PPM <sup>†</sup> (Before → After)
Hydraulic Motors	57	13	15,000 → 500
Seat Assemblies	80	16	50,000 → 500
Hydraulic Valves	93	14	50,000 → 1,500
Wiring Harnesses	94	20	3,000 → 500

\*Data taken from Ericksen and Suri (2001).

<sup>†</sup> PPM measures Parts Per Million defects observed after delivery of product.

*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 expensive 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 (Meyer, 2002). 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, simple fixes to the accounting system involving strategic pools created by management can go a long way towards making the accounting system support QRM (Suri, 1998).

This concludes our summary of the main QRM principles. Now we describe the new material control strategy devised as a part of the QRM approach for companies serving 21st century markets.

## The Need for an Alternative to Pull for Material Control

Pull (or Kanban) systems are a key technique used in Lean Manufacturing, and have been very successful at many companies. In particular these pull systems have helped to overcome some of the problems of MRP such as excessive WIP, constantly changing dispatch lists, and the need for expediting. If pull has worked so well, why do we need another material control system?

The answer lies in looking toward the future, specifically at the 21st century markets which we described earlier as markets that require a large variety of products and have variable demand for each type of product, and/or markets that require custom-engineered products. We will see that pull systems have multiple drawbacks in serving such markets. To understand this, let us review three key concepts in Lean (from Womack and Jones, 1996), and at the same time we will 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.

### ***Waste in Lean Versus Waste in QRM***

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 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 ten key forms of waste that needed to be eliminated. They posted 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.*

### **Strategic Variability and Where Lean Concepts of Flow and Takt Time Break Down**

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), and also, multiple cells can be combined in different sequences to make custom products. 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, consider a company serving 21st century markets, which require that products be customized with very different specifications, or require the delivery of a large number of products with highly variable demand for each one. In this situation, the variability in processing needs, as well as the variability in the 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. In addition, given these variations as well as the potential for customized products requiring unique processing sequences, 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.

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 21st century markets’ context, the variability 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 serving 21st century markets. Instead, the QRM approach tackles this variability in requirements while still achieving 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).
- Implementation of novel constructs such as time-slicing and POLCA.
- Understanding and exploiting system dynamics that result from this type of interaction and variability, and then making appropriate capacity and lot sizing decisions that help to achieve performance targets in the presence of 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.

An insight into "variability" may help to sharpen this point and also to clarify the differences between Lean/Flow and QRM. We will define two types of variability. The first we call *dysfunctional variability*, which is caused by errors, ineffective systems and poor organization. Examples of dysfunctional variability are: rework; changing priorities and due dates; and "lumpy" demand due to poor interfaces between sales and customers. The second type of variability we call *strategic variability*, which an organization uses to maintain its competitive edge in the market. Examples of strategic variability are: the ability to cope with unexpected changes in demand without degradation of service; offering a large number of options to customers; or even offering to custom-engineer products for individual customers. The way that Lean/Flow work is by attempting to eliminate *all* variability in the manufacturing system. This is good as far as eliminating dysfunctional variability, since it leads you to work on the root causes of this and eliminate them. However, you may not want to eliminate strategic variability, particularly if it is the basis of your competitive advantage. The QRM approach agrees with the Lean/Flow approach in trying to get rid of all dysfunctional variability. However, in QRM we do not try to eliminate strategic variability, instead we try to design the organization and systems so that we can cope with this variability and serve those customer markets effectively. These markets are, in fact, the 21st century markets described earlier.

