

An Extension to the Tactical Planning Model for a Job Shop: Continuous-Time Control

Chee Chong. Teo, Rohit Bhatnagar, and Stephen C. Graves
Singapore-MIT Alliance, Nanyang Technological Univ., and Massachusetts Institute of Technology

Abstract — We develop an extension to the tactical planning model (TPM) for a job shop by Graves [1]. The TPM is a discrete-time model in which all transitions occur at the start of each time period. The time period must be defined appropriately in order for the model to be meaningful. Each period must be short enough so that a job is unlikely to travel through more than one station in one period. At the same time, the time period needs to be long enough to justify the assumptions of continuous workflow and Markovian job movements. We build an extension to the TPM that overcomes this restriction of period sizing by permitting production control over shorter time intervals. We achieve this by deriving a continuous-time linear control rule for a single station. We then determine the first two moments of the production level and queue length for the workstation.

Index Terms—job shop, tactical planning model, moments of production quantities and queue lengths, production smoothing

I. INTRODUCTION

THIS paper considers an extension to the tactical planning model of a job shop by Graves [1]. A job shop is a process structure in which there is a wide variety of jobs and a jumbled work flow through the shop. Due to the large variety of jobs and the diverse processing requirements of each job, there is no distinct workflow through the shop. Because of the wide job variety and thus a lack of prevailing work flow, production control is difficult and can be very complex.

A job shop often represents the most complex and generic form of a manufacturing environment. Therefore, the ability to plan a job shop will provide useful insights for production control of other process structures. Graves [1] develops an analytical model to support tactical planning in job shops. The model characterizes the

Manuscript received 19th November 2004.

C.C. Teo is with the Singapore-MIT Alliance (SMA) at Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. (email: ps7516907h@ntu.edu.sg)

R. Bhatnagar is with the Nanyang Business School, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. (email: ARBHATNAGAR@ntu.edu.sg)

S. C. Graves is with the Sloan School of Management and the Engineering System Division at MIT, Cambridge MA 02139 USA. (email: sgraves@mit.edu)

The second and third authors are Fellows of the Singapore-MIT Alliance.

interrelationship of variability, production smoothness and work-in-process inventory in a job shop. As such, the model provides a foundation for the understanding for the trade-offs inherent in the planning of a job shop.

The tactical planning model (TPM) is a discrete-time model in which all transitions within the model are governed by an underlying time period. The model assumes that all movement of jobs occurs at the start of each time period. As such, one must set the time period to be short enough so that it is unlikely for one job to travel through two successive stations in one time period.

The TPM does not explicitly model the flow of discrete jobs, but rather models the flow of work due to the jobs. Work completed in the current period flows to downstream stages in the next period. However, in discrete manufacturing, each job is only transferred to the downstream station upon completion. So in order to accurately model the job movement, the time period for the TPM should preferably be long to increase the “fluidity” of the flow of the discrete jobs.

Furthermore, the TPM assumes a Markovian workflow. The validity of this assumption depends on whether each workstation in the shop produces a stable mix of jobs. If many jobs can be completed in one period, then it is more likely that there is a stable output; but this also argues for a longer time period.

In this paper we extend the model in [1] to address the model’s limitation due to the setting of the period length. In the next section, we give a literature review of work related to the TPM. We then present a brief review of the TPM in section III. Next, we illustrate in section IV the limitation of the TPM due to the restriction of sizing the time periods. In section V, we derive a new linear control rule that will remove this restriction for a single-station system, and also determine the first two moments of the production quantity and queue length. We conclude in section VI with some thoughts on future research directions.

II. RELATED WORK

Graves [1] develops the TPM as a tactical planning tool for job shop operations. The TPM is a discrete-time linear-system model that determines the first two moments of the production and queue levels, given the planned lead times of the workstations. The model tracks the workload at each station rather than the individual jobs; the model assumes

that the volume of work arrivals at a station are in fixed proportions of the work completed at upstream stations.

To date, there are several extensions to the TPM. Parrish [2] proposes a framework for modeling work releases to meet the delivery due date for a finished product. In addition, he also shows how to adjust the control parameters of the TPM to change service measures in meeting demand.

