Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Kolmogorov equations based approximate analysis and sizing of constant work in process unreliable manufacturing system loops.

Fatima Zahra Mhada and Roland Malhamé

Abstract— **CONWIP or constant work in process is an important manufacturing systems production discipline whereby within a CONWIP loop, there is a cap on the maximum amount of work in process that is permitted at any time. This allows for some mobile storage within the loop, albeit a bounded amount. Enforcement of the discipline is carried out at the entrance of the loop. The presence of a loop wide constraint creates indirectly a significant degree of solidarity among the machines within the loop. This property is exploited to develop a model of storage dynamics involving a number of (virtual) macro machines having some common states and interacting through some unknown parameters which are then estimated. Numerical results are presented and an application in minimal CONWIP loop storage sizing for a given demand rate and service level requirement is reported.**

INTRODUCTION

There is a rich literature on decomposition methods for the analysis and optimization of unreliable manufacturing lines [3] ,[4], [5],[6],[7],[8]. Most of these papers have analyzed so-called Kanban (card in japanese; introduced by Toyota) related production disciplines whereby one machine cannot get ahead of the next machine downstream by more that a certain maximum amount of parts designated as the Kanban parameter at that stage. In effect, Kanban parameters define the permissible buffer spaces between machines; buffering partially alleviates starvation (absence of parts) or blocking problems (no space for moving a finished part) that can be experienced by operational machines in the vicinity of a failed machine, thereby enhancing line productivity at the cost of some system storage. While the Kanban principle leads to a a highly decentralized production control architecture, CONWIP, or Constant Work in Process (wip) introduces an alternative more centralized competing architecture by treating a string of machines as a unique Kanban cell within which the wip level is bounded above by some constant z, machines otherwise always producing at their maximum allowable rate. When machines never fail, and demand rate is less than line production rate, total wip saturates at z and remains constant. However, in practice, machine failures will induce a fluctuating total wip. It is also possible to combine Kanban and CONWIP as in [1], [2] to generate high performance hybrid policies. The above policies are special instances of the lean manufacturing paradigm and the optimization of their defining parameters has been the subject of much research [15]. While Monte-Carlo simulation and perturbation analysis can be useful optimization tools in that respect, the associated computation times tend to become prohibitive in long transfer lines. Alternatively, analysis based, approximate decomposition/aggregation methods can lead to efficient optimization schemes.

Except for the notable exception of the decomposition approach in [9], based on which extensions for analyzing hybrid CONWIP/Kanban architectures were developed in [1], [2], there is a paucity of published results on decomposition/aggregation methods for CONWIP controlled transfer lines. In [9], the decomposition is achieved first by assuming that, despite machine failures, total internal wip is always at saturation level z, and then replacing that strong system wide storage levels coupling constraint over time by a much weaker one stating that long term average wip is equal to z. The method is accurate for large loops with wip limits that are neither "too large", nor" too small". Our objective here is to propose alternative and hopefully improved approximate methods for the performance evaluation of CONWIP policies, and which could serve as building blocks for the evaluation of more complex hybrid production control architectures.

One of the important performance measures of CONWIP controlled transfer lines is the mean amount of total work in process for a given fixed rate of parts production. Our approximate computation of mean total work in process flows from treating the CONWIP controlled portion of the transfer line as a Kanban controlled macro cell, thus situating the approximation within the class of aggregation methods. The input/output behavior of the CONWIP macro cell is affected on the one hand, by the reliability statistics of the first machine, and on the other hand by the probability, also called coefficient of availability, that work in process be available at the output of that macro cell. The latter quantity is complex to compute as it involves having to model every machine and intermediary work in process within the line. The so-called machine decoupling approximation, and demand averaging principle, both of which form the basis of the approximate Kanban performance analysis in [8], are liberally used in the process.

The rest of the paper is organized as follows: in Section I, we define our CONWIP controlled model of an n machine homogeneous unreliable transfer line (machines with identical reliability statistics and identical maximum production rates) and provide examples of some of the Monte Carlo simulation trajectories which have enhanced our understanding and inspired our analysis. In Section II, we focus on building a

Fatima zahra Mhada, Department of Electrical Engineering, École Polytechnique de Montréal and GERAD, Montreal, Canada, fatima-zahra.mhada@polymtl.ca

Roland P. Malhamé, Department of Electrical Engineering, École Polytechnique de Montréal and GERAD, Montreal, Canada, roland.malhame@polymtl.ca

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

Markovian model of the last machine in the CONWIP macro cell with the aim of computing the coefficient of availability of wip at its output. Unfortunately, this model is affected by the unknown coefficients of availability conditional on the state of the line, for all other intermediary stocks. We are then led to develop in Section III the Markovian models of each one of these stocks. They are associated with n^{th} order linear differential equations (Forward Kolmogorov equations) which are coupled through a total of $(n-1)^2$ unknown coefficients of availability at the various storage bins in the loop, and conditional on the various transfer line operational states . In Section IV, we develop our aggregate Markovian model of the total wip in the system. It can produce an answer once provided with the coefficient of availability at the last bin in the CONWIP macro cell. In Section V, we give the details of our solution steps. In Section VI, we compare our theoretical results against Monte Carlo simulations for a number of transfer lines and illustrate the application of the analysis to the problem of minimally sizing the storage parameter of a CONWIP loop to secure a given service level at a certain demand. In Section VII, we summarize our conclusions and discuss directions for future work.

