Applicability aspects of workload control in job shop production

Henrich, P.

Chapter 4 Grouping machines for effective workload control

Accepted for publication in the International Journal of Production Economics, with Martin Land and Gerard Gaalman as co-authors.

Abstract

Workload control (WLC) allows the release of new orders to the shop floor as long as workload norms for capacity groups, generally a number of functionally similar machines, are not exceeded. Effective WLC requires a profound decision on the grouping of machines as well as on the norm levels for the respective capacity groups. Also a routeing decision has to be taken in case of several machines allowing to perform the same kind of operations. In practice, some intuitive rules are used to define the capacity groups for WLC and to make a routeing decision; the norm levels are often determined by a trial and error approach. This paper aims at providing a theoretical starting point for these decisions.

Queuing theory provides some insights in possible pooling synergies and losses when grouping machines, but only with rather restrictive assumptions. Up to now, little attention has been paid in the literature to machine grouping within WLC. Additionally, the question of how to relate norm levels to the composition of the capacity groups and the appropriate routeing decision rules still remain unanswered.

Supported by a simulation study, this paper points out that pooling synergy insights can be translated to situations with controlled workloads. Absolute performance can be strongly improved by appropriately defining capacity groups in combination with suitable routeing decision rules. Besides, the choice of norm levels appears critical for both grouped and non-grouped parallel machines, and affects other control parameters as well.

4.1 Introduction

Workload control (WLC) is a production planning and control concept, especially developed for the requirements of small make-to-order job shops (Kingsman 2000). The term job shop is used to indicate a type of manufacturing situation where a large number of different products are produced according to a customer specification with highly variable routeings and processing times.

The WLC concept is based on principles of input/output control. Input control relates to accepting orders and releasing them to the shop floor. Order release is a main control element within WLC. The release of new orders to the shop floor is allowed as long as workload norms for capacity groups are not exceeded. Each operation is related to a specific capacity group (Henrich et al. 2004B)[⊗]. Capacity groups are considered as the smallest unit to be controlled centrally. Once released the orders remain on the shop floor. Simple priority dispatching rules will direct the orders along their downstream operations.

As production planning and control is quite complex even in small make-to-order job shops, early research on WLC focuses on fundamental insights in the basic mechanisms of input/output control mainly supported by simulation studies. The majority of the simulation studies utilised hypothetical models to characterise a job shop (Wisner 1995). Pure job shop models show the most extreme type of routeing variety. The routeing sequences of the orders are completely random and the flows through the shop are undirected. The same operational characteristics (i.e. operation processing time, capacity) for all the machines, lead to a balanced shop floor (i.e. same average utilisation and throughput time per machine).

Later studies have investigated the functioning of controlled order release within shop floor models that differ from pure job shops. For instance Oosterman et al. (2000) consider an explicit flow structure. Park and Salegna (1995), Salegna and Park (1996), Enns and Prongué Costa (2002) look at specific bottlenecks within job shops. Bertrand and Van de Wakker (2002) include assembly operations within their simulation model; Sabuncuoglu and Karapinar (1999) include transportation times. Missbauer (1997) investigates the effect of sequence dependent set up times. Henrich et al. (2004C)[⊕] analyse shops consisting of sub-departments.

All these approaches in modelling job shops have one thing in common: the capacity groups itself are seen as given. Capacity groups are mostly considered to be a single machine, and sometimes they contain a group of several machines using one queue of waiting orders. Nevertheless, an explicit decision on the grouping of machines, i.e. the decision about which machines should be clustered in the same capacity group, has not been made (Perona and Miragliotta 2000). Only Nyhuis and Wiendahl (1999) give some rules on grouping machines, but the grouping decision itself is not addressed explicitly.

The decision on grouping machines into capacity groups, though not considered in previous research, is very relevant and important. Within companies, where the WLC concept is going to be implemented, grouping decisions have to be always made for machines with the same or at least similar process characteristics (e.g. sawing, drilling, milling, etc.).

[⊗] Chapter 2.
[⊕] Chapter 3.

The grouping decision, being of basic importance for the functioning of WLC, can influence the control task and shop floor performance in many ways. Grouping several similar machines into a single capacity group may lead to performance improvements (e.g. a decrease in throughput time – also known as 'pooling synergy') or a less complex control task (e.g. less norms have to be considered during release). On the other hand negative effects on overall performance might arise, especially if machines are clustered that are not completely identical and interchangeable. Then an unbalanced queue of orders might emerge at a capacity group. The possible benefits arising from routeing flexibility might pay off, for instance, by additional efforts in deriving alternative process plans.

Within this paper we investigate the impact of grouping machines on the effective use of workload control within make-to-order job shops. This paper aims to provide a starting point for a profound decision on grouping machines as well as on the control of the respective capacity groups. In Section 4.2, we distinguish several possibilities to group machines within a workload controlled environment. The implications of those possibilities for the relevant control elements within WLC are discussed systematically. To get insight into the effects of different grouping- and control-alternatives within WLC on overall shop performance a set of representative simulation experiments is defined in Section 4.3. The simulation study is based on a shop floor model that is confined to the basic shop floor elements, necessary for investigating above described implications. The outcomes of the simulation study are discussed in Section 4.4.

4.2 Grouping machines

In this section we discuss in detail the possible implications of grouping machines within a workload controlled environment. First we describe the function of capacity groups within WLC: Machine characteristics, workloads and norms are the main determinants of the capacity groups. Based on the different machine characteristics we discuss the most common grouping choices and show that the grouping decision cannot be seen independently from the resulting routeing alternatives, based on several 'similar' machines on the floor. We draw to a close by relating the prior discussion to the question of parameter setting, within WLC.

Capacity groups within workload control (WLC)

The concept of capacity groups within WLC is embedded into the control mechanism of releasing orders from the pool to the shop floor. Within this section we describe the WLC aspects related to capacity groups and their influence on order release. For a more complete and detailed description of WLC we refer to Kingsman (2000), Henrich et al. (2004B) and Land (2004B).

Before WLC allows orders to enter the shop floor they are collected in a so-called order pool. The decision to release an order from the pool to the shop floor is based on its influence on the momentary shop floor situation as defined by the workload per capacity group. Orders in the pool are allowed to be released as long as its release will not cause any workload norm to be exceeded. This controlled order release process guarantees a stable throughput time per capacity group. The capacity groups are considered to be the smallest units to control during release. The shop floor view can only be as detailed as defined by the capacity groups. Capacity groups consist of one or more machines or workplaces to perform operations. For each capacity group one workload and one workload norm is defined. They are normally expressed in time units. Workloads are calculated as an aggregate of individual processing times per capacity group. Most workload definitions count up the processing times of orders waiting in front of a capacity group (direct load) and those of orders upstream (indirect load) as shown in Figure 4.1.