Graves [3] presents three extensions to a single-station model of the TPM. First, he models a station that fails according to a Bernoulli process. Second, he incorporates variability due to lot-sizing, and finally, he presents the mathematical bounds on a station with capacity constraint.

Mihara [4] extends the work of Graves [3] when he looks at an unreliable multi-station TPM. Similar to Graves' work, the stations fail according to a Bernoulli process.

Fine and Graves [5] test the TPM on a real-life job shop when they apply the TPM to a shop that manufactures thermal conduction modules for mainframe computers [5]. Here, the model is extended to allow consideration of features such as release policies. The model is then used to study the impact of various planning policies and the effect of changes in product mix.

Hollywood [6] demonstrates how to calculate approximations for the steady-state moments of TPM with general non-linear control rules. His model allows for the modeling of machine congestion due to capacity loading.

Other efforts adapt the TPM to pull systems. Leong [7] models a Kanban control system and other pull systems using the TPM in which work is produced at a station whenever there is a downstream inventory shortfall. More recently, Graves and Hollywood develop a constant-inventory TPM in which the release of work into the shop is regulated to maintain a constant inventory level [8].

III. REVIEW OF THE TACTICAL PLANNING MODEL

The tactical planning model (TPM) is a discrete-time, continuous flow model. All transitions within the model occur at the start of each time period, and the jobs are modeled as workload measured in time units (e.g. hours). The workflow is assumed to have a Markov property: that is, the processing requirements at a station do not depend on how work got to the station. As such, each individual job has no identity.

Central to the TPM model is the linear control rule, which is stated as

$$P_{it} = \alpha_i Q_{it} \quad (1)$$

where P_{it} is the amount of production completed by work station i in time period t , Q_{it} is the queue level at the start of period t , and the parameter $\alpha_i, 0 < \alpha_i \leq 1$, is a smoothing parameter. This rule states that the production P_{it} at workstation i is a fixed portion (α_i) of the queue of work Q_{it} at the start of the period. In particular, $1/\alpha_i$ represents the number of periods, on average, the work requires to

move through the work station. We interpret $1/\alpha_i$ to represent the planned lead time. We can view the control rule in (1) as a prescriptive equation, i.e. to preserve the integrity of the planned lead time, we must shift capacity to heavily loaded stations. But (1) can also be considered as a descriptive equation where production resources are naturally flexed to accommodate the varying workloads at the stations.

The queue level Q_{it} satisfies the standard inventory balance equation

$$Q_{it} = Q_{i,t-1} - P_{i,t-1} + A_{it} \quad (2)$$

where A_{it} is the amount of work that arrives at workstation i at the start of period t . By substituting (1) into (2), we obtain a first-order smoothing equation with α_i as the smoothing parameter:

$$P_{it} = (1 - \alpha_i)P_{i,t-1} + \alpha_i A_{it} \quad (3)$$

Each workstation can receive two types of arrivals; one type of arrival consists of jobs that have their first processing step at the station, while the other type of arrival consists of in-process jobs that have just completed processing at an upstream station. We model the arrivals to station i from another station j by the equation:

$$A_{ijt} = \phi_{ij} P_{j,t-1} + \varepsilon_{ijt} \quad (4)$$

A_{ijt} is the flow of work arriving at station i from station j at the start of period t , ϕ_{ij} is a positive scalar and ε_{ijt} is a random variable. We assume that one unit (e.g. hour) of work at station j will trigger, on average, ϕ_{ij} time units of work at station i . The variable ε_{ijt} is a noise term that models uncertainty between production at j and arrivals to i , and is assumed to be an i.i.d. random variable with zero mean and a known variance.

The arrival to station i is given by

$$A_{it} = \sum_j A_{ijt} + N_{it} \quad (5)$$

where N_{it} is an i.i.d. random variable for the workload from new jobs that enter the shop at station i at time t . Substituting for A_{ijt} , we obtain

$$A_{it} = \sum_j \phi_{ij} P_{j,t-1} + \varepsilon_{it}, \text{ where } \varepsilon_{it} = N_{it} + \sum_j \varepsilon_{ijt} \quad (6)$$

The term ε_{it} represents arrivals that are not predictable from the production levels of the previous period, and consists of work from new jobs and noise in the flow. By assumption, the time series ε_{it} is independent and identically distributed over time.

