

# INTEGRATION OF PLANNING AND EXECUTION OPERATIONS: THEORY AND CONCEPTS

Bharat T. Doshi and Dan C. Krupka

**Bharat T. Doshi and Dan C. Krupka** are with AT&T Bell Laboratories. Mr. Doshi is a supervisor in the Performance Analysis Department at the Holmdel, New Jersey facility. He and his group develop general methods and tools for performance analysis and analyze and improve the performance of computer, communication, and manufacturing systems. He received a B.Tech. in mechanical engineering from the Indian Institute of Technology, Bombay, and an M.S. and Ph.D. in operations research from Cornell University. He joined AT&T in 1979. Mr. Krupka is head of the Manufacturing Systems Engineering Department at Murray Hill. He works to improve manufacturing operations through quantitative analyses and by directly participating in improvement programs. His department also works on methods to control (continued on page 98)

On the surface, material requirements planning (MRP) systems, generally associated with push manufacturing, appear to be incompatible with pull manufacturing, often identified with just-in-time (JIT) techniques. In this paper, we clarify these concepts and demonstrate how MRP and pull can be made to coexist in a system that we have dubbed *integrated pull manufacturing* (IPM). In IPM, MRP is used for planning, while pull is used for shop-floor execution. IPM is also used to synchronize the operation of feeder lines having disparate manufacturing intervals. The conditions at AT&T's Denver Works that allowed the implementation of IPM are described and a high-level description is given of day-to-day IPM operations.

## Introduction

Other papers in this issue describe advances made at the Denver Works in manufacturing planning and production. For planning, a material requirements planning system was installed; for shop floor execution, "pull" systems were instituted. Now MRP—or, more correctly, manufacturing resources planning (MRP II)—is generally associated with "push" when it is used to control shop floor operations. At first sight, therefore, there seems to be a need to reconcile the apparent incompatibility of MRP's push and the pull used for execution. Indeed, in the early phases of shop-floor improvements, suggestions were made for transforming the materials planning process to a pull mode. It happens, however, that MRP for planning and pull systems for execution can coexist. In Denver, for manufacturing AT&T's Definity® telecommunications system products, they complement each other very successfully through integrated pull manufacturing. In fact, papers have appeared since IPM was introduced at Denver that describe other approaches for reconciling MRP and pull systems.<sup>1,2</sup>

In the next section, we discuss the apparent incompatibility between MRP and pull. We draw a distinction between just-in-time manufacturing and pull, and point out that applying MRP in a highly predictable but dynamic environment can lead to JIT operation

**Panel 1. Terms and Acronyms in This Paper**

BOM	bill of materials
Definity system product line	AT&T's mid- to high-range digital customer-premises switching and telecommunications products
IPM kanban	integrated pull manufacturing literally, card; system in which a fixed number of cards are used to control production and inventory simultaneously
JIT lead-time-based scheduling	just in time same as look-ahead policy
look-ahead policy	production scheduling based on knowledge of future demands (or forecasts of demand)
MRP	material requirement planning; system for planning production, material requirements, and purchasing based on a mixture of near-term hard orders and long-term forecasts
MRP II	manufacturing resource planning; combines MRP functionality with planning of manufacturing resources, tracking of jobs through manufacturing steps, and providing closed-loop control via feedback from the shop
TQC	total quality control
WIP	work in process

throughout the manufacturing system. In a real environment, such as that encountered in Denver, execution using pull is preferred.

If the Definity system manufacturing process consisted merely of equipment assembly, the integration of MRP and pull would be relatively straightforward. However, in reality, two major subassembly flows must be synchronized: assembly and test of circuit packs and assembly of equipment. The manufacturing interval for the former is measured in days; that for the latter is

measured in hours. Thus, in addition to resolving the MRP-pull issue, we needed to devise a method to synchronize the operations of all shops feeding final assembly when these shops differ significantly in manufacturing interval. The solution to this problem is also presented in later sections.

To create a full understanding of the system, it is also necessary to describe how it operates week by week, day by day, and hour by hour. Such a description, with a discussion of major concepts, can be found in later sections.

