Chapter 2
Production Control Strategies (PCS)

2.1 Fundamental Concepts and Coherences

2.1.1 PCS in the Broader Context of Production Planning and Control

The following discussions on PCS can be associated to the field of production systems engineering (PSE). "PSE is an emerging branch of Engineering intended to uncover fundamental properties of production systems and utilize them for analysis, continuous improvement, and design" (Li and Meerkov 2009). Other than classical manufacturing engineering, PSE is not concerned with the operation technology of machines or material handling devices, but with the parts flow through a production system (Li and Meerkov 2009). In this context, by production control strategy (PCS), it is referred to the information flow and the logic behind it that controls the movement of material within a factory (Hopp and Spearman 2008). Thus, it is focused rather on the information flow that controls the material flow, than on the actual physical material flow, which is the center point in the science of material flow (Arnold and Furmans 2009). In Fig. 2.1, the following work is positioned within the context of overall corporate operations planning. In the planning framework of Stadtler and Kilger (2008), PCS are part of the short term production planning.

Thereby, the focus of the presented PCS engineering framework lies on shopfloor control, or more specifically WIP control, the OPP allocation, and order release. Scheduling and lot-sizing problems,[1] which can also be relevant in short-term production planning, are not explicitly covered. The task of warehouse replenishment planning found in the short term distribution planning field can be seen as part of the following work. Neighboring planning tasks like personnel

[1] The interested reader is referred to Hopp and Spearman (2008)

C. Karrer, *Engineering Production Control Strategies*, Management for Professionals, DOI 10.1007/978-3-642-24142-0_2, © Springer-Verlag Berlin Heidelberg 2012

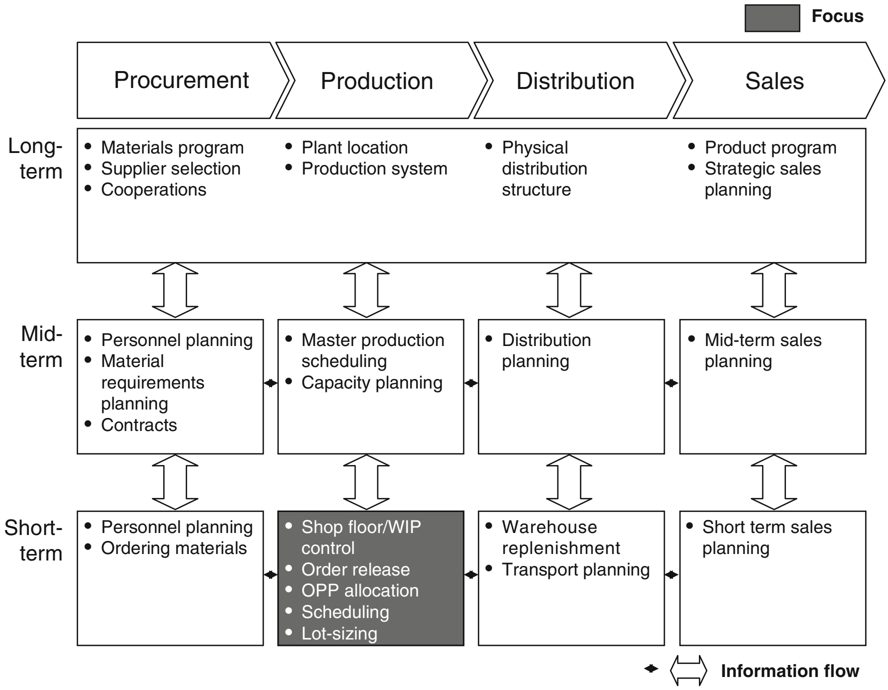

Fig. 2.1 Scope of PCS within overall planning (Adapted from Stadtler and Kilger 2008)

planning, capacity planning, outbound logistics planning, or the ordering of raw materials determine important operating conditions and thus parameters for PCS design.

2.1.2 The Push/Pull Enigma and Their Basic Implementations

PCS are often classified to be either of push- or pull-type. This distinction has caused lots of confusion and dissent among practitioners and researchers (Benton and Shin 1998; Bonney et al. 1999).

One reason for this is the large variety of often contradicting definitions used in literature and practice. Another reason for confusion originates from the fact that in practice, neither push nor pull are found in their purest form (Pyke and Cohen 1990).

The definition of Hopp and Spearman (2008) is found to be the most useful one to discuss PCS from a systems engineering point of view. According to them, the distinguishing feature is how the movement of work is triggered. In a push system, work orders are scheduled based on actual or forecasted demand by a central

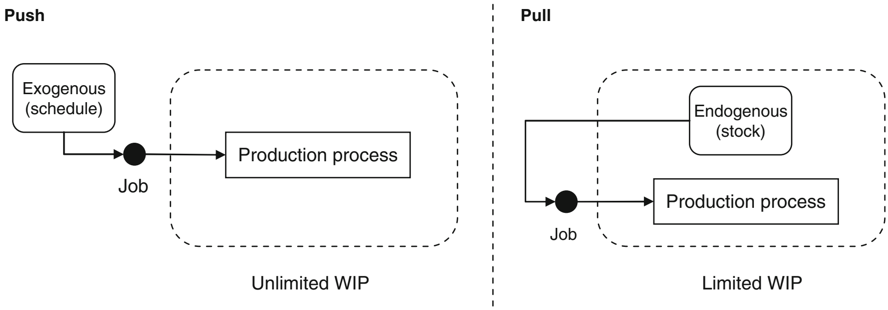

Fig. 2.2 Push and pull mechanics (Adapted from Hopp and Spearman 2008)

system. In a pull system, work is authorized based on the current system status. Figure 2.2 illustrates the two concepts.

Hopp and Spearman (2008) refine this definition based on the distinguishing effect of the two principles and state “A pull system establishes an a priori limit on work-in-process, while a push system does not.” However, push systems are able to proactively operate based on forecasted demands what is not the case in pull systems, which only react to the actual status of the system. In practice, ‘pull’ is often used to refer to three different principles. The first two will be relevant in the following work. First, as described in the definition above, pull refers to the fact that WIP is limited between process steps, and a preceding process is only allowed to produce if sufficient space is available in the input buffer of the next process. Second, pull is used to describe a make-to-stock replenishment system, also called ‘supermarket’, in which different variants are stored to fulfill customer orders. Whenever a variant is removed from the supermarket, the same quantity is reordered and reproduced to fill up the empty spot. Third, with “pull” it is referred to a concept in internal production logistics, in which the production line is supplied with raw materials based on actual demands (Hopp and Spearman 2008; Womack et al. 2007).

Prominent implementations using mainly the push principle are material requirement planning (MRP) systems. Prominent implementations of the pull system are Kanban and Basestock. They will be briefly introduced in the following.

The idea of MRP systems was developed in the early 1960s by Joseph Orlicky (1975) as computer technology started to be used commonly by companies. The basic function of MRP is to calculate quantities and process start times for intermediate products (or ordering times for raw materials) based on actual or forecasted demands for final products. Thus, each process in the production system is planned and scheduled by a central system. The production orders are then ‘pushed’ into the system.

Due to the ability of MRP systems to process actual and forecast-based orders alike, MRP, and thus push, are often equalized with make-to-forecast (MTF). To translate the demand for final products into demands for raw materials and intermediate products, the so-called bill-of-material (BOM) is used. The BOM is a tree explaining on different levels the composition (type and quantity) of end and intermediate products (Orlicky 1975). Soon, operating problems of MRP were discovered. A general problem is the contradiction between the MRP's deterministic nature and the uncertainty of operations where actual lead times can seldom be predicted accurately. This leads for instance to long planned lead times to safeguard timely deliveries what then causes high levels of work-in-process in the system and again, longer and more variable lead times (Hopp and Spearman 2008).

Another issue is that production capacity is not considered what leads to infeasible production schedules and again increased variability in lead times. A further problem is called system nervousness and refers to the effect that small changes in the master production schedule can lead to large changes in planned order releases. Some of these problems could be mitigated by the introduction of manufacturing resources planning (MRP II), however, in shop floor reality, most of them remained (Hopp and Spearman 2008; Benton and Shin 1998). Today, MRP II software is usually part of comprehensive software packages called enterprise resource planning (ERP) (Jacobs and Bedoly 2003).

A popular pull-type implementation is Kanban. The Kanban system originates from the Toyota Production System (TPS) where it has been implemented as control mechanism for a production line in the mid-1970s. Kanban is Japanese for "card" and refers to the information carrier used to convey production authorizations between consecutive process steps (Ohno 1988; Kimura and Terada 1981). Even though a large variety of articles has been published on the topic, the definition of a Kanban system remains ambiguous. A summary of different definitions is provided by Berkley (1992). The mechanism is explained along the unified framework for pull control mechanisms developed by Liberopoulos and Dallery (2000) in Fig. 2.3. Each manufacturing stage MF_{i} has an input buffer I_i and an output buffer PA_i. Arriving Kanban cards are collected in queue DA_i.