We can restate the equations for production (3) and for arriving work (6) in matrix-vector form:

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{D})\mathbf{P}_{t-1} + \mathbf{D}\mathbf{A}_t, \quad (7)$$

$$\mathbf{A}_t = \mathbf{\Phi}\mathbf{P}_{t-1} + \boldsymbol{\varepsilon}_t \quad (8)$$

where $\mathbf{P}_t = \{P_{1t}, \dots, P_{nt}\}'$, $\mathbf{A}_t = \{A_{1t}, \dots, A_{nt}\}'$, and $\boldsymbol{\varepsilon}_t = \{\varepsilon_{1t}, \dots, \varepsilon_{nt}\}'$ are column vectors of random variables, n is the number of workstations, \mathbf{I} is the identity matrix, \mathbf{D} is a diagonal matrix with $\{\alpha_1, \dots, \alpha_n\}$ on the diagonal, and $\mathbf{\Phi}$ is an n -by- n matrix with elements ϕ_{ij} . By substituting equation (8) into equation (7), we find that

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{D} + \mathbf{D}\mathbf{\Phi})\mathbf{P}_{t-1} + \mathbf{D}\boldsymbol{\varepsilon}_t \quad (9)$$

By iterating this equation and assuming an infinite history of the system, we rewrite \mathbf{P}_t as an infinite series

$$\mathbf{P}_t = \sum_{s=0}^{\infty} (\mathbf{I} - \mathbf{D} + \mathbf{D}\mathbf{\Phi})^s \mathbf{D}\boldsymbol{\varepsilon}_{t-s} \quad (10)$$

The mean and the covariance for the noise vector $\boldsymbol{\varepsilon}_t$ are denoted by $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_n\}'$, and $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$ respectively. The first two moments of \mathbf{P}_t are given by

$$\begin{aligned} E[\mathbf{P}_t] &= \sum_{s=0}^{\infty} (\mathbf{I} - \mathbf{D} + \mathbf{D}\mathbf{\Phi})^s \mathbf{D}\boldsymbol{\mu} \\ &= (\mathbf{I} - \mathbf{\Phi})^{-1} \boldsymbol{\mu} \end{aligned} \quad (11)$$

and

$$\mathbf{S} = \text{var}(\mathbf{P}_t) = \sum_{s=0}^{\infty} \mathbf{B}^s \mathbf{D}\boldsymbol{\Sigma}\mathbf{D}\mathbf{B}^{s'} \quad (12)$$

where $\mathbf{B} = \mathbf{I} - \mathbf{D} + \mathbf{D}\mathbf{\Phi}$

We note that \mathbf{S} provides the variance of the production requirements for each station, as well as the covariance for each pair of workstations. In addition, we determine the first two moments of the queue levels. From (1), we note that

$$\mathbf{Q}_t = \mathbf{D}^{-1}\mathbf{P}_t \quad (13)$$

Therefore we have

$$E[\mathbf{Q}_t] = \mathbf{D}^{-1}E[\mathbf{P}_t] \quad (14)$$

and

$$\text{Var}(\mathbf{Q}_t) = \mathbf{D}^{-1}\mathbf{S}\mathbf{D}^{-1} \quad (15)$$

The infinite series in equations (11), (12), (14) and (15) converge provided that $\rho(\mathbf{\Phi}) < 1$, where $\rho(\mathbf{\Phi})$ denotes the spectral radius of $\mathbf{\Phi}$ (see [1]).

IV. LIMITATION OF TPM

In this section, we examine the limitations of the TPM due to the sizing of the discrete time period. This provides the motivation for our work, which is presented in the next section.

In the TPM, all job movements can only occur at the start of each time period. In order to model movement of jobs in the actual shop, we are restricted to set the period length to be short enough so that it is highly improbable for one job to travel through more than one station in a single time period.

However, due to the continuous-flow assumption, the time period should be long relative to the workload of the individual discrete jobs. This will increase the ‘‘fluidity’’ of the discrete jobs and will make the continuous-flow assumption more reasonable. Now we use an example to illustrate the discrepancies between the model and the actual system due to the above contradictory objectives in period sizing.