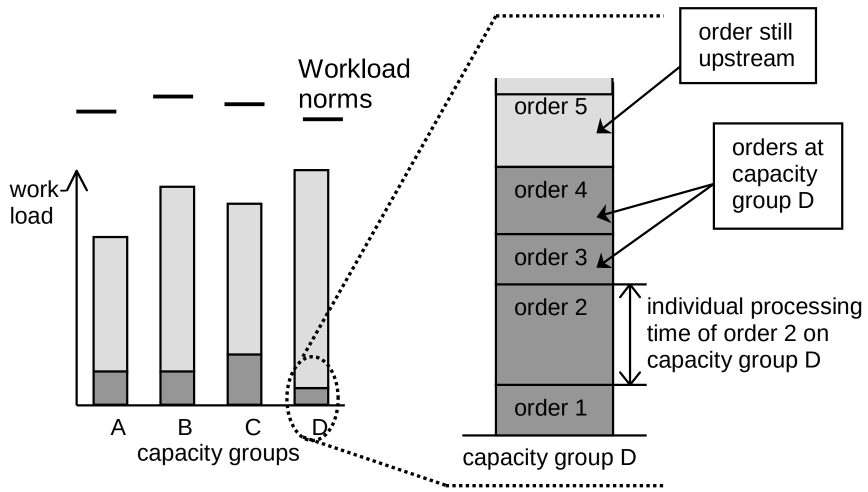


Figure 4.1 Calculating workloads per capacity group

Within our example in Figure 4.1 the momentary shop floor situation is defined by the workloads across the four capacity groups A to D. Per capacity group one workload norm is given.

Before we can investigate the effect of grouping machines on the effective use of WLC we will discuss the different machine characteristics, being the fundamental characteristics for common grouping decisions in practice.

Machine characteristics

Three types of machine characteristics are relevant while discussing grouping alternatives: process, functional and operational characteristics.

Machines can be distinguished by their different *process characteristics*. Process characteristics (as defined in DIN 8580[1]) are defined by the way the product is operated (e.g. reshaping, separation, changing material compositions, layering).

Even while showing the same process characteristics machines can differ within their *functional characteristics.* For instance, a laser and a saw show the same process characteristics (i.e. separation). But they have different functional characteristics: while the saw is used to cut large metal sheets with low complexity, the laser could be used to cut thin metal plates within small tolerances. By comparing different machines with the same process characteristics, three different alternatives relating to functional characteristics can be distinguished:

(a) Interchangeable machines: The machines are completely interchangeable. All orders on, e.g. machine A can be operated on machine B, and vice a versa.
(b) Semi-interchangeable machines: The machines are not completely interchangeable. Some orders only can be operated on a specific machine while others can be operated on, e.g. both machines A and B.
(c) Non-interchangeable machines: Non-interchangeable machines have such specific characteristics that no orders can be operated on, for instance, machine A and B.

Functionally interchangeable machines may differ by their *operational characteristics.* Operational characteristics can be defined by the operation processing times necessary to perform an operation. Those operational characteristics can be:

(a) Identical (if several machines need exactly the same operation processing time to perform the same operation on a specific order), or
(b) Different (here all kind of relationships might be considered, for a more complete description, we refer to Cheng and Sin (1990)).

[1] DIN = Deutsche Industrie Norm (German): German Industry Norm.

Common grouping choices

It is possible to cluster machines into capacity groups that might vary in all of the three above described aspects. All alternatives can be observed in practice. For instance, the whole shop floor could be seen as one capacity group as defined within the CONWIP concept (Spearman et al. 1990, Hopp and Spearman 2001). Another possibility is to cluster machines with different process characteristics but with the same relative position in the order routeing, for instance all machines that perform preparing or finishing operations. This grouping choice can be found in flow shops. Common practice within make-to-order shops is that machines with the same process characteristics are grouped together.

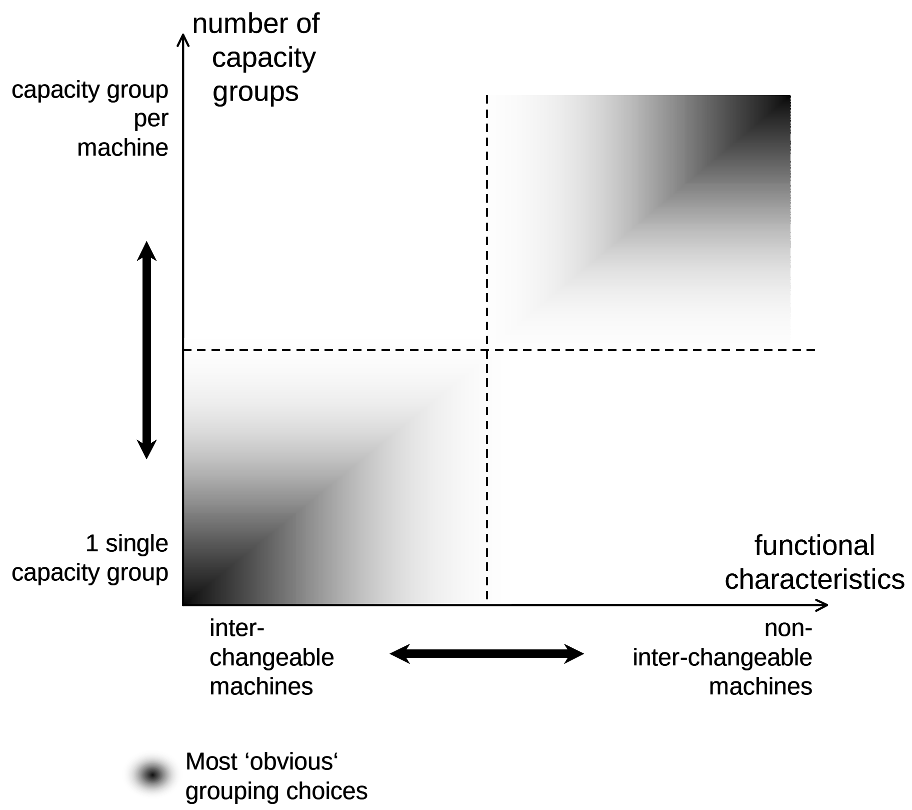


Figure 4.2 Most common grouping choices related to functional interchangeability

Machines with the same process characteristics can be functionally interchangeable or functionally non-interchangeable. While non-interchangeable machines normally are grouped into separate capacity groups, interchangeable machines mostly are clustered into one single capacity group. The grey-shadowed areas in Figure 4.2 show the most obvious grouping choices made within make-to-order shops. Nevertheless all other grouping choices are possible. Nyhuis and Wiendahl (1999), for instance, discuss the alternative to group non-interchangeable machines within a single capacity group to reduce, respectively, the number of norms and workloads that have to be considered during order release. In contrast, for interchangeable machines, it is also possible to define a separate capacity group for each machine. In practice, this can be found if the interchangeable machines, for instance, are placed on different locations on the floor (Tavana and Rappaport 1996). Even if some grouping choices are more common than others, little is known about their performance implications within a workload controlled environment.

Within functionally interchangeable machines there is a stronger tendency to group machines with similar operational characteristics than with different characteristics. In this paper, we only consider machines with the same process characteristics for grouping, as machine grouping with different process characteristics is quite uncommon in job shop practice. Therefore, we will focus on functional characteristics when considering grouping alternatives. We do not consider grouping based on operational characteristics in this paper, neither.