Another requirement when implementing flow is the use of *flex fences* to deal with components from 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 waste (*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-

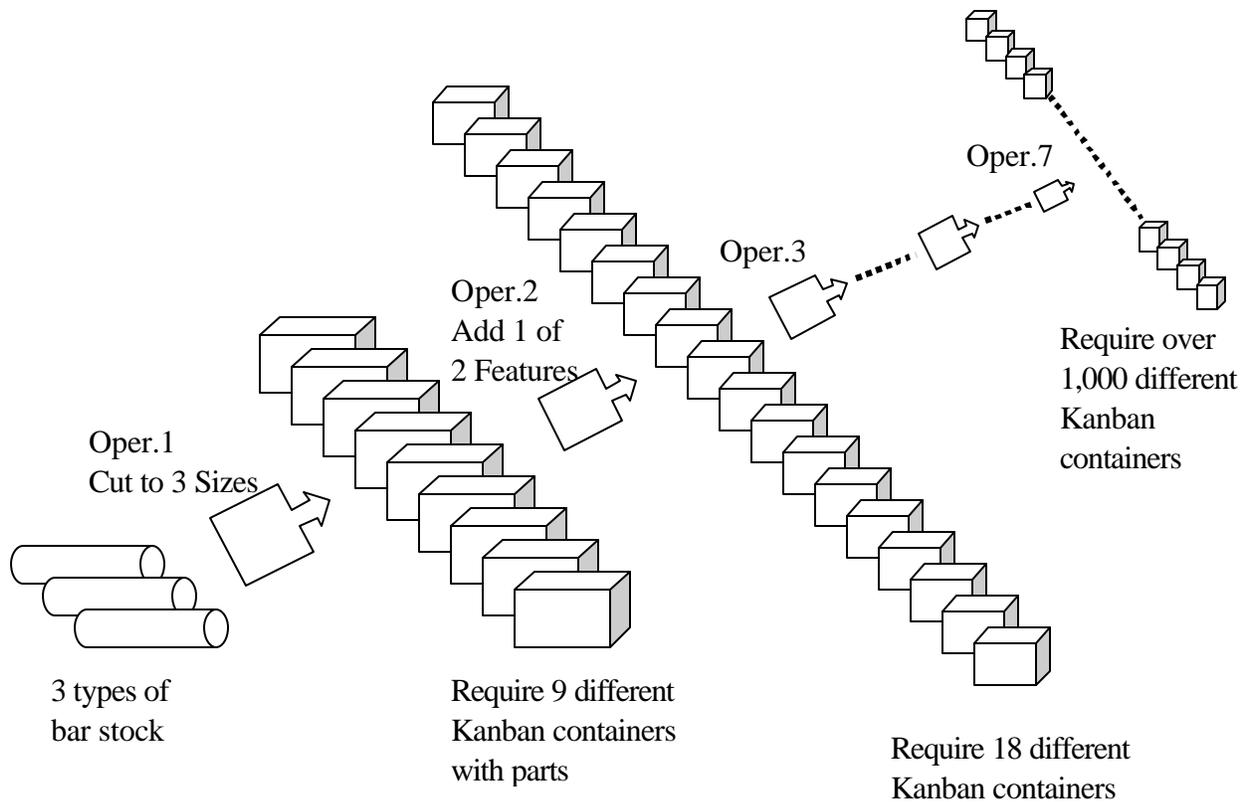
engineered products, the company often will not know the design of a component until after an order is received and engineered. In either case, the flex fence approach is not practical. The QRM approach instead changes both the operation of the suppliers (Erickson and Suri, 2001), as well as the very structure of the interaction between a company and its suppliers (see the details in Suri, 1998).

### ***Why Pull Might Result in Proliferation of Inventory***

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.

While pull systems have been widely touted, their disadvantages are less widely publicized. Consider the implications of the above system for a company serving 21st century markets. 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 a 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 4). This is a vast amount of inventory: is this not again *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 and design engineering completed. 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.