Before presenting our solution, we summarize the situation in Denver that permitted *implementation* of IPM. Although the concepts were created early in the improvement cycle, implementation had to await the establishment of an environment with sufficiently short intervals and sufficiently high predictability. Perhaps the most important achievement was successfully implementing MRP. For example, as a result of the implementation, storeroom balance accuracies were high, bills of materials were accurate, and the production control organization determined what was to be built each week. On the shop floor, through the techniques described in Reference 3, manufacturing intervals for circuit packs had dropped to a few days. More importantly, the standard deviation was approximately half a day. The final systems were scheduled—and are currently scheduled—in weekly “buckets.” That is, it was immaterial whether a system was completed on a Monday or Friday. This freedom, as is demonstrated in detail in the following section, allows smoothing of the production schedule for all subassemblies.

**Developing a Solution**

Before describing our development of a solution, we distinguish between a material requirements planning system and pull-type manufacturing.

**MRP versus Pull Manufacturing.** In a totally predictable production environment, demand forecasts would always be realized; bills of materials would be absolutely

---

accurate; suppliers would ship their orders on time and with total accuracy; nothing would be misplaced or miscounted in the storeroom; machines on the shop floor would never fail; all manufacturing personnel would be present when expected; and all intervals in the process would be fully predictable. In such environments, all activities from materials planning to shop execution could be scheduled by using a push or MRP II system; every component would reach its destination just in time and every subassembly would be completed on schedule—just in time.

However, our environment is not so predictable. To ensure product delivery in the face of the inevitable problems that arise throughout the chain, from materials ordering to product manufacture, slack is built into the schedule of an MRP II system. Since this slack is demanded by the system, there is little incentive to reduce it. Moreover, MRP II systems inevitably suffer from incorrect lead times because they do not recognize that lead times depend on the congestion or load within the shop.

Given the real environment, could a system that is based totally on pull perform better? For a product that is built to order, such as the Definity system, the answer is “no.” It is “no” because it would be impossible to know what components and subassemblies to place upstream in the pipeline to satisfy this week’s demand. A pull system knows nothing of the future; it merely initiates production to replenish what was withdrawn. A simple pull system in an environment with dynamic demand will result in excess inventory of finished goods and/or in failure to meet delivery requirements. This is contrary to the goals of a JIT system.

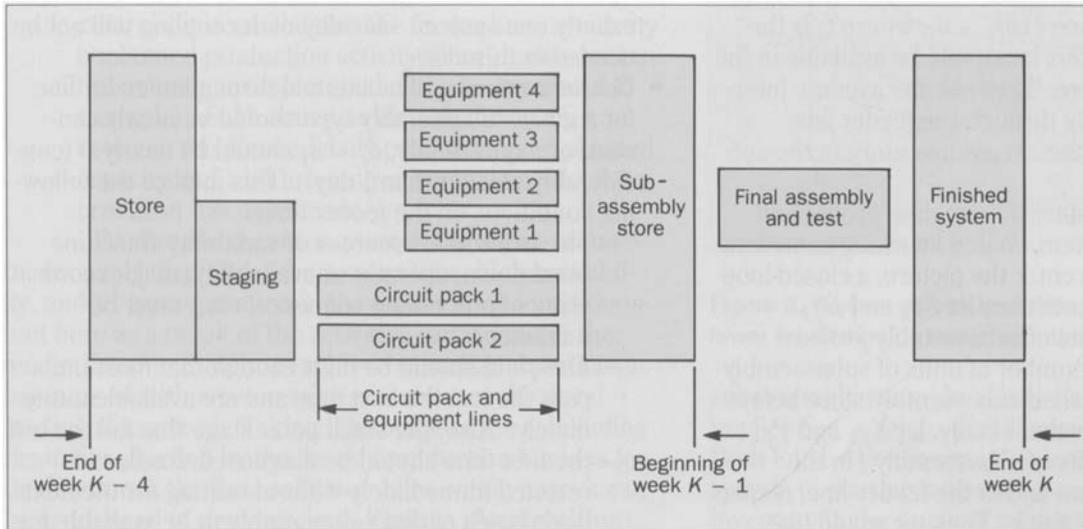
And yet, since the advantages of a pull system are apparent from the accompanying papers, it is appropriate to summarize them. A pull system, using kanban cards, has the virtue of combining the control of production with control of inventory: the number of kanban cards determines the inventory level and the issuance of a card triggers production. As the number of cards is lowered, the corresponding reduction in lead time

becomes very visible, as do problems with the process. This provides an incentive to improve operations by reducing setup times and equipment downtimes and by striving for shorter service times and higher quality. A pull system also filters out demand surges. A fixed set of kanban cards between two stages sets an upper limit on the number of subassemblies that can be drawn from the preceding stage. Overall, a pull system works well in an environment with little demand variability. MRP II type systems, on the other hand, work well in the presence of high demand variability but with highly predictable intervals.

