Integrating Release and Dispatch Policies in Production Models Based on Clearing Functions

Dieter Armbruster, Michael Herty, Xinping Wang and Lindu Zhao

Preprint No. 407 **August 2014**

Key words: production planning, dispatch control, partial differential equations, reentrant production

AMS Subject Classifications: 35L60

Institut für Geometrie und Praktische Mathematik
RWTH Aachen

Templergraben 55, D-52056 Aachen, (Germany)

Dieter Armbruster is with the School of Mathematical and Statistical Sciences at Arizona State University, Tempe, AZ, USA; email: armbruster@asu.edu

Michael Herty is with the RWTH Aachen University, Aachen, Germany

Xinping Wang and Lindu Zhao are with Southeast University, Nanjing, PRC

Integrating release and dispatch policies in production models based on clearing functions

Dieter Armbruster, Michael Herty, Xinping Wang, and Lindu Zhao

***Abstract*—Aggregate production planning for highly re–entrant production processes is typically generated by finding optimal release rates based on clearing function models. For production processes with very long cycle times, like in semiconductor production, dispatch policies are used to cover short term fluctuations. We extend the concept of a clearing function to allow control over both, the release rates and priority allocations in re-entrant production. This approach is used to improve the production planning problem using combined release and the allocation dispatch policy. The control parameter for priority allocation, called the push-pull point (PPP), separates the beginning of the factory which employs a push policy from the end of the factory, which uses a pull policy. The extended clearing function model describes the output of the factory as a function of the work in progress (wip) and the position of the PPP. Its qualitative behavior is analyzed. Numerical optimization results are compared to production planning based only on releases. It is found that controlling the PPP significantly reduces the average wip in the system and hence leads to much shorter cycle times.**

Note to Practitioners: ***Abstract*—This study is focussed on the semiconductor production industry where multiple passes through the same machines (re-entrant production) are common. We show that changing priority rules is essentially a re-allocation of spare production capacity which may lead to significantly reduced lead times. In addition, a control scheme that determines the changes in the priority rules is shown to lead to smaller safety stocks for large demand fluctuations compared to a static priority rule.**

Primary and Secondary Keywords ***Index Terms*—Production planning, dispatch control, partial differential equations, re-entrant production.**

I. Introduction

In high-technology capital-intensive industries such as semiconductor manufacturing, many machines are repeatedly used for similar processing operations resulting in a re–entrant product flow. In such a re-entrant production line, semiconductor wafers return to the same set of machines many times. The associated cycle times are typically on the order of months. However, the demand fluctuates on a smaller timescale (typically weeks), strongly impacting the problem of production planning and generating the need for large safety stock in inventories. Besides the possibility to vary the inflow of the factory the only other option to influence the output of the factory is via dispatch policies. Specifically, re-entrant production creates the opportunity to set priority rules for the various stages of production competing for capacity at the same machines. This dispatch policy is also called priority allocation and typically allows for two modes of operation. A push policy, also known as first buffer first served policy, gives priority to earlier production steps over later production steps, and is typically assigned to the front of the factory. A pull policy, also known as shortest-expected-remaining-process-time policy, gives priority to later production steps over earlier production steps and is used at the end of the factory. The step where push policy switches to pull policy is called the push-pull point (PPP). Moving the PPP is a change in dispatch rules that can in principle be done instantaneously without affecting the work in progress (wip) in the factory at that moment. However, it has significant short term effects on the wip as well as the total output.

Push and pull policies are a special cases of dispatch policies which are widely used methods for shop-control problems [20], specifically also in the semiconductor industry [7], [22]. We refer to [25], [11] for survey articles. Typically dispatch policies are First-In, First-Out (FIFO), Shortest Processing Time (SPT), Earliest Due Date (EDD), Shortest Remaining Processing Time (SRPT), Least Slack (LS), Least Setup Cost (LSC), or combinations thereof [27], [18]. Usually a dispatch policy is fixed over the time interval of interest. However, some researchers consider dynamic characteristics of the shop-floor system and develop heuristic dispatching policies following the change of shop-floor systems [24], [26], [28], [14], [19]. A PPP dispatch policy is a highly simplified version of policies used in practice at INTEL to deal with short term fluctuations of the demand. A simulation study based on a discrete event simulation model of a semiconductor factory [23] showed that a heuristic PPP dispatching policy, coupled with a CONWIP starts policy, significantly reduces the difference between production output and demand for high demand with high variance.

The production planning problem, i.e. to generate a specific output within a specific time frame, has recently been summarized for example in the Operations Research Tutorial [4]. In general, it refers to two closely related problems: The forward problem determines the production rate depending on the current wip and a given release signal. The backward problem determines the releases required to produce a desired output pattern over time. The simplest approach involves deterministic linear programming (LP) models based on discrete time periods. These models divide the planning horizon into discrete time buckets, and determine the output in each time bucket under a set of constraints representing system capacity and dynamics at an aggregate level [13]. More

Dieter Armbruster is with the School of Mathematical and Statistical Sciences at Arizona State University, Tempe, AZ, USA; email: armbruster@asu.edu
Michael Herty is with the RWTH Aachen University, Aachen, Germany .
Xinping Wang and Lindu Zhao are with Southeast University, Nanjing, PRC .

sophisticated approaches involved the concept of a clearing function [16]. Fundamentally, a clearing function describes the expected aggregate output in a planning period as nonlinear function of the wip. A general relation between cycle time and wip for multi-product production is described using (allocated) clearing functions in [6] and in [21] those functions are estimated from data. Further details can be found in [4].

In [4] it was also shown that a clearing function approach can be viewed as a discretization of a continuous flow model, describing the flow of products through a factory like a fluid flow. Introducing a variable x to describe the production stage $x \in [0,1]$ a continuous model based on the clearing function F may be written as

$$\partial_t \rho + \partial_x F = 0,$$

where ρ describes the product density at stage x and time t. Here, $x = 0$ refers to the point of raw material and $x = 1$ to the finished product. Equations of this type have been studied for example in [3], [9], [12], [1].

Refining a model proposed in [2] we derive an extended continuum model based on clearing functions to integrate the impact of the dispatch policies and in particular to allow for changes in the PPP. We follow the presentation in Asmundsson et al. [6] and consider the production planning problem as a nonlinear optimization problem. In section II we present a detailed derivation of the extended model, its numerical discretization is discussed in section III. Section IV discusses the control action while the optimization results for simultaneously optimal PPP dispatch policies and release policies are presented in section V.

II. Clearing function for re–entrant production line with push-pull dispatch policies.

Before introducing the model we give a definition of a push-pull dispatch policy.