To simplify our illustration, we consider a simple system that consists of two stations in series, namely Station i and Station j . We further assume that the planned lead time of Station i is 2 hours, while that of Station j is 1 hour. We set the length of the time period to be 1 hour, as we assume that it is unlikely for a job in Station i to travel beyond Station j in 1 hour. Now consider a job that enters the empty system and arrives at Station i at the start of period t . We suppose that the job has a processing time of 2 hours at Station i , and 1 hour at Station j . We illustrate and compare the sequence of events for the actual system and the TPM from period t to $t + 2$.

Fig. 1 shows the actual system from the start of period t to $t + 2$. The job arrives at Station i at the start of period t . Since the planned lead time is 2 periods, the job is processed at Station i till the end of $t + 1$. The job is then transferred to Station j at the start of $t + 2$. It is then processed at Station j till the end of period $t + 2$ since the planned lead time is 1 period.

Now we look at the same scenario in the context of the TPM. Suppose that both Stations i and j produce according to the TPM control rule in equation (1). In this case, given the planned lead time of each station, we have $\alpha_i = 1/2$ and $\alpha_j = 1$. Job movements between the two stations are modeled by equation (4). We assume that the term $\phi_{ij} = 0.5$ given the processing times of the job, and $\varepsilon_{ijt} = 0$. Fig. 2