**Figure 4. Potential for Proliferation of WIP in a Pull System with Many Products**

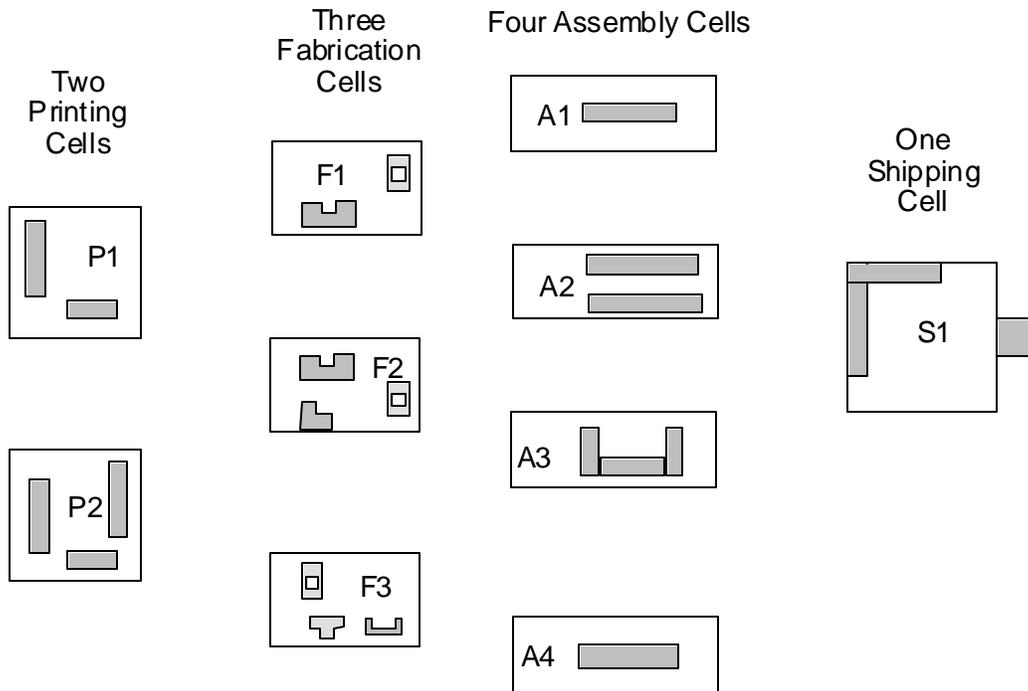
The preceding discussion makes it clear that there are two prerequisites for success of a Lean strategy. One is a limited degree of customization. Specifically, Lean systems are designed to make products that involve minor customization of a main product, or products where the customization involves choosing from a set of predefined options (e.g. assemble-to-order), as opposed to totally custom-engineered products. The second prerequisite for Lean 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 some 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% in three years! 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. It is clear that 21st century markets will contain more and more opportunities for such 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 companies that are trying to serve 21st century markets. 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. In order to support the overall QRM approach for companies serving 21st century markets, there was a need to develop a whole new material control method. We have devised such a method called *Paired-cell Overlapping Loops of Cards with Authorization*, or POLCA.

To understand the basis for POLCA, let us review the types of companies serving 21st century markets. These are: (i) companies that make custom-engineered products in small batches (or even one-of-a-kind), and (ii) companies that don't custom engineer 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. HL/MRP, described earlier, is used to provide high level planning and coordination of materials from external suppliers and across these internal cells, but not micromanage workcenters in the cells. The lead times for each cell are set in the HL/MRP system, which uses this information to back-schedule from ship dates and determine start dates for an order at each cell that it will visit (but this will be qualified below).

To proceed from this high-level structure to the next level, which involves shop floor material movement and POLCA, we will consider an example of a company called CFP Corp. This company 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, and 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 5. Organization of Cells at CFP Corporation**

CFP has created several cells to serve its highly varied markets via a QRM strategy (Figure 5). First there are two Printing Cells: P1 focuses on screen printing and 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 fasteners, 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