Definition 2.1: Let $x = \xi$ be the PPP and $p(y) > p(z)$ define a higher priority for step y over step z. Then, for two parts at positions y and z competing for machine capacity,

- if $y > z > \xi$ then $p(y) > p(z)$. This is the pull regime of the factory.
- if $y < z < \xi$ then $p(y) > p(z)$. This is the push regime of the factory.
- if $z < \xi < y$ then $p(y) > p(z)$. Every stage in the pull regime has priority over every stage in the push regime.

A sketch of the factory under consideration including the different regimes is given in Figure 1.

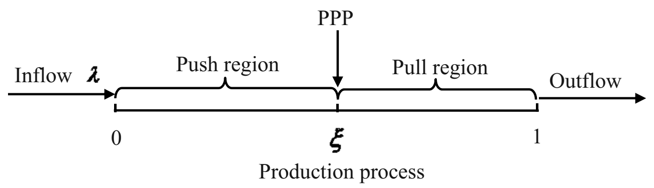

Fig. 1. Schematic view of a re–entrant production line with inflow λ and PPP being ξ.

We describe the factory using a clearing function approach using Kamarkar's clearing function [16], [15]

$$F^{ss} = \alpha \left(1 - \exp(-\beta W^{ss})\right). \tag{1}$$

Here W^{ss} is the average number of parts in the production unit reflecting the wip, F^{ss} is the corresponding average output measured in parts per minute and $\alpha, \beta > 0$ are two constant parameters: α is the maximal throughput (the capacity of the factory) and $\alpha\beta$ is the the slope of $F^{ss}(W^{ss})$ at $W^{ss} = 0$, i.e. the marginal increase of the throughput for small wip. Note that the clearing function is a relationship between steady state quantities, i.e. for a stationary start rate $F^{ss} = \lambda$, equation (1) describes the relationship between steady state output and steady state wip. Inverting the clearing function gives us the steady–state wip W^{ss} as a function for a constant start rate of $\lambda < \alpha$,

$$W^{ss}(\lambda) = \frac{1}{\beta} \log(\frac{\alpha}{\alpha - \lambda}). \tag{2}$$

By Little's law the mean cycle time is given as $\tau = \frac{F^{ss}}{W^{ss}}$. Defining a velocity $v = \frac{1}{\tau}$, mass conservation on the average is reflected in the relation for an equilibrium output.

$$F = Wv. \tag{3}$$

¿From equation (3) and (2) we obtain the steady–state velocity $v(\lambda) = \beta\lambda \frac{1}{\log(\frac{\alpha}{\alpha-\lambda})}$. Note that this velocity is constant in x since the wip is an integral quantity.

In order to obtain a stage–dependent model we need to specify the relationship between the clearing function for a full factory and a clearing function for a factory that has *length* or size y. The most simple modeling assumption is that in steady–state the function $y \to F^{ss}(W^{ss}(y))$ is linear in y. This leads to $F^{ss}(W^{ss}(y)) = \lambda y$ or equivalently, still assuming Karmarkar's functional form of the clearing function,

$$W^{ss}(y) = \frac{1}{\beta} \log(\frac{\alpha}{\alpha - \lambda y}). \tag{4}$$

Derivation of a clearing function for a pull policy

Since we are interested to address the changes in the output that are generated by changing dispatch policies the production rate at position x depends on its distance from the PPP ξ. Hence in the pull regime, parts at a position $x > \xi$ always have priority over parts at a position $y < x$ and hence wip upstream from x does not matter. Therefore the relevant wip becomes the wip downstream from a position x. We therefore define the wip in the pull regime as

$$W^{pull} = W^{pull}(x) = \int_x^1 \rho(x)dx. \tag{5}$$

In equation (5) $\rho(x)$ is the product density at stage x. Since priority affects the speed of moving through the production line, the velocity $v(x)$, even in equilibrium, is not constant any more. Hence, the fundamental hydrodynamic relationship between flux (here the production rate) and velocity becomes

$$F = \rho v \tag{6}$$

and thus *replacing* $F = Wv$. Rewriting equation(5) by $\rho(x) = -\frac{d}{dx} W^{pull}(x)$ and using the steady state wip equation

(4) with $W^{ss}(1-x)$ for the wip in the pull regime, we get the wip profile in the pull regime in terms of the steady state product density $\rho^{pull}(x)$ as

$$\rho^{pull}(x) = \frac{\lambda}{\beta\left(\alpha - \lambda(1-x)\right)}. \tag{7}$$

Hence for a constant start rate of λ, and the definition of a flux as in equation (6) we obtain the steady state velocity in the pull region

$$v^{pull}(x) = \beta(\alpha - \lambda(1-x)).$$

Derivation of a clearing function for a push policy

It has been shown in [23] that the cycle time of a production line, run completely in push mode, is higher than the cycle time of a production line run completely in pull mode. Specifically, given the same wip, the output and hence the velocity of a factory in push mode is lower than a factory in pull mode. Assuming a PPP at ξ, the velocity at $x = 0$ can be written as

$$v^{push}(x=0) = K \min_{\xi \le x \le 1} v^{pull}(x) = K\beta(\alpha - \lambda(1-\xi)).$$

with $K < 1$. With the same arguments as for the pull regime we then get a spatially dependent equilibrium velocity and equilibrium density in the push regime, i.e. for $0 < x < \xi$ of the form:

$$\begin{aligned} v^{push}(x) &= K\beta(\alpha - \lambda(1-\xi) - \lambda x), \\ \rho^{push}(x) &= \frac{\lambda}{v^{push}(x)} = \frac{\lambda}{K\beta(\alpha - \lambda(1-\xi) - \lambda x)}, \\ W^{push}(x) &= \frac{1}{K\beta}\log\left(\frac{\alpha - \lambda(1-\xi)}{\alpha - \lambda(1-\xi) - \lambda x}\right). \end{aligned} \tag{8}$$

Note that in steady–state

$$W^{pull}(x) = \frac{1}{\beta}\log\left(\frac{v^{pull}(1)}{v^{pull}(x)}\right), \tag{9a}$$

$$W^{push}(x) = \frac{1}{K\beta}\log\left(\frac{v^{push}(0)}{v^{push}(x)}\right), \tag{9b}$$

$$v^{pull}(1) = \alpha\beta, \tag{9c}$$

$$v^{push}(0) = K\beta(\alpha - \lambda(t)(1-\xi)). \tag{9d}$$

Summarizing, the time–dependent model for the part density $\rho(t,x)$ at time t and stage $x \in [0,1]$ including a PPP at position $\xi \in [0,1]$ and using the Heaviside function $H(x)$ reads

$$\begin{aligned} 0 &= \partial_t \rho(t,x) + \partial_x F(\rho, x, \xi) \qquad (10) \\ F(\rho,x,\xi) &= \rho(t,x) v(\rho,x,\xi), \\ v(\rho,x,\xi) &= H(x-\xi) v^{pull}(1)\exp(-\beta W^{pull}(x)) \\ &+ H(\xi - x) v^{push}(0)\exp(-K\beta W^{push}(x)), \\ W^{pull}(x) &= \int_x^1 \rho(t,y)dy, \;\; W^{push}(x) = \int_0^x \rho(t,y)dy. \end{aligned}$$

with boundary conditions

$$\rho(t,x)v(\rho,x,\xi)|_{x=0} = \rho(t,0)v^{push}(0) = \lambda(t). \tag{11}$$

Note that the flux (production rate) at the position x in the factory, $F(\rho,x,\xi)$, depends explicitly on the specific PPP ξ and via the functional dependence on the density on position x and on time t. A related model has been studied theoretically in [2].