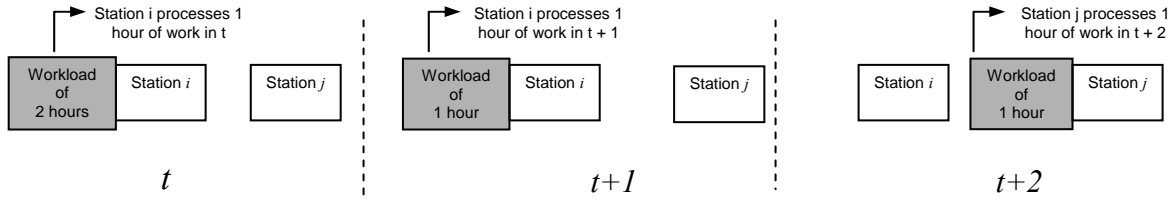


Fig 1. Actual system from start of period t to $t+2$

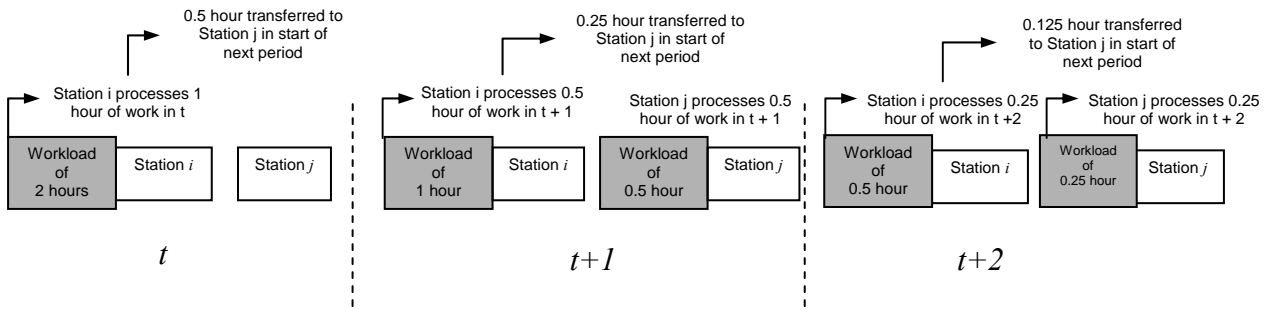


Fig. 2. TPM from start of period t to $t+2$

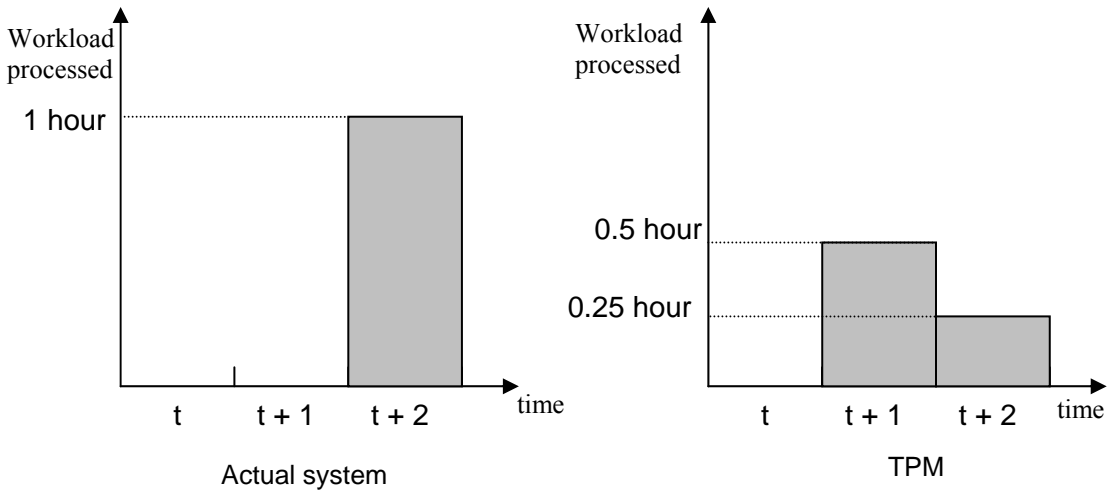


Fig. 3. Production levels at Station j in actual system and TPM

illustrates the workflow in the TPM from the start of period t to $t + 2$. As shown in the figure, Station i processes half of the in-queue workload at the start of each period ($\alpha_i = 1/2$), while Station j processes the entire workload ($\alpha_j = 1$). And the workload processed by Station i in each period generates half the workload at Station j at the start of the next period ($\phi_{ij} = 0.5$).

In the actual system, the discrete job moves to the downstream station only upon completion. However, in the TPM, work flows as a fluid to the downstream station even if the discrete job is not completed. In this example, the discrete job is “split” up and moved in parts to the downstream station. This is due to the fact that the period length is short (1 hour) compared to the workload of the job at Station i (2 hours). As a result, a workload from the job is generated at Station j even though it is still in-process at the upstream station.

Fig. 3 shows the production levels at Station j over period t to $t + 2$ in both the actual system and the TPM. The production level in the actual system consists of a “spike” of 1 hour in period $t + 2$. In the TPM, the production is “smoothed”, with production levels at 0.5 hour and 0.25 hour in period $t + 1$ and $t + 2$ respectively.

This example considers a system of only two stations. To model a complex job shop using the TPM, one must set the period length by considering the job movements between all stations.

On the one hand, the period length should be such that it is unlikely that a job completes processing at more than one work station in a time period. In a shop with many stations and where jobs move quickly between stations, this implies that the period length be set on the order of the average job workload.

On the other hand, the accuracy of the TPM depends on assumptions of continuous workflow. We prefer to set a long time period relative to the workload of the jobs so that the discrete jobs will be “more fluid.” Furthermore, the TPM assumes a Markovian workflow such that transitions do not depend on the history of the system. In essence, the model assumes that each station processes a relatively stable mix of jobs in each time period, so that subsequent flow to downstream stations is stable as well. The validity of this assumption also depends on the length of the time period. If only a few jobs are completed in each period, then it is unlikely that there is a very stable output. Therefore, this assumption will be more valid if we are able to set a longer time period.

In addition, the restriction of period sizing may hinder the application of the TPM to production planning and scheduling. In some job shops, it takes only a short time (e.g. less than an hour) for a job to travel through more than one station, and thus the discrete time period has to be short. However, the parameters in most planning systems, such as the demand requirements, are defined in daily or weekly time units. Thus the ability to set a longer time period will facilitate the application of the TPM to production planning and scheduling.

In the next section we describe an approach to extend the TPM to be less dependent on the choice of time period.

V. MODEL

In this section, we develop a single-station model to overcome the limitations of the TPM discussed in section IV. We will derive a linear control rule that accommodates more frequent arrivals to the work station; as a consequence, we can now permit work to flow through more than one work station within a time period. We find the first two moments of the production and queue length variables using the derived control rule.