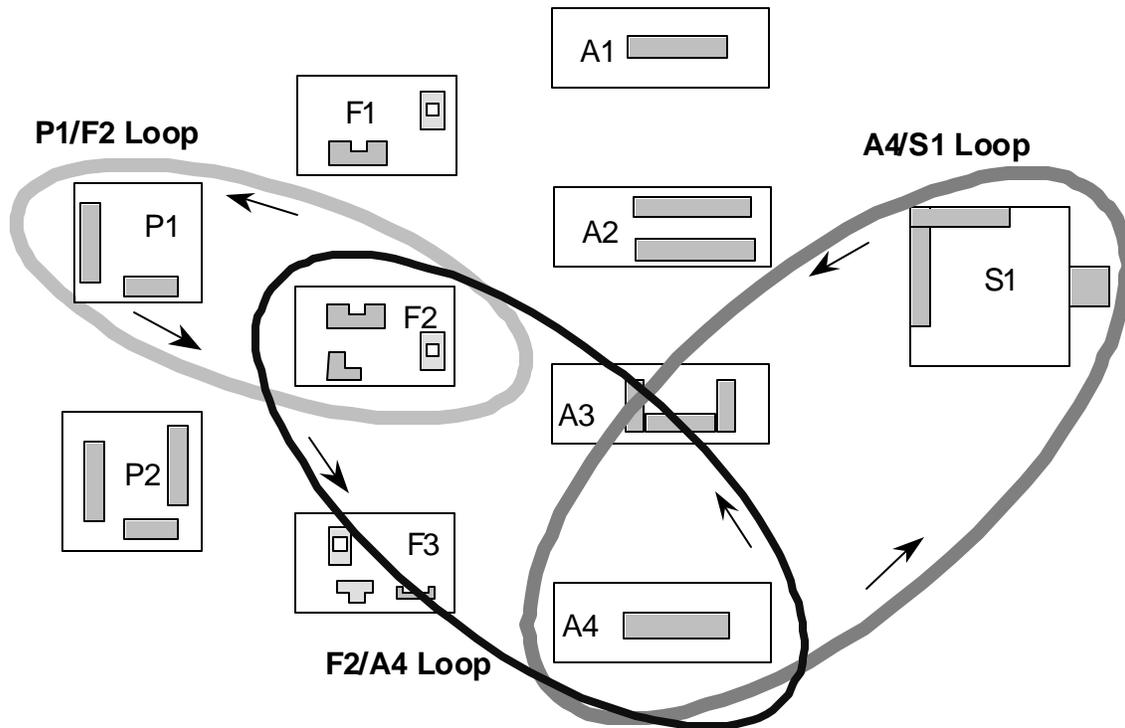
different combinations of cells, as needed by a given order; (2) within a cell, the ability for products to use machines in different sequences; and (3) a good deal of flexibility in terms of capacity requirements for each operation in a cell. We should note that the original intent of MRP systems was to enable a company such as CFP to achieve these requirements, but for reasons that are by now well-documented 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 POLCA (*Paired-cell Overlapping Loops of Cards with Authorization* – each of these terms will be explained below). This system 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 POLCA. We now describe the key features on which POLCA is based.

First, for each order, *Release Authorization times* are created via HL/MRP. Similar to start dates in an MRP system, the HL/MRP system generates times when each cell may begin work on a particular order as explained above. 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, which 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 several 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, are not specific to a product, as in a pull system, but are assigned to *pairs of cells*, and apply to all products going from the first cell to the second cell in the pair. Figure 6 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.

Third, and this is a key difference, whereas a Kanban card is an *inventory* replenishment signal, a POLCA card is a *capacity* signal. Specifically, a POLCA card returning from a downstream cell signals that the cell has available capacity to process more work. Thus, when a workcenter reviews its list of jobs whose start has been authorized, it can only work on a job if it has a POLCA card from the destination cell. For example, if cell P1 has a job authorized that is going to F3 next, then a P1/F3 card must be available at P1 in order for it to begin that job. If a P1/F3 card is not available, that means that there is a bottleneck at F3 and working on that job will only add to the work-in-process at F3. Instead, it would be better for P1 to put its resources into a job that is needed by another cell that is not backlogged. So the cell team at P1 skips the P1/F3 job for now, and looks at the next authorized job to see if a card is available for that job, and so on.