I. Mathematical model of the transfer line: definitions and assumptions

We consider a manufacturing transfer line consisting of n machines, M_i $i = 1,\cdots,n$, each associated with a work in process variable x_i. All machines in the line, except the last one, are assumed to be involved in a single CONWIP (constant work in process) loop, in essence a loop control discipline attaching a single Kanban level to the group of machines in the loop. We assume that the transfer line produces a single type of parts and it is subjected to a constant rate of demand of d parts per unit time. Under these circumstances, and for a given CONWIP level z, we are interested in two indices of performance: mean total wip in the loop, and coefficient of availability of wip at the $(n-1)^{th}$stage.

Given a CONWIP loop, the CONWIP discipline is such that the first machine in the loop and the last one become strongly correlated when the total amount of wip enclosed reaches the set maximum level. At that stage, the first machine can only process the exact amount of parts that leave the last machine. Thus, much more correlated machine behavior can result as opposed to plain Kanban policies, and it is at the entry point of the loop, i.e. at the first machine, that total work in process regulation is enforced. This largely explains our intuition that an aggregate Kanban controlled macro cell model with adequately defined parameters stood a good chance of yielding a fairly accurate characterization of mean total work in process in the CONWIP loop.

A. Machines and policy parameters

Machines M_i $i = 1,\cdots,n$, are assumed to have an operational and a failure state, and to randomly evolve between these states according to a two state continuous time Markov chain. For machine M_i, maximal production rate, failure rate and repair rate are respectively given by: k, p_i and r_i. Production rates and number of parts produced are treated as continuous variables (fluid model). In addition, no backlog is permitted anywhere in the flow line except at the buffer associated with the last machine. Furthermore, a CONWIP policy is enforced on the first $n-1$ machines whereby total work in process in the loop cannot exceed z. Given work in process $x_i(t)$ in buffer i, we can make the following observations:

- if $x_i(t) > 0$, then machine M_{i+1} has access to parts.
- if on the other hand $x_i(t) = 0$, this can mean one of two things: either machine M_i is off, and in that case machine M_{i+1} is starved and consequently its production ceases, or M_i is operational, in which case M_{i+1} and M_i share the same production rate.

B. A few helpful simulation based observations

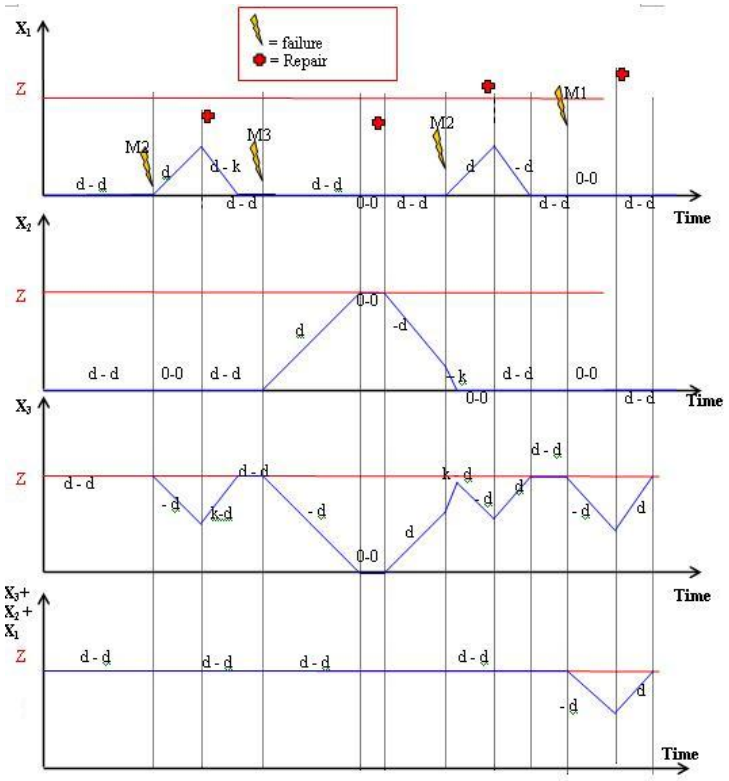

Fig. 1. Simulation trajectories of work in process for a three machine CONWIP loop

Understanding of CONWIP loop dynamics can be enhanced through some Monte Carlo simulation based observations as shown in Fig. 1. One notices that under our production rate assumptions, if machines remain operational for a long time, all work in process (wip) tends to accumulate at the $(n-1)^{th}$ stage of the loop, while the first (n-2) storage bins remain empty (but active!). Also, once the loop is blocked (total amount of wip has reached the maximum z), it will remain so, even through any single machine failure, unless that machine happens to be M_1. As a result, the

loop will be blocked with high probability; internal wip levels will be zero most of the time, resulting in a high degree of *coupling* of internal machines and thus pointing at aggregation as a promising direction for approximation. Also, the reliability statistics of M_1 appear to play a special role. Furthermore, when an internal machine fails, wip starts immediately accumulating upstream of it, while the wip at the $(n-1)^{th}$ stage starts emptying until it eventually finishes (loop blocked and $(n-1)^{th}$ buffer empty). At that stage, the whole line ceases production. Thus CONWIP in effect provides for mobile storage as failures occur, while transfer line sections on either side of a failure, appear to act as aggregates.