Without loss of generality, we suppose that each discrete time period t has a length of one time unit. We sub-divide each time period t into m equal subintervals of sub-period s , where $s = 1, 2, \dots, m$. We define $m = 1/\Delta$, where Δ is the length of each sub-period.

We assume that work flow can arrive at the start of each sub-period. We also assume that we set the production in each sub-period according to the linear control rule (1). Thus, we control the production according to a finer time grid, and we allow for more fluid arrivals to the work station.

We restate the control rule (1) for each sub-period s as

$$Y(\Delta, s) = \alpha \Delta X(\Delta, s) \quad \text{for } s = 1, 2, \dots, m \quad (16)$$

where $Y(\Delta, s)$ is the production level in sub-period s of length Δ , $X(\Delta, s)$ is the queue length at start of sub-period s of length Δ and α is the smoothing parameter. We interpret $1/\alpha$ as the planned lead time; however, we now permit α to assume any positive value, and thus, we permit the planned lead time to be less than one time period. Equation (16) is analogous to (1) such that the production $Y(\Delta, s)$ in each sub-period s is a fixed fraction $\alpha \Delta$ of the queue length $X(\Delta, s)$ at the start of each sub-period.

Now we proceed to develop the linear control rule for P_t in terms of Q_t and A_t . These variables have the same definition as in TPM: P_t is the production completed in period t , Q_t is the queue length at the start of period t , and A_t is the arrival of work to the station in period t . However, we now assume that A_t does not arrive at the start of the period, but rather arrives uniformly over period t . In particular, we assume that in each sub-period, the arrival amount is equal to A_t/m .

We have the following boundary condition for the queue length for the first sub-period:

$$X(\Delta, s = 1) = Q_t + A_t/m$$

For $s > 1$, we model the queue length in the sub-period s by the standard inventory equation

$$X(\Delta, s) = X(\Delta, s - 1) - Y(\Delta, s - 1) + A_t/m \quad (17)$$

By substituting (16) into (17), we obtain

$$X(\Delta, s) = (1 - \alpha\Delta) X(\Delta, s - 1) + A_t/m \quad (18)$$

Now to get an expression for P_t , we note that

$$P_t = \sum_{s=1}^m Y(\Delta, s) = \alpha\Delta \sum_{s=1}^m X(\Delta, s) \quad (19)$$

We sum the above expressions for $X(\Delta, s)$ to find

$$\sum_{s=1}^m X(\Delta, s) = Q_t + (1 - \alpha\Delta) \sum_{s=1}^{m-1} X(\Delta, s) + A_t \quad (20)$$

From (20), we observe that

$$\begin{aligned} \sum_{s=1}^m X(\Delta, s) &= Q_t + (1 - \alpha\Delta) \sum_{s=1}^m X(\Delta, s) + A_t - (1 - \alpha\Delta) X(\Delta, m) \\ \Rightarrow \alpha\Delta \sum_{s=1}^m X(\Delta, s) &= Q_t + A_t - (1 - \alpha\Delta) X(\Delta, m) \end{aligned} \quad (21)$$

We now combine (19) and (21) to get

$$P_t = Q_t + A_t - (1 - \alpha\Delta) X(\Delta, m) \quad (22)$$

From (22), in order to get an expression for P_t , we need to find $X(\Delta, m)$. From (18) and repeated substitution, we can then write

$$\begin{aligned} X(\Delta, m) &= (1 - \alpha\Delta)^{m-1} \times Q_t \\ &\quad + \left(1 + (1 - \alpha\Delta) + \dots + (1 - \alpha\Delta)^{m-1}\right) \times \frac{A_t}{m} \end{aligned} \quad (23)$$

We can use (23) to re-write (22) as

$$\begin{aligned} P_t &= \left(1 - (1 - \alpha\Delta)^m\right) \times Q_t \\ &\quad + \left(1 - \left(\frac{1 - \alpha\Delta}{\alpha}\right) \cdot \left(1 - (1 - \alpha\Delta)^m\right)\right) \times A_t \end{aligned} \quad (24)$$