Whenever a Kanban card is present in DA_{i+1} and the corresponding material is available in PA_i, the processing of this part type is initiated by launching it in the input buffer I_{i+1} of the next process. At the same time, the Kanban card is detached

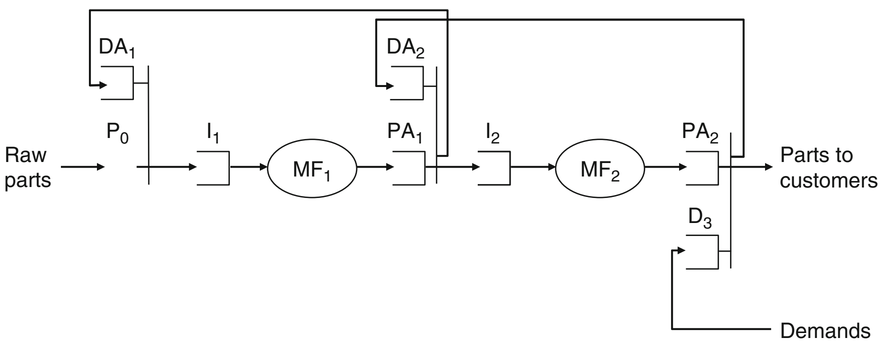

Fig. 2.3 Illustration of a two-stage Kanban system (Liberopoulos and Dallery 2000)

from the material and sent back to DA_i to reproduce the consumed part. The WIP in this system equals per definition the number of Kanban cards and is thus limited. The customer demand for end products is communicated stepwise upstream (Liberopoulos and Dallery 2000). It can be distinguished among Kanban systems that perform the blocking and production authorization per part type, and systems that block by total queue size (Berkley 1992). The first ones are in practice sometimes referred to as supermarket systems, the latter as sequential pull systems. The question how to set the number of Kanban cards has been extensively addressed in literature. A survey on this question can be found in Berkley (1992).

To close the introduction of the Kanban system, a few remarks about its relation to Just-in-Time (JIT) should be made. JIT stands for an approach that ensures that it is only produced what the customer needs, at the right point in time, in the right quantity, and with minimal lead time. To achieve this, JIT resorts to the tools of continuous flow, takt time, production leveling, and pull systems. Kanban is one option to implement a pull system. Thus, JIT is a superordinate concept to Kanban (Drew et al. 2004).

A similar way to implement the pull principle that even appeared earlier in literature than Kanban is Basestock (Clark and Scarf 1960). Applying the same framework and two-stage production system as used to illustrate the Kanban system, Basestock can be described as displayed in Fig. 2.4. In its initial state, the output queues P_i of the manufacturing processes MF_i contain a certain initial amount of stock, the so-called ‘basestock’ that gave the system its name. Whenever a demand event for an end-product occurs, it is instantly communicated to the demand queues D_i of all processes. Given that the needed inputs are available in P_{i-1}, production is started.

The distinctive feature when comparing Kanban and Basestock is that in Basestock, the demand information is immediately communicated to all processes, whereas in Kanban, is travels stepwise upwards against the material flow. The Basestock system is equivalent to the Hedging Point Control System (Liberopoulos and Dallery 2000).

A large variety of enhancements and combinations of MRP, Kanban, and Basestock were developed. An overview on them will be given later in this chapter in the course of the review of current PCS method design research.

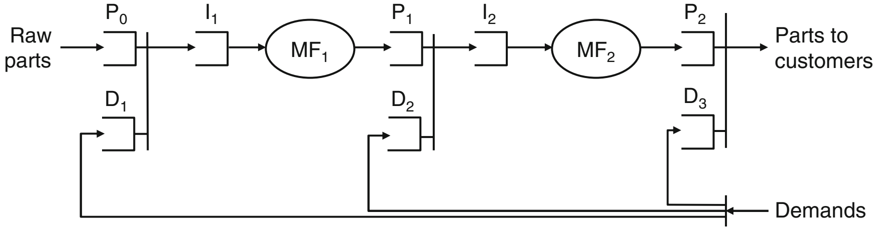

Fig. 2.4 Illustration of a two-stage Basestock system (Liberopoulos and Dallery 2000)

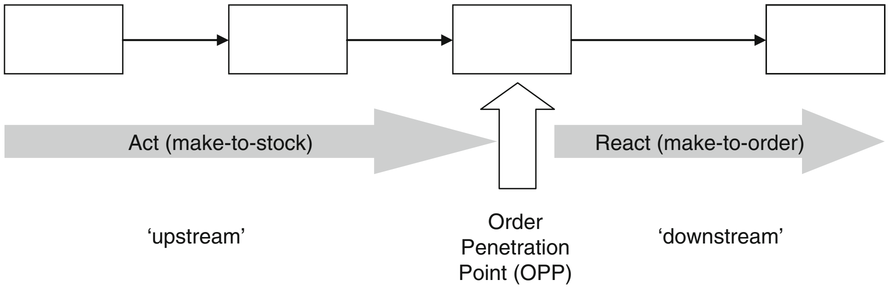

Fig. 2.5 The order penetration point (Adapted from Alicke 2005)

2.1.3 The Order Penetration Point

A concept that will be important within the PCS engineering framework constructed in the following is the order penetration point (OPP), sometimes also referred to as customer order point. "The order penetration point (OPP) defines the stage in the manufacturing value chain, where a particular product is linked to a specific customer order" (Olhager 2003). Figure 2.5 illustrates this idea.

To the processes left of the OPP it is referred to as 'upstream', to the processes to the right of the OPP as 'downstream'. For the upstream and downstream part, different production strategies have to be considered (Olhager 2003). Before the OPP, a way of dealing with uncertain demands needs to be found. After the OPP, a make-to-order system is feasible which does not have to hold any inventory to cover for uncertain demands. Therefore, moving the OPP as far upstream as possible, saves inventory (Alicke 2005). How far the OPP can be moved upstream depends on the comparison of the customer lead time, which is the time period the customer is willing to wait from order to delivery, with the production lead time, which is the time needed to complete the order from the potential OPP location to delivery. To determine the OPP location, also other criteria can play an important role. Examples include the customization options provided to the customers or the product structure in general. For a detailed discussion of these strategic considerations the reader is referred to Alicke (2005) or Olhager (2003).

2.1.4 The Influence of the PCS on Operational Performance

The following section has the objective to summarize some main influences of the PCS on operational performance as discussed in literature and observed in practice. Thereby, important cause-and-effect relationships that are relevant for the subsequent discussions are described. However, due to the large variety of potential influences, only the most prominent ones can be considered and the overview is not exhaustive. Operational performance is commonly measured in three dimensions:

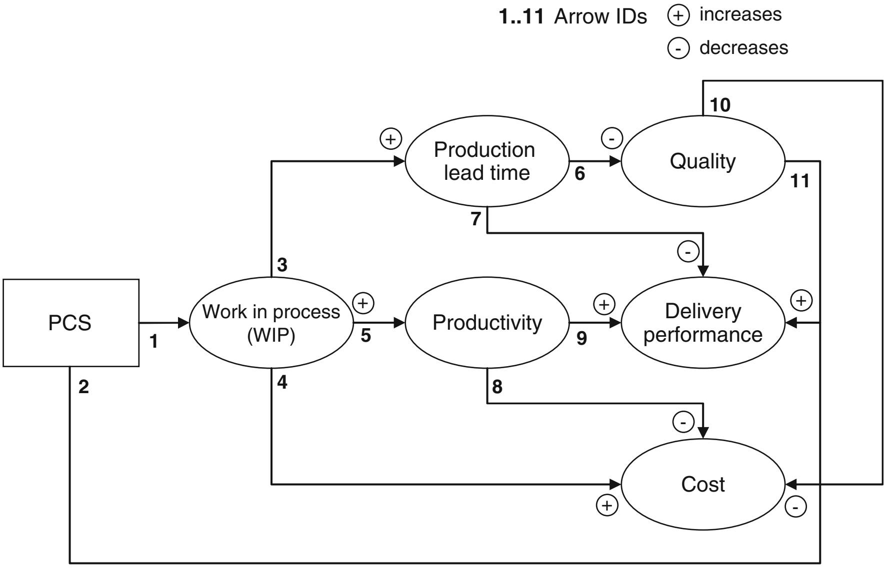

Fig. 2.6 Influence of the PCS on operational performance

quality, cost, and delivery performance (Drew et al. 2004). With the help of the illustration in Fig. 2.6, the most important causal chains from PCS over system characteristics to operational performance will be explained. On the arrows of the influence diagram it is indicated whether there is an influence that increases or decreases the impacted characteristic. Moreover, an identifier (ID) is assigned to each arrow to ease the following discussion.

Per definition, the PCS has a direct influence on WIP in the production system, i.e. its location, type, and amount {1}. The PCS has also an obvious direct influence on the delivery performance {2}. It needs to trigger material movements such that timely delivery is ensured (Hopp and Spearman 2008).

Besides the actual processing time, WIP is closely linked to production lead time, since it causes waiting time in front of processes {3} (Arnold and Furmans 2009).

The inversely proportional impact of production lead time on quality performance {6} can be explained with the idea of quality feedback loops. In many production systems, the quality of parts cannot be or is not directly assessed until a later process step. The longer the lead time to this process step is, the longer it takes until a potential error is discovered and solved. This also means that more parts are produced, potentially containing the error that would lead to scrap or rework (Drew et al. 2004).

Delivery performance is, besides the direct influence of the PCS, driven by the production lead time {7} and, under the assumptions of insufficient capacity, by productivity {9}. A longer production lead time has a negative influence on delivery performance for multiple reasons. First, the flexibility to react no short notice changes or orders is limited. Second, the longer the lead time is, the bigger is also its absolute variation and thus the probability to fail to deliver on time and in full. Third, assuming that a certain lead time is given by the customer, the shorter the production lead time is, the more time is available to react to external disturbances like for instance poor delivery reliability of suppliers. The impact of productivity on delivery performance {9} is the bigger, the closer the plant operates at its capacity limit. In a plant with low utilization, the effect of low or variable productivity is weakened. Along the same logic, the impact of quality problems {11} can be argued. Under the assumption that the plant operates at the capacity limit, quality losses lead to a reduced output and impact the ability to deliver (Lödding 2008; Alicke 2005; Drew et al. 2004).