Implications of grouping for WLC

The machine grouping choices will be reflected by the shop floor control concept. Figure 4.3 shows the most important grouping implications on WLC. The figure is based on the example of two machines A and B that allow for a grouping decision. In that case we have just two grouping possibilities. More machines would successively lead to more grouping possibilities. The two machines can be grouped into one single capacity group. This implies that only one common workload norm and one workload is considered for both machines during order release. Segregating the machines into two capacity groups means that an individual workload (WL) and workload norm (WLN) have to be calculated for each machine.

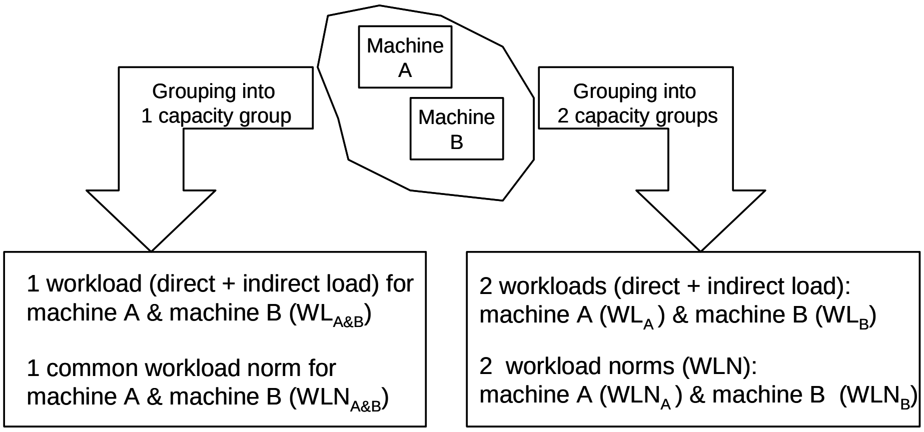
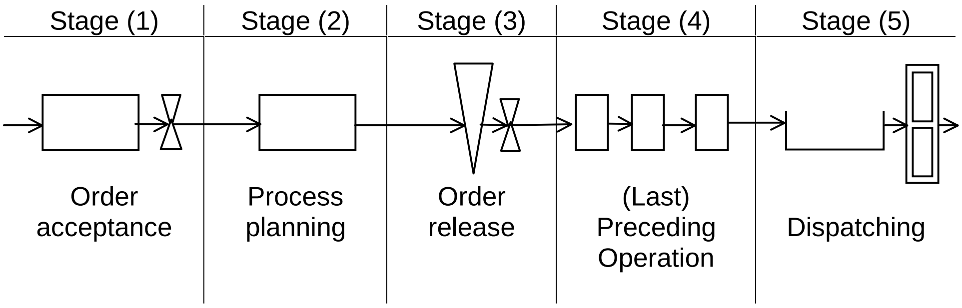


Figure 4.3 Grouping the machines A and B

Combining the 'Grouping decision' with a 'Routeing decision'

Two machines A and B that can perform the same (or at least similar) operations ask for a routeing decision. Taking a routeing decision means deciding on what machine an order has to be operated. This routeing decision may distribute the workload across the different machines. There are different possible stages within the order process to make a routeing decision (or better workload allocation decision) (see Table 4.1):

Table 4.1 Stages that allow a routeing decision



The first stage that allows for a routeing decision is the order acceptance stage (1) (see e.g. Corti et al. 2004). Kingsman et al. (1996), and Kingsman and Hendry (2002) show that controlling the workload during the order acceptance stage can be beneficial for overall performance. In practice the routeing decision often is made in stage 2: process planning. Here the process planner may choose always the machine with the shortest processing-, or set up time. During order release (stage 3) it is possible to combine the routeing decision with the decision of releasing orders to the shop floor. This may lead to a well-balanced workload across the capacity groups as two different mechanisms for workload allocations are combined. After the order release there are two possible stages on the floor left to decide on the order routeing. The routeing decision can be made at stage 4, before the order is going to be operated on one of the alternative machines. A decision during stage 4 makes sense, if for instance the alternative machines A and B are placed on different locations. A routeing decision has to be done at least on dispatching level in front of the machines (stage 5) before the operation starts. In fact this leads to a common queue for both alternative machines. The use of a common queue (pooling) may lead to a decrease in total throughput time, also known as pooling synergy (see e.g. Kleinrock 1975, 1976). To decide on the dispatching level it is necessary to group the machines A and B into the same capacity group.

The later the routeing decision is taken, the longer it is possible to react on unexpected events like disturbances, or machine breakdowns. On the other hand, postponing the routeing decisions leads to additional efforts on information processing. For instance, all routeing alternatives have to be processed and handled (i.e. alternative process plans, routeings, operation processing times, etc.) up to the moment the definitive routeing decision has been made.

The different positions within the order flow to decide on the order routeing only hold for interchangeable machines. The more obvious routeing decision for non-interchangeable machines has to be done during process planning (stage 2).

Parameter setting within WLC

The grouping and routeing decisions have consequences for the parameter setting in WLC. The workload norms (WLNs) that are closely related to the calculation of the workloads (WLs) and the chosen WLC approach are important parameters within WLC (Land 2004A):

The workload norms (WLNs) determine per capacity group (together with the actual workload (WL) per capacity group) the maximum amount of work to release from the order pool to the shop floor. In practice mostly intuitive guidelines are used to find WLNs, though the WLN level has an important impact on shop performance (Perona and Portioli 1998). Land and Gaalman (1996B) show that it is not trivial to find 'good' WLN levels analytically, even for capacity groups containing only single machines in a pure and balanced job. It will become even more difficult to find workload norms in situations where several machines are grouped into capacity groups.

With non-identical machines it may even become (technically) difficult to calculate the workloads (WLs). Within WLC it is only possible to release an order if its processing time fits into the WLN of the relevant capacity group. But it is not clear what operation processing time has to be considered, when the routeing decision itself and thus the determination of the operation processing times, is postponed to a moment after order release (stage 4 or 5).

The grouping and routeing choices may influence the WL in front of the machines in many ways.

It becomes difficult to predict the influence of the different choices according grouping, routeing and parameter setting on the WL in front of the machines. For instance, we group two machines A and B and take the routeing decision on the dispatching level (stage 5). This leads to a common queue in front of the two machines. Without controlled WLs it is possible to estimate the amount of waiting work in front of the machines and the average throughput time. From classical queuing literature (e.g. Kleinrock 1975, 1976) we for instance know that grouping machines generally performs better than defining a separate queue per machine. But little is known about the resulting workloads and throughput times in a situation within a workload controlled environment.

We have shown that the grouping decisions the routeing decision and the parameter setting interact and cannot be seen independently. Possible interactions and side effects are yet unknown. Up to know little is known about possible implications of those decisions on shop floor performance. In the next section we describe a discrete event simulation model to be a starting point in analysing the interactions in between the above-mentioned influences.

4.3 Simulation study

The previous section has shown that grouping multiple machines into capacity groups for the effective use of WLC is not trivial at all. To investigate machine grouping, in combination with the routeing decision and parameter setting within WLC we choose to set up a discrete event simulation. We start testing different combinations in making a grouping and routeing decision based on an elementary shop floor model.