Thus we can write (24) as

$$P_t = \beta(\Delta) Q_t + \gamma(\Delta) A_t \quad (25)$$

where

$$\beta(\Delta) = 1 - (1 - \alpha\Delta)^m$$

and

$$\begin{aligned} \gamma(\Delta) &= 1 - \left(\frac{1 - \alpha\Delta}{\alpha}\right) \left(1 - (1 - \alpha\Delta)^m\right) \\ &= 1 - \beta(\Delta) \left(\frac{1 - \alpha\Delta}{\alpha}\right) \end{aligned}$$

Now we proceed to determine the continuous-time limits for $\beta(\Delta)$ and $\gamma(\Delta)$ as the length of the sub-period goes to zero. This corresponds to a continuous-time control in which the production level will satisfy (16) at every instant in time. We use the formula $\lim_{x \rightarrow 0} (1 - x)^{\frac{1}{x}} = e^{-1}$ to obtain the continuous-time limit of $\beta(\Delta)$:

$$\begin{aligned} \beta &= \lim_{\Delta \rightarrow 0} \beta(\Delta) = \lim_{\Delta \rightarrow 0} [1 - (1 - \alpha\Delta)^m] \\ &= \lim_{\Delta \rightarrow 0} [1 - (1 - \alpha\Delta)^{\frac{1}{\Delta}}] \\ &= 1 - e^{-\alpha} \end{aligned}$$

For $\gamma(\Delta)$, we find that

$$\begin{aligned} \gamma &= \lim_{\Delta \rightarrow 0} \gamma(\Delta) = \lim_{\Delta \rightarrow 0} \left[1 - (1 - (1 - \alpha\Delta)^m) \left(\frac{1 - \alpha\Delta}{\alpha}\right)\right] \\ &= 1 - \lim_{\Delta \rightarrow 0} \left[\frac{1}{\alpha} (1 - (1 - \alpha\Delta)^m)\right] + \lim_{\Delta \rightarrow 0} \left[\Delta (1 - (1 - \alpha\Delta)^m)\right] \\ &= 1 - \lim_{\Delta \rightarrow 0} \left[\frac{1}{\alpha} (1 - (1 - \alpha\Delta)^m)\right] \\ &= 1 - \frac{1}{\alpha} (1 - e^{-\alpha}) = 1 - \frac{\beta}{\alpha} \end{aligned}$$

We can now restate (25) for the continuous-time control as:

$$P_t = \beta Q_t + \gamma A_t \quad (26)$$

where β and γ are given above.

The balance equation for the queue length for the single station is now given by:

$$Q_t = Q_{t-1} - P_{t-1} + A_{t-1}. \quad (27)$$

This balance equation differs from (2) in the TPM, due to the new assumption that arrivals occur continuously throughout a period. Hence, we define Q_t to be the queue length at the start of period t , prior to any arrivals in period t .

By substituting (26) into (27) and repeated substitution, we obtain:

$$Q_t = (1 - \gamma) \sum_{i=1}^{\infty} (1 - \beta)^{i-1} A_{t-i}. \quad (28)$$

If we assume that the arrivals are i.i.d. with mean μ and variance σ^2 , then we find the two moments for the queue length from (28):

$$E[Q_t] = \left(\frac{1-\gamma}{\beta} \right) \mu = \frac{\mu}{\alpha}$$

$$\text{Var}(Q_t) = \frac{(1-\gamma)^2 \sigma^2}{2\beta - \beta^2}$$

Similarly, we obtain the two moments for the production variable:

$$E[P_t] = \mu$$

$$\text{Var}(P_t) = \left(\frac{\beta}{2-\beta} (1-\gamma)^2 + \gamma^2 \right) \sigma^2$$

Thus, we have analytical expressions for the moments of these two variables. We have expressed these in terms of the parameters β and γ , both of which are functions of our smoothing parameter α . However, it is not obvious how the variances of production and the queue length depend on the smoothing parameter α . To provide some insight into this, we graph each variance as a function of the smoothing parameter α in Fig. 4 (with $\sigma = 1$). We see from these graphs that the variance of production drops as we do more smoothing, i.e., small values of α , or equivalently longer planned lead times. However, as we do more smoothing of production, both the expected queue length and its variance grow.

