

Sizing of Heijunka-controlled Production Systems with Unreliable Production Processes

Christian R. Lippolt, Kai Furmans
Institut für Fördertechnik und Logistiksysteme (IFL),
Universität Karlsruhe (TH)
Kaiserstraße 12, D 76128 Karlsruhe, Germany
WWW home page: <http://www.ifl.uni-karlsruhe.de>

Heijunka is the notion to level a production system by removing ups and downs in volume caused by batch processing and customer order fluctuation in order to reach a mixed model production system with a constant flow of parts. We show how to implement lean production principles in systems with unreliable production processes. Process unreliabilities occur because tool machines may have small overall equipment effectiveness. Our present results were derived during performed implementation projects, where supermarket-pull-systems had to be dimensioned. In particular, the calculation of required inventory levels is presented which uses analytical mathematical models on the basis of discrete time queuing systems. By considering variable capacities we essentially extend the content of reference [1]. The application of our model is demonstrated by an example.

1 Introduction

Production unreliability is a frequent problem when lean production systems have to be installed. The problems occur because unreliability leads to variability in processing times and consumption, which must be compensated by inventory buffers. Production orders for the relevant process, which would guarantee appropriate refill of the process exit buffer supplying goods for the next process step are behind schedule and the next step is starved. This effect is more noticeable with increasing process utilization.

In cases where all processes are highly reliable with low variability, the standard Toyota formulas can be used for sizing the buffers. Otherwise the relevant literature suggests concentrating first on the process itself (point kaizen) prior to implementing kanban cycles. Unreliable processes are causing constraints. Special tool machines often have an overall equipment effectiveness (OEE) lower than 70%. OEE

Please use the following format when citing this chapter:

Lippolt, C.R. and Furmans, K., 2008, in IFIP International Federation for Information Processing, Volume 257, Lean Business Systems and Beyond, Tomasz Koch, ed.; (Boston: Springer), pp. 11–19.

comprehensively indicates the relative productivity of a piece of equipment compared to its theoretical performance. Low reliability may be caused by technological complexity, which is not under control. Therefore it is the question whether lean production can only be implemented and the production be levelled if all production processes are highly reliable or if it is possible to begin with levelling before making improvements. [2] dealt with the problem of determining the number of circulating kanbans for a manufacturing system with machine breakdowns utilizing perturbation analysis. In order to find the optimal number he estimated gradients and performed stochastic approximations.

We approach the problem by developing a method for the calculation of buffer sizes. A path towards an implementation of kanban cycles and levelled production for cases where the production processes are unreliable is shown by calculating the required inventory levels which are necessary to guarantee a certain service level. The following sections provide insights into the impacts of variability in production systems and into production levelling and present the derived model and its assumption as well as the analytical method to compute the required inventory levels. We finish the text with an illustrative example.

2 Basic concepts

2.1 The effect of variability on Lean Manufacturing

Lean Manufacturing is a management philosophy focusing on reduction of the seven types of waste (over-production, waiting time, transportation, processing, inventory, motion and scrap) in manufacturing or any type of business [3]. [4] note that it is possible to achieve the same throughput either with long cycle times and large work or short cycle times and small work in process. The difference between both cases is variability. It exists in all production systems and affects significantly production throughput, delivery, quantity, costs and customer satisfaction [5]. The most common causes of variability in manufacturing environments according [4] are: natural variability (including minor fluctuations in process time due to differences in operators, machines, and material), random outages, setups, operator availability, rework and scrap. [6] distinguish between process and flow variability. After [4] both kinds of variability in production systems will be buffered by some combination of inventory, capacity and time and we can conclude that variability is the root cause for waste. Because in a perfect balanced system no buffering is needed, increasing variability always degrades the performance of production systems [4].