Within the set of operational metrics considered above, cost is mainly driven by productivity {8}, WIP {4}, and quality {10} (Simchi-Levi et al. 2007). In some business models, also delivery performance could be added as cost driver (e.g. special freight cost, penalties) (Alicke 2005). Even though not indicated in the diagram above, in some industries, an influence from lead time to cost is present. In the apparel industry for instance, where fast reactions to trends are indispensible, a long lead time can cause opportunity costs in form of lost sales (Simchi-Levi et al. 2007).

To conclude the analysis, the influence of WIP on productivity {5} will be explained in more depth. In a production system with variable cycle times, buffers ensure a smooth operation by preventing processes from starving if the supplier does not deliver in time, or from blocking if the subsequent process is not ready to accept a new job. Thus, depending on the variability level, increasing WIP leads, under certain assumptions, to increased productivity. The productivity gain is diminishing with increasing WIP level. This basic coherence has been extensively explored by Nyhuis and Wiendahl (1999) within their operating curves approach (Fig. 2.7).

However, as mentioned above, this relation assumes that there are no other influences of WIP on productivity, which is not necessarily true. For instance, in space constraint environments, additional WIP can also reduce productivity by inducing waste (e.g. more motion required) (Drew et al. 2004).

Moreover, a further perspective can be added to this trade-off. From a time perspective, the improvement speed of the whole system, which is the speed in which variability and thus the need for WIP can be reduced, is the faster, the lower the WIP level is. On the one hand, low WIP levels derogate productivity, but on the other hand, make problems more visible. With low WIP levels, more processes are

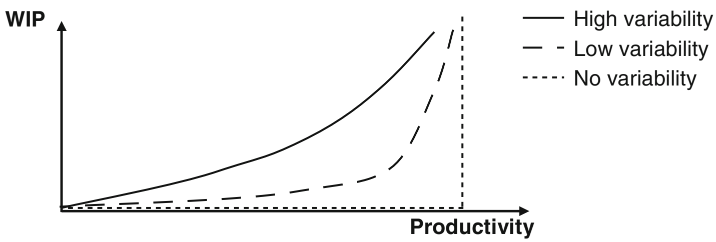

Fig. 2.7 Illustration of the trade-off between WIP and productivity (Nyhuis and Wiendahl 1999)

affected faster by problems and thus, more ‘pain’ is caused within the organization. This enables ultimately a more consequent root-cause-problem-solving of issues. This feedback loop is one of the core ideas in Lean Manufacturing (Womack et al. 2007).

2.2 Review of Current Research on PCS

2.2.1 Segmentation of Literature

In the following, an overview of current research in the field of production control strategies (PCS) will be given. Therefore, the existing literature is clustered and the main findings within each cluster are summarized and interpreted in the context of the following work. The literature review is fitted towards first, giving a broad and comprehensive overview on current research in the field, and second, towards introducing the groundwork on which the PCS engineering framework will be built on.

PCS related research is mainly addressed by publications from the field of Operations Research (OR) and production engineering. It is hard to overlook the vast body of literature. Thus, it is proposed to cluster the publications in the field according to a logical order from PCS method development, over PCS selection, to PCS implementation, and related design questions as depicted in Fig. 2.8. The discipline of PCS engineering, to which this investigation wants to contribute, spans across all four fields. Getting a broad overview is essential in order to develop a practically applicable and holistic PCS engineering framework.

Publications in the first area, PCS method development, create new, enhance existing, or help to parameterize PCS. Starting from basic push and pull approaches, a large variety of enhancements has been developed, also unifying characteristics of push and pull, leading to so-called hybrid systems. This cluster is well penetrated and can be considered as fundamental PCS research. The second field addresses the question of selecting an appropriate production control strategy for a given production system. Studies in this area compare the performance of selected PCS. The third cluster comprises implementation studies. Having chosen and customized the right PCS, implementation studies deal with how to turn these strategies into shopfloor reality. The fourth cluster bundles publications addressing design questions closely related to PCS design. Examples include scheduling, lot sizing, or inventory control. The literature review has its emphasis on the first three areas. Figure 2.9 shows a taxonomy of the relevant literature along which the remainder of this chapter will be structured.

The cluster of PCS method development can be split up further relying on the distinction between push and pull systems. The developed methods can be classified as either ‘advanced push-type’, ‘advanced pull-type’, or ‘hybrid’, which combine push and pull features. Hybrid systems can be distinguished further into ‘horizontally integrated systems’ and ‘vertically integrated systems’ (Cochran and Kaylani 2008). Vertically integrated systems consist of a higher-level push system

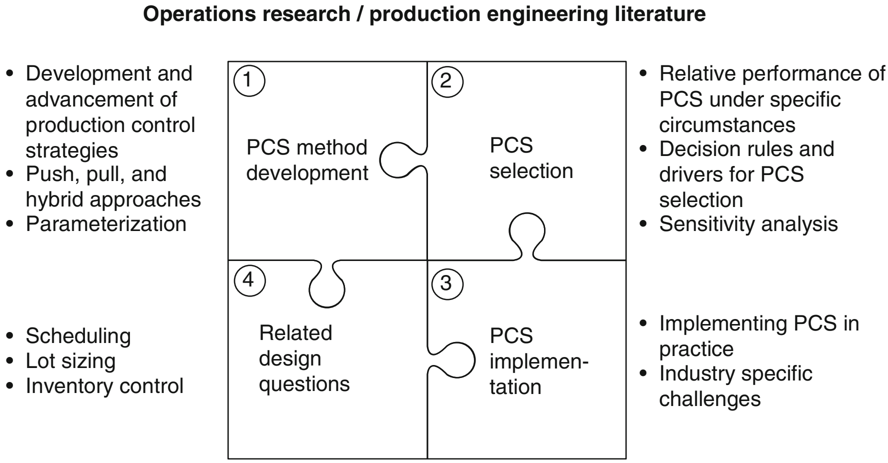

Fig. 2.8 Segmentation of PCS related literature

superimposed on a lower level pull system. In horizontally integrated hybrid strategies, some stages are controlled by the push principle and others by the pull principle. Figure 2.10 illustrates the two types of hybrid strategies.

Maes and VanWassenhove (1991) argue qualitatively that hybrid systems should be superior to pure push or pull systems in many application cases. In the following, important contributions to vertically and horizontally integrated hybrid systems, as well as to advanced pull- and advanced push systems are presented. The table in Appendix 8.6 gives an overview of the subsequently mentioned publications related to method design, thereby comparing their solution approach and major assumptions. A comprehensive survey of early publications on hybrid PCS is compiled by Benton and Shin (1998).

2.2.2 PCS Method Development

2.2.2.1 Vertically Integrated Hybrid PCS

One of the first contributions in this field comes from Hall (1986). He presents the “Synchro MRP” system used at Yamaha plants. Synchro MRP combines a classic MRP system with two card Kanban loops between all consecutive processes. Each process step is only authorized to produce, if an MRP production order and a Kanban card are present for the specific variant. Suri (1998) developed a similar system, the Polca (*P*aired-cell *O*verlapping *L*oops of *C*ards with *A*uthorization) control system. In the Polca system, a central MRP system determines the start date of each production order in every process by backward scheduling. Polca cards rotate between two consecutive processes and authorize production. Only if the start date of a production order is reached, and a Polca card of the subsequent process