C. Definitions

a_i (*Coefficient of availability of wip* x_i) : it is the probability that wip x_i be available for machine M_{i+1} (i=1..n-1). Note that unlike the Kanban case for which this coefficient is equivalent to the probability that wip x_i be positive, in the case of CONWIP, it also includes the case where x_i is zero and all machines upstream of x_i are operational.

D. Assumptions

- *Common machine repair rate*: r_1=r_2=....=r_n=r.
- *Common machine maximum production rate*: The maximum production rate of all machines is k.
- *Exclusion of coincident machine failures* : In order to minimize the impact of that approximation, we replace the original failure rates p_i by $\tilde{p}_i$ rates, properly adjusted so that the transfer line remains to falls operational with the same probability (see [14] for details).
- *CONWIP loop always blocked except when* M_1 *fails*: we make this assumption based on our observation over simulations, that a blocked CONWIP loop remains so through any machine failure except machine M_1. It will be quite crucial when modeling total wip dynamics in Section IV.

II. A Markovian model of wip x_{n-1}

In our approximate modeling, we shall aim at representing the part of the transfer line upstream of wip x_{n-1} as an aggregate multi state Markovian machine, called the *effective* machine upstream of x_{n-1}, and denoted $\tilde{M}_{n-1}$. As described below, it has n discrete states or equivalently modes, which evolve according to a Markov chain and interact with the continuous wip state x_{n-1} to generate combined hybrid Markovian dynamics. The dynamics will be associated with particular Forward Kolmogorov differential equations which will constitute our predictive model of that part of wip behavior. We shall denote by $\tilde{\alpha}_{n-1}$ the discrete state of effective machine $\tilde{M}_{n-1}$.

$\tilde{\alpha}_{n-1}$ = 1: Here all machines in the CONWIP loop are operational. From that mode, one can jump to single machine failure situation, i.e. **states 0i** below with rate $\tilde{p}_i$.

$\tilde{\alpha}_{n-1}$ = $0i$, $i = 1..(n-1)$: State $0i$ of $\tilde{M}_{n-1}$, is defined by the situation where all machines except M_i are operational. $\tilde{M}_{n-1}$ leaves from that state to state 1 upon repair of machine M_i, i.e. with rate r_i.

A. Defining the hybrid state $\tilde{x}_{n-1}$ dynamics

In what follows, we build on the effective Markovian machine model $\tilde{M}_{n-1}$, to obtain the Kolmogorov differential equations associated with the hybrid state Markov process $\tilde{x}_{n-1}$. The latter is constructed by concatenation of continuous state x_{n-1} with the discrete $\tilde{M}_{n-1}$ machine state $\tilde{\alpha}_{n-1}$. At this point however, we must recognize that the time evolution of x_{n-1} is dictated by both the random production rate of $\tilde{M}_{n-1}$, and the random instantaneous demand on x_{n-1}, i.e. the production rate of machine M_n. The latter is quite complex to describe and would involve another random continuous variable namely x_n. This is where the *demand averaging principle* [8] is invoked. It stipulates that as far as computing mean quantities, it is sufficient to replace this complex random demand process *by any other one which would result in the same long term demand*. In particular, we find it convenient to replace the demand by a constant $\tilde{d}_{n-1}$, only on these intervals of time on which stock x_{n-1} is available. This leads to $\tilde{d}_{n-1} = \frac{d}{a_{n-1}}$, where it is recalled that d is the constant rate of demand for parts, while a_{n-1} is obtained through a provably convergent iterative calculation, on the associated Forward Kolmogorov equations (see [8] for further details). The evolution of x_{n-1} must occur in accordance with the assumed (CONWIP) production discipline and the demand for parts function (assumed constant at the final inventory/ backlog bin), and as a result its dynamics will also depend on the region of wip space in which current wip lies. Three regions with distinct dynamics can be defined as follows:

$0<x_{n-1}<z$: In this region, x_{n-1} is constrained to evolve with a piecewise constant rate of change dictated by the mode of $\tilde{M}_{n-1}$. We shall designate by $v_{n-1}^{\tilde{\alpha}_{n-1}}$, the associated velocity. More specifically:

- $\tilde{\alpha}_{n-1}$ = 1: $dx_{n-1}/dt = v_{n-1}^1 = (k - \tilde{d}_{n-1})$.
- $\tilde{\alpha}_{n-1} = 0i$, $i = 1,..,(n-1)$:

$dx_{n-1}/dt = v_{n-1}^{0i}$ = $(-\tilde{d}_{n-1})$+ k*Pr$[\bigcup_{l=i}^{n-2}(x_l > 0)|M_{ioff}]$.