2.2 Production levelling strategy

The basis of a lean or just-in-time production is to level the work flow for optimizing the manufacturing line [7]. Customer orders may arrive relatively constantly in the long run but they appear to be inconstant and unpredictable in short intervals. The aim of levelling is the reduction of the customer order variability by analysis of the orders in a given time span resulting in a pattern which fits into

a smaller time scheme. It creates a constant flow of parts in a mixed model production and reduces or eliminates the need for spare capacity or stocks to cope with peaks of demand. [8] distinguishes two phases of levelling: I. Of the total production volume, II. Of the product mix. Phase I specifies both kind and quantity of product variants (model types) which are to be produced in an individual manufacturing shift. The computation starts with the monthly order of each product variant. The quantity of a given levelling horizon per variant is divided by the number of available manufacturing shifts. Thus the levelled outputs per shift of each variant and the cycle times for all products are found. The aim of the continuous improvement is to produce every-part every-day or better every-shift [9]. Because in today's production systems the number of variants can be very high, [10] suggests to level in a first step only the large-volume variants (high runners), which are to be produced following the average demand in a given time interval (every-part-every-interval, EPEI). The length of the EPEI is an indicator for the capability of the production process. Its determination is described in detail in [9].

The many small-volume products (low runners) are scheduled upon need in a reserved period of the day. If a production order for a low runner cannot be fulfilled completely within a reserved time block, then it is scheduled into the next cycle [11]. The reduction of flow variability by smoothing the orders leads to a decoupling of production and demand which must be compensated by an inventory buffer for finished goods. A short levelling horizon facilitates the implementation and minimizes the required inventory buffer for finished goods, however the desired levelling effect is lost, if the period is chosen too short. The optimal level depends on the fluctuation of the customer orders and the price of finished goods.

In phase II the production sequence of the individual orders per shift is determined, which leads to a finer levelling of the product mix. Numerous relevant algorithms exist in the literature [e.g. 8, 12]. The final result of levelling is a production sequence with a continuous flow adjusted to the customer demand and leading to an even utilization of the production stages.

3 Buffer sizing model for unreliable systems

3.1 System Model and Control Policy

A manufacturing structure, where each process has only one predecessor and one successor [12] is the one with the lowest possible complexity. The complexity increases with the number of the possible preceding and following processes. Evidently an even utilization of all resources is much more difficult in a system with high material flow complexity. To improve transparency and to reduce system variability, lean manufacturing tools introduce and steadily improve flow production. The flow production method arranges the production processes complying with the material flow of the products. The parts flow after each production process to the subsequent process without buffering and are processed immediately. There is no over production, because each stage produces exactly what the next stage needs at a given time. Abandoning of buffers results in short throughput times and low work-in-process inventory levels. To realise pure flow production, the production times

must be balanced perfectly. This is often only a vision but nevertheless the aim of the continuous improvement process. A useful tool to improve flow is value-stream design initiated by [8]. One important design principle is to arrange directly connected production processes in a flow and manage the flow between the sections by continuous flow with pull control loops [8].

Each stage consists of a manufacturing process, which is a sub-part of the production system, and an output buffer (supermarket). The manufacturing process contains parts that are currently being processed in the stage (either waiting for, or receiving service at the different machines), referred to as the work-in-process (WIP) of the stage. The output buffer contains the finished parts of the stage, referred to as the finished good inventory of the stage.

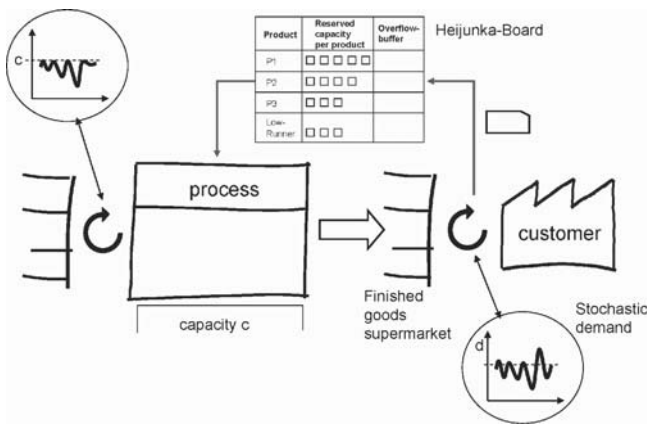


Fig. 1. Scheme of a Kanban cycle: Single stage Heijunka-controlled production

The coordination between the stages is achieved by a pull mechanism. Each material withdrawal generates a production order for the previous process with the goal of filling up the gap as soon as possible. As information medium between consumers and the previous production process in practice a kanban is used.