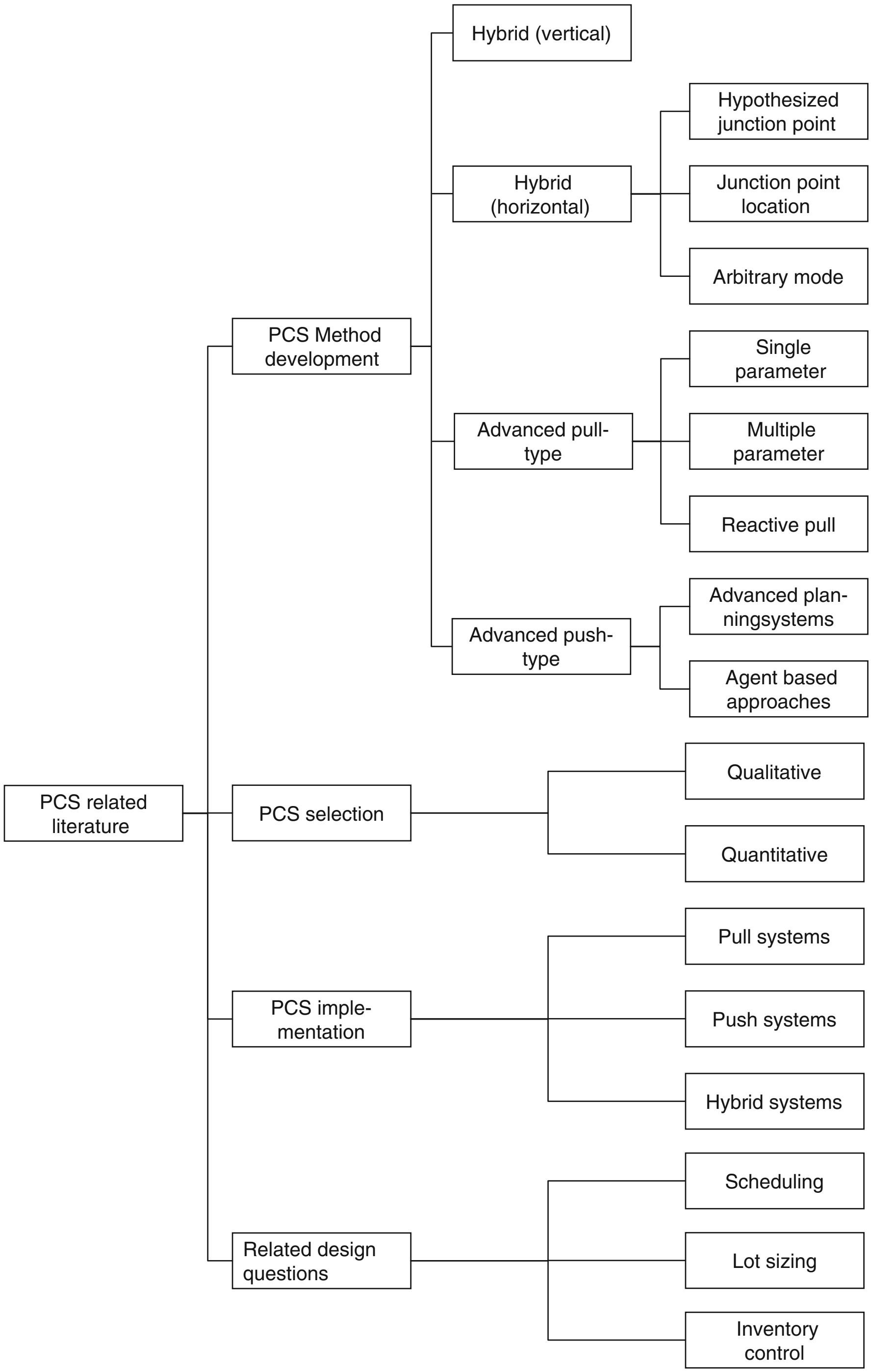

Fig. 2.9 Taxonomy of PCS related literature

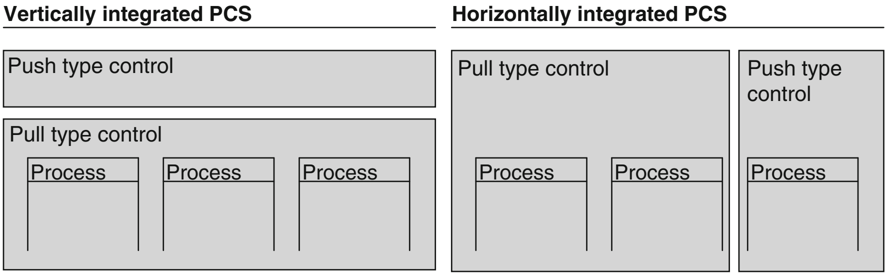

Fig. 2.10 Illustration of vertically and horizontally integrated PCS

is present, the production order is executed. Unlike Kanban cards in Synchro MRP, Polca cards are not variant-specific. A different approach is taken by Bertrand and Wortmann (1981) and their system called "workload control." A high-level MRP system generates a list of prioritized production orders. For each process, the system maintains a workload account and a workload threshold. The workload account contains the workload of all orders in the system that still need to pass this process. A production order is only released into the system, if each process the order needs to pass, would not exceed the workload threshold. Using this mechanism, the system establishes a pull-type characteristic and limits the total WIP.

A similar approach has been proposed 1984 by Bechte.[2] He introduces the concept of load-oriented manufacturing control. It works according to the same basic principle as workload control. However, when an order is released into the production system, not the full workload is assigned to the succeeding processes, but a discounted workload $Tbooked_j$, depending on the distance of the order to the considered process j. The calculation of the booked time is illustrated in (2.1) (Bechte 1984).

$$Tbooked_j = \begin{cases} Torder_j & \text{if } currentProcess = \mathrm{j} \\ Torder_j \cdot \prod\limits_{i=currentProcess}^{j-1} \dfrac{DiscountFactor_i}{100} & \text{if } currentProcess < \mathrm{j} \end{cases} \quad (2.1)$$

$Tbooked_j$: Time booked on account of process j
$Torder_j$: Processing time in process j
$currentProcess$: Index of process that currently works on the order
$DiscountFactor_i$: Discount factor of process i

Bechte (1984) suggests to calculate the discount factor as the reciprocal of the load limit divided by the planned throughput per planning period of a process.

[2] A description in English language of the concept can be found in (Bechte 1988)

This approach leverages the obvious coherence that the probability to actually complete a job within a planning period at a process is the bigger, the smaller the total waiting and processing time for a process is. This approach has been approved in practice but also been criticized for several reasons. The most important drawback occurs in production systems with low utilization. Here the completion probability of jobs is underestimated. An improved method in which the discount factor is independent from the load level is delivered by Perona and Portioli (1996).

2.2.2.2 Horizontally Integrated Hybrid PCS

Publications in this field address the question, which stages of a production system to control using the push principle, and which stages of a production system to control with the pull principle in order to create a hybrid system. The problem is solved either by modeling it as a Markov Decision Process (MDP) or by using discrete-event simulation. A distinctive feature of horizontal integration studies is the considered solution space. Along this criterion, three basic categories can be identified. In the first category, the location of a junction point[3] at which the control mode changes is directly hypothesized (for instance at the bottleneck). In the second category, the existence of one junction point is assumed and its location is determined via optimization. In the third category, each stage is allowed to either push or pull and the optimal control mode is determined for each stage via optimization.

A set of publications that directly hypothesizes the junction point location proposes to locate it at the bottleneck. Thereby, pull control is used from the bottleneck upstream and push control from the bottleneck downstream. This intuitive logic is used for example by the "Drum-Buffer-Rope" concept as described by Goldratt and Fox (1986), the "Starvation Avoidance" concept (Glassey and Resende 1988) or by the approach developed in Huang (2002). Beamon and Bermudo (2000) suggest a system that locates the junction point between sub and final assembly lines. Push logic is used for subassembly lines and pull logic within the final assembly line.

Olhager and Ostlund (1990) identify and describe further potential locations of the junction point. They propose to locate it according to the customer order point, the bottleneck, or the product structure. However, they do not provide guidance how to choose among the three options.

The problem of optimally locating the junction point and not hypothesizing it has been addressed by Takahashi and Soshiroda (1996) with the help of a set of difference equations. They allow the first processes to consistently either push or pull up to the junction point where the control mode alternates. They establish a relationship between the autocorrelation of the demand with the value of the

[3] Also known as 'push-pull-boundary' (Alicke 2005)

integration parameter. A similar problem is investigated by Hirakawa (1996) with means of simulation. Also Cochran and Kaylani (2008) picked up the junction point location problem. They focus on the question, whether each part type should have its own junction point or if a common junction point should be preferred. They minimize inventory holding and tardiness costs. Therefore, they optimize the junction point location, the safety stock level for the push stages, and the number of Kanban cards in the pull stages. The number of feasible solutions for a system with m stages, n parts, Q different counts of Kanban cards, and S different levels of safety stock equals according to Cochran and Kaylani (2008)

$$(S \cdot \sum_{i=0}^{m} Q^{m-i})^{n} \tag{2.2}$$

The underlying optimization problem is *NP*-hard and a genetic algorithm is applied to solve it. From simulation experiments and the application to a tube shop of an aerospace manufacturer, the following main conclusions are drawn:

- Horizontally integrated strategies can create value compared to pure push or pull strategies.
- If a bottleneck exists, the push-pull barrier should be located at the bottleneck process.
- Lower variability in parts arrival leads to lower safety stock.
- One junction point should be preferred compared to several product specific ones unless two parts sharing equipment have largely differing ratios of inventory holding cost to late cost.

Hodgson and Wang (1991a) studied the problem with a completely open solution space, e.g. each station can either push or pull. They developed an MDP for a four-stage iron and steel works production system. For the observed case example, they conclude that pushing in the first two stages and then pulling in the last two stages is a strategy with superior operational characteristics. They later extended their work (Wang and Hodgson 1992) to general parallel and/or serial multistage production systems. In the observed convergent material flow, they propose to push until the flows merge and to pull afterwards. Geraghty and Heavey (2004) later build on their model and show that in the way they modeled the pure push logic, it still has a WIP cap in each stage and thus their model is equivalent to a Kanban/CONWIP system, which will be presented as an advanced pull system later.

Hodgson and Wang's work has been extended by Pandy and Khokhajaikiat (1996) who introduce uncertain demand, production, and raw material supply. They then study a four-stage hair dryer production system. An overview of contributions to horizontally integrated hybrid systems can also be found in Geraghty and Heavey (2005).

2.2.2.3 Advanced Pull-Type Systems

In Sect. 2.1.1, the two most common pull-type production control strategies, Kanban and Basestock, were introduced. According to the pull definition of Hopp and Spearman (2008), they share the commonality that WIP is limited within them. In the following, extensions of these systems developed in current research will be presented. First, advanced pull-type systems that have one parameter per control loop are covered. Next, generic systems with multiple parameters per control loop will be investigated. The section concludes with what will be referred to as 'reactive pull-type systems' that adjust one or more of their parameters during operation to changing environmental conditions. To illustrate the advances in the field of one parameter pull-type systems, a three-step production line as displayed in Fig. 2.11 is used.

In Fig. 2.11, a classical Kanban system is displayed, which constraints the amount of WIP for each variant between two consecutive processes. A Kanban signal authorizes the reproduction of one unit as soon as an entity leaves the inventory. The single control parameter is the number of Kanbans per loop. This classic approach has been integrated with a variety of related optimization problems like the lot sizing problem (Li and Liu 2006). An overview of studies related to the classic Kanban system is provided by Berkley (1992).

