Trading Off Between Heijunka and Just-in-Sequence

Andreas Hüttmeir
Suzanne de Treville
Ann van Ackere
University of Lausanne

Johann Prenninger
BMW

1 Corresponding author: Professor Suzanne de Treville, HEC – University of Lausanne, 1015 Lausanne-Dorigny, Switzerland. Telephone +41 21 692 3448, Telefax +1 831 302 6641, e-mail suzanne.detreville@unil.ch.
Introduction

Lean production’s “Factory Physics” (de Treville & Antonakis, 2006; Hopp & Spearman, 1996) calls for high capacity utilization combined with relatively low inventories, requiring that system variability (concerning both arrival and service rates) be minimized. The key lean production practice that reduces variability in arrival of jobs to be processed is **heijunka**, in which production is scheduled such that the production line produces the same sequence of products throughout a given time period, with that sequence alternating between demanding and less demanding products.

As demand variability increases, however, the practice of heijunka becomes more challenging. An alternative approach, referred to as just-in-sequence (JIS), has begun to play an increasing role in lean production, even though production lines operating JIS will clearly be less lean than those operating according heijunka.

The choice between heijunka and JIS illustrates the tension between leanness and agility that increasingly must be sorted out if manufacturing is to be competitive (Narasimhan, Swink, & Kim, in press). In this paper, we use a case study of a BMW engine plant to gain insights into the nature of the trade-off and explore some ways to combine limited heijunka with partial JIS.

Heijunka

The objective of heijunka is to avoid peaks and valleys in the production schedule. Consider, for example, a workstation that produces two products, A and B, with A requiring 1.5 minutes, and B requiring 1 minute of processing, respectively. Suppose that the company receives an order for 100 units of both A and B. A naïve schedule would be to produce 100 units of one product and then 100 units of the other, resulting in a situation in which the demand faced by the workstation would vary considerably. Transport this workstation to a production line with a cycle time of 1.4 minutes, and the workstation is overloaded (and a
bottleneck for the entire line) for 100 cycles and underloaded for another 100 cycles. 

Accommodating this schedule requires increasing the cycle time for the entire line, at least during the period when A is being produced. It is not realistic to change the cycle time to accommodate such workload fluctuations on many production lines (this would not be possible at the BMW engine plant described in this paper, for example), hence this workstation might well be obliged to operate at a relatively low average capacity utilization.

Heijunka calls for distributing the jobs requiring more labor input throughout the production schedule to permit higher average utilization assuming that the cycle time is held constant over time. In our simple example, products A and B would be alternated, so that the workstation could either work in lots of one unit of A and one unit of B, with cycle time determined based on the combined work content of $1.5 + 1.0 = 2.5$ minutes, or allowing the workstation to get a bit behind during the cycle when A is produced, catching up during the cycle when B is produced.

Heijunka (also referred to as production smoothing or leveling the production schedule) has played an integral role in just-in-time (JIT) and lean production since its inception (Chase & Aquilano, 2004; Coleman & Vaghefi, 1994; Fujimoto, 1999; Hall, 1983; Hopp et al., 1996; Horn & Cook, 1997; e.g., Monden, 1983; Schonberger, 1982; Womack, Jones, & Roos, 1990). Teece, Pisano, and Shuen (1997) referred to heijunka as an example of a dynamic manufacturing capability or routine that might increase firm competitiveness when operating in configuration with other routines or capabilities.

Rinehart, Huxley, and Robertson (1997), however, described difficulties faced by one lean auto assembly plant in leveling the production schedule. Although heijunka was a stated company policy, variation in demand meant that it was constantly violated. Unfortunately, production was scheduled and capacity allocated under the assumption that heijunka was functioning correctly, so surges of "overburdened jobs" resulted in extreme worker stress,
repetitive strain injuries, and potential quality problems. In labor disputes that eventually emerged, overburdened jobs resulting from a failure to maintain heijunka were a major point of conflict.

McLachlin (1997: 287) listed "uniform plant loading" as a JIT flow element, operationalized as "(a) There is a fixed and level schedule; (b) they produced the same mix of end items or families each day, and possibly each hour (to match daily demand rates); (c) there is a reduction in upstream inventory swings and panic reactions to [changes in] schedule[d] demand; and (d) there is little or no expediting." In six JIT implementations evaluated using a Likert scale ranging from 1 to 6, with 3 as neutral, two rated their level of uniform plant loading at 4, two at 3, one at 2, and one at 1, indicating that for these plants leveling the production schedule was not a top priority.