There is a flow of parts moving from upstream to downstream stages, and a flow of demands is going from downstream to upstream stages. Fig. 1 illustrates a single stage production unit as a segment of the whole process chain. The customers request parts in each period n according to a stochastic density function d . The kanbans attached to the requested products move to the Heijunka Board and are loaded into the corresponding hole to create a production order. For each product a certain capacity of the overall capacity is reserved according to the average demand. If the amount of kanbans for a given product is less than the reserved capacity, the kanban is put in the next empty hole for the product. Otherwise, the kanban is put in the overflow box. The overflow box contains backorders that will be produced as soon as the demand during one period is lower than the capacity. The peaks of the stochastic demand are thus cut and the production is levelled. If the production process with capacity c is unreliable, a stochastic maximum production

output C_n of the stage exists. Because of machine breakdowns and other unexpected difficulties the real output may deviate from the planned output. For modelling the unreliability of the relevant process, its temporal losses of capacity are evaluated retrospectively and the probable daily distribution of working capacity c_i is calculated and is assumed to be identically each day. On the basis of this distribution the process exit buffer (supermarket) is filled with parts in a way that the following process is always supplied.

Reserved capacity which can not be filled remains empty in order to avoid producing of goods, which are not requested. The amount of consumed parts in one period is equivalent to the stochastic demand of the stage given in form of a kanban to the pervious stage.

3.2 Mathematical model formulation

A methodology for sizing the buffers (supermarkets) by using the G|G|1-queueing system on discrete time is developed. The number of items in the supermarket (buffer) results from the superposition of two stochastic processes: Filling up by the (limited) production with capacity C_n and emptying by selling the quantity D_n . The inventory level X_n changes in each interval with:

$$X_n = Q_n - C_n.$$

W_n , the number of kanbans (cards) on the Heijunka board is given by:

$$W_{n+1} = \max\{ (W_n + X_n); 0 \} \quad (1)$$

The number of items in the supermarket-buffer is the difference between the number of kanbans in the system and the last value of W_n . Equation (1) is equivalent to Lindley's equation for queueing systems in discrete time. The production capacity corresponds with the arrival vector, the demand rate with the service time vector. Methods of [15] can be used to determine the numbers of items in the supermarket and of kanbans on the Heijunka board. [15] determine waiting time and idle time distributions of a G|G|1-queue where the inter-arrival time A is distributed by a_i and the service time S is distributed by s_i . The inter-arrival time between two customer orders specifies the time period with decreasing workload, which corresponds to the production capacity in our model. The service time of an arriving customer order increases the workload, which is equivalent to the sold quantity, because this quantity must be reproduced.

The algorithms of [15], which are based on a Wiener-Hopf factorization using ladder height distributions, can be used to solve equation (1). Therefore, the same methods which are used to compute the distribution of the waiting time, can be used to compute the distribution of the number of waiting orders at the end of a period (allocation of the overflow box), which is equivalent to the number of kanbans on the Heijunka board. This results in a vector \vec{w} , where w_i is the probability that the number of waiting orders in the overflow box is exactly i . Additionally, the idle capacity distribution can be computed by the algorithm of [15], because it is equivalent to the idle time distribution.

We assume that the replenishment time is one period (EPEI = 1). Then the probability distribution \vec{h} of the number of items not yet replaced in the finished goods stock, which is equivalent to the allocation of the Heijunka board at the end of a period, is given by

$$\vec{h} = \vec{d} \otimes \vec{w} \tag{2}$$

In order to reach the desired alpha service level SL_α , the base stock of finished goods and hence the number of required kanbans at the considered production stage k is the absolute value of the smallest integer k , where the relation

$$SL_\alpha \geq \sum_{i=-\infty}^k h_i \tag{3}$$

is satisfied. The number of waiting production orders at the end of a period is equivalent to the amount of kanbans in the overflow box. For each kanban in the overflow box, there must be one part in the finished goods supermarket to fulfil the customer order. The allocation probability of the supermarket s_j depends on the constant maximal overall inventory I (WIP cap), which is the sum of items in the supermarket and amount of kanbans on the Heijunka board. Thus the allocation of the supermarket at the end of a period is

$$s_{I-j} = h_j \quad 0 \leq j \leq I \tag{4}$$

If the EPEI is longer than one period, we suggest to add for each additional period the amount of one maximal request per period extra in the supermarket to guarantee the desired service level.