The rightmost term above is the probability that any of the wips past failed machine M_i be strictly positive (so as to be able to supply machine M_{n-1}), conditional on machine M_i being off. When multiplied by k, it becomes the mean production rate conditional on machine M_i being off. In this region, one defines the following hybrid probability density functions, for $i = 1,..,(n-1)$

$f_1^{n-1}(\lambda,t)d\lambda$ = Pr$[(\lambda < x_{n-1}(t) \leq \lambda + d\lambda) \cap (\tilde{\alpha}_{n-1} = 1)]$.

$f_{0i}^{n-1}(\lambda,t)d\lambda$ = Pr$[(\lambda < x_{n-1}(t) \leq \lambda + d\lambda) \cap (\tilde{\alpha}_{n-1} = 0i)]$.

x_{n-1}= z: Under our modeling assumptions, the only way to remain at this level is if all machines in the CONWIP loop are on, i.e. if $\tilde{\alpha}_{n-1}$=1. This corresponds to a blocked loop with all wip concentrated at the last stage of the loop. Wip remains fixed at z until the next jump of $\tilde{\alpha}_{n-1}$, which happens at rate $\sum_{i=1}^{n-1}\tilde{p}_i$. Associated with that region is the probability mass: $P_z^{n-1}(t)$ =Pr$[x_{n-1}(t) = z]$.

$x_{n-1} = 0$: Under our modeling assumptions, the only way to reach this situation is if $\tilde{\alpha}_{n-1}$=$0i$, for some $i = 1,..,n$. It corresponds to an empty inactive last buffer in the CONWIP loop. Wip remains at zero until the next jump of $\tilde{\alpha}_{n-1}$ towards state 1, which happens at rate r_i (machine M_i is

repaired). Associated with that region are the probability masses:
$P_{0i}^{n-1}(t)$= $\Pr[(x_{n-1}(t)=0)\cap(\tilde{\alpha}_{n-1}=0i)]$, $i=1..n-1$.

B. Forward Kolmogorov equations for hybrid state $\tilde{x}_{n-1}$

Based on the dynamics of hybrid state $\tilde{x}_{n-1}$ developed in A , and the associated functions/ variables, one can develop the corresponding Kolmogorov differential equations. For details on the rules for setting up boundary conditions for this class of hybrid state non diffusion Markov processes, please refer to ([12], [11]). These equations are quite similar to those corresponding to the intermediate storage dynamics in Section III below, and thus are omitted here for lack of space.

III. A MARKOVIAN MODEL OF INTERMEDIATE WIP x_j, $j=1..n-2$

We designate by $\tilde{M}_j$ for $j=1..n-2$ the effective macro machine upstream intermediate wip x_j. In the following, we define its states for some fixed j, $j=1..n-2$.
Here parallel to the definitions in section 2, we define the macro machine modes $\tilde{\alpha}_j$ = **1** and $\tilde{\alpha}_j$ = **0i**, i=1..(n-1). Also, we designate by $v_{\tilde{\alpha}_j}^j$ the velocity of x_j in mode α_j, and we introduce hybrid probability densities $f_{\alpha_j}^j(\lambda,t)$.