The shop floor model

The simulation model is built in *e*M-Plant (Tecnomatix 2001). The model is derived from the shop floor model of a pure job shop as e.g. used in Melnyk et al.

(1991), and Land and Gaalman (1998). The original model is discussed first. Later the necessary changes are presented.

The inter-arrival time of orders at the pool follows a negative exponential distribution. The original model consists of a job shop with six machines, with one machine for each operation. Six capacity groups are defined to present the current shop status during order release. Each machine is a capacity group. The number of operations per job is uniformly distributed between 1 and 6, resulting in an average of 3.5 operations per job. The processing times at each machine follow a 2-Gamma distribution with a mean of 1 time unit. The set up times are supposed to be sequence independent and are modelled as a part of the operation processing times. Orders are processed at the capacity groups on a first-come first-served (FCFS) basis. The resulting utilisation is 90% for all machines. Only one order can be operated at each machine at the same time. The externally set due dates of the orders are known upon arrival.

Order release

For all orders that are waiting in the order pool the process plan is known. The process plan includes the routeing, the operation processing times and the externally set due dates. To start the order release the orders in the pool are ordered by a planned release date. The planned release date of an order is determined by subtracting a constant allowance (8 time units) per operation from the due date. Once every 5 time units all the orders in the pool are considered for release by comparing the WLs with the WLNs of the relevant capacity groups. Two different WLC approaches for calculating WLs are considered: (1) the ‘aggregate workload’ developed by Bertrand and Wortmann (1981), and (2) the ‘corrected aggregate load’ developed by Land and Gaalman (1996A). For a more detailed description of the order release procedure based on the two different WLC approaches we refer to Oosterman et al. (2000).

Adaptation of the ‘original shop floor model’ to allow for machine grouping

Within the original shop floor model described above, each of the six machines was considered to be a single capacity group. In our adapted model 2 machines replace one of those six machines to test the grouping of multiple machines into capacity groups. Now two alternative machines (A and B) exist for one of the operation types. In terms of WLC these two machines can be grouped into one or two capacity groups. The other five machines always form individual capacity groups. For the machines A and B we consider two alternatives: The machines are (1) completely *interchangeable* or (2) *non-interchangeable*. The processing times at the replaced machine follow a 2-Gamma distribution with a mean of 2 time units. This preserves the average machine utilisation of 90% for all machines. All other settings remain identical.

Routeing decision rules

In case of A and B being interchangeable, different alternatives arise in deciding on the definitive order routeing. As described in Section 4.2 there are various possibilities to decide on the order routeing (Table 4.1). Within our simulation we consider three different stages to decide on the order routeing: process planning, order release, and dispatching.

Table 4.2 Routeing decision rules (see also Table 4.1)

Stage	Routeing decision rules	
(2) Process planning	(1) Shop floor status independent	(A: 50%, B: 50%)-rule
	(2) Balancing numbers	(A/B/A/B)-rule
(3) Order release	(3) Balancing workloads	(LLGF)-rule
(5) Dispatching	(4) FCFS with a common queue	(First-Come First-Serve)-rule

At the process planning stage (2) we implemented two different routeing decision rules: The first rule represents the shop floor status independent rules. Before an order is sent to the pool a shop floor status independent decision is made in between machine A and B. On average half of the orders is randomly selected to be operated on machine A, half for machine B. This status independent rule at the process planning stage is indicated as (A: 50%, B: 50%)-rule (Table 4.2).

The second routeing decision rule also divides the arriving orders rather naïvely across the two different machines A and B during process planning: before an order enters the order pool, every second order is send to machine A, every other order to machine B. This leads, during a constant period of time (e.g. release period), to the same number of orders (+/-1) that enters the pool for machine A and for machine B. This routeing rule is indicated as (A/B/A/B)-rule.

Our third routeing decision rule makes the routeing decision during order release (stage 3). To release an order, the operation processing times have to fit under the workload norm levels of the relevant capacity groups. Orders that can be operated on machine A or machine B might better fit the WL for capacity group 1A (containing machine A) or respectively 1B. The difference in between WLN level and actual WL is determined for both alternative capacity groups 1A and 1B (see Figure 4.1). Actually the order is send to the machine with the largest difference – the largest load gap, if it obeys that norm. This routeing rule is indicated as (largest-load-gap-first)-rule. The (LLGF)-rule is only implemented under the assumptions of machine A and B not being grouped into a single capacity group, because it needs an exact workload overview per machine.

The fourth and last rule is used at the dispatching level (stage 5). By introducing a common queue (pooling) for both machines A and B the routeing decision can be done on a FCFS bases. In this case it is logical that the machines are grouped.

For machine A and B being non-interchangeable it is not possible to make a routeing decision. Two independent order streams for both machines arrive at the production system. Actually this results in the same outcomes as for the shop floor status independent (A: 50%, B: 50%)-rule for interchangeable machines.

Workload norms and norm ratio

To determine the best performing workload norm level (WLN) it is common practice in simulation studies to define it as an experimental variable. This variable is varied stepwise down from infinity. Within a real live job shop nearly every capacity group will show different characteristics (utilisation, operation processing time, stream of arriving orders, position within the order flow, average throughput time, etc.). Therefore for each capacity group a workload norm (WLN) has to be determined.

In our model we distinguish 6 capacity groups, but only one capacity group (split up in A and B) differs from the other five capacity groups on the floor. The remaining five capacity groups do not differ. Consequently we have to distinguish two different WLNs: one workload norm level ($WLN_{A/B}$) for capacity group 1 (or: 1A and 1B) containing machine A and/or B, and one 'general' workload norm level (WLN_{rest}) for the remaining 5 capacity groups.

This is modelled by using one 'general' norm level in combination with the norm ratio for the two different group types. The norm ratio (NR) is defined as follows:

$$NR = \frac{WLN_{A/B}}{WLN_{rest}}$$

NR Norm ratio
$WLN_{A/B}$ Workload norm at the capacity group(s) containing machine A and/or B
WLN_{rest} Workload norm at the five remaining capacity groups

Both, the general norm level and the NR are experimental variables.

To reduce the number of experiments it becomes necessary to think about a range of NR values, within which the best performing NR might be found. Reasonably this should be deducted from the capacity groups characteristics.

One approach is to relate the NR to the WLs at the different capacity groups under infinite workload norms ($WLN = \infty$). This can be calculated as follows:

$$(2)\ NR = \frac{WL_{A/B}}{WL_{rest}} \text{ with } WLN_{A/B} = WLN_{rest} = \infty$$

$WL_{A/B}$: Workload (=direct load) at capacity group(s) containing machine A and/or B
WL_{rest}: Workload (=direct load) at each of the five remaining capacity groups

Table 4.3 shows the different NRs. For a more detailed description of the calculations we refer to the appendix A. The calculations are based on the comparisons of GI/G/m and based on Land (2004B).