A. Steady–states

By definition, in steady state the flux (production rate) is constant along the production line even across the PP and hence equal to start rate and to the output. However, since the priorities change discontinuously at the PPP, the velocities change discontinuously and hence the steady state wip-profile does too. Figure 2 plots a typical wip-profile given by equations (7) and (8).

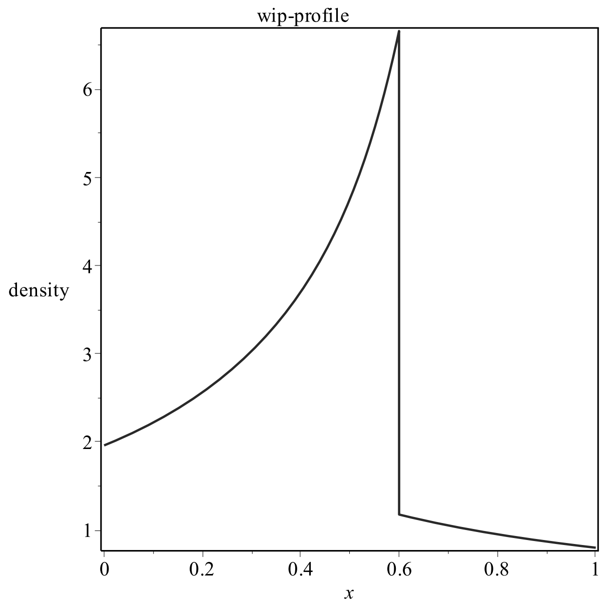

Fig. 2. wip in push and pull region for the parameters $K = 0.6, \alpha = 1, \beta = 1, \lambda = 0.8$ and a PPP at $\xi = 0.6$.

The total wip in steady state is given as

$$W = \int_0^1 \rho(x)dx = W^{pull}(\xi) + W^{push}(\xi). \tag{12}$$

Using equation (9) this evaluates to $W = -\frac{1}{\beta}\log(\frac{\lambda\xi+\alpha-\lambda}{\alpha}) + \frac{1}{K\beta}\log(\frac{\lambda\xi+\alpha-\lambda}{\alpha-\lambda})$ and $\frac{d}{d\xi}W = \lambda\frac{1-K}{K\beta(\lambda\xi+\alpha-\lambda)} > 0$. Therefore, the minimal total wip W is attained for $\xi = 0$. This reflects a pure pull policy and is consistent with the fact that a pure pull policy leads to the shortest cycle time and the lowest wip in the factory. A graph of W as function of ξ is given in Figure 3.

III. A BUCKET MODEL OF THE CLEARING FUNCTION

An analytical solution for the partial differential equation (PDE) (10) is in general not possible to obtain. Also, an accurate numerical solution of equation (10) requires a full discretization of stage and time of the partial differential equation using possibly a large number of discretization points. However, when the solution of an optimization problem requires the iteration of solutions of the PDE a simplified model is necessary to solve the optimization problem with finite

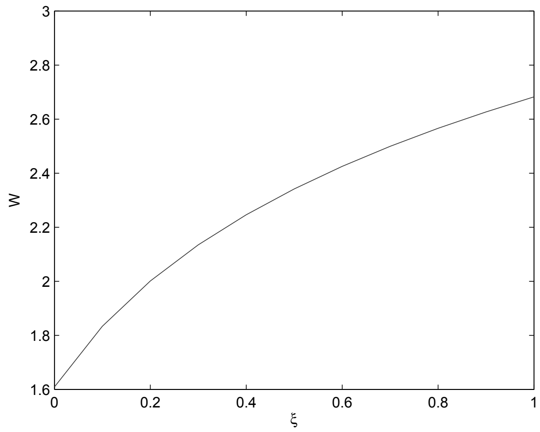

Fig. 3. Total wip in steady state as a function of the PPP ξ for parameters $K = 0.6, \alpha = 1, \beta = 1, \lambda = 0.8$.

resources. Our model is motivated by finite–volume methods for equation (10). It uses only three spatial grid points that are naturally suggested by the model and is designed to be similar to the way clearing functions are typically used for example in [5].

We assume for the moment that $\xi \in (0,1)$ is fixed, dividing the interval $[0,1]$ into two cells $[0,\xi]$ and $[\xi,1]$, respectively. We denote by $\rho^{push}(t)$ and $\rho^{pull}(t)$ the corresponding cell averages of $\rho(x,t)$,

$$\rho^{push}(t) = \frac{1}{\xi}\int_0^{\xi} \rho(t,y)dy, \quad \rho^{pull}(t) = \frac{1}{1-\xi}\int_{\xi}^{1} \rho(t,y)dy. \tag{13}$$

Note that the following relations always hold by definition of W

$$W^{push}(t,\xi) = \xi\rho^{push}(t), \quad W^{pull}(t,\xi) = (1-\xi)\rho^{pull}(t). \tag{14}$$

The time–evolution of the cell averages $\rho^{push}(t)$ and $\rho^{pull}(t)$ are obtained by integration of equation (10) on the corresponding intervals and using the inflow condition (11).

$$\begin{aligned} \partial_t \rho^{push}(t) &= \tfrac{1}{\xi}\left(\lambda(t) - F(\rho,\xi^-,\xi)\right), \\ \partial_t \rho^{pull}(t) &= \tfrac{1}{1-\xi}\left(F(\rho,\xi^+,\xi) - F(\rho,1,\xi)\right), \end{aligned} \tag{15}$$

where $\xi^{\pm}$ indicates the right and left limit at $x = \xi$. The previous formulation is still exact. Within the flux (production rate) function, the velocity v only depends on the wip given by equation (14) but the flux F itself depends on the value of ρ at position $x = \xi$ and $x = 1$, respectively. In steady state, in order to conserve mass across $x = \xi$ we have to have continuity in the flux at $x = \xi$,

$$F(\rho,\xi^-,\xi) = F(\rho,\xi^+,\xi). \tag{16}$$

Since products flow across the boundary $x = \xi$ from the push to pull region, the production rate at $x = \xi$ is given by the output of the push region which is its production rate at this point. Hence, the fluxes in equation (15) depend on the point values of $\rho(x = \xi, t)$ and $\rho(x = 1, t)$, respectively.