A. Defining the hybrid state $\tilde{x}_j$ dynamics

$\tilde{x}_j$ is constructed by concatenation of continuous state x_j with the $\tilde{M}_j$ machine discrete state $\tilde{\alpha}_j$. Three regions with distinct dynamics can be defined as follows:
0<x_j<z:
- for $\tilde{\alpha}_j$ = 1: dx_j/dt =v_1^j = $(\tilde{d}_{n-1}-k)$.
The above is in view of the fact that unless machine M_1 is down, we are always assuming that the loop is blocked and production at M_1 proceeds at the rate at which wip is being drawn out of buffer x_{n-1}. This rate is set to be a constant set at $\tilde{d}_{n-1}$, in accordance with the demand averaging principle, whenever x_{n-1} is active (which is always the case when all machines are operational) . Finally, M_{j+1} pulls parts from M_j at rate k.
- for $\tilde{\alpha}_j$ =0i, $i=1,..,j-1$:
$dx_j/dt = v_{0i}^j = -k+k*Pr[\bigcup_{l=i}^{j-1}(x_l>0)|M_{ioff}]$.
This is in view of the fact that machine M_j will pull wip at rate k provided that any one of the buffers between the failed machine M_i and machine M_j is non empty, while machines in the loop past M_i are all operational, and thus will pull parts at rate k. The rightmost term corresponds to the mean rate of production of machine M_j *conditional* on M_i being off.
- for $\tilde{\alpha}_j$ =0j: dx_j/dt =$v_{0j}^j = -k$
- for $\tilde{\alpha}_j$ =0$(j+1)$:
dx_j/dt =$v_{0(j+1)}^j$= $k*Pr[\bigcup_{l=1}^{j}(x_l>0)|M_{(j+1)off}]+(\tilde{d}_{n-1})*$
$Pr[(x_{n-1}>0)|M_{(j+1)off}]$ $*Pr[\bigcap_{l=1}^{j}(x_l=0)|M_{(j+1)off}]$.
The right hand side term is the mean rate of production at machine M_j. Indeed, if any of the buffers upstream of it are non empty, it will be k; if on the other hand, the latter are *all empty*, and given that we assume for all failures except that of machine M_1 the CONWIP loop remains blocked, it will be by virtue of the demand averaging principle, $\tilde{d}_{n-1}$, provided also buffer $(n-1)$ is non empty. Note that the calculation of probability $\Pr[\bigcap_{l=1}^{j}(x_l=0)|M_{(j+1)off}]$ is based on that of its complement $\Pr[\bigcup_{l=1}^{j-1}(x_l>0)|M_{ioff}]$ under assumptions of events independence: $Pr[\bigcup_{l=1}^{j-1}(x_l>0)|M_{ioff}]=\sum_{l=1}^{j-1}Pr[(x_l>0)|M_{ioff}]$
$-\sum_{q=1,s=2,s\neq q}^{j-1}(Pr[(x_q>0)|M_{ioff}]*Pr[(x_s>0)|M_{ioff}])$
Note also that since machine M_{j+1} is down in this mode, the rate of extraction of wip from buffer j is zero.
- for $\tilde{\alpha}_j$ =0i, $i=(j+2),..,(n-1)$:
dx_j/dt =$v_{(0i)}^j = k*Pr[\bigcup_{l=1}^{j-1}(x_l>0)|M_{ioff}]+\tilde{d}_{n-1}$
$*Pr[x_{n-1}>0)|M_{ioff}]*Pr[\bigcap_{l=1}^{j-1}(x_l=0)|M_{ioff}]-k$.
The same arguments as for the above case hold here, except that there is a non zero rate of extraction of parts k which needs to be accounted for.
$x_j=z$: Under our modeling assumptions, the only way to remain at this level is if all machines in the CONWIP loop except M_{j+1} are operational , i.e. $\tilde{\alpha}_j$=0$(j+1)$. This corresponds to a blocked loop with all wip concentrated at the j^{th} buffer. Wip remains fixed at z until the next jump of $\tilde{\alpha}_j$ which occurs at rate $r_{(j+1)}$, as machine M_{j+1} gets repaired. Associated with that region is the probability mass: $P_z^j(t)$ =Pr[$x_j(t)$ = z].
$x_j=0$: Here, it is crucial to distinguish the case of an empty but *active* j^{th} buffer from that of an empty *inactive* (no supply of parts) buffer. The first situation can be reached from macro machine $\tilde{M}_j$ state $\tilde{\alpha}_j=1$, or $\tilde{\alpha}_j=0i$, $i=(j+2),..,(n-1)$. The second situation can be reached from states $\tilde{\alpha}_j=0i$, $i=1,..,j$. It is a behavior peculiar to CONWIP loops , that once x_j has reached the zero level, it cannot leave it unless machine M_{j+1} fails. Under our modeling assumptions, this means that $x_j=0$ can only be left at rate $\tilde{p}_{(j+1)}$ from discrete state $\tilde{\alpha}_j=1$.
Associated with that region are the probability masses:
$P_{0i}^j(t)$= $\Pr[(x_j(t)=0)\cap\tilde{\alpha}_j=0i]$, $i\neq(j+1)$
$P_1^j(t)$= $\Pr[(x_j(t)=0)\cap\tilde{\alpha}_j=1]$.

B. Forward Kolmogorov equations for hybrid state $\tilde{x}_j$

0<x_j<z:

$$\frac{\partial f_1^j(x,t)}{\partial t} = -v_1^j * \frac{\partial f_1^j(x,t)}{\partial x} - (\sum_{i=1}^{n-1}\tilde{p}_i)*f_1^j(x,t)+(\sum_{i=1}^{n-1}r_i*f_{0i}^j(x,t)) \quad (1)$$

$$\frac{\partial f_{0i}^j(x,t)}{\partial t} = -v_{0i}^j * \frac{\partial f_{0i}^j(x,t)}{\partial x} - r_i*f_{0i}^j(x,t)+\tilde{p}_i*f_1^j(x,t)\forall i=1,..,(n-2) \quad (2)$$

x_j = 0:

$$\frac{d}{dt}P_{0i}^j(t) = -v_{0i}^j*f_{0i}^j(0^+,t)+\tilde{p}_i*P_1^j(t) - r_i*P_{0i}^j(t),\forall i=1,..j,j+2,..,(n-2) \quad (3)$$

$$\frac{d}{dt}P_1^j(t) = -v_1^j * f_1^j(0^+,t) - (\sum_{i=1}^{n-1}\tilde{p}_i) * P_1^j(t) + \sum_{i=1,i\neq j+1}^{n-1}(r_i * P_{0i}^j(t)) \quad (4)$$

x_j = z:

$$\frac{d}{dt}P_z^j(t) = v_{0(j+1)}^j * f_{0(j+1)}^j(z^-,t) - r_{j+1} * P_z^j(t) \quad (5)$$