**Figure 6. POLCA Card Flows for a Particular Order at CFP Corp.**

The characteristics of this custom-engineered order require it to be processed in cells P1, F2, A4 and S1. Thus, this order will flow through the company by using the POLCA loops P1/F2, F2/A4 and A4/S1 as shown above.

The fourth 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 6.

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 Push (MRP) and Pull (Kanban) Systems***

Now we discuss how, for companies serving 21st century markets, the POLCA system overcomes the drawbacks of both push and pull systems. Industry has already recognized that the use of cells is a prerequisite to competitive manufacturing. Cells also form a key building block of QRM strategy. Note that POLCA builds on the cellular structure in an organization,

and provides a simple mechanism to enable the cells to work together effectively. In a sense, POLCA combines the best of push and pull systems, while at the same time avoiding their disadvantages. Let us elaborate on this.

First, POLCA helps in managing short-term fluctuations in capacity and also assists in reducing congestion on the shop floor. Essentially, 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 (e.g. this cell could work on another job that is needed soon by a different downstream cell).

Second, the use of HL/MRP as a driver in POLCA has two benefits: first, it allows a make-to-order environment through flexible routings that use cells as needed, and second, the use of authorization times generated by HL/MRP prevents build-up of unnecessary inventory. As we showed in earlier examples, for the typical 21st century markets context pull systems have the disadvantage of filling intermediate stages with inventory. By coupling the routing and authorization procedure using HL/MRP and the POLCA scheme, we ensure that the company does not make products just because those products have a pull signal; it makes products only when there is explicit demand for them.

Third, POLCA cards are not linked to part numbers. This ensures that there is no proliferation of inventory for companies that make a large variety of products or components. Since pull is essentially a replenishment system, it requires a base stock level of each component, which is replenished when it is used. In a high mix environment this could result in a very large amount of inventory, particularly for low volume products.

Fourth, unlike a pull 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 21st century markets 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, but the actual rates and bottlenecks will vary from day to day. This is one 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. There are additional reasons for having the overlapping loops, see Suri (1998) and Suri and Krishnamurthy (2003).

### **Case Studies on POLCA Implementation**

In partnership with its member companies, the Center for Quick Response Manufacturing (QRM) has implemented POLCA at several factories. For example, Rockwell Automation has implemented POLCA at factories in the US and Canada that make motor control centers (Honerlaw and Cronce, 2001; Gilson, 2002). These factories already had cells in place and had obtained substantial lead time reductions prior to implementing POLCA. The authors report that POLCA led to additional 15-30% reductions in lead time and work-in-process. At one factory, on-time deliveries between cells increased from under 40% to over 90%, verifying that POLCA truly helps the coordination of material flow across cells. At another factory, POLCA helped to increase output by 18%. It also eliminated the time previously spent by team leaders and schedulers on expediting. The system has been accepted – even praised – by shop floor employees and schedulers. A second example is provided by Olsen Engineering, a contract manufacturer that supplies hardened and precision ground steel pins, bushings, miscellaneous CNC parts, and tube bending parts. The manufacturing facility produces over 5000 different part numbers with highly variable demand, so pull did not seem appropriate, and thus the management team decided to implement POLCA (Dawson, Hansel, and Miller, 2002). The authors report that POLCA resulted in lead time reductions ranging from 22% to 68%. Work in process and stock inventories were also reduced significantly, in some cases by as much as 90%! In addition, the POLCA system significantly improved the operator morale and instilled a culture of continuous improvement at the facility.

The experiences with these implementations have led us to develop a step-wise procedure for implementing POLCA in a factory. These steps address several practical issues such as establishing the prerequisites for POLCA, determining the POLCA loops, computing the number of POLCA cards, determining the quantum of work a POLCA card represents, and addressing part shortages. Details of this step-wise approach can be found in Krishnamurthy (2002) and Suri and Krishnamurthy (2003).