In order to obtain a closed equation we need to express those values in terms of the cell averages $\rho^{push}(t)$ and $\rho^{pull}(t)$. This is known in finite volume schemes as the reconstruction of $\rho(t,x),\ \forall x \in (0,1)$. Typically, in a finite volume method, a piecewise constant reconstruction, i.e., $\rho(t,x) = \rho^{push}(t)(1 - H(x-\xi)) + \rho^{pull}(t)H(x-\xi)$ is used. However, as can be easily seen from Figure 2 the steady state wip profile is far from being piecewise constant. In particular, a piecewise constant reconstruction will not preserve the steady–state wip equation (12) and hence violates the clearing function model assumption of a relationship between steady state wip and steady state flux.

Therefore we reconstruct $\rho(t,x)$ based on the explicit formulas for the steady–states. Specifically, assuming a cell in the push region is in steady state, the values of the cell average ρ^{push} is determined by $\frac{1}{\xi}\int_0^{\xi}\rho_{push}(x)dx = \rho^{push}$ where $\rho_{push}(x)$ is given by the steady state equation (8). Hence by eliminating λ we can reconstruct $\rho(t,x=\xi)$ as a function of the cell average

$$\rho(t,x=\xi) = \frac{1}{K\beta\xi}\left(\exp(K\beta\xi\rho^{push}(t)) - 1\right). \tag{17}$$

Using $W^{push}(\xi)$ from the steady state relation equation (9b) and with $W^{push}(t,\xi) = \xi\rho^{push}(t)$ we find the flux $F(\rho,\xi^-,\xi)$ as

$$F(\rho,\xi^-,\xi) = \frac{v^{push}(0)}{K\beta\xi}\left(1 - \exp(-K\beta\xi\rho^{push}(t))\right). \tag{18}$$

Similarly, in the pull region $[\xi,1]$, using the steady state (7) and the steady state relation equation (9a) we reconstruct $\rho(t,x = 1)$ and the corresponding flux $F(\rho,1,\xi)$ as

$$\begin{aligned} \rho(t,x=1) &= \tfrac{\alpha}{\beta\alpha(1-\xi)}\left(1 - \exp(-\beta(1-\xi)\rho^{pull}(t))\right), \\ F(\rho,1,\xi) &= v^{pull}(1)\tfrac{\alpha}{\beta\alpha(1-\xi)}\left(1 - \exp(-\beta(1-\xi)\rho^{pull}(t))\right). \end{aligned} \tag{19}$$

Now, we have closed form of the model given by (15) together with (17) and (19). The only approximation applied is the reconstruction of the cell densities ρ using the cell averages ρ^{push} and ρ^{pull}. The presented reconstruction preserves the original steady states. This is summarized in the following Lemma.

Lemma. *The approximate model given by equation (15),(17),(18) and* (19) *are a closed system of equations for the evolution of the cell averages* $\rho^{push}(t)$ *and* $\rho^{pull}(t)$. *Its steady–state solution* $(\rho^{push}, \rho^{pull})$ *lead to wips in the push and the pull section given by equation* (14) *that coincide with the steady–state wip of the continuous model* (9) *evaluated at* $x = \xi$ *and* $x = 1$ *respectively.*

Note that by using the definitions of $v^{push}(0)$ and $v^{pull}(1)$ we may reformulate the fluxes as

$$\begin{aligned} F(\rho,\xi^-,\xi) &= F(\rho,\xi^+,\xi) \\ &= \frac{\alpha - \lambda(1-\xi)}{\xi}\left(1 - \exp(-K\beta\xi\rho^{push}(t))\right), \\ F(\rho,1,\xi) &= \frac{\alpha}{(1-\xi)}\left(1 - \exp(-\beta(1-\xi)\rho^{pull})\right). \end{aligned}$$

For the numerical treatment of equation (15) it is advantageous to consider the formulation in terms of $W^{push}(t,\xi)$

and $W^{pull}(t,\xi)$ given by (14) instead of the cell averages. Therefore, we multiply by ξ and $1-\xi$, use the previously computed flux and obtain the following equivalent set of equations.

$$\partial_t W^{push}(t,\xi) = \lambda(t)$$
$$-\tfrac{\alpha-\lambda(t)(1-\xi)}{\xi}\left(1-\exp(-K\beta W^{push}(t,\xi))\right),$$
$$\partial_t W^{pull}(t,\xi) = \tfrac{\alpha-\lambda(t)(1-\xi)}{\xi}\left(1-\exp(-K\beta W^{push}(t,\xi))\right)$$
$$-\tfrac{\alpha}{(1-\xi)}\left(1-\exp(-\beta W^{pull}(t,\xi)\right).$$

Next, we discretize the system in time to advance the push and pull wip from time t^n to time $t^n+\Delta t$ for some Δt. An explicit in time discretization would require to fulfill the CFL condition [8] i.e., $\Delta t \le \frac{\min\{\xi,1-\xi\}}{\max\{v^{pull}(1),v^{ps}(0)\}}$. In order to avoid possibly small time–steps we instead discretize equation (10) *implicit* in time with time step $\Delta t = 1$. We further introduce the following notation

$$\begin{aligned}
W_t^- &:= W^{push}(\xi,t^n),\ W_t^+ := W^{pull}(\xi,t^n),\\
R_t &:= \lambda(t^n),\\
Y_t &:= Y(t^n) \equiv \frac{\alpha-R_t(1-\xi)}{\xi}\left(1-\exp(-K\beta W_t^-)\right)\\
X_t &:= X(t^n) \equiv \frac{\alpha}{(1-\xi)}\left(1-\exp(-\beta W_t^+)\right).
\end{aligned}$$

Using the new notation the fully discretized model finally reads for $t=0,1,\dots,$

$$W_t^- = W_{t-1}^- + (R_t - Y_t)\,, \tag{20a}$$
$$Y_t = \frac{\alpha-R_t(1-\xi)}{\xi}\left(1-\exp(-K\beta W_t^-)\right) \tag{20b}$$
$$W_t^+ = W_{t-1}^+ + (Y_t - X_t), \tag{20c}$$
$$X_t = \frac{\alpha}{(1-\xi)}\left(1-\exp(-\beta W_t^+)\right). \tag{20d}$$

Comparing (20) with the model equation (10) which is continuous in space and time we observe that the complexity of the partial differential equation is reduced to a coupled set of four difference equations. Not coincidentally, these difference equations have the same structure as the mass balance equations set up in standard production planning models based on nonlinear clearing function models [5], [6].