1) Boundary conditions: The following equations are associated with probability flow balancing at $x_j = z$ and $x_j = 0$.
At $x_j = z$:

$$-v_1^j * f_1^j(z^-,t) = r_{j+1} * P_z^j(t) \quad (6)$$

$$f_{0i}^j(z^-,t) = 0 \forall i = 1..n-1, i \neq (j+1) \quad (7)$$

At $x_j = 0$:

$$\tilde{p}_{j+1} * P_1^j(t) = v_{0(j+1)}^j * f_{0,(j+1)}^j(0^+,t) \quad (8)$$

2) Total probability mass =1
In the steady state, the above system of Kolmogorov partial differential equations becomes a system of coupled differential equations which can be solved to obtain the relevant performance measures. The coupling of the differential equations is weak, and takes place through a number of unknowns equal to $2(n-1)^2$, i.e., a contribution of $2(n-1)$ unknowns per stage in the CONWIP loop. In section V, we capitalize on this particular structure to propose an efficient solution procedure.

IV. A Markovian model of x_T

We designate by $\tilde{M}_T$ the effective aggregate Markovian machine upstream total wip x_T=$\sum_{i=1}^{n-1}(x_i(t))$, and as in section 2 and 3, the macro machine modes $\tilde{\alpha}_T = 1$, $\tilde{\alpha}_T$ =**0i**, $(i = 1..(n-1))$, and associated x_T velocities $v_T^{\tilde{\alpha}_T}$. Also we introduce hybrid probability $f_{\tilde{\alpha}_T}^T(\lambda,t)$.

A. Defining the hybrid state $\tilde{x}_T$ dynamics

Three regions with distinct dynamics can be defined as follows:
0<x_T<z: We designate by $v_T^{\tilde{\alpha}_T}$, the velocity associated with mode $\tilde{\alpha}_T$. More specifically:
- $\tilde{\alpha}_T = 1$: dx_T/dt =v_T^1 = (k -$\tilde{d}_{n-1}$).
The above is a mathematical expression of the fact when all machines in the CONWIP loop are operational, and the loop is not blocked, total wip increases at maximum rate k and decreases at the rate at which x_{n-1} decreases. The latter rate is considered to be constant at $\tilde{d}_{n-1}$ (this is from the demand averaging principle combined with the fact that x_{n-1} is always active when all machines are active).
-$\tilde{\alpha}_T$ =01: dx_T/dt =v_T^{01} = $(-\tilde{d}_{n-1} * Pr[(x_{n-1} > 0)|M_{1off}])$.
-$\tilde{\alpha}_T$ =0i, $i = 2,..,(n-1)$:
dx_T/dt =v_T^{0i} = $(k - \tilde{d}_{n-1} * Pr[(x_{n-1} > 0)|M_{ioff}])$.
The rightmost term above is the mean wip rate of extraction from x_{n-1} conditional on machine M_i being off, according to the demand averaging principle. In this region, one defines the following hybrid probability.

$x_T = z$: Under our modeling assumptions, this level is reached from all states of effective machine $\tilde{M}_T$ except state $\tilde{\alpha}_T = 01$, i.e. whenever machine M_1 fails. It corresponds to a blocked CONWIP loop. x_T will remain at this level unless effective machine state $\tilde{\alpha}_T = 1$ and jumps to $\tilde{\alpha}_T = 01$. This occurs at rate $\tilde{p}_1$. Associated with that region is the probability mass: $P_z^T(t)$ =Pr$[x_T(t) = z]$.
$x_T = 0$: Under our modeling assumptions, the only way to reach this situation is if $\tilde{\alpha}_T$=01. It corresponds to a completely empty CONWIP loop with all machines but M_1 operational. x_T remains at zero until repair of M_1 is achieved in which case $\tilde{\alpha}_T$ jumps from 01 to 1; this occurs at rate r_1. Associated with that region are the probability masses: $P_0^T(t)$= Pr$[(x_T(t) = 0)]$.