In 1990, Spearman et al. introduced and studied the CONWIP (*Con*stant-*W*ork-*I*n-*P*rocess) pull system that puts a total WIP cap on the whole production line as displayed in Fig. 2.12.

Since its invention, the CONWIP system received lots of attention from research and got attested superior operational characteristics compared to other approaches by various studies. Framinan et al. (2003) provide an exhaustive survey of CONWIP related publications.

Bonvik and Gershwin (1996) combined the Kanban and CONWIP system as displayed in Fig. 2.13 to the Kanban/CONWIP system. Here, processes that are part of more than one Kanban loop, like the first process in Fig. 2.13, are only allowed to produce, if a card from each loop is present. It is shown that in a certain operating environment, this policy can achieve almost the same output with less WIP compared to pure Kanban or CONWIP systems. Kleijnen and Gaury (2003) attest this system a superior performance when robustness and risk are considered.

Finally, optimization models that allow installing arbitrary pull loops as displayed in Fig. 2.14 were developed. Within these systems, the challenge is to determine the optimal number of Kanban cards (allowed WIP) for each control loop. Control loops with an infinite number of cards are not implemented.

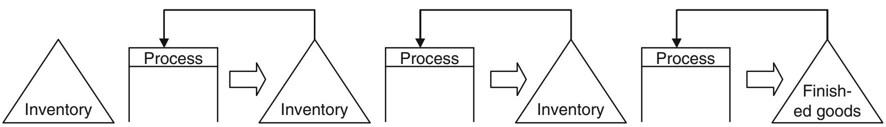

Fig. 2.11 Illustration of a Kanban system within a serial three-step production line

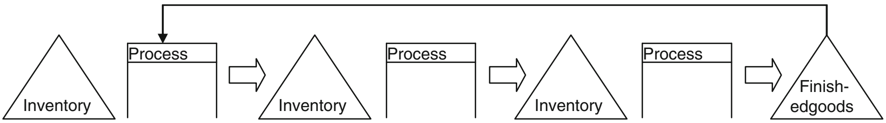

Fig. 2.12 Illustration of a CONWIP system

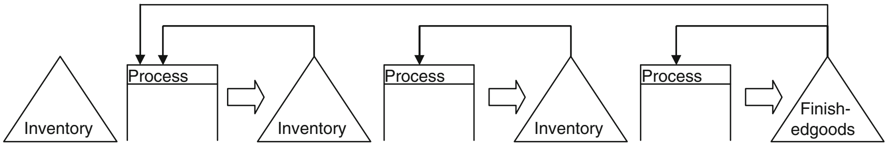

Fig. 2.13 Illustration of a combined Kanban/CONWIP system

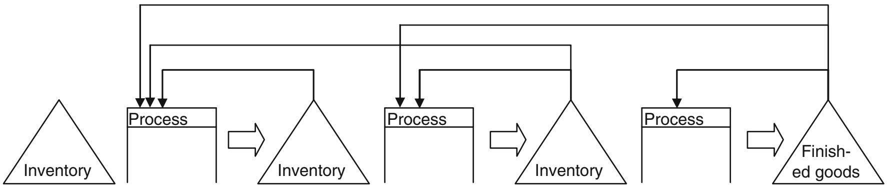

Fig. 2.14 Illustration of a production line with arbitrary pull loops

Gaury et al. (2001) apply discrete-event simulation and an evolutionary algorithm as heuristic to study system configurations. They perform a Plackett-Burman design (Montgomery 2009) with ten parameters varied on two levels and compare the resulting configurations. The observed parameters are displayed in Table 2.1 and include measures of the production system's structure, process variability, and demand variability.

They conclude that there is no dominant one-fits-all solution according to the considered WIP/delivery performance trade-off. However, they identify two important patterns that characterize most solutions. One links each stage to the first stage, and the other links the last stage to each preceding stage.

For the same problem, Masin and Prabhu (2009) apply a simulation-based feedback control algorithm called Average-Work-In-Process (AWIP) to derive solutions for the allowed WIP in each loop. In each simulation step, the number of Kanban cards in every loop is either increased or decreased. The number of cards $\mu_{ij}(t)$ at time t in the loop between stages i and j can be expressed as (Masin and Prabhu 2009)

$$\mu_{ij}(t) = \int_0^t k_{ij}(\tau) \cdot \xi_{ij}\left(\mu^* - \mu(\tau), \frac{1}{i} - w_{0i}(\tau), 0 - B_{ij}(\tau)\right) d\tau + u_{ij}(0) \tag{2.3}$$

Table 2.1 Experimental design in Gaury et al. (2001)

Factor	Two factor-levels	
	+	–
Line length	4	8
Line imbalance	0	0.5
Imbalance pattern	Funnel	Reverse funnel
Processing time coefficient of variation	0.1	0.5
Machine reliability	Perfect	Exponential breakdown
Demand coefficient of variation	0	0.5
Demand rate/capacity	0.9	0.8
Service level target [%]	99	95
Inventory value ratio	1	2
Customers' attitude	Lost sales	Backorders

$\mu_{ij}(0)$ represents the initial number of Kanban cards. The integral accounts for the changes in each adaption round from 0 to t. $k_{ij}(t)$ is a gain function that determines the magnitude of the change. $\xi_{ij}(t)$ can take values -1 or 1 and thus determines whether the number of cards is increased or decreased. It is a function of actual and required times between departures of units from the process step, WIP, and blocking characteristics. A further discussion would go beyond the scope of this survey. By applying the feedback control algorithm above, Masin and Prabhu (2009) show that they can save up to 50% stock compared to classical Kanban systems.

In the pull systems discussed above, the demand information is propagated together with the Kanban cards in opposite direction of the material flow. Thereby, the number of Kanban cards is the only parameter of each control loop. The following systems have more than one parameter and allow for more sophisticated information flows. They are generic, meaning that depending on their configuration, they can emulate different PCS. The Extended Kanban Control System (EKCS) (Dallery and Liberopoulos 2000) and the Generalized Kanban Control System (GKCS) (Buzacott 1989) separate the demand information from the Kanbans. Both are a combination of the Kanban and the Basestock system. However, Dallery and Liberopoulos positioned the EKCS as an enhancement of the GKCS since it exhibits two advantages over it. First, the functions of the roles of the parameters are clearly separated, which enables easier optimization. Second, the demand information is propagated upstream faster in the EKCS (Liberopoulos and Dallery 2000).

The EKCS will be explained in more detail in the following. For an N stage serial production system with one product type, Fig. 2.15 illustrates the operation of the EKCS. Production on the manufacturing process MP_i is triggered if input material is available in queue PA_{i-1}, a Kanban is present in queue A_i, and a is demand present in queue D_i. $J_{i,j}$ denotes the synchronization station between processes i and j that triggers production if all prerequisites are met. The two parameters of each control loop are the number of Kanban cards K_i and the initial base stock level S_i. S_i Kanban cards are attached to the finished goods of MP_i and $(K_i - S_i)$ Kanban cards

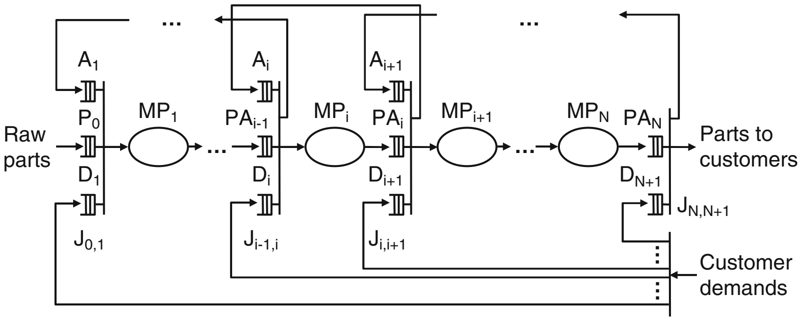

Fig. 2.15 Illustration of the EKCS (Dallery and Liberopoulos 2000)

are in queue A_i. Thus, during operation, the WIP level will be somewhere between S_i and K_i. The system therefore adapts its WIP level to varying demand and demand information is instantly passed on to all stages. It can be shown that the EKCS includes Kanban and Basestock as special cases (Dallery and Liberopoulos 2000).

An application and extension of the EKCS to non-serial flows in assembly manufacturing systems is demonstrated by Chaouiya et al. (2000). An extension to multiple products is provided by Baynat et al. (2002). Further possible enhancements of Kanban, EKCS, and the GKCS identified in Liberopoulos and Dallery (2000) are the introduction of WIP stage control, which limits the WIP of one single stage, and segmented systems, which nest different types of pull systems. Moreover, it is shown for several blocking mechanisms, like for instance minimal blocking (Mitra and Mitrani 1989) that they can be emulated by the approaches mentioned above. Comprehensive surveys of advanced pull-type systems can be found in Geraghty and Heavey (2005) and Liberopoulos and Dallery (2000).

A further generalization of the GKCS is the Production Authorization Cards (PAC) system (Buzacott and Shanthikumar 1992). The PAC system operates with a large variety of 'tags' among which the most important ones are production authorization cards, order tags, requisition tags, process tags, and material tags. Its basic operation will be explained using Fig. 2.16.

Production cells can request parts from stores using requisition tags. Process tags and order tags eventually form a production authorization (PA) card. The PA card is processed by the cell management. Thereby, order and requisition tags are generated and distributed to the preceding store. The cell management might introduce a delay between sending the process tag and the requisition tag. The operation of the cell management is not specified in detail in order to keep the system as flexible as possible. The PAC system is able to emulate a large variety of existing PCS approaches, among them Make-to-Order, Basestock, Kanban, MRP, and CONWIP (Buzacott and Shanthikumar 1992). The PAC system is a powerful approach but due to its high number of different tags, also a very complex system. The definitive paper on the PAC system is held very general in order to ensure a wide applicability and does not propose an approach for its customization. A comprehensive guideline for its customization has been delivered later by Rücker (2006).