Table 4.3 Determination of start values for norm ratios (NRs)

<table>
<tr><th rowspan="2">Routeing decision rules</th><th colspan="2">Norm ratio (NR)</th></tr>
<tr><th>Grouping/
1 capacity group
(1 workload norm)</th><th>Non-grouping/
2 capacity groups
(2 workload norms)</th></tr>
<tr><td>FCFS-rule</td><td>2.28</td><td></td></tr>
<tr><td>(LLGF)-rule</td><td></td><td rowspan="3">2</td></tr>
<tr><td>(A/B/A/B)-rule</td><td rowspan="2">4</td></tr>
<tr><td>(A: 50%, B: 50%)-rule</td></tr>
</table>

Experimental design

Table 4.4 sums up the five different experimental variables within the simulation study as described above: (1) grouping decision, (2) routeing decision rule, (3) norm ratio, (4) workload norm, and (5) WLC approach. Additionally it shows the used settings of the variables.

Obviously it is not possible to combine all experimental variables independently. This holds especially for the grouping decision (1) and the routeing decision (2). Grouping the two machines A and B cannot be combined with a routeing decision at order release (stage 3). Non-grouping both machines will not be combined with a routeing decision on dispatching level (stage 5).

The functional characteristics, i.e. the machines A and B being interchangeable or non-interchangeable, are not included as additional experimental variable. As described above the results of the (A: 50%, B: 50%)-rule can also be interpreted as the results for non-interchangeable machines.

To test the influence of the WLC approach, two different approaches are used: ‘aggregate load’ and ‘corrected aggregate load’. To find the optimal norm ratio (NR) we test a broad range of values. For each of the resulting combinations different 'general' workload norm levels (WLN) have been tested, going stepwise down from infinity.

Table 4.4 Experimental Design

Experimental variables	Settings			
(1) Grouping decision	Grouping (1 capacity group)		Non-grouping (2 capacity groups)	
(2) Routeing decision rule	Shop floor independent/ (A:50%,B:50%)-rule	Balancing numbers/ (A/B/A/B)-rule	Balancing workload/ (LLGF)-rule	FCFS-rule with common queue
(3) Norm ratio (NR)	$NR \in \{x \mid 0.5 \le x \le 6, x = 50\}$			
(4) Workload norm (WLN)	Stepwise down from infinity			
(5) WLC approach	Corrected aggregate load		Aggregate load	

For each experiment 100 independent replications are performed. The replication length is 13000 time units, with observations from the first 3000 time units being deleted to avoid start up effects. Common random numbers are used as a variance reduction technique across all experiments.

4.4 Results and analysis

Before we can investigate the effect of grouping machines for effective workload control we have to understand the underlying 'behaviour' of the simulated environment. Therefore we start our analysis by comparing results at infinite workload norms ($WLN = \infty$), i.e. unrestricted periodic release. As described in the next section, these outcomes are independent from the chosen norm ratio (NR), the WLC approach and the grouping decision. They show some basic influences of the different routeing decision rules as presented in Table 4.2 and 4.4. The second step is to find the appropriate norm ratios (NRs). We introduce several performance indicators directly related to the different workload norm (WLN) levels. After the discussion on the parameter setting, we can start to investigate the upcoming effects by comparing grouping with non-grouping. We conclude this section by discussing the constraints of our experimental settings and conduct a sensitivity analysis.

Unrestricted periodic release

Discussing the simulation results under infinite workload norms ($WLN = \infty$), i.e. periodic unrestricted release, helps us to understand the routeing choice mechanism in more detail, as the differences in between grouping and non-grouping vanishes: Grouping the machines A and B means to calculate one WLN and one WL for both machines A and B. Non-grouping means to distinguish a single WLN and a single WL per machine (Figure 4.3). At periodic unrestricted release all the orders that are collected in the order pool are released immediately at the beginning of a release period without any restriction on exceeding WLNs. The differences in shop floor performance that still arise are independent from the chosen WLC approach, the norm ratio (NR) or the grouping choice. They only depend on the routeing decision rules, that means on the way the orders are allocated across the machines A and B.

Table 4.5 shows the system behaviour at periodic unrestricted release by presenting the different simulated station throughput times (STT) at capacity group 1 (containing machine A and/or B). The station throughput time is calculated as the sum of the average waiting and average operation processing time at machine A (or B). The shortest station throughput time (8.1 time units) can be realised with a routeing decision on dispatching level, with introducing a common queue in front of machine A and B. The longest station throughput time (15.1) is the result of the shop floor independent routeing decision ((A: 50%, B: 50%)-rule) during process planning. These two simulation outcomes are not unexpected and differ not much from the analytically made estimates of respectively 8.4 and 15.5 time units. The analytically derived approximations of station throughput times are based on Buzacott and Shanthikumar (1993). For further explanations we refer to the appendix A.

Table 4.5 Periodic unrestricted release

Position of routeing decision	Stage (2) process planning		Stage (3) order release	Stage (5) dispatching
Routeing decision rule	(A:50%,B:50%)-rule	(A/B/A/B)-rule	(LLGF)-rule	FCFS
$STT_{A/B}$ (sim.)	15.1	11.5	9.4	8.1
$STT_{A/B}$ (calc.)	15.5	15.5	15.5	8.4
STT_{rest} (sim.)	7.5	7.5	7.6	7.6
STT_{rest} (calc.)	7.8	7.8	7.8	7.8
Floor Time (simulated)	30.8	28.7	27.5	26.8
Common Idle Time	1.0	1.7	2.7	5.2

$STT_{rest:}$	Station throughput time at the remaining five capacity groups
$STT_{A/B}$:	Station throughput time at capacity group containing machine A and/or B
Common Idle Time:	Fraction of time [in %] that machine A and B are idle at the same time
Floor Time:	Average time of an order in between order release and order completion
sim.:	Simulated
calc.:	Calculated (derived analytically)

Unexpected were the station throughput times that result from the two decision rules: balancing numbers/(A/B/A/B)-rule and balancing workloads/(LLGF)-rule. The routeing mechanism appeared not to influence the coefficient of inter-arrival times at the machines A and B, so the same station throughput times (15.5) would be calculated. Nevertheless the differences in simulated station throughput time are quite large.

For the (A/B/A/B)-rule (balancing numbers) always the same number (+/-1) of A-orders than B-orders arrives in the pool during the release period. As for unrestricted periodic release always all the orders waiting in the pool are released, every 5 time units exactly (+/-1) the same number of A and of B orders is released. Using the (A: 50%, B: 50%)-rule leads to a less balanced set of orders that is sent to the machines.

The (LLGF)-rule (balancing workloads) even leads to a better balanced arrival of WL for the machines A and B, because a comparable amount of work is sent to both machines per release period. This leads to the relatively short station throughput time of 9.4 time units.

One reason for the shorter station throughput time is the better coordination of idle times. Actually every machine on the shop floor has a utilisation level of 90%. That means, that each machine is idle 10% of its time anyway. If idle times per machine overlap average throughput times must become shorter. This especially is the case with the routeing decision on the dispatching level, leading to a common queue in front of the both machines A and B. This is the best way to realise synchronised idle times (5.2%).

The different routeing decision rules only influence the station throughput time at capacity group 1 containing the machines A and/or B ($STT_{A/B}$). The other station throughput time at the other capacity groups (STT_{rest}) are not influenced. Obviously the differences in floor time (containing operations on all machines) based on the different routeing decision rules are less strong than in between the different $STT_{A/B}$ values.