Shah and Ward (2003) did not include heijunka or leveling the production schedule in their extensive list of lean manufacturing practices. Production smoothing was mentioned in the context of bottleneck removal rather than schedule leveling, and there was a general reference to planning and scheduling strategies, which could refer to either heijunka or JIS scheduling.

De Smet and Gelders (1998) noted that implementation of heijunka was only possible in situations where there were few schedule disturbances, that is, demand was relatively stable and predictable. Hines, Holwe, and Rich (2004) observed that heijunka reduces manufacturing flexibility (see also Cusumano, 1994). In cases where manufacturing flexibility must be added to leanness, heijunka may be replaced by a JIS production planning and control approach, which we describe in the following section.

**Just-in-sequence (JIS) production scheduling**

JIS is a production scheduling approach that calls for delivery—and some production—to be carried out in the order that parts will be needed by the customer. JIS
delivery requires either an inventory of parts that can be picked in the right sequence, or producing (or doing assembly of) parts in the order needed (e.g., Kempfer, 2005; Sayer, 1986). JIS delivery is now required by companies such as DaimlerChrysler (Feare, 2003), Saturn, and Dell (Trebilcock, 2006). Liker and Wu (2000) described the increasing importance of JIS production to effective manufacturing.

There is an increasing comprehension that making production as lean as possible is not always the best strategy. Narasimhan, Swink, and Kim (in press) compare lean and agile manufacturing practices, noting that although there is considerable overlap between these practice sets, there are also significant differences. Leanness can hinder agility and vice versa. Adding JIS to the lean production skill set that used to refer solely to heijunka is a recognition of the tension between leanness and agility.

One of the earliest success stories of JIS production and delivery is told in the Toyota Motor Manufacturing case (Mishina & Takeda, 1994), describing how the seat supplier developed the capability to produce and deliver seats in the order required. Toyota had originally suggested that the seat supplier produce to stock and deliver JIS (as was done for all other Toyota plants). The seat supplier argued, however, that the production and transportation lead times for the seat were short enough to permit seat production to begin after the car left painting (the point at which the final sequence was established due to variability in painting operations) and be delivered to the factory floor in time to be installed. Also, the number of seat variations was sufficiently high to require a quite large warehouse, and the number of variations was increasing over time. The seat supplier was located 30 minutes away from the Toyota plant, making JIS production easier. The case also describes, however, some of the problems arising with JIS when quality problems occurred, as JIS leaves much less latitude for rework. JIS also forced the seat supplier and Toyota to work together to fix the quality problem much more aggressively than would have occurred.
otherwise, so there was value in the disruptions. In other words, just as removal of inventory
has been demonstrated to lead to learning under the right conditions (de Treville, 1987), so
JIS production can also create powerful motivation for improvement.

Other authors have described use of a postponement strategy to permit JIS delivery
and assembly from a decoupling inventory, allowing an emphasis on high utilization
upstream, and responsiveness downstream (Feitzinger & Lee, 1997; Pagh & Cooper, 1998;
van Hoek, 2001).

**BMW**

Along the lines presented in the previous two sections, BMW faced a trade-off
between heijunka and JIS production in one of their European engine factories in 2003.
Historically the engine plant had been organized around the concept of heijunka, permitting
extremely stable production and a high utilization (ranging between 95 and 99%). The
smoothing of the production schedule was carried out by a module of the SAP/R3 system.
Engines were delivered to the customer approximately in the order in which they were
produced.

In the fall of 2003, three major assembly plants in Germany began to request JIS
delivery as part of a move toward offering increasingly customized cars to consumers, with a
goal that each customer should receive the car within ten days of signing the contract.
Responding to this request required a) a warehouse to permit engines to be resorted prior to
shipment, b) a decoupling point to allow standard engines to be produced to stock and then
finalized in the right sequence, or c) that the engine plant move to a production schedule that
matched the required sequence. Although the idea of JIS was attractive, there was great
concern that its implementation would result in an unacceptably low capacity utilization.
Furthermore, the long distance of the engine plant from customer plants and the complexity of
the product implied that JIS implementation would be relatively difficult. The question came
down to where in the process engines should be put into sequence, that is, at what point in the process did need for flexibility become greater than need for efficiency? The first author of this paper was brought in by the company to aid the engine plant in this analysis under the supervision of the second and third authors.