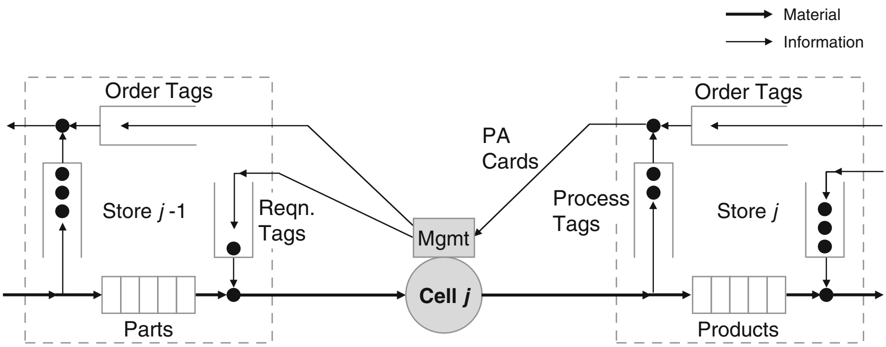

Fig. 2.16 Illustration of the PAC system (Buzacott and Shanthikumar 1992)

An interesting stream of publications examines the effect of advance demand information (ADI) on pull systems. Claudio and Krishnamurthy (2009) provide a survey on this field and examine the effect of perfect ADI on single and multiple product Kanban systems. The effect of imperfect ADI on a Basestock system in a single product single stage system is analyzed by Gayon et al. (2009). In their work, the ADI is imperfect in the sense that customers may cancel orders or order prior or later than expected. A similar scenario is analyzed by Liberopoulos (2008). He shows that in a basestock system, in which the ADI lead time is long enough, the basestock level drops to zero and that, under certain conditions, a linear relation between the basestock decrease and ADI lead time increase exists. Liberopoulos and Koukoumialos (2008) study the effects of variability and un-certainty on a Make-to-Stock supplier. Thereby, single items are ordered one at a time by two types of customers. The first type requires immediate delivery whereas the second type provides uncertain ADI in the form of cancellable reservations. The authors draw conclusions regarding the necessary stock in the capacitated and uncapacitated case. Babai et al. (2009) suggest a new dynamic reorder point policy that considers externally given and known forecast uncertainties in a single-stage single-product production system. Benjaafar et al. (2010) examine a single-product production system with stochastic processing times. ADI is provided and updated with variable lead time. Orders can be cancelled. They suggest for future research to “consider systems where order sizes are variable and where the actual number of units in each order is not exactly known until the order becomes due”, an aspect picked up in the model developed in the next section. Further work examining the impact of ADI on PCS has been provided by Tan et al. (2007), Karaesmen et al. (2004), and Liberopoulos et al. (2003). The groundwork for integrating ADI in Basestock systems can be found in Karaesmen et al. (2002).

A reactive Kanban system has been proposed by Takahashi and Nakamura (2002) and will be described in the following. The demand and its variability is one of the drivers for the number of Kanban cards within a loop. Therefore, they monitor the inter-arrival time of customer orders in order to detect the need to recalculate the number of Kanban cards. It is distinguished between stable and unstable demand changes. A stable demand change comes from the random nature of the observed variable and does not indicate that the structure (mean, variance, or distribution type) of the underlying random variable changed. Unstable demand changes indicate that the underlying distribution changed and thus, the number of Kanban cards should be adapted to the new situation. The differentiation between stable and unstable changes is done with the help of an exponentially weighted moving average (EWMA) control chart. If an unstable demand change is detected, the number of Kanban cards is adapted to the new mean and variance of the inter-arrival times. A survey of further reactive Kanban approaches is provided by Tardif and Maaseidvaag (2001).

2.2.2.4 Advanced Push-Type Systems

To overcome the weaknesses of classical MRP push systems, they were not only combined with pull systems, but also efforts to improve the underlying push logic itself were made. The integration of algorithms from Operations Research, mostly heuristics, lead to so-called Advanced Planning Systems (APS) (Günther and Tempelmeier 2012). They still rely on the basic centralized push logic, but are able to solve more complex planning or sequencing problems. A detailed survey on the extensive amount of literature available on APS will not be given here. The interested reader is referred to Günther and Tempelmeier (2012), Stadtler and Kilger (2008), or Tempelmeier (2001).

During the last decade, also agent-based production control systems were investigated, for instance by Gelbke (2008), Mönch (2006), and Khoo et al. (2001). Mönch (2006) proposes a framework for a distributed hierarchical control system for the semiconductor industry, called FABMAS (Fab-multi agent system), whose high level architecture is shown in Fig. 2.17. FABMAS performs planning on three levels: productions system, production area, and process group.

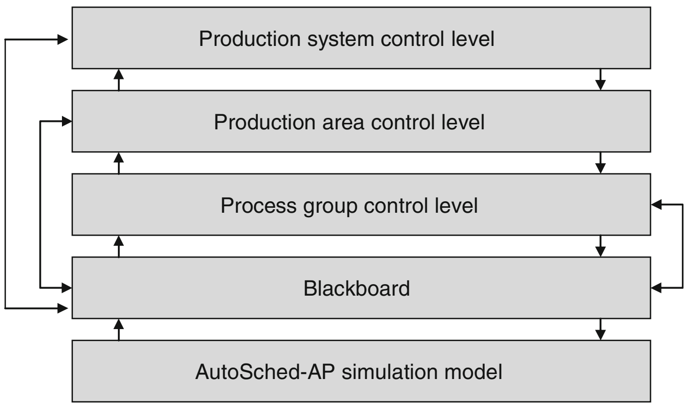

Fig. 2.17 High-level architecture of the FABMAS prototype (Translated from Mönch 2006)

The blackboard represents a data layer that is used by agents to exchange information. The bottom layer is a discrete-event simulation model used to test the system. All entities in the production system are modeled as agents (e.g. product agents, batch agents, service agents, decision maker agents, and so on). The three planning levels are supplemented with heuristics that resort to central information. The production system control level uses for instance a beam-search-algorithm, the production area control level a distributed-shifting-bottleneck-heuristic, and the process group control level relies on a machine hours-based resource allocation algorithm. These algorithms are also executed by designated agents. Production plans evolve through the performed optimizations and interactions among agents (Mönch 2006).

Agent-based approaches are promising due to their decentral, flexible, and complexity reducing character. However, up to today, they are embedded in systems following the push logic. It still needs to be investigated how they perform facing the typical pitfalls of push/MRP systems in practice. An integration with pull systems, for instance by using an agent-based approach for scheduling, could be worth investigating.

Comprehensive surveys of earlier PCS design literature are given by Geraghty and Heavey (2005), or Benton and Shin (1998).

2.2.3 PCS Selection

The following section focuses on publications that aims at deriving information on the relative performance of PCS, thus leading to drivers and decision rules for PCS selection. In the presented studies, usually two or more PCS are placed into a hypothetical production system and their performance is evaluated according to one or more metrics. In some cases, also the influence of environmental factors (e.g. presence of emergency orders) and characteristics of the production system are analyzed. Most studies argue on a quantitative basis by modeling the system with the help of Markov Decision Processes (MDPs), Petri nets, or discrete-event simulation models. Even though also qualitative studies (e.g. Razmi et al. 1996; Razmi et al. 1998) exist, the focus will be on quantitative studies here.

In the following, an analysis of 21 recent PCS comparison studies identified during a literature review is provided. Further surveys of PCS comparison literature are contained in Geraghty and Heavey (2005) and Benton and Shin (1998). The following dimensions are considered and represent the column headings in Table 2.2:

- *Compared PCS,* denotes the solution space, lists the PCS considered in the study
- *Factor variation,* denotes the factors whose impact on the PCS performance and decision were analyzed
- *Performance comparison,* describes how the performance comparison is made (metrics, approach, production setting)
- *Conclusion,* summarizes the conclusions drawn from the study

Table 2.2 Analysis of PCS selection studies

Reference	Compared PCS	Factor variation	Performance comparison			Conclusion
			Metrics	Approach	Production setting	
Bonvik and Gershwin (1996)	Kanban, Minimal Blocking, Basestock, Kanban/CONWIP	N/A	WIP, delivery performance	Discrete-event simulation (C++ implementation)	Serial production line, single part type, four stations	Hybrid dominated followed by CONWIP and then basestock
Geraghty and Heavey (2005)	Kanban/CONWIP, Kanban Hybrid (horizontal), Basestock, GKCS, EKCS	N/A	WIP, delivery performance	Discrete-event simulation (EM Plant)	Five stage parallel/serial line, one product produced out of two subassemblies, random process failures	Kanban consistently the worst performer, Kanban/hybrid consistently best performer
Grosfeld-Nir et al. (2000)	Push, Pull	Uncertainty in processing times, number of stations	WIP, throughput	Discrete-event simulation (SIMAN)	Serial production with 1–20 stations and stochastic processing times	If the number of stations is larger than seven, push dominates else pull
Gstettner and Kuhn (1996)	Kanban, CONWIP	N/A	WIP, throughput	Discrete-event simulation	One product serial flow line, exponentially distributed processing times	Kanban dominates
Hoshino (1996)	Push, Pull	Demand variation, variation of forecast error	WIP	Analytical approach	Single process step examined	Push dominates if variation of forecast error is small relative to variation of demand
Huang et al. (1998)	MRP, Kanban, CONWIP	N/A	WIP, throughput, raw material consumption rate, machine utilization	Discrete-event simulation	Cold rolling plant, six stages, convergent and divergent flows	CONWIP dominates
Kilsun et al. (2002)	Push, Pull (Kanban)	Demand variation, emergency orders	Cost (WIP holding and setup cost)	Discrete-event simulation (WITNESS 7.0)	Two stages serial production, ten products, only demand uncertainty	With low demand variation and no emergency orders push dominates, else pull dominates
Kleijnen and Gaury (2003)	Kanban, CONWIP, Kanban/CONWIP, Generic Kanban (see Gaury et al. 2000)	N/A	WIP, short-term delivery performance	Discrete-event simulation, Monte Carlo simulation	Serial production line, single product, four stations	Hybrid system dominates