Determining appropriate norm ratios (NRs)

Before the differences in between grouping and non-grouping can be discussed appropriate norm ratios (NRs) have to be found. This has to be done for each possible combination of the three experimental variables: grouping decision, routeing decision rule and WLC approach. Before we discuss the outcomes of best performing norm ratios as presented in Table 4.6 for all possible experiments we explain how we determine best performing norm ratios. Our explanation is based on a single example. Grouping the machines A and B is combined with the (A/B/A/B)-rule and ‘corrected aggregate loads’. We use corrected aggregate loads. The workload norms (WLNs) and the norm ratios (NRs) are varied within the ranges as described in Table 4.4. Figure 4.4 shows the results of these experiments. The representation method used is similar to that used in related research, such as the work of Land and Gaalman (1998), Sabuncuoglu and Karapinar (1999), Oosterman et al. (2000), Enns and Prongué Costa (2002), and Henrich et al. (2004C).

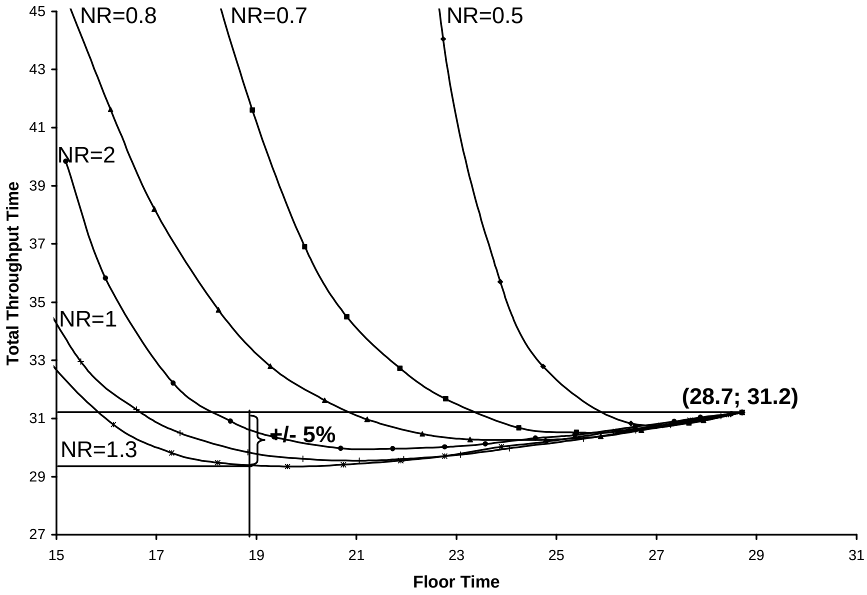


Figure 4.4 Finding an appropriate norm ratio

In Figure 4.4 the horizontal axis depicts the floor time. The floor time is the average time an order stays on the shop floor, it is the time between the order release and the order completion. It is used as an instrumental variable to indicate norm tightness. The vertical axis shows the total throughput time. To derive the total throughput time, the pool waiting time, i.e. the average time an order is waiting in the order pool, is added to the floor time. The total throughput time is used to indicate overall performance, though other measures such as tardiness have been recorded as well. For each norm ratio different workload norms (WLNs) starting from infinity have been tested. The resulting points that refer to the same norm ratio (NR) are connected by the curves, called performance curves. Obviously, all curves start at the point (28.7; 31.2) at 'infinite' norms. Here the total throughput time is 2.5 time units longer than the floor time, exactly halve the length of a release period. By tightening the norms the depicted points per performance curve move from right to left. Moving too far left, the norms are so tight that orders have to wait extremely long in the pool. This results in short floor times based on the low workload on the shop floor combined with long total throughput times due to long waiting times within the order pool. Each point on such a performance curve presents a realisable combination of floor time and total throughput time. Points that lay the most ‘left under’ combine short total throughput times (as an indicator for overall shop floor performance) with short floor time (as an indicator for a low workload level). In Figure 4.4 the curve based on the norm ratio NR=1.3 uniformly shows the best overall performance. Other simulated norm ratios lead to worse combinations of total throughput and floor time.

For the other experiments (as defined in Table 4.4) we also investigated the norm ratios. Table 4.6 shows the best performing norm ratios for corrected aggregate loads.

Table 4.6 Best performing norm ratios (NRs)

<table>
<tr><th rowspan="2">Routeing decision rules</th><th colspan="2">Norm ratio (NR)</th></tr>
<tr><th>Grouping/
1 capacity group
(1 workload norm)</th><th>Non-grouping/
2 capacity groups
(2 workload norms)</th></tr>
<tr><td>FCFS rule</td><td>2.2</td><td></td></tr>
<tr><td>(LLGF)-rule</td><td></td><td rowspan="3">1.3</td></tr>
<tr><td>(A/B/A/B)-rule</td><td rowspan="2">4 (or higher)</td></tr>
<tr><td>(A: 50%, B: 50%)-rule</td></tr>
</table>

The results show that the analytically derived workloads (Table 4.3) might be used to derive possible ranges for norm ratios. The simulated results for the grouping alternatives are close to the calculated outcomes with 2.2 for the FCFS-rule and 4 for the (A/B/A/B)-rule and the (A: 50%, B: 50%)-rule. The latter two routeing decision rules are quite insensitive against norm ratios higher than 4. The determined norm ratios for the non-grouping alternatives differ from the pre-calculated ones. But comparing the total throughput time as indicator for overall shop floor performance (Figure 4.4) the difference within a large part of the curves (comparing NR=2 with NR=1.3) are smaller than 5%. We see that it is possible by ‘fine-tuning’ the norm ratio to improve overall performance slightly. Thus similar to finding the right workload norm level (WLN) the question how to derive the best performing norm ratio (NR) analytically remains unanswered.

Nevertheless we see that too tight norm ratios can lead to unacceptable performance losses, while too wide norms show less dramatically results. In practice it might still be difficult to find the norm for ‘different’ capacity groups, but our simulation study shows that it is safer to keep the norms for those groups high.

Additionally we can conclude that the best performing NR is independent from the chosen WLN level, since the comparable performance curves for different ratios

(Figure 4.4) do not cross. This shows that it is possible to use fix NRs for defining the WLN levels at the different capacity groups.

Routeing mechanisms and the influence of WLC

The results in Table 4.5 were necessary to understand the basic influences of the routeing mechanisms on the floor time. Table 4.6 showed the best performing NRs. Within this subsection we analyse the effects of grouping (non-grouping) machines within a workload controlled environment. We only present the experimental settings as defined in Table 4.4 for corrected aggregate loads and the NRs as defined in Table 4.6.

Figure 4.5 shows all performance curves that are related to distinguishing two different capacity groups (non-grouping). We tested the (A: 50%, B: 50%)-rule and (A/B/A/B)-rule for a routeing decision at stage 2 (process planning) and the (LLGF)-rule for a routeing decision at stage 3 (order release). This results in three different curves. The starting points for unrestricted periodic release have already been discussed in Table 4.5. The curves run rather ‘parallel’ and do not cross each other. This means that the vertical distances in between the curves do not change. The shape of the curve depends on the WLC release mechanism and is nearly the same for all the three curves. This shows that the release mechanism works independently from the three different routeing decision rules. The different routeing decision rules just lead to a parallel shift of the performance curve towards the different starting points at periodic unrestricted release.