The challenge, therefore, is to marry the inherent strengths in planning of an MRP system with the responsiveness of a pull execution system. In the rest of this section we discuss the tradeoffs in detail and develop the overall approach at a conceptual level. Recall that the final system deliveries are scheduled in weekly buckets, the interval in the circuit-pack lines is slightly under a week, and in the equipment lines it is a few hours. Also recall that the final assembly and test for a system can be completed in one week. These conditions permit a simple solution in which the whole operation is decomposed into three steps of 1 week each.

#### **A Simple Solution**

Given the above conditions, the following is a possible approach to production management (Figure 1): at the end of week  $k - 4$ , look at the hard orders for week  $k$ , translate the system bills of materials (BOMs) to create the subassembly needs and translate the subassembly BOM again to create the raw material needs (grouped into “picks”) in the storeroom. The required picks are then completed during week  $k - 3$ . The feeder lines then use the material (staged in the storeroom) as the capacity becomes available. At the end of week  $k - 2$ , all the subassemblies needed for the systems to be delivered in week  $k$  are available in the finished subassembly store. The final assemblies will be started in week  $k - 1$ . By the end of week  $k$ , all the systems to be delivered in week  $k$  will be ready for shipping. This is



**Figure 1. Overall operation of a simple solution to production management. Orders for week  $K$  are examined 4 weeks earlier (that is, in week  $K - 4$ ). The operation resembles MRP, but is in weekly buckets.**

an MRP-like approach between major staging areas but is in weekly buckets. It is simple to implement and it decouples the feeder line operations from the final assembly operation during a week. Moreover, it does not put any constraints on the subassembly feeder shops regarding the order in which the material withdrawn from the storeroom should be processed within a week. Each feeder line can then select the order most suitable for its own operation. On the other hand, this strategy builds an extra interval (of a week or so) into the whole process. In particular, the finished subassembly store (as well as the storeroom staging area) will carry a significant amount of inventory. This has obvious inventory cost and customer delivery interval implications. In addition, with the potential of design changes, this may lead to occasional large amounts of unwanted inventory.

An ideal solution would be to predetermine the order in which the final assemblies will be started in week  $k$  and then schedule the whole feeding operation so that all the needed subassemblies for a given system arrive just in time for the beginning of the final assembly. As we will see in the next section, this ideal is hard to

achieve in the environment under consideration. Even allowing up to 1 day's (or one shift's) worth of staging requires that many conditions be met. We first discuss the issues that should be addressed and prerequisites on the operation that must be met before attempting to reduce the staging interval to less than 1 day.

#### **Prerequisites for a Preferred Solution**

Suppose the sequence in which the final assemblies have to be carried out in week  $k$  has been determined somehow. This effectively determines what should be built on a given day of that week (we allow some flexibility within a day). All the subassemblies needed for the systems to be built on Monday should be available on Monday (at the start of the day), and so on.

Figure 1 represents the feeder lines providing the needed subassemblies. Note that some feeder lines (circuit-pack assembly and testing, for example) provide more than one type of subassembly. Suppose a feeder line has enough capacity and line discipline to make a subassembly lot of type  $j$  in less than  $d_j$  days with very high probability. Then the latest time a lot of type  $j$

should enter the feeder line  $i$  is  $t_j - d_j$ , where  $t_j$  is the day by the end of which this lot should be available in the finished subassembly store. Let  $\bar{d}_j$  be the average interval of a type  $j$  subassembly through the feeder line. Then  $d_j - \bar{d}_j$  represents the excess inventory in the subassembly store.