(continued)

Table 2.2 (continued)

Reference	Compared PCS	Factor variation	Performance comparison			Conclusion
			Metrics	Approach	Production setting	
Koh and Bulfin (2004)	CONWIP, DBR (Drum-buffer-rope, horizontally integrated hybrid system with junction point at bottleneck)	N/A	WIP, throughput	Markov process model	Serial production line, three stations, unbalanced, exponential processing times	DBR dominates
Lee and Lee (2003)	Push, Pull	N/A	WIP, throughput	Discrete-event simulation	TFT-LCD production facility	Pull operates with much less WIP at the price of a minor decrease in production
Ozbayrak et al. (2004)	Push, Pull (Kanban)	N/A	Activity based costing	Discrete-event simulation (SIMAN)	MTO system (6 month demand freeze), six cells, 35 part types for ten products, scrap, rework, breakdowns	Pull dominates
Ozbayrak et al. (2006)	Push, Kanban, CONWIP	N/A	WIP, delivery performance, mean flow time, responsiveness	Discrete-event simulation	10 components assembled into one final product, different routing per component, processing times normally distributed	Depending on the metric, either push, Kanban or CONWIP dominated
Papadopoulou and Mousavit (2007)	Push, CONWIP	Dispatching rule (first-come-first-served, shortest-imminent-processing-time, earliest-due-date, work-content-in-the-queue-of-next-operation)	Average WIP, mean flow time, deviation from due date (earliness, tardiness), total time in queue	Agent-based simulation	Job shop, eight workstations, ten job types, revisiting of processes possible, batch size of 10–50 units	CONWIP dominated push under most dispatching rules and in most metrics
Persentili and Alptekin (2000)	Push, Pull	Product flexibility	WIP, throughput, average flow time, backorder level	not specified	Two products, five stages, convergent and divergent material flows	No clear dominance found
Razmi et al. (1996) and (1998)	Push, Pull (Kanban), hybrid (not specified further)	Variety of cost, flexibility and market factors	Utility function	Analytical hierarchical process	N/A	N/A

(continued)

Table 2.2 (continued)

Reference	Compared PCS	Factor variation	Performance comparison			Conclusion
			Metrics	Approach	Production setting	
Sarker and Fitzsimmons (1989)	Push, Pull (Kanban)	Cycle time variation	WIP, throughput	Discrete-event simulation	Three stations, serial flow, one product, exponentially distributes process times	The higher the coefficient of variation of the processing time is, the better performs push compared to pull
Sharma and Agrawal (2009)	Kanban, CONWIP, Hybrid	Demand distribution	Utility function	Analytical hierarchical process	Multistage serial production, single part	Kanban dominates
Tsubone et al. (1999)	Push, Pull (Kanban)	Breakdowns, processing time, flow sequence	Lead time	Discrete-event simulation	Serial flow, different jobs with different process sequences	Only small differences between push and pull, however push is more sensitive to internal variations
Wang and Xu (1997)	Push, Pull (Kanban), Hybrid (horizontal, each stage can either push or pull)	N/A	Inventory, delivery performance	Discrete-event simulation	Several production systems investigated from linear to an arbitrary convergent material flow, random demand and processing capacities	Hybrid dominates
Weitzman and Rabinowitz (2003)	Push, Pull (CONWIP)	Information updating rate, failure characteristics of machines	WIP, delivery performance	Discrete-event simulation	Serial flowshop, eight machines	Pull dominates, the worse the information updating the better performs pull

The objective of the analysis shown in Table 2.2 is to give an overview of the state of the art and to identify commonalities in the studies' conclusions, but also their weaknesses with regard to practical applicability.

The different conclusions drawn by the individual studies indicate that no per se dominant solution exists. The preferred approaches differ on a case by case basis and seem to be sensitive to changes in the assumptions or the production setup. However, it looks like that in most cases, especially if uncertainty is accounted for in the underlying models, pull-type or hybrid systems dominate pure MRP systems. Bonney et al. (1999) raise the question, whether the frequently argued superiority of pull-type systems compared to push-type systems is rather caused by their prerequisites, like small batch sizes, than by the different control of the material flow. Spearman and Zazanis (1992) suggest that the effectiveness of pull systems (i.e. less congestions) does not result from "pulling," but from the effect of limiting WIP and thus the variability of the amount of WIP. Moreover, they argue that in practice, pull systems are inherently easier to control since they focus on controlling WIP, other than push systems, which focus on controlling throughput, which cannot be visually observed like WIP. By examining the attributes in the columns of Table 2.2, further observations can be made.

Examining the applied factor variations, it can be concluded that the factor variation is seldom analyzed and thus the actual drivers for a decision remain unknown. However, this could be of interest especially for practitioners since they would know, which factors to monitor in order to get a signal to adapt the applied PCS to a constantly changing environment. The analyzed studies judge the PCS performance according to different metrics. WIP level and delivery performance are however frequently used metrics. To provide decision support in practice, the different objective functions and risk attitudes of the companies and their managers need to be taken into account. Trade-offs between metrics should be part of a utility function used for performance evaluation. Looking at the production systems in which the PCS are compared, the studies usually resort to very simplistic production setups with mostly serial flows, few products, and sometimes even deterministic parameters (e.g. cycle times, demands). Variability in the production system and within demand and forecasts is mostly addressed very rudimentary. Still, this helps to provide a directive value. Hence, the influence of complexity on the PCS decision could be explored in more depth.

An interesting approach is taken by Kleijnen and Gaury (2003). They apply a risk analysis-based approach in order to judge the robustness of the system. They combine discrete-event simulation, heuristic optimization, risk analysis, and bootstrapping and come up with a different ranking of the systems compared to a study (Gaury 2000), in which risk is not considered.

Geraghty and Heavey (2005) perform an exhaustive study in which they compare Kanban, Basestock, Generalized Kanban Control Strategy (GKCS), Extended Kanban Control Strategy (EKCS), and the horizontally integrated hybrid PCS presented in Hodgson and Wang (1991A), which equals the Kanban/CONWIP advanced pull system according to Geraghty and Heavey (2004). They conclude that Kanban/CONWIP consistently dominates all other PCS, and that Kanban consistently performs the worst. The dominance compared to Kanban is significant, the dominance compared to the other approaches is rather small, especially in the case of low system utilization. It is argued that the dominance of Kanban/CONWIP and CONWIP is due to the direct communication of demand information to the first stage that is inherent the CONWIP system and not done by pure Kanban, where the demand information is stepwise propagated upstream the material flow. However, the positive side of this information delay is not pointed out. It can also be interpreted as option to decrease the customer lead time and react to demand uncertainty since inventory of finished goods or semi-finished that are not yet dedicated to customer orders are held. A similar performance like in the Kanban/CONWIP system is shown by the EKCS. The main difference is that through its base stock levels, the EKCS tends to hold more of the inventory within the line whereas the Kanban/CONWIP systems tends to hold more inventory as finished goods. The details of these studies are presented in Geraghty (2003).

2.2.4 PCS Implementation

After choosing and customizing an appropriate production control strategy in theory, the next challenge is to implement it sustainably on the shopfloor. Therefore, several aspects need to be addressed appropriately to enable a sustainable implementation. They comprehend suitable

- Hardware
- IT systems
- Organization and processes
- Mindsets of involved employees

Only if all aspects above are addressed and potential industry specific challenges are solved, a sustainable implementation is possible. The available literature in this field can be categorized into publications that treat the implementation of pull-systems, publications dealing with the implementation of push systems, and publications dealing with hybrid systems.

Implementation studies in the first category, pull systems, focus mainly on the Kanban system. They are found in practice-oriented Lean Manufacturing publications like Womack et al. (2007), Drew et al. (2004), Womack and Jones (2003), or Ohno (1988). They treat the problem as part of holistic Lean transformations. Lee-Mortimer (2008) provides an implementation study of a Kanban system at an electronic manufacturer and stresses the aspects of the IT implementation, but also cultural aspects. Schwarzendahl (1996) describes the introduction of Kanban in a Siemens plant producing digital electronic switching equipment. One key challenge encountered in almost every implementation of a pull system is how to integrate the new control logic in the existing MRP system of the company. Flapper et al. (1991) present a three-step solution approach for this issue. A comprehensive survey of further IT-focused implementation studies can be found in Benton and Shin (1998). Slomp et al. (2009)

demonstrate in a case study how CONWIP control can be implemented together with other Lean Manufacturing tools in high-variety/low-volume make-to-order environments.

In the second category, push systems, literature focuses mostly on IT implementations of MRP systems or APS. Here, a large body of literature exists in the area of business computing. It will not be surveyed in detail here. A good introduction to software implementations of MRP systems is given by Scheer (2007) and for APS by Stadtler and Kilger (2008).