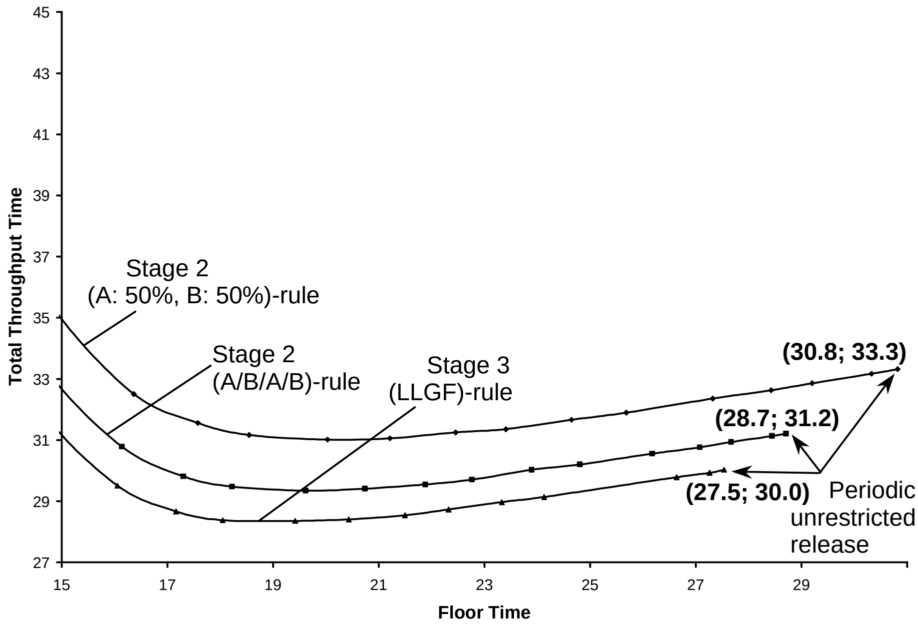


Figure 4.5 Non-grouping machines

Figure 4.6 shows all performance curves that are related to grouping the machines A and B into a single capacity group. We tested the (A: 50%, B: 50%)-rule and (A/B/A/B)-rule for a routeing decision at stage 2 (process planning) and the FCFS-rule for a routeing decision at stage 5 (dispatching). This results in three different curves. The solid lines are all performance curves that are related to grouping the machines A and B. The dashed lines show the performance curves as described in Figure 4.5. Under periodic unrestricted release ($WLN = \infty$) again the performance curves start at respectively (30.8; 33.3), (28.7; 31.2), and (26.8; 29.3) as discussed above. The two upper curves based on the (A: 50%, B: 50%)-rule and (A/B/A/B)-rule do not allow for a simultaneous reduction of total throughput time and floor time. The lowest performance curve, based on the routeing decision at dispatching level (stage 5), leads to the best overall performance. As could be seen from the calculations in Table 4.5 it is not unexpected that under infinite workload norms (periodic unrestricted release) the floor time and the total throughput time are lower than at all other depicted alternatives, but these experimental outcomes show that this effects also works through under tighter workload norms. Our simulation shows that it is beneficial to take the routeing decision as late as possible.

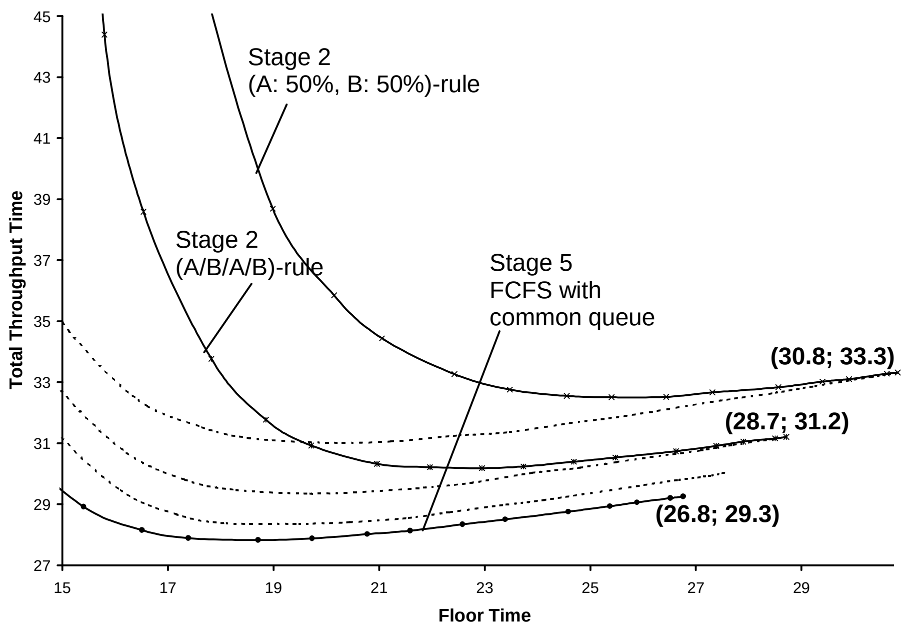


Figure 4.6 Grouping machines

To compare the absolute effects of grouping versus non-grouping within WLC means comparing the dashed lines with the solid lines. The curves that start on the right hand side within the same points are based on the same kind of routeing decisions. Under tighter norm, thus moving from right to left on the specific performance curve the differences between grouping and non-grouping become obvious. Grouping the two machines A and B (while remaining a single queue per machine each) performs worse than distinguishing different capacity groups. Only the routeing decision on dispatching level (lowest curve), which requires grouping, outperforms all other performance curves. We can conclude that if it is not possible to take the routeing decision on the dispatching level, non-grouping performs better than grouping. In such a situation the routeing decision should be made at stage 2 or stage 3, as late as possible.

Our analysis is based on interchangeable machines. Basically the same holds for non-interchangeable machines unless fewer alternatives are possible: The curves based on the (A: 50%, B: 50%)-rule also depict the performance curves of non-interchangeable machines.

To compare the direct impact of workload control on the overall performance we have to be able to compare the different curves. Therefore we established Figure 4.7. Here again all the curves as defined in the Figures 4.5 and 4.6 are depicted. Again the dashed lines are based on grouping, the solid lines on non-grouping. This is because the position of the characteristic curves in the Figures 4.5 and 4.6 are strongly influenced by its starting points at periodic unrestricted release. In Figure 4.7 we 'neutralise' those starting effects. We see the same performance curves as above, but now all the performance curves are normalised relative to periodic unrestricted release (1; 1).