Further, we conclude that the capacity will be completely used, if the number of items not yet replaced in the finished goods stock, given by h , is at least the available capacity; Otherwise, exactly the capacity will be used to fill up the items in the finished goods supermarket. The production vector is denoted by q and represents the amount of required material in one period. Therefore q is the demand vector to the preceding production level. We obtain

$$q_i = c_i \cdot \sum_{j=i}^{\infty} h_j + h_i \cdot \sum_{j=i}^{c^{\max}} c_j \quad \text{for } c^{\min} \leq i \leq c^{\max} \tag{5}$$

To explain the usage of the modelled production system and the G|G|1-queue in discrete time we give an example in the next section.

3.3 Explicatory application

The required amount of kanbans in the kanban cycle to ensure a given service level has to be calculated. At first, the data for the calculation has to be provided. The demand function can be determined easily by taking the customer orders or forecasted demands. The effective production capacity is the difference between total capacity and the sum of all breakdown times of one shift or day. Often only the mean length of a downtime of a machine (MTTR) and the mean time between two successive down times (MTBF) are used to describe the availability A of a machine ($A = MTBF/(MTTR + MTBF)$). But the performance of a production system is

heavily depending on the higher moments of the distribution. It is Important, whether the breakdowns are frequent and short-lived, or seldom and long-lasting. First of all, the cause of the output-gap has to be determined and classified as e.g. breakdowns or changeover. The effective production capacity needed to get a discrete density function can be determined as an excerpt from the shift-book and displayed as illustrated in Fig. 2.

With the algorithms from [15] the allocation of the Heijunka Board, the overflow-box and the demand function of the process are calculated. Table 1 illustrates the performance of the calculation from data input to result. The result is the service level with a given amount of kanbans, which is taken from the record.

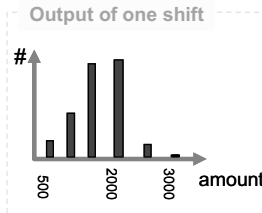


Fig. 2. Example of a histogram with shift-output

Table 1. Illustration of the computational steps of the model

Demand distribution		Production capacity		Number of allocation of kanbans				Effective Production	
				Overflow-Box	Heijunka-Board	Servicelevel	SL		
0	0%	0	0,0%	0	37,8%	0,0%	0,0%	0	0,0%
1	17%	1	0,0%	1	13,2%	6,5%	6,5%	1	6,5%
2	17%	2	0,0%	2	12,2%	8,4%	14,9%	2	8,4%
3	17%	3	30,0%	3	9,6%	10,5%	25,5%	3	32,7%
4	17%	4	40,0%	4	6,8%	12,1%	37,5%	4	33,8%
5	17%	5	30,0%	5	5,2%	13,0%	50,6%	5	19,3%
6	17%	6	0,0%	6	3,9%	14,2%	64,8%	6	0,0%
7	0%	7	0,0%	7	2,9%	8,4%	73,2%	7	0,0%
8	0%	8	0,0%	8	2,2%	6,7%	80,0%	8	0,0%
9	0%	9	0,0%	9	1,6%	5,1%	85,0%		
10	0%	10	0,0%	10	1,2%	3,7%	88,8%		
				11	0,9%	2,8%	91,6%		
				12	0,7%	2,1%	93,7%		
				13	0,5%	1,6%	95,3%		
				14	0,4%	1,2%	96,5%		
				15	0,3%	0,9%	97,3%		
				16	0,2%	0,7%	98,0%		
				17	0,2%	0,5%	98,5%		
				18	0,1%	0,4%	98,9%		
				19	0,1%	0,3%	99,1%		
				20	0,1%	0,2%	99,3%		

4 Summary and outlook

Our result is a new analytical calculating method for the required inventory levels in the buffers (supermarkets) of a manufacturing system and is applicable when unreliable machines are involved. For that purpose, the manufacturing system is treated as a multi-stage production system. The application of standard formulas for lean production systems is limited because they require reliable processes. The

benefit of the new approach is that the buffers can be sized without performing laborious and time consuming simulations, which were used so far.