B. Forward Kolmogorov equations for hybrid state $\tilde{x}_T$

0<x_T<z:

$$\frac{\partial f_1^T(x,t)}{\partial t} = -v_1^T * \frac{\partial f_1^T(x,t)}{\partial x} - (\sum_{i=1}^{n-1}\tilde{p}_i) * f_1^T(x,t) + (\sum_{i=1}^{n-1} r_i * f_{0i}^T(x,t)) \quad (9)$$

$$\frac{\partial f_{01}^T(x,t)}{\partial t} = \tilde{d}_{n-1} * \frac{\partial f_{01}^T(x,t)}{\partial x} - r_1 * f_{01}^T(x,t) + \tilde{p}_1 * f_1^T(x,t) \quad (10)$$

$$\frac{\partial f_{0i}^T(x,t)}{\partial t} = -v_{0i}^T * \frac{\partial f_{0i}^T(x,t)}{\partial x} - r_i * f_{0i}^T(x,t) + \tilde{p}_i * f_1^T(x,t) \quad ,\forall i = 2..n-1 \quad (11)$$

x_T = 0:

$$\frac{d}{dt}P_0^T(t) = \tilde{d}_{n-1} * f_1^T(0^+,t) - r_1 * P_0^T(t) \quad (12)$$

x_T = z:

$$\frac{d}{dt}P_z^T(t) = -\sum_{i=2}^{n-1}(v_{0i}^T * f_{0i}^T(z^-,t)) + (v_1^T) * f_1^T(z^-,t)) - \tilde{p}_1 * P_z^T(t) \quad (13)$$

1) Boundary conditions: The following equations are associated with probability flow balancing at $x_T = z$ and $x_T = 0$.
At $x_T = z$:

$$\tilde{p}_1 * P_z^T(t) = \tilde{d}_{n-1} * f_{01}^T(z^-,t) \quad (14)$$

At $x_T = 0$:

$$r_1 * P_0^T(t) = (-v_1^T) * f_1^T(0^+,t) \quad (15)$$

$$f_{0i}^T(0^+,t) = 0, \forall i = 2..n-1 \quad (16)$$

2) Total probability mass =1 :

V. ITERATIVE CALCULATION OF SYSTEM SOLUTION

Our proposed solution technique relies on the following observations on the structure of the Kolmogorov equations of the various subsystems: except for the particular case of the coefficient of availability $a_{(n-1)}$ that we shall discuss separately, provided probabilities $Pr[x_j > 0|M_{ioff}]$, and $Pr[x_j = 0|M_{ioff}]$, $\forall i = 1,..,(n-1), j = 1,..,(n-1)$, are assumed to be known, all subsystems, namely those associated with $\tilde{x}_j$, $j = 1,..,(n-1)$, and x_T can be solved *independently*, i.e. as systems of n^{th} order differential equations. Thus, if the vector of these unknown probabilities is initialized sufficiently close to its true value, one could hope that through successive iterations, one would converge to a fixed vector associated with the overall solution.

Further remarks are that the x_T subsystem only depends on $a_{(n-1)}$, and has by itself *no impact* on the rest of the subsystems,while the calculation of $a_{(n-1)}$ can be carried out via a fixed point (provably convergent) algorithm valid for stocks fed through isolated multi state machines and subjected to a constant demand: $a_{(n-1)}$ is initialized at 1 and will uniformly decrease towards its true value (see [8] for further details). Thus summarizing, the idea of the proposed algorithm is as follows: Start with an initial guess of the unknown vector of coupling probabilities; solve for a_{n-1} through the fixed point algorithm applied to the $\tilde{x}_{(n-1)}$ dynamics; feed the result to the *separate* x_j subsystems, $j = 1,..,(n-2)$, to generate a new candidate vector of unknown probabilities; repeat the process until convergence is (hopefully) achieved.

Once convergence in the x_j subsystems, $j = 1,..,(n-1)$, is achieved, use the result to compute mean total storage in the system by solving for the $\tilde{x}_T$ dynamics. Further details can be found in [14].

VI. NUMERICAL RESULTS

In the following, we report on a comparison of predictions of our CONWIP loop approximate aggregate model against results obtained from Monte Carlo simulations. Percentage errors are given for the corresponding quantities. Three transfer lines subjected to a demand rate of $d = 1$ part per unit time,including a 3 machine, a 4 machine and a 5 machine CONWIP loop with an extra machine outside the CONWIP loop have been studied. The lines are completely homogeneous with failure rate $p = 0.05$, repair rate $r = 0.75$, and maximum production rate $k = 2.8$. In Subsection VI-A below, we report on the main performance indicators of the CONWIP loops; more specifically, mean total wip, and coefficient of availability of wip at last loop stage (a measure of service level) are reported. In Subsection VI-B, we illustrate the application of this analysis in the minimal sizing of CONWIP parameter z for achieving a given service level $(a_{(n-1)})$ at a given demand rate.