The above procedure for deciding production starts is an open-loop system. When lot sizing considerations and potential losses enter the picture, a closed-loop schedule is needed. In particular, let  $S_{j,k}$  and  $D_{j,k}$  denote the number of units of subassembly  $j$  to be started on day  $k$  and the number of units of subassembly  $j$  needed to be in the finished subassembly store before the end of day  $k$ , respectively. Finally, let  $X_{j,k}$  and  $P_{j,k}$  denote the number of units of subassembly  $j$  in the finished subassembly store and in the feeder line, respectively, at the beginning of day  $k$ . Then we would like

$$94 \quad S_{j,k} = \text{LOT} \left[ D_{j,k} + D_{j,k+1} \cdots + D_{j,k+d_j} - X_{j,k} - P_{j,k} \right] \quad (1)$$

where  $\text{LOT}[x]$  is the value of  $x$  rounded up to meet the lot size constraints. The above discussion implies the following prerequisites on the feeder lines for implementing nearly stockless operation of the finished subassembly store:

- **Lot sizes.** The minimum lot sizes for most, if not all, subassemblies should be no more than the typical daily requirement. If the minimum lot size for a subassembly is, say, about 1 week's requirement, then equation (1) implies cyclical production—that is, a large lot started on day 1 followed by 4 days of no starts for this subassembly, followed by another large lot, and so on. Clearly, this would add roughly half a week of inventory and interval in the finished subassembly store and would effectively preclude stockless operation. On the other hand, if the daily requirement is one or more lots of minimum sizes, the starts can be timed to match the demand exactly on a daily basis. Of course, if the minimum possible lot size is 1, then much tighter control can be exercised by having a lot correspond to the demand for

exactly one system. One day of decoupling will not be needed in this case.

- **Constant interval.** The interval through a feeder line for a given subassembly type should be nearly constant or, equivalently,  $d_j - \bar{d}_j$  should be nearly 0 (considerably smaller than 1 day). This implies the following conditions on the feeder lines:
  - The extraneous sources of variability (machine breakdown, operator unavailability, major reordering of processing sequence, etc.) must be minimized.
  - The yield should be high enough that most units pass the test the first time and are available immediately. Also, the small percentage that fail the test the first time should be diagnosed, fixed, and retested immediately without waiting for the next day's batch of that subassembly to be available for testing.<sup>4</sup> This requirement means that setup time for changing test fixtures must be short and test/diagnostics/repair stations must be collocated.
  - The total load of work sent to any work area in a feeder line during a day should be less than the daily production capacity of that work area. In general, the work areas are sized so that each has enough capacity to meet the *average* daily requirement in a day. Typically, it will have some reserve capacity. To keep the interval through the line nearly constant, it is necessary to make sure that the requirement on a given day does not exceed the average requirement plus the reserve capacity. Thus, either the variation in demand should be kept to the minimum or additional reserve capacity should be provided in the potential bottleneck work areas.
  - While the latest start for the subassembly type  $j$  needed on day  $t_j$  should be  $t_j - d_j$ , it should not be much earlier either. This is especially true for subassemblies that are expensive to hold in inventory and that are likely to be made obsolete by, for example, design changes. This implies very tight controls on production starts. Given these tight

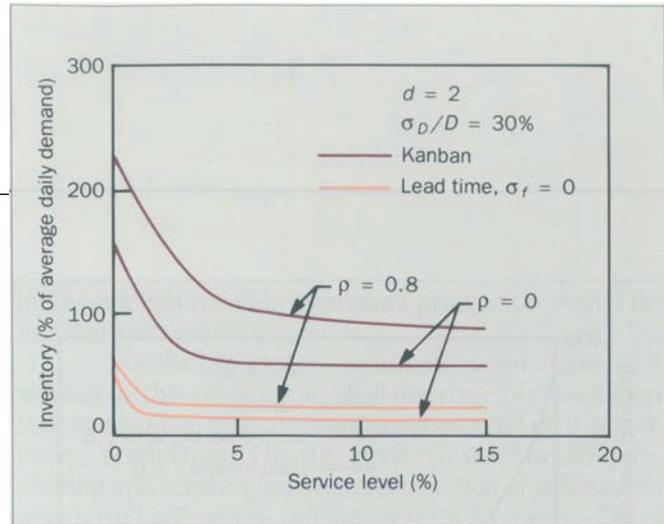
controls, it is crucial that the feeder lines avoid any unplanned production activity ("hot" orders within the manufacturing interval of the feeder line, for example). Such activities disturb the balance between the planned daily production and planned capacity and lead to a longer interval for which no provision has been made.

As discussed earlier, most of the above requirements imply discipline, short setup time, high availability, and JIT-type operation in the feeder lines (conditions met here as a result of the activities carried out in the feeder lines<sup>3</sup>). Finally, to achieve a constant manufacturing interval without excessive reserve capacity, level loading is required. Taking advantage of the scheduling flexibility afforded by the weekly production buckets for final systems, the final subassemblies within a week can be sequenced to generate level loading. An algorithm designed to achieve this is described in Reference 5.