IV. Moving the push–pull point

The model discussed in the previous section requires the PPP to be fixed over time. We are, however, interested to use a moving PPP as a controller for the output of the production system. A change in the position of the push–pull point, leads to additional waves traveling through the production line, see [2], [23], that can be tracked only by solving the fully time dependent PDE model. Instead, even though equation (20) is only derived from equation (10) for constant λ and ξ, we intend to use this model also in the case when λ and ξ change over time by approximating the impact of the change in the PPP through the differences in the total wip in the production line. The underlying assumption is that the system has returned to steady state much faster than the timescales that govern our control changes.

To determine the impact of such a control change, we assume the push–pull point changes within one time unit from $\xi_{t-1} = \overline{\xi}$ to $\xi_t = \xi$. The resulting steady–state wip-profile will be of the type shown in Figure 2 but with a discontinuity at ξ instead of $\overline{\xi}$. If λ (resp. R) is constant, the difference in the work–in–progress in the push and pull region induced by the different steady states can be computed as

$$W^- = \begin{pmatrix} \bar W^- \ \frac{\xi}{\bar\xi} & \xi \le \overline{\xi} \\ \bar W^- + \bar W^+\ \frac{\xi-\bar\xi}{1-\bar\xi} & \xi > \overline{\xi} \end{pmatrix}, \tag{21}$$

$$W^+ = \begin{pmatrix} \bar W^+ + \bar W^- \ \left(1-\frac{\xi}{\bar\xi}\right) & \xi < \overline{\xi} \\ \bar W^+\ \frac{1-\xi}{1-\bar\xi} & \xi \ge \overline{\xi} \end{pmatrix}, \tag{22}$$

where the wip before and after the jump is called $\bar W^\pm$ and $W^\pm$, respectively. Then, the complete model approximating the dynamics of (10) reads

$$\begin{aligned}
W_t^- &= [W_{t-1}^- - Y_t + R_t][\tfrac{\xi_t}{\xi_{t-1}}H(\xi_{t-1}-\xi_t) + H(\xi_t-\xi_{t-1})]\\
&+ [W_{t-1}^+ + Y_t - X_t]\tfrac{\xi_t-\xi_{t-1}}{1-\xi_{t-1}}H(\xi_t-\xi_{t-1}),\\
W_t^+ &= [W_{t-1}^- - Y_t + R_t](1-\tfrac{\xi_t}{\xi_{t-1}})H(\xi_{t-1}-\xi_t)\\
&+ [W_{t-1}^+ + Y_t - X_t][\tfrac{1-\xi_t}{1-\xi_{t-1}}H(\xi_t-\xi_{t-1}) + H(\xi_{t-1}-\xi_t)].
\end{aligned} \tag{23}$$

Here, X_t and Y_t are given as before in equation (20). We may further add an inventory to the model buffering the outgoing parts X_t and comparing with a given demand D_t. In discretized form the inventory evolution I_t over time reads

$$I_t = I_{t-1} + X_t - D_t. \tag{24}$$

Non-negativity constraints complete the model

$$0 < \xi_t < 1,\ 0 \le W_t^+,\ 0 \le W_t^-,\ 0 \le R_t,\ 0 \le I_t. \tag{25}$$

Contrary to [5] the final model is nonlinear in the dynamics. However, by relaxing the equality constraints on X_t and Y_t to inequalities

$$0 \le X_t \le \frac{\alpha}{(1-\xi_t)}\left(1-\exp(-\beta W_t^+)\right). \tag{26}$$

and

$$0 \le Y_t \le \frac{\alpha-R_t(1-\xi_t)}{\xi_t}\left(1-\exp(-K\beta W_t^-)\right), \tag{27}$$

we significantly reduce computational time. This is possible, since the flux is monotone in $W_t^\pm$.

The final production planing problem is therefore given by the nonlinear programming problem

$$\min_{R_t,\xi_t} \sum_t R_t + X_t + Y_t + W_t^+ + W_t^- + I_t \tag{28}$$

$$\text{subject to } (23), (24), (25), (26), \text{ and } (27). \tag{29}$$

Notice that the nonlinearity only enters through the movement of the PPP ξ_t and through the clearing function. If we fix $\xi = 0$ we obtain the model presented in [6] (using Karmarkar's clearing function) where it also has been shown that problem (28) can be solved efficiently. The variables and relations are depicted in Figure 4.

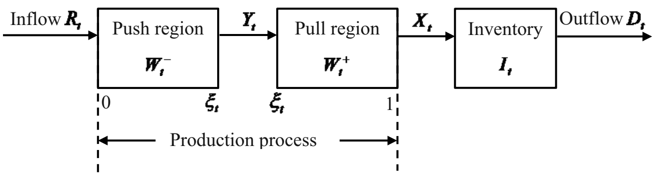

Fig. 4. Schematic view of the production planning problem equation (28) with inflow R_t and PPP ξ_t.

V. Numerical results

To understand the interplay between the dispatch policy (moving the PPP) and the starts policy we study the production planning model introduced in Section 4 numerically. Using $K = 0.6$, $\alpha = 1$ and $\beta = 1$ parametrizes a typical clearing function model which leads to a dependence of the total wip on the PPP ξ as depicted in Figure 3.

The nonlinear optimization problem (28) is solved using an interior–point method implemented in Matlab's *fmincon* function. Convergence for all scenarios shown below has been observed. Matlab's optimization method is used as black–box solver with no further modifications. The initial value for the optimization is set as the steady state computed in Section II-A and for $\xi = 0.01$.

We consider two scenarios: As **reference scenario** we optimize a model with the PPP set equal to zero. Then, the optimization problem (28) reduces to finding the optimal inflow over time R_t for $t = 1, \ldots, T$ which is an allocation problem for wip, inventory and inflow rate, given a desired demand D_t over a time horizon T. For a deterministic and constant demand, the optimal solution is the steady state with a pure pull policy and constant inflow rate.

Our **test scenario** involves solving the full optimization problem (28) for the unknown inflow R_t as well as the time–dependent PPP position ξ_t. Due to the relatively small number of unknowns in the model we do not observe any computational disadvantage when optimizing with variable PPP compared to the case of $\xi_t = 0$. Since the equations (28) are formally not defined for $\xi_t = 0$ we bound the possible values of ξ by $\xi \in [0.005, 0.995]$.