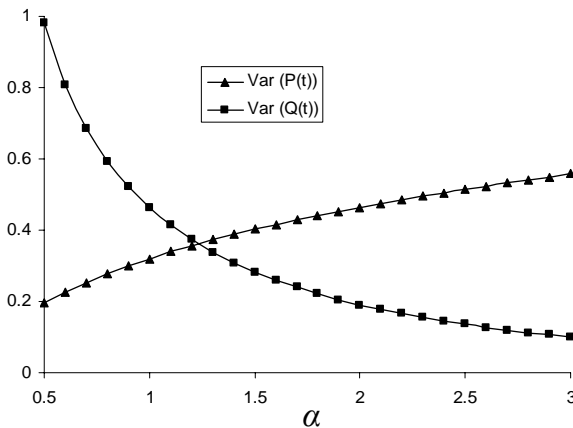


Fig. 4. Variance and P_t and Q_t as a function of α for $\sigma = 1$

These moments provide a simple way to see the fundamental trade-offs across the three elements of time, capacity and variability. For given level of demand variability (given by the arrival process), as we reduce the planned lead time, production becomes more variable and more capacity is required; alternatively, as we smooth

production, we need less capacity but more time in terms of the planned lead time.

These insights are the same as for the TPM. But the model given here is more general in that we permit continuous arrivals to the work station and continuous production control.

VI. CONCLUDING REMARKS

In this paper, we extend the TPM to address its dependency on the choice of time period; in particular, we permit work to arrive throughout a period, as opposed to at the start of each period and we permit continuous production control. We derive a linear control rule by assuming that the production is controlled according to a finer time grid. We then obtain the continuous-time limits of the control rule, and find the first two moments of the production and queue random variables based on the control rule.

Our next step is to extend the single-station model to a network of workstations as would exist in a job shop. We will need to determine how best to model the flow of work between stations, so as to capture actual behavior and yet still retain analytical tractability. We will need to investigate the stability of this model; in particular, we wish to find the conditions of convergence for the first two moments of production and queue vectors. In the TPM in [1], the first two moments converge provided that the spectral radius of workflow matrix Φ is less than 1. This implies that a unit of work at any station cannot eventually result in more than one unit of work for the same station. It is of interest to evaluate the convergence conditions of this model as compared to the TPM in [1].

In addition, we have not established the benefits of this extension over the original TPM under different job shop conditions. It is worthwhile to carry out a study, perhaps through a computational experiment, to compare how this extension performs relative to the TPM in different operating conditions, such as the stability of job mix and speed of job movement between stations.

Another opportunity for future research is to extend this model to incorporate congestion effects due to capacity loading. This would involve relaxing the linear control assumptions for each station. Hollywood [6] has developed an approximation for the TPM with nonlinear control rules. It may be possible to integrate his approximation into this model.

REFERENCES

- [1] S.C. Graves, "A Tactical Planning Model for a job shop", *Operations Research*, 34, 1986, pp 522-533.
- [2] S.H. Parrish, "Extensions to a model for tactical planning in a job shop Environment", S.M. Thesis, Operations Research Center, Massachusetts Institute of Technology, June 1987.
- [3] S.C. Graves, "Extensions to a Tactical Planning Model for a job shop", *Proceedings of the 27th IEEE Conference on Decision and Control*, Austin, Texas, December 1988.

- [4] S. Mihara, "A Tactical Planning Model for a job shop with unreliable work stations and capacity constraints", S.M. Thesis, Operations Research Center, MIT, Cambridge MA, January 1988.
- [5] C. H. Fine and S. C. Graves, "A Tactical Planning Model for manufacturing subcomponents of mainframe computers", *J. Mfg. Oper. Mgt.*, 2, 1989, pp 4-34.
- [6] Hollywood, John S. 2001. "An Approximate Planning Model for a Job Shop with Nonlinear Production Rules," working paper, 32 pp.
- [7] T.Y. Leong, "A Tactical Planning Model for a mixed push and pull system", Ph.D. program second year paper, Sloan School of Management, Massachusetts Institute of Technology, July 1987.
- [8] S. C. Graves and J. S. Hollywood, "A constant-inventory Tactical Planning Model for a job shop", Working paper, January 2001, revised March 2004, 35 pp.