Figure 1: Production Using Heijunka
A first question concerned the extent of heijunka-induced sequence changes. Engine plant management discovered that no data was available on how many changes were carried out per time period, and what kind of capacity buffer would be required to permit production without such changes. The first step was to collect production data to determine how many engines changed place, and by how many positions. We discovered that well over half of the engines changed places, but the vast majority remained within 60 places of their original position, and only a very small percentage were delayed by more than 200 places, with larger shifts in position due to rework, and the highest values due to extra testing for special engines.

Let us first consider a move to pure JIS. Increasing the cycle time enough to permit all engines (irrespective of labor content) to be produced in the sequence requested by the customer would probably require a capacity buffer on the order of 10% for the majority of workstations given the fact that the line is fairly balanced and the more difficult engines have
about 10% more work content. This 10% needs to be compared to the cost of a final goods
warehouse and the labor required to sort engines manually before shipment. Such a buffer
would also allow the production line to be more flexible in preparing for vacation shutdowns:
Given the high capacity utilization, the plant had to begin preparing for vacation shutdowns
almost a year in advance. Overall, however, a capacity buffer of this magnitude was unlikely
to be accepted by company management, especially as capacity utilization is one of the most
important performance measures.

To produce engines following pure heijunka, however, is no longer compatible with
being fully responsive to customers. Furthermore, the costs of building responsiveness at the
end of an inflexible production line (i.e., the costs of rigidity) are high as well. Therefore,
rather than making a naïve choice between pure JIS production and pure heijunka, it is worth
exploring a finer trade-off. For example, if the parameters of the SAP/R3 system could be set
to be somewhat less drastic in seeking to smooth the production schedule, perhaps a smaller
capacity buffer (say, on the order of 2-3%) would suffice. Or, it might be possible to agree a
general heijunka-based schedule with customers such that the highest content jobs would
already be distributed throughout the schedule in the sequence requested by the customer,
allowing 20% of the heijunka to reap 80% of the leanness. This might be reasonable given
that large engines tend to go into large cars.

We observed that the vast majority of engines went out of sequence by 60 places or
less. Assuming a cycle time of very approximately one minute (common in the automotive
industry), this would imply that a buffer of about an hour's production at the end of the line
would give space to resort almost all engines without need for a warehouse. Such a buffer
could be designed with racks so that each engine could be put into its correct place upon
leaving the production line. Including the remaining 25% of engines that go out of sequence
by 60 to over 200 places would require a much larger space (holding 3 to 8 hours of
production). Therefore, this solution becomes increasingly attractive as limited heijunka is used to eliminate the position shifts that cause the most disruption. If we consider the choice between heijunka and JIS, it becomes clear that the choice is not either/or, but can be considered choosing a position on a continuum (Figure 3) with decreasing returns as we move to extreme points (Figure 4).

Orders are produced so that machines and workforce are optimally utilized and so that delivery date/time can be meet

Orders are produced according to delivery sequence and delivery date/time

Utilization optimized scheduling

Fully Sequenced Scheduling

Figure 3: A Continuum of Heijunka and JIS

Cost

Total cost

Resorting and rigidity costs

Production cost

JIS

Heijunka

Figure 4: Trading off between JIS and Heijunka

It is important to note that consideration of JIS production brought new attention to rework. Although the primary motivation for heijunka was capacity utilization, pure JIS
production would require a substantial reduction in rework on a line that lean production has already brought to a high level of performance. In other words, JIS (representing a combination of leanness and agility) may have the ability to motivate process improvement to a degree that exceeds lean production alone.

**Summary and Conclusions**

Is it better for a manufacturing plant to use heijunka to maximize its leanness, or to use JIS to maximize its responsiveness? Our work at BMW indicates that the answer may lie somewhere in the middle, with heijunka used to smooth out the most extreme production values with the remainder of production carried out JIS. It is important to understand this trade-off, as it gives essential insights into the bigger picture of trading off between leanness and agility. We have blindly accepted lean production practices—and leanness itself—as good for competitiveness, but it is now time to reach a finer understanding of these dynamics. Over time, we may find that the essential dynamic capabilities for competitive operations will not come from configurations of routines that drive leanness, but from configurations of routines that permit essential flexibility in spite of leanness. Also, it may be that the combination of leanness and agility will result in motivation to improve processes that goes even beyond the improvement motivation of lean.
References