To average over the stochastic demand realizations we solve 3000 realizations of the optimization problem (28) for each fixed ϵ and each scenario and average to compute mean and variance of all variables. The following figures show the means and Table I and II present standard errors. We choose to optimize over a time horizon of $T = 20$ but restrict our figures to always show only the time interval $t \in [5, 15]$ to avoid initial or terminal transition layers.

Case of deterministic demand

For a steady state influx of $\lambda = 0.75$, a constant deterministic demand $D_t = \lambda$ and $\xi = 0.01$, based on Section II-A we obtain the steady–state wip of $W^- = 0.0493$ and $W^+ = 1.3567$, respectively with an optimal input $R_t = \lambda$, optimal inventory of $I_t = 0$ and optimal PPP at $\xi = 0.01$. Both scenarios converge to this solution. Figure 5 shows this for the optimization of our test and reference scenario.

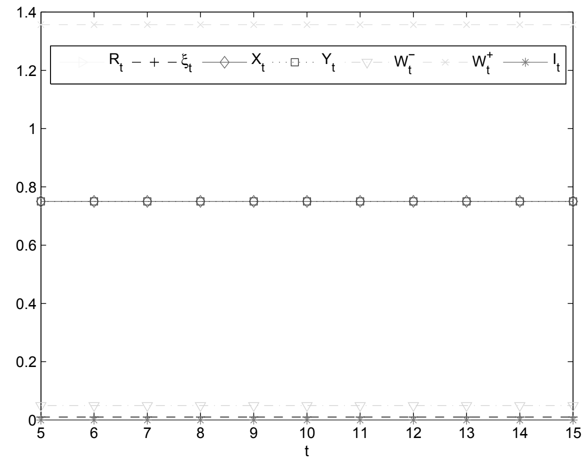

Fig. 5. Optimal values for different state variables under PPP policy with deterministic customer demand $D_t = 0.75$.

Case of stochastic demand

We define stochastic demands as $D_t = 0.75 + \epsilon Z_t$ where Z_t is a random variable with uniform rectangular distribution in $[-\frac{1}{2}, \frac{1}{2}]$ at every point in time and ϵ is a parameter that controls the variance of the demand. We vary $\epsilon \in (0.25, 1.5)$. This in particular implies that for $\epsilon = 1$ the demand might be as low as 0.25 and also possibly higher than the maximum production capacity of one. Note that for the given sets of parameters using Little's law, the cycle time in steady state for a pure pull policy is $\tau^{pull} = 1.87$ and for a pure push policy is $\tau^{push} = 3$. Since the demand variations occur on a timescale that is significantly shorter than the cycle time, the reference scenario will not be able to follow these fluctuations via changes in the inflow. As a result, the inventory increases approximately linearly with ϵ for the reference scenario (Figure 6).

For the test scenario, the variation in ξ creates additional variability that need to be covered by an increased inventory. Hence for small noise levels, the inventory for the moving PPP is higher than for the reference scenario. Only for noise levels $\epsilon > 1$ is the inventory of the reference scenario higher (Figure 6).

In Figure 7 we present the mean of all variables in the two scenarios for the noise level $\epsilon = 1$ as a function of time. Figure 7 (a) shows that for the reference scenario the average values of fluxes and wip converge well to the steady–state values. Here the optimal strategy is to buffer the variable demand via inventory and to not adjust the inflow.

Figure 7 (b) shows some residual variance for the average over 3000 runs in the test scenario. We find that on average, the PPP sits at $\xi = 0.4$. The major difference between the two scenarios is the wip-level. In the test scenario, the total wip on average is $W = W^+ + W^- \approx 0.65 + 0.1 = 0.75$ compared to $W \approx 1.4$ for the reference scenario.

Figure 8 shows that the reduction in wip is mostly due to a strong reduction in the wip in the pull region (W^+). The

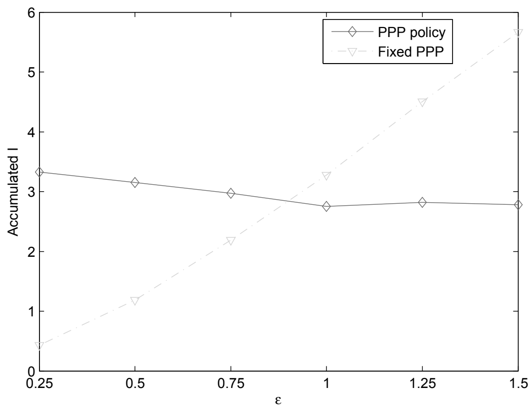

Fig. 6. Mean inventory levels as a function of the noise strength ϵ.

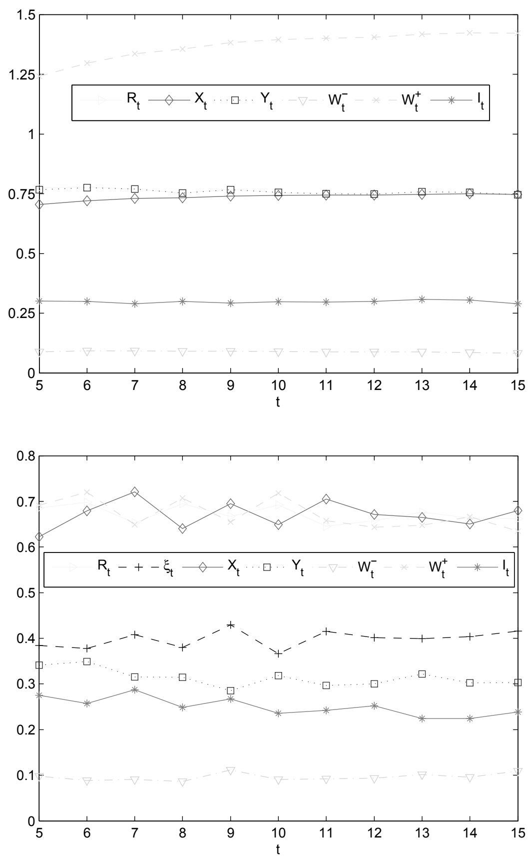

Fig. 7. Optimal values for different state variables under fixed PPP (left) and variable PPP (right) with stochastic demand $D_t = 0.75 + \epsilon\ Z_t$ where Z_t is a random variable with rectangular distribution in $[-\frac{1}{2}, \frac{1}{2}]$ and $\epsilon = 1$.

reduction is strong for all noise levels and increases slightly with increasing noise amplitude. By Little's law this wip reduction leads to a significant reduction in the cycle time: For a noise level $\epsilon = 1$, a total wip of $W \approx 0.75$ and an average influx of $\lambda = 0.75$, the cycle time $\tau \approx 1$ compared to, as discussed above, $\tau^{pull} = 1.87$ for a pure pull policy and $\tau^{push} = 3$ for a pure push policy.