#### Feeder Line Scheduling

With the final assembly sequence determining the order in which the systems are to be assembled and tested (with some flexibility of reordering within a day), the demand for subassemblies is determined, at least in daily buckets. Effectively, this creates a master production schedule for the subassemblies in daily buckets.

The next issue is related to the operations of the storeroom and the feeder lines to meet the demands so generated. Equation (1) is a simple representation of a closed-loop look-ahead policy (like a closed-loop MRP) for scheduling the feeder lines. (In the specific case under consideration, the critical elements that permitted this approach of building to *actual* demand, rather than building to *anticipated* demand, were the short manufacturing intervals and their low variability.) As the discussion in the previous section indicates, this approach, properly implemented, represents a true just-in-time operation. On the other hand, the approach requires more information than a simple kanban operation. In our case, the use of kanban corresponds to the production starts of a subassembly to match its withdrawal from the finished

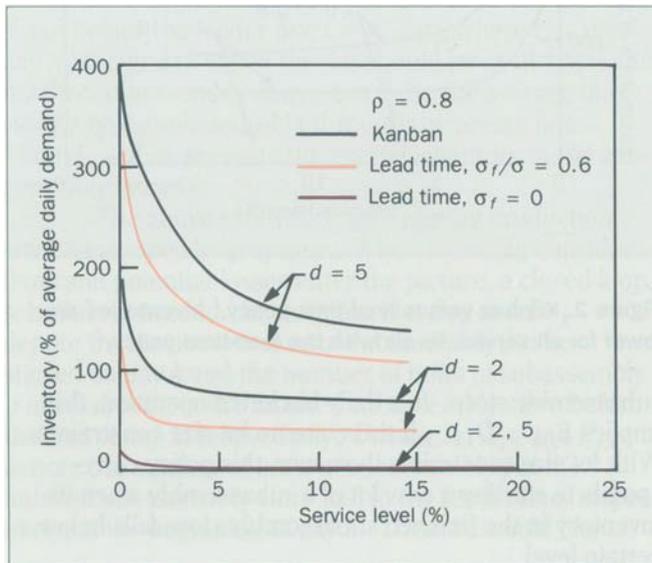


**Figure 2. Kanban versus lead-time policy. Inventories are lower for all service levels with the lead-time policy.**

subassembly store. In a daily bucketed operation, this implies  $S_{j,k} = D_{j,k-1}$  if there are no lot size constraints. With lot size constraints, however, this policy corresponds to starting a new lot of a subassembly when its inventory in the finished subassembly store falls below a certain level.

The choice between lead-time-based (look-ahead) and kanban-based production scheduling depends on many factors. Two studies at Bell Laboratories<sup>6,7</sup> shed some light on this question.

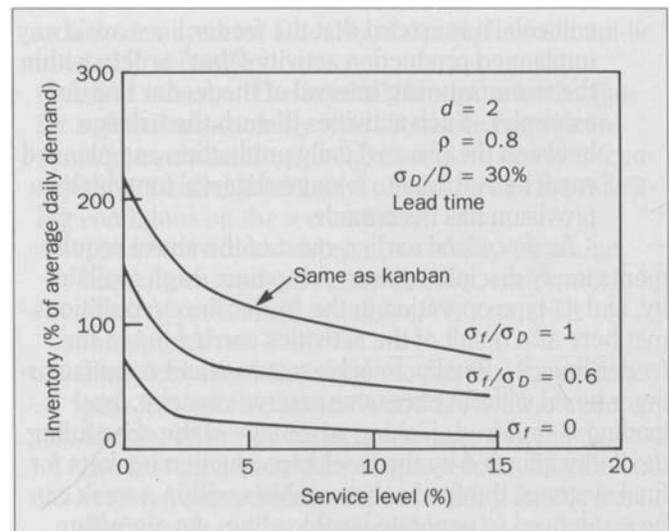
The study by H. Heffes<sup>6</sup> deals with one subassembly on one production line and considers the effects of demand variability, manufacturing interval and its variability, yield, and forecast errors (when the demands are not known with certainty). Both look-ahead and kanban policies require some finished subassembly inventory to achieve a desired service level. Figures 2 to 4 show some typical comparisons. The parameter  $d$  represents the average interval through the feeder line and  $\sigma_d^2$ , its variance.  $\bar{D}$  represents the mean demand per period and  $\sigma_D^2$ , its variability;  $\sigma_f^2$  represents the forecast error. In general, the demands during successive periods could be correlated;  $\rho$  represents the magnitude of this correlation (assumed nonnegative). In all cases, the inventory requirement under the look-ahead policy is smaller than under kanban policy. The difference strongly depends on the manufacturing environment. Overall, the inventory requirement under kanban-type policy is considerably larger than under the look-ahead policy when the