## **Prerequisites for Successful Implementation of QRM**

We conclude this paper with a perspective on how to successfully implement the QRM strategy in an organization. As we 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 enterprise, 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 eventually 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. To get the maximum lead time reduction and the most benefit from the QRM program, all these policies need to be made consistent with the QRM ideal. However, note below that this should be achieved in stages, not all at once.
- *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 previously in this 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 should be reorganized for QRM, eventually spreading the QRM program to the whole enterprise, as discussed above.
- *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 detailed roadmap for successful implementation (Suri, 1998). It is important for management to review the entire map early on in the initiative, so that they buy in to the whole plan.

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 companies 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.

As the enterprise implements QRM, it will find itself more and more effective at serving 21st century markets. While the competition struggles to find efficient ways of dealing with the myriad of options or customizations demanded by 21st century customers, this company – using the winning combination of QRM and POLCA – will not only satisfy these customers, but it will do so with higher quality, lower price, and shorter lead time than the competition. In so doing, this firm will substantially increase its market share and gain long-term competitive advantage.

## Bibliography

- Blackburn, J.D. (Ed.) 1991. *Time-Based Competition: The Next Battle Ground in American Manufacturing*, Business One Irwin, Homewood, IL.
- Becker, T. 2001. “It’s the People that Make the QRM Cell Successful,” in Suri et al. (2001) cited below.
- Charney, C. 1991. *Time to Market: Reducing Product Lead Time*, Society of Manufacturing Engineers, Dearborn, MI.
- Dawson, M., T. Hansel and B. Miller. 2002. “POLCA Implementation in the Manufacture of Pins with a High Variety of Dissimilar Routings and Order Quantities”, in Suri and Rath (2002) cited below.
- Ericksen, P.D. and R. Suri. 2001. “Managing the Extended Enterprise,” *Purchasing Today*, Vol.12, No.2, February, pages 58-63.
- Gilson, S. 2002. “Results of Implementing POLCA at Rockwell Automation, Richland Center,” in Suri (2002) cited below.

- Honerlaw, S. and L. Cronce. 2001. "Implementation of POLCA in Assemble-to-Order/ Engineer-to-Order Applications at Three Rockwell Automation Manufacturing Facilities," in Suri et al. (2001) cited below.
- Hopp, W.J. and M.L. Spearman. 1996. *Factory Physics: Foundations of Manufacturing Management*, Richard D. Irwin, Burr Ridge, IL.
- Krishnamurthy, A. 2002. "How to Plan and Implement POLCA – The Material Control Strategy for High-Mix/Low-Volume or Custom Products", in Suri and Rath (2002) cited below.
- Meyer, J. 2002. "Navigating the Journey to QRM from Traditional Functional Based Manufacturing," in Suri and Rath (2002) cited below.
- 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.
- Suri, R. (Ed.) 2000. *Proceedings of the Quick Response Manufacturing 2000 Conference*, Society of Manufacturing Engineers, Dearborn, MI,
- Suri, R., F. Rath and T. Dewar (Eds.) 2001. *Proceedings of QRM 2001: The Second Annual Conference on Quick Response Manufacturing*, Center for Quick Response Manufacturing, Madison, WI.
- Suri, R. (Ed.) 2002. *Proceedings of the 2002 POLCA Implementation Workshop*, Center for Quick Response Manufacturing, Madison, WI.
- Suri, R. and F. Rath (Eds.) 2002. *Proceedings of QRM 2002: The Third Annual Conference on Quick Response Manufacturing*, Center for Quick Response Manufacturing, Madison, WI.
- Suri, R. and A. Krishnamurthy. 2003. "How to Plan and Implement POLCA – A Material Control System for High Variety or Custom-Engineered Products," Technical Report, Center for Quick Response Manufacturing, University of Wisconsin-Madison, WI.
- Tubino, F. and R. Suri. 2000. "What Kind Of Numbers Can A Company Expect After Implementing QRM?" in Suri (2000) cited above.
- 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.