There is one additional puzzling result in Figure 7(b): The

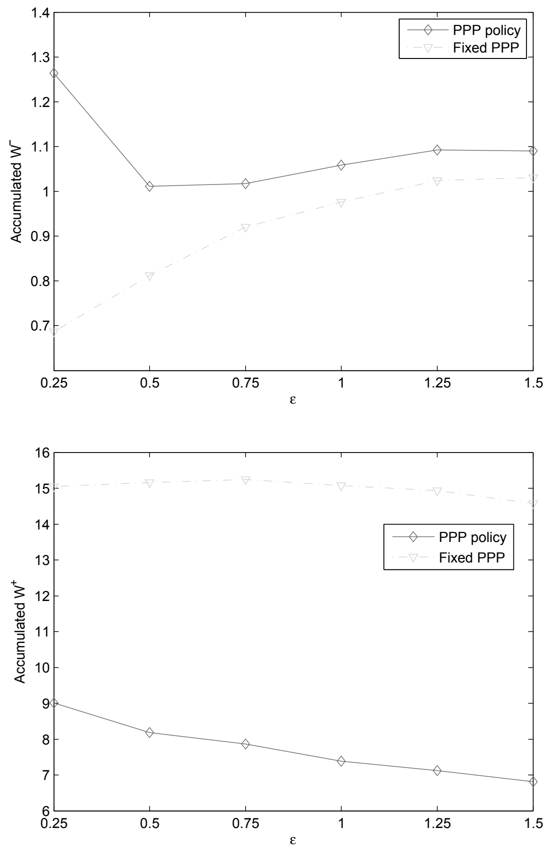

Fig. 8. Average accumulated wip in the push region (left), in the pull region (right) as a function of the noise level ϵ for the reference scenario (fixed PPP policy) and the test scenario (variable PPP policy).

average flux at the PPP (Y) is significantly less than the steady state flux. This seems to contradict mass conservation since clearly we have to have on average a total flux of $\lambda = 3/4$ everywhere. The reason why mass conservation is valid even for a lower mean flux at the PPP, is that the PPP is changing at every time step and on average there will be an equal number of jumps upstream and downstream. Looking at the wip-profile in Figure 2 we see that the gain in W^+ for a jump upstream is much larger than its loss in wip for a jump downstream. Hence the jump action of the PPP acts like a pump, moving wip downstream making up for the missing flux at the PPP.

Finally, in Table I and II we show the standard errors for some of the variables, accumulated for $t = 5 \ldots 15$. We observe that for the reference scenario the mismatch between demand and outflow has a standard error of less than 3% for all noise levels, indicating that all 3000 samples pretty much behave the same way. For the test scenario the standard error for the same variable lies between 6 and 9% with an outlying maximal value of 12% occurring for small noise $\epsilon = 0.25$. Similar observations with similar standard error values can be made for cumulative wips and inventories.

VI. Summary

Extending the standard model of approximating the production rate of a factory via a clearing function [16], we

ϵ	$\sum_{t=5}^{15} \lvert X_t - D_t \rvert$	$\sum_{t=5}^{15} W_t^-$	$\sum_{t=5}^{15} W_t^+$	$\sum_{t=5}^{15} I_t$
0.25	0.0033	0.0027	0.0224	0.0037
0.5	0.0075	0.0064	0.0498	0.013
0.75	0.0124	0.0112	0.0761	0.03
1.0	0.0177	0.0142	0.0964	0.0537
1.25	0.0233	0.0171	0.1194	0.0817
1.5	0.0294	0.0192	0.1366	0.1093

TABLE I
STANDARD ERROR, ϵ, OF SELECTED ACCUMULATED VARIABLES FOR DIFFERENT STOCHASTIC DEMANDS AND USING A FIXED PUSH–PULL POINT.

ϵ	$\sum_{t=5}^{15} \mid X_t - D_t \mid$	$\sum_{t=5}^{15} W_t^-$	$\sum_{t=5}^{15} W_t^+$	$\sum_{t=5}^{15} I_t$
0.25	0.1213	0.0527	0.1281	0.0734
0.5	0.074	0.0308	0.0877	0.0515
0.75	0.0659	0.0275	0.0615	0.0487
1.0	0.0673	0.0309	0.0688	0.0545
1.25	0.0788	0.0361	0.076	0.0726
1.5	0.0893	0.0375	0.0792	0.0746

TABLE II
STANDARD ERROR, ϵ, OF SELECTED ACCUMULATED VARIABLES FOR DIFFERENT STOCHASTIC DEMANDS FOR VARIABLE PUSH–PULL POINT.

have developed a clearing function model for a re-entrant factory that includes a push dispatch policy (first buffer first served policy) at the front of the factory and a pull policy (shortest-expected-remaining-process-time policy) at the end of the factory. The transition point between the two dispatch policies is called the PPP and is used as an additional control parameter besides the start rate into the factory to optimize the production planning problem.

For constant PPP we have developed a partial differential equation based on the notion of continuous transport that models the progress of wip through the factory under the two dispatch policies. The resulting PDE serves as the bases for a linearized approximation of the influence of jumps in the PPP.

The accurate production planning problem would require to optimize a cost function, under the constraint that wip and fluxes are related through the PDE for this model. Such an approach has been done for a uniform FIFO dispatch policy, optimizing the start rate, based on the use of adjoint calculus for constraint optimization problems [17]. In this paper, to keep things simple, we have discretized the PDE into two buckets, the push bucket and the pull bucket using a numerical finite volume scheme. While the resulting numerical errors are necessarily high due to the use of large step sizes, the scheme looks exactly like traditional discretizing of production planning as discussed in [4], [6].

We have studied the resulting nonlinear optimization problem for the time dependent influx and the time dependent PPP in a representative case, where the change in the demand occurs on a time scale that is faster than can be controlled by changing the releases into the factory. As a result, without the control action of the dispatch policy, the inventory needed to satisfy the demand increases linearly with the variance of the demand signal.

Using the changing PPP as a dispatch control leads to increased variance in the lead time and in the outflow of the factory, hence requiring a higher inventory than for the pure influx control for small variation in the demand signal, entirely consistent with standard theory on safety stock [10]. However, the inventory stays essentially constant as the variance of the demand signal increases, leading to inventories that are lower than those for pure influx control for large changes in the demand. The main benefit of the extra control action is the reduced average wip in the factory which translates via Little's law into a reduced cycle time: The optimized test scenario, based on variable releases and variable PPP has a cycle time that is less than half of the cycle time obtained by a fixed PPP in the middle of the factory. In hindsight this is intuitive: Since we were running the factory at a utilization of 0.75, 25% of the factory's capacity is unused in steady state. Moving the PPP moves regions of high wip into the high priority regions allowing productive use of that extra capacity leading to a reduced cycle time.