Some publications also study the implementation of hybrid PCS, the third category. Wang and Chen (1996) suggest an IT implementation of a horizontally integrated PCS by integrating a pull mechanism into the master production scheduling of MRP II. Goncalves (2000) studies the horizontally integrated hybrid PCS used in Intel's microprocessor fabrication facilities with the help of a system dynamics model. His focus is determining the influence of endogenous customer demand. Therefore, he assumes that the inventory levels influence demand. He models the behavior of several stakeholders (e.g. planners, warehouse managers, marketing) and studies the company's resulting production and service levels. Thereby, also insights on the behavior of the system in presence of failure (e.g. stock outs) are gained. Besides classical simulation analysis, he applies eigenvector analysis and derives stabilizing policies. Another example for a hybrid implementation study comes from Xiong and Nyberg (2000) who describe the implementation of a horizontally integrated hybrid system in refineries. The implementation of a push/pull production system for a single product production system in the food industry is described in Claudio et al. (2007). Here, different strategies are applied for different customers, depending on whether the customer provides forecast information or not.

2.2.5 Related Design Questions

Scheduling, lot-sizing, and inventory control are further design questions from the field of production planning and control, which are closely linked to the development of PCS. To conclude the literature review, they will be briefly addressed.

Scheduling is a discipline mainly addressed within Operations Research and computer science. "It deals with the allocation of resources to tasks over given time periods and its goal is to optimize one or more objectives" (Pinedo 2008). In the context of PCS, scheduling determines, for a process and its preceding buffer, which job to process next. It addresses thus an integral design question of every PCS and may influence its operational performance.

It is distinguished between single machine or multi-machine problems, in which schedules for several parallel processes are derived. Moreover, it is differentiated between flow-shop and job-shop problems. In a flow-shop, all goods flow along the same route through production (e.g. steel milling). In a job shop, goods follow, usually in batches, different routes (Pinedo 2008).

In practice, either no rules or standards exist, or very simple dispatching rules are applied. Common dispatching rules are First-In-First-Out (FIFO), Earliest Due Date (EDD), Shortest-Processing-Time (SPT), Shortest-Remaining-Processing-Time (SRPT), Longest-Processing-Time (LPT), Longest-Remaining-Processing-Time (LRPT) (Günther and Tempelmeier 2012; Framinan et al. 2000). An investigation of the impact of different dispatching rules on a CONWIP flow-shop is presented in Framinan et al. (2000). In their experiment, SPT outperformed FIFO and SRPT. Also Vinod and Sridharan (2009) present a simulation analysis of different scheduling rules. They investigate them in a job shop and consider sequence dependent setup times. They find the setup oriented version of SPT to perform best.

Another prominent approach to scheduling is earliness-tardiness scheduling. Here, the objective is to dispatch the orders in a sequence such that the deviation from the planned completion date is minimized (Baker and Scudder 1990). Lately, also optimization approaches from the area of artificial intelligence became popular in solving scheduling problems. Jensen (2001) presents an approach to solve stochastic scheduling problems in order to create robust and flexible systems with the help of evolutionary algorithms. Framinan et al. (2001) present a performance comparison of different heuristics. A high-level summary of the field can be found in Stuber (1998). However, also the argument exists that more complex scheduling rules, or the combination of simple scheduling rules, cannot significantly improve the performance of a production system, but rather generates a vicious circle by continuously reprioritizing production orders (Sabuncuoglu and Bayiz 2000; Stuber 1998; Kosturiak and Gregor 1995). For a general introduction to scheduling, the reader is referred to Pinedo (2008), Brucker (2007), or Blazewicz et al. (2001).

A second PCS design related area is lot-sizing. However, it will not be treated in detail here. The interested reader is referred to Günther and Tempelmeier (2012) or Hopp and Spearman (2008). Other than for instance Li and Liu (2006), this work does not integrate lot sizing in the optimization model but assume that optimal batch sizes are in place.

A third area closely related to PCS design questions is inventory control. "Inventory control determines which quantity of a product should be ordered when to achieve some objective, such as minimizing cost" (Kleinau and Thonemann 2004). Inventory control problems can be categorized depending on whether their demand is stationary, and whether it is fully observed (Treharne and Sox 2002). Inventory control can be delimited from the field of production control by the scope. Inventory control focuses on one single inventory, often of finished goods. In contrast, the scope of production control is broader. Production control looks at the whole production system and is in addition concerned with question where to set up inventories and the overall material and information flow in the production system. However, the terms are not consistently used in literature. A good survey on inventory control approaches is provided by Babai et al. (2009).

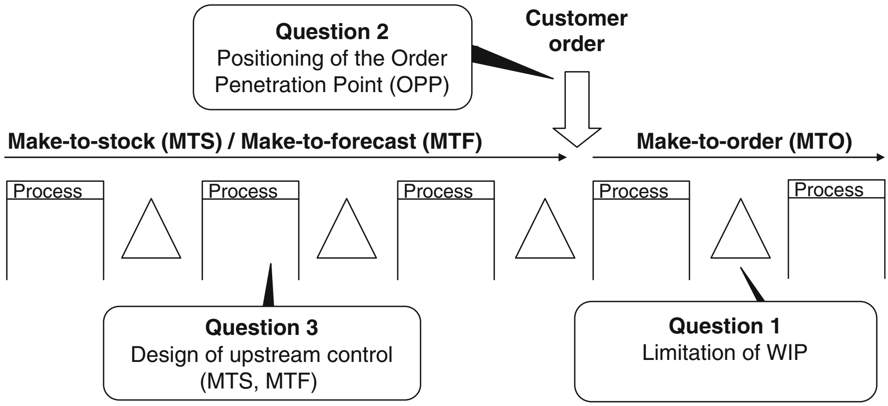

Fig. 2.18 Key questions of PCS engineering in practice

2.3 Synthesis and Positioning of the Following Work

After understanding basic concepts of PCS, their influence on operational performance, and current research in the field, now conclusions will be drawn and the following work be positioned.

The analysis of the PCS selection literature (Sect. 2.2.3) with regard to the creation of a PCS engineering framework shows that currently, a systematic and comprehensive guide to engineer a PCS for a complex discrete real world production system is missing. This need has also been recognized by Lödding (2008) who surveys different PCS but notes that further research is necessary to decide among them. Existing comparison studies with strong assumptions and limited solution spaces can provide directions, but are not able to provide concrete answers for complex real world production systems. The following PCS engineering framework seeks to contribute to closing this gap. Thereby, it is essential to provide an integrated answer to the three most important questions of PCS engineering as asked by industrial practice. Inspired by a Lean Manufacturing view on PCS, they can be structured as shown in Fig. 2.18 (Rother and Shook 1998).

On the first question, a large amount of literature has been published as outlined in the review of PCS method development literature (Sect. 2.2.2.3). Most papers conclude that limiting WIP in the buffers is beneficial, in other words that pull systems dominate pure push systems, which has also been confirmed by a large number of the analyzed PCS selection studies (Sect. 2.2.3). And even if this would not be true, the case of unlimited buffers can be emulated by setting a sufficiently high WIP cap. For determining how much WIP is needed, the operating curve concept of Nyhuis and Wiendahl (1999) provides a practically applicable approach, which is conform with the Lean Manufacturing philosophy and is closely related to

Little's Law.[4] The physics and strategic considerations behind the second question can be found in Olhager (2003). The basic idea is to move the OPP as far upstream as permitted by the customer lead time. For the third question, usually Lean Manufacturing would recommend building a supermarket-based MTS system (Rother and Shook 1998), whereas MRP-based systems would resort to an MTF strategy (Orlicky 1975).

It is aimed at developing a PCS engineering framework that provides an integrated answer to all three questions. As it is obvious from the literature review, there is no one fits all solution and the choice of the PCS and its parameters is strongly influenced by its operating environment (Sect. 2.2.3). Therefore, using a generic PCS model, like the EKCS or the PAC system that can emulate different relevant PCS, seems to be the most promising approach. Hence, a queuing network based meta-model that provides, by optimization of its parameters, an answer to the three questions formulated above is developed. It should be described in a simple and intuitive way and the essence of the PCS that have proven successful in the past should be synthesized into it. In contrast to many existing PCS selection studies, it is tried to gain a holistic systems engineering perspective on the design drivers. Moreover, through explicitly puzzling out complexity reduction techniques, the approach should be able to handle the complexity of real life production systems. Emphasis is put on selecting the right objective function. Relevant operational metrics and trade-offs among them should be taken into account from a production management point of view. The objective function should thus be oriented at the analysis of the influences of PCS on operational performance as developed in Sect. 2.1.4.

The focus is then put on the third question. Here, a new perspective is provided by introducing a hybrid MTF/MTS approach, which is enabled by the proposed generic model. The approach will be combined with limited buffers between processes. It will be shown that under certain conditions, this approach is able to significantly outperform the pure strategies. Thereby, other than in existing studies that incorporate advance demand information (ADI), forecast error is explicitly addressed as part of the dynamic complexity in the PCS design.

From a methodological standpoint, the following work resorts to systems engineering (Sage and Rouse 1999), Operations Research (Neumann and Morlock 2002), decision analysis (Howard 1983), and simulation (Law and Kelton 2008). The relevant methods will be introduced when they are needed within the following chapters.

[4] For an introduction to Little's Law see Stocker and Waldmann (2003)

Springer

http://www.springer.com/978-3-642-24141-3

Engineering Production Control Strategies
A Guide to Tailor Strategies that Unite the Merits of
Push and Pull
Karrer, C.
2012, X, 178 p., Hardcover
ISBN: 978-3-642-24141-3