Uncertainty of the results may be due to inevitable limitations set by assumptions: The availability of raw material has to be assumed to be 100%, the distribution of daily demand has to be independent and identical, and the demand pattern cannot be modelled. Hopefully further research will relieve these restrictions.

References

1. Furmans, K. (2005): Models of Heijunka-levelled Kanban-Systems. In: C.T. Papadopoulos (Ed.): 5th International Conference on Analysis of Manufacturing Systems - Production and Management 2005, Zakynthos Island, Greece, pp. 243-248. Publishing Ziti, Thessaloniki, Greece, <http://www.icsd.aegean.gr/aic2005/Papers/Furmans.pdf>.
2. Yan, H. (1995): The optimal number of kanbans in a manufacturing system with general machine breakdowns and stochastic demands, *Internat. Journal of Operations & Production Management*, Vol. 15, Nr. 9, pp. 89-103.
3. Ohno, T. (1988): *The Toyota Production System: beyond large scale production*, Productivity Press, Portland, OR, Cambridge, Mass.
4. Hopp, W. J., Spearman, M. L.(1996): *Factory Physics, Foundation of Manufacturing Management*. Irwin/McGraw-Hill, Boston, Mass., 2nd edition 2000, 668 p.
5. Adams, M., Schroer, B. J., Gunter, L. (2001): Impact of process variability on lean manufacturing systems. *Proceedings of the Huntsville Simulation Conference, 2001*, Huntsville, AL.
6. Koch, T., Kornicki, L. (2003): Minimizing variability in manufacturing systems in the context of lean manufacturing implementation. *Modern trends in manufacturing. 2nd Internat. CAMT Conference*. Wrocław, Poland.
7. Liker, J. K. (2004): *The Toyota way: 14 management principles from the world's greatest manufacturer*. McGraw-Hill, New York. 350 p.
8. Monden. Y. (1983): *Toyota Production System, practical approach to production management*. Industrial Engineering and Managment Press, Institute of Industrial Engineers, Norscross, GA. USA.
9. Rother, M., Shook, J. (1999): *Learning to see: Value-Stream Mapping to Create Value and Eliminate Muda*, Lean Enterprise Institute (LEI), Brookline, Mass., 112 p.; ISBN: 0966784308.
10. Takeda, H. (1996): *Das System der Mixed Production, flexibel, rationell, kundenorientiert*. Verlag Moderne Industrie, Landsberg, Deutschland.

11. Smalley, A. (2004): Creating level pull: a lean production-system improvement guide for production-control, operations, and engineering professionals. Lean-Management-Institut (LMI), Aachen, Deutschland. ISBN: 0-9743225-0-4 (englisch); 0-9763152-4-6 (deutsch). c..110 p.
12. Miltenburg, J., Sinnamon, G. (1989): Scheduling mixed-model multi-level just-in-time production systems. Internat. Journal of Production Research 27, pp.1487-1509.
13. Lödning, H. (2004): Verfahren der Fertigungssteuerung: Grundlagen, Beschreibung, Konfiguration. Buch der VDI-Reihe, Springer, Heidelberg, Berlin. 540 p.
14. Günther, H.-O., Tempelmeier, H. (2005): Produktion und Logistik (6. Auflage; Erstauflage 1995), Springer-Lehrbuch. Springer, Heidelberg.
15. Grassmann, J. L., Jain, W. K. (1989): Numerical solutions of the waiting time distribution and idle time distribution of the arithmetic GI|G|1 queue, Operations Research 37 (1989), No 1, pp. 141-150.