**Figure 3. Effects of interval length and forecast error. Generally, inventory decreases for lower values of interval  $d$ .**

interval is relatively long (large  $d$ ) but nearly constant (small  $\sigma_d^2$ ), and the demands in the near future are known with near certainty (small  $\sigma_f^2$ ) but are quite variable (relatively large  $\sigma_b^2$ ). The difference is even larger when the demands have high positive serial correlation. On the other hand, with small interval or small demand variability, the difference is very small. If the forecast error  $\sigma_f^2$  is large, then both kanban and look-ahead policies require large finished subassembly inventory with relatively small difference between them.

The study by A. Kumar<sup>7</sup> actually models an operation that purchases components from many independent vendors. In the manufacturing environment, this corresponds to the operations of many feeder lines. The production starts in different lines are synchronized by appropriate lead-time offsets [equation (1) applied individually for each subassembly type and each feeder line] to meet the final assembly needs, but otherwise the



**Figure 4. Effects of forecast error. The more certain the forecast, the lower the inventory.**

lines are operated independently. The results show that the inventory requirement of each subassembly increases as the number of feeder lines increases if the interval in each feeder line has significant variability. If the intervals in most feeder lines are nearly constant, then the inventory requirement for all subassembly types remains small. In the former case, the inventory requirement can be reduced if the operations of the feeder lines are further synchronized. That is, the sister lots required together in the subassembly store produced in possibly different feeder lines are tracked in their feeder lines and are brought into the subassembly store almost together by using a control and priority structure.

In the production environment for Definity products, the demands for various subassemblies in the equipment (non-circuit-pack) shops have very little temporal variability and the interval is very short—a few hours in most cases. It is clear that a simple kanban procedure between the subassembly store and the feeder

---

line input, and between the feeder lines and the storeroom, would be quite effective and simple to implement. Circuit-pack lines, on the other hand, contain many different codes. Some of these are high runners (over 200 circuit packs per week) while the others are low-volume types (under 10 circuit packs per day). Even after the demand smoothing provided by the final assembly sequencer, the demand for individual codes has significant variability. Minimum lot sizes (10 to 40 depending on the code) also add some demand variability. Moreover, circuit packs are relatively expensive to hold in the finished subassembly inventory, and they are more likely to be subjected to design changes that could make the excess inventory obsolete. These observations and the results of our analysis suggest that the circuit-pack lines for Definity products should use a look-ahead policy, with the finished subassembly store being the demand generator. Since all equipment shops can run with very short and almost constant interval, we do not have to worry about many feeder lines with significant interval variability. This allows a simple conceptual implementation of the look-ahead policy in the circuit-pack line without an elaborate information system to track and control individual lots in the line.

Effectively, what should enter the line on a given day can be determined by a look-ahead policy given by equation (1). A work-in-process (WIP) limit imposed on the circuit-pack lines<sup>3</sup> would guarantee small variability in interval within the line. The lots to be started on a given day can now be pulled into the shop as other lots leave the line and enter the finished subassembly store. If the demand smoothing is reasonably effective and if the line has sufficient capacity, then enough pulls should be generated during the day to move the daily planned starts into the line during that day. With commitment from the shop management to not disturb the order of production significantly once the lots are in the line, very little tracking is needed. Also, with the commitment from the store to not make much more than the daily planned starts available as completed picks, the demand

pulls could not result in premature production even if the line had extra capacity.