We expect that the increased variance in the outflow due to the changes in the dispatch control are to a large extend a result of the two bucket approximation of the PDE, and to a smaller extend to the linear approximation of the changes in wip due to PPP jumps. Hence a more accurate solution of the optimization problem, solving the PDE with high accuracy in every optimization step, should lead to a better match between outflow and demand for every time step and hence to a reduced inventory. We will pursue this research direction in the future.

ACKNOWLEDGMENT

This work has been supported by DFG Cluster of Excellence Integrative production technology for high-wage countries, BMBF Kinetic models and optimization for production processes KinOpt, DAAD 55866082 and National Natural Science Foundation of China (No. 71390333). D.A wishes to acknowledge travel support from KI-Net, NSF RNMS grant No. 1107291.

REFERENCES

[1] D. ARMBRUSTER, P. DEGOND, AND C. RINGHOFER, *A model for the dynamics of large queuing networks and supply chains*, SIAM J. Applied Mathematics, 66 (2006), pp. 896–920.

[2] D. ARMBRUSTER, M. HERTY, AND C. RINGHOFER, *A continuum description for a des control problem*, in Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, IEEE, 2012, pp. 7372–7376.

[3] D. ARMBRUSTER, D. MARTHALER, C. RINGHOFER, K. KEMPF, AND T.-C. JO, *A continuum model for a re–entrant factory*, Operations Research, 54(4) (2006), pp. 933–950.

[4] D. ARMBRUSTER AND R. UZSOY, *Continuous dynamic models, clearing functions, and discrete-event simulation in aggregate production planning*, in New Directions in Informatics, Optimization, Logistics, and Production, J. C. Smith, ed., vol. TutORials in Operations Research, INFORMS, 2012.

[5] J. ASMUNDSSON, R. L. RARDIN, C. H. TURKSEVEN, AND R. UZSOY, *Production planning with resources subject to congestion*, Naval Res. Logist., 56 (2009), pp. 142–157.

[6] J. ASMUNDSSON, R. L. RARDIN, AND R. UZSOY, *Tractable nonlinear production planning: Models for semiconductor wafer fabrication facilities*, IEEE Transactions on Semiconductor Wafer Fabrication Facilities, 19 (2006), pp. 95–111.

[7] J. H. BLACKSTONE, D. T. PHILIPS, AND G. L. HOGG, *A state-of-the-art survey of dispatching rules for manufacturing job shop operations*, International Journal of Production Research, 20 (1983), pp. 27–45.

[8] R. COURANT, K. O. FRIEDRICHS, AND H. LEWY, *Über die partiellen differenzengleichungen der mathematischen physik*, Mathematische Annalen, 100 (1928), pp. 32–74.

[9] C. D'Apice, S. Göttlich, M. Herty, and B. Piccoli, *Modeling, simulation, and optimization of supply chains*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. A continuous approach.

[10] G. D. Eppen and R. K. Martin, *Determining safety stock in the presence of stochastic lead time and demand*, Management Science, 34 (1988), pp. 1380– 1390.

[11] J. W. Fowler, G. L. Hogg, and S. J. Mason, *Workload control in the semi-conductor industry*, Production Planning and Control: The Management of Operations, 13 (2002), pp. 569–578.

[12] S. Göttlich, M. Herty, and A. Klar, *Modelling and optimization of supply chains on complex networks*, Commun. Math. Sci., 4 (2006), pp. 315–330.

[13] S. T. Hackman and R. C. Leachman, *A general framework for modeling production*, Management Science, 35 (1989), pp. 478–495.

[14] K. Itoh, D. Huang, and T. Enkawa, *wofold look-ahead search for multi-criterion job shop scheduling*, International Journal of Production Research, 31 (1993), pp. 2215–2234.

[15] N. B. Kacar, *Fitting clearing functions to empirical data: Simulation, Optimization and Heuristic Algorithms*, North Carolina State University, 2012.

[16] U. S. Karmarkar, *Capacity loading and release planning with work-in-progress (wip) and lead-times*, Journal of Manufacturing and Operations Management, 2 (1989).

[17] M. La Marca, D. Armbruster, M. Herty, and C. Ringhofer, *Control of continuum models of production systems*, IEEE Trans. Automat. Control, 55 (2010), pp. 2511–2526.

[18] Y. H. Lee, K. Bhaskaran, and M. A. Pinedo, *A heuristic to minimize the total weighted tardiness with sequence dependent setups*, IEEE Transactions on Design and Manufacturing, 29 (1997), pp. 45–52.

[19] R. K. Li, Y. T. Shyu, and S. A. Adiga, *A heuristic rescheduling algorithm for computer-based production scheduling systems*, International Journal of Production Research, 31 (1993), pp. 1815–1826.

[20] K. N. McKay, F. R. Safayeni, and J. A. Buzacott, *Job shop scheduling theory: what is relevant?*, Interfaces, 18 (1988), pp. 84–90.

[21] H. Missbauer and R. Uzsoy, *Optimization models for production planning*, in Planning Production and Inventories in the Extended Enterprise: A State of the Art Handbook, K. Kempf, P. Keskinocak, and R. Uzsoy, eds., New York, 2010, Springer-Verlag, pp. 437–508.

[22] S. S. Panwalkar and W. Iskander, *Capacity planning and control*, Operations Research, 25 (1997), pp. 45–61.

[23] D. Perdaen, D. Armbruster, K. Kempf, and E. Lefeber, *Controlling a re-entrant manufacturing line via the push-pull point*, International Journal of Production Research, 46 (2008), pp. 4521–4536.

[24] V. Subramaniam, G. K. Lee, G. S. Hong, Y. S. Wong, and T. Ramesh, *Dynamic selection of dispatching rules for job shop scheduling*, Management of Operations, 11 (2000), pp. 73–81.

[25] R. Uzsoy, C. Y. Lee, and L. A. Martin-Vega, *A review of production planning and scheduling models in the semiconductor industry part : Shop-floor control.*, IIE Transactions, 26 (1994), pp. 44–55.

[26] R. Vancheeswaran and M. A. Townsend, *Two-stage heuristic procedure for scheduling job shops*, Journal of Manufacturing Systems, 12 (1993), pp. 315–325.

[27] L. M. Wein, *Scheduling semiconductor wafer fabrication*, IEEE Transactions on Semiconductor Manufacturing, 1 (1988), pp. 115–129.

[28] M. J. Zeestraten, *The look ahead dispatching procedure.*, International Journal of Production Research, 28 (1990), pp. 369–384.