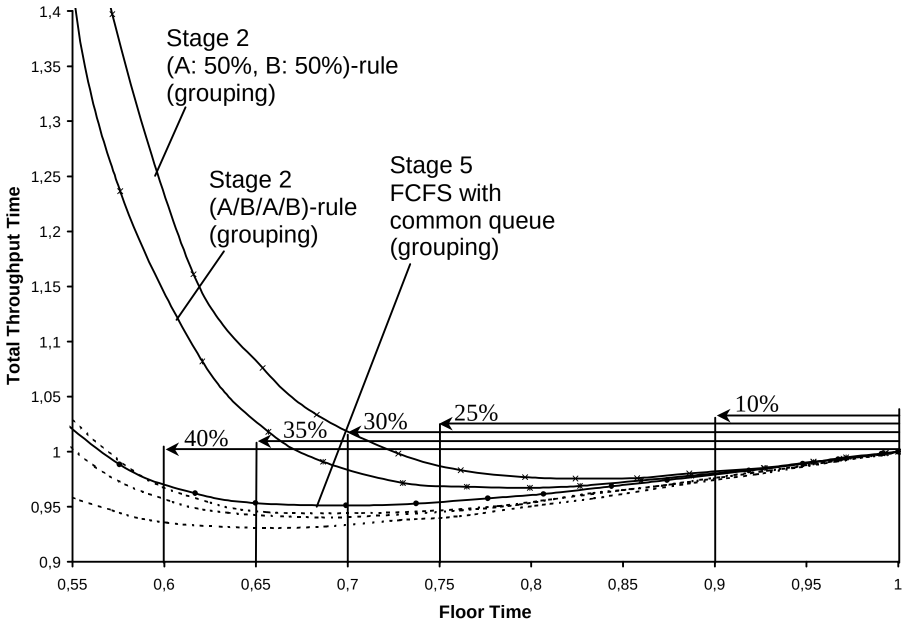


Figure 4.7 Normalised effects of grouping (dashed lines) vs. non-grouping (solid lines)

No large differences in total throughput time changes arise for the different curves when tightening norms up to a decrease in floor time of about 10% relative to unrestricted periodic release. When trying to reduce the floor time by a higher percentage, the differences become more visible. After a decrease in floor time from about 25% to 30% it is not possible to decrease the floor time without increasing the total throughput time anymore for the two (solid) upper curves based on grouping and a stage 2 routeing decision. For the other four curves this just happens after a decrease in floor time by 35% to 40%.

Especially interesting is the position of the normalised performance curve based on a routeing decision on dispatching level. Its shape is quite similar to the dashed lines based on non-grouping. Those situations are based on the control of one queue of waiting orders per capacity group. The upper two curves are based on 2 queues within the same capacity group. This may indicate that not the grouping versus non-grouping decision influences the relative benefits of WLC but the number of queues to control within the same capacity group.

Sensitivity of results

To compare the different grouping alternatives we had to made several assumptions before we were able to investigate the consequences of grouping machines on effective workload control.

The presented results are based on the *WLC approach* using corrected workloads. The same experiments have also been done using aggregate loads. The results are not presented here, because similar conclusions would be derived, though the overall performance is worse.

The *job shop size*, that means the number of machines on the floor may also influence the simulation outcomes. This analysis has been confirmed by additional experiments. Considering a job shop routeing with 6 capacity groups. 28.6% of all orders that arrive at a distinctive capacity group directly come from the pool (see appendix B). With a larger job shop this percentage decreases. This has mainly two effects. On the one hand the orders will arrive more regularly, as the influence of the periodical order release (that disturbs the inter-arrival time of order) will become less strong. On the other hand the effect of choosing the right routeing decision rule (especially the (A/B/A/B)-rule and the (LLGF)-rule) diminishes. This behaviour is mirrored by the same best performing norm ratios for experiments that are based on the same stage to take a routeing decision.

4.5 Conclusions

The overall research perspective was to investigate the impact of machine grouping on the effective use of WLC within make-to-order job shops. This paper is a starting point for a profound decision on grouping machines as well as on the control of the respective capacity groups.

Machines that are considered for being grouped together often allow for different routeing alternatives. Our simulation study has shown that deciding on the order routeing as late as possible (with a common queue in front of the machines) results in the best performance in a workload controlled environment. If this is not possible the routeing decision should be combined with the order release decision. This may lead to similar results. If the routeing decision is made during process planning the actual shop status should be considered.

Within WLC not only the determination of the right workload norm but also the relation in between the different workload norms is important. Up to now it is not possible to determine the best performing norm ratios beforehand. Though we could see that too tight norms for machines that differ from the remaining ones perform worse than looser ones. We showed that the best performing norm ratio always outperforms all other norm ratios (given a floor time).

The known effects of grouping machines on overall performance in environments without WLC do not change systematically within a workload controlled environment. Nevertheless we conclude that WLC itself performs better, if the design of capacity groups allows for a shop floor view that is as detailed as possible. This can be realised by defining a single capacity group for each machine on the floor.

While the grouping of machines and the internal structure of capacity groups influences the shop performance, the choice of the right WLC approach remains important. The WLC approach based on ‘corrected loads’ always outperforms the use of ‘aggregate loads’ but the arising patterns resulting from the same type of grouping decision show strong similarities.

This paper can be seen as a starting point in investigating the influences of grouping machines on effective WLC. The simulation model was based on (non-)interchangeable machines with identical operation processing times. Future research could consider the wider spectrum of semi-interchangeable machines with different operation processing times.

4.6 Appendix A

The calculations for the station throughput time (*STT*) and the workloads (*WL*) are based on approximations according *GI/G/m* queues. The approximations of station throughput times are based on Buzacott and Shanthikumar (1993):

(1) $STT = p\left(1 + \frac{\pi_m}{(m - m\rho\ )} \frac{(C_a^2 + \rho^2 C_s^2)}{(1 + \rho^2 C_s^2)} \frac{(1 + C_s^2)}{2}\right)$

with

$\rho = \frac{1}{m}\lambda p$ utilisation factor

π_m probability that an arriving order is forced to join the queue

$1/\lambda$ inter-arrival time

p average operation processing time

C_a^2 squared coefficient of variation of the inter-arrival times

C_s^2 squared coefficient of variation of the operation processing times

For all calculations we assume:

m=1: $\pi_1 = \rho$

m=2: $\pi_2 = \frac{4\rho^2}{(2 + 4\rho^2)}$

$C_a^2 = 1$

$C_s^2 = 0.5$

$p_{A/B} = 2$ average operation processing time on machine A and B

$p_{rest} = 1$ average operation processing time on the remaining 5 machines

$\rho = 0.9$ average utilisation for all machines on the shop floor

The workloads can be calculated analogously (Land 2004B):

(2) $WL = m\rho\left(STT + C_s^2\, p\right)$

4.7 Appendix B

The percentage of orders that arrive at a chosen capacity group directly from the pool can be derived as follows:

M: total number of orders arriving at the production system
n: number of machines (capacity groups) in the system

$\frac{1+n}{2}$ average number of operation per order (=average routeing length)

$M_i = M\frac{1+n}{2n}$ total number of orders arriving at machine *i*

$M_{i\_first} = \frac{M}{n}$ total number of orders that visit machine *i* for start operation (directly from the pool)

$P_{i\_first} = \frac{M_{i\_first}}{M_i}100\% = \frac{2}{1+n}100\%$ percentage of orders that arrives at machine *i* for start operation (relative to all orders that arrive at machine *i*)