For the production environment under consideration, this provides an ideal marriage of closed-loop MRP and kanban pull concepts to achieve the goal of just-in-time production. The MRP system is used to plan purchasing and create a long-term production plan based on near-term hard orders and longer-term forecasts. The production in near-term weekly buckets is planned from the known hard orders and intervals. The flexibility available within a week is used to generate a smooth demand profile during that period. The resulting final assembly sequence becomes a local demand generator for subassemblies. MRP concepts are again applied to translate these demands into production starts in daily buckets for relatively long-interval feeder lines. All the feeder lines operate with strict WIP control and pull signals to take advantage of the simplicity and effectiveness of kanban concepts at the shop level. Components in the storeroom, completed picks in the storeroom waiting for pull signals from the lines, completed subassemblies waiting for the start of final assembly, and completed systems waiting for shipment are four "regulator" areas in the sense discussed by Hartland-Swann.<sup>2</sup> The movement between regulator areas is swift and reliable because of the shop discipline and exercise of JIT and total quality control (TQC) concepts throughout. This overall approach to production is what we call *integrated pull manufacturing* (IPM).

It should be noted that a different production environment could have produced a different ideal solution. For example, as shown in our analysis and by extensive experience in the Japanese automobile industry, with low-option products made in large volume and with nearly constant demand, an overall kanban operation could result in an ideal mix of simplicity and effectiveness. On the other hand, with inherently long intervals, greater fluctuation in demands, greater optionality in the subassembly mix for a given final system, and multiple feeder lines with relatively long intervals, the cost of

---

introducing additional regulator areas within the feeder lines and extending some of the MRP tracking concepts between the regulator areas could be justified. Fortunately, the Definity system production environment is not like a true job shop in that sense.

#### MRP-Pull Interface

An important requirement for the shop execution strategy to work effectively with daily (or smaller) buckets is that the components required to assemble picks for a given day be available in the storeroom prior to the beginning of the day. Guaranteeing component availability when needed is related to appropriate purchasing procedures and to the verification procedure, typically done at MRP level. MRP systems tend to operate in weekly buckets. Thus, there is a need to reconcile the weekly bucketed planning and daily bucketed execution from the component availability perspective.

98

A simple solution here is to add a feature called *bank account*. At the end of week  $k - 1$ , the planning system determines the pick needs for week  $k$  based on the subassembly needs in week  $k + 1$ . The bill of materials translates this into the component needs. The system then checks the availability electronically and assigns the components to the feeder lines. This electronic lock provides a "bank account" against which the shops can draw during week  $k$ . The actual sequence in which picks are performed in week  $k$  is then governed by the procedure discussed above. More details of the IPM operation can be found in Reference 8. (There it is stressed that a critical ingredient in reconciling the apparent conflict between MRP and pull is the elimination of the launching of shop orders by the master scheduler.)

The above procedure would identify at the planning stage any problems likely to arise during execution, when it is easier and less disruptive to provide alternatives.

#### Acknowledgments

The concepts presented in this paper are based on the work and brainstorming of many people. Particular mention should be made of the contributions by G. Brosowsky, G. Foo, T. R. Hsueh, S. A. Hendryx, and L. E. Kinney.

#### References

1. U. Karmarkar, "Getting Control of Just-in-Time," *Harvard Business Review*, No. 5, September-October 1989, pp. 122-131.
2. J. R. Hartland-Swann, "MRP or JIT: Which is Best?," *Proceedings of Second International Conference on JIT Manufacturing*, C. A. Voss (ed.), IFS Publications, Bedford, England, 1987, pp. 81-92.
3. R. E. Albano et al., "Manufacturing Execution: Circuit Packs," *AT&T Technical Journal*, Vol. 69, No. 4, July/August 1990, pp. 64-80.
4. B. Doshi, "Scheduling Strategies in Test/Repair Areas for Repairable Items," submitted to *International Journal on Production Research*.
5. H. Luss, M. B. Rosenwein, and E. T. Wahls, "Integration of Planning and Execution: Final-Assembly Sequencing," *AT&T Technical Journal*, Vol. 69, No. 4, July/August 1990, pp. 99-109.
6. H. Heffes, "Manufacturing Operating Process Characteristics," submitted to *Journal of Flexible Manufacturing*.
7. A. Kumar, "Analysis of Stockroom Inventory When Supplier Lead Times Are Uncertain," *Transactions of the Institute of Industrial Engineers*, June 1988.
8. L. E. Kinney, E. H. Denning, and G. Foo, "High-Option Manufacturing: The Denver Works Information Architecture," *AT&T Technical Journal*, Vol. 69, No. 4, July/August 1990, pp. 110-116.

Biographies (continued)

*problems resulting from electrostatic discharge in component design and in factories and telephone installations. He has a B.S. in engineering physics from McGill University and a Ph.D. in experimental physics from Cornell University. He joined AT&T in 1967.*

*(Manuscript received May 8, 1990)*

---