A. Main performance indicators of the CONWIP loop

The table below summarizes the performance of hybrid dynamic model $\tilde{x}_T$ construed as a predictor of total wip dynamics, for 3, 4, or 5 machine CONWIP loops, with a variable total available storage specified level z. Notice that for 5 machine loops, Monte Carlo simulations on a Pentium(R)4 (CPU 2.60 GHz), took about two days of cpu time, while our approximate mathematical model requires just a few minutes to produce an answer. Both mean total wip level and service level as computed from theory and Monte Carlo simulations are presented with percentage errors. One notices in general

TABLE I

MEAN TOTAL WIP AND SERVICE LEVEL

		MC Simulation with tolerance within 0.009		Theory based estimate		Percentage Error	
$n-1$	z	a_{n-1}	$E[x_T]$	a_{n-1}	$E[x_T]$	a_{n-1}	$E[x_T]$
3	1	0.8580	0.9354	0.9002	0.9391	-4.92	-0.40
3	2	0.9242	1.9012	0.9494	1.9016	-2.73	-0.02
3	3	0.9566	2.8762	0.9702	2.8786	-1.42	-0.08
3	5	0.9881	4.8696	0.9915	4.8625	-0.34	0.15
3	10	0.9995	9.8666	0.9996	9.852	-0.01	0.15
4	2	0.9038	1.9004	0.9350	1.9319	-3.45	-1.66
4	8	0.9972	7.8645	0.9987	7.8625	-0.15	0.03
5	1	0.7998	0.9368	0.8392	0.9560	-4.93	-2.05
5	9	0.9971	8.8661	0.9988	8.8545	-0.17	0.13

an excellent agreement between predictions and the results of Monte Carlo simulations, in particular for mean total wip. Worst case percentage relative errors are on the order of 2, and only so in rare cases corresponding to a combination of small available storage and long lines. These cases also coincide with more significant relative errors in the Monte Carlo simulations themselves. Worst case percentage errors for coefficient of availability are on the order of 5 with a mean around 2. The worst cases combine small storage with long lines. Furthermore, one notices that all errors have the same sign pointing at some small bias, probably due to the way joint probabilities are computed in the dynamics.

B. Minimal sizing of CONWIP parameter z for securing a given service level

In the following, we give an example of how one can use the above analytical model for minimally sizing the storage parameter z so as to secure a given level of service (coefficient of availability of work in process feeding last machine, $a_{(n-1)}$), for a given level d of required parts production per unit time. The idea is simple: $a_{(n-1)}$ is a monotone increasing function of z. Therefore, one can perform a dichotomic search on some segment of z values for which it is known that the minimum value is too small for the required $a_{(n-1)}$ whereas the maximum value is too large. Fig.2 summarizes the successive values of $a_{(n-1)}$ obtained by sequentially modifying the candidate z until convergence occurs within a tolerance of 10^{-6} in 15 steps , for a desired service level of 0.95, in the 4 machine line of Table **??**. Note of course that as soon as the iterated values of the required z had started oscillating between 2 and 3, one could have immediately relied on the integrality constraint to declare that 3 was the smallest integer wip that could meet the required service level. On the other hand, the exact value

of the minimum z computed via the fluid model sitting at 2.37 could be suggestive of a cyclic wip limit periodically moving between 2 and 3 as the minimal storage strategy.

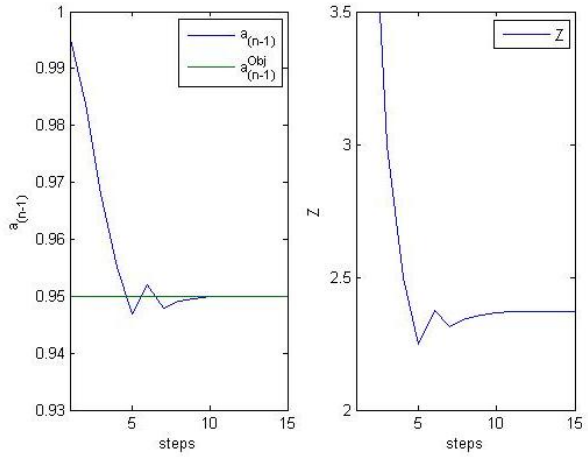

Fig. 2. Required Z for a given coefficient of availability $a_{(n-1)}$

VII. Conclusion

We have developed an approximate mathematical model, amenable to computations, for the evaluation of important performance indicators of CONWIP controlled transfer lines, which can incorporate an arbitrary number of failure prone machines. The model for $a_{(n-1)}$ machine loop corresponds to a concatenation of interconnected aggregate (or macro) machines, with one machine associated to each of the buffer dynamics within the loop, and a machine dedicated to the dynamics of total wip within the loop. The macromachines are coupled each with an individual buffer state to produce a hybrid state vector associated with its set of Forward Kolmogorov equations. In our aggregate modeling, the total wip dynamics is seen to be essentially affected by the reliability statistics of machine M_1, and the service level at buffer (n-1), thus reflecting the view of CONWIP as a form of Kanban imposed on a collection of machines.
By building on our previous modeling work [13], one can show that the results remain valid even for non uniform p_i's but common repair rate, modulo a change in the calculation of $\tilde{p}_i$. The case of non uniform repair rates will be considered in future work. The availability of the current modeling tool, makes it possible to compute minimal storage requirements in a CONWIP controlled loop given a required service level and a fixed demand rate. It can also become part of a tool for the optimization of hybrid Kanban/ CONWIP ([2]) architectures in transfer lines.

ACKNOWLEDGMENTS

This research was funded through a discovery grant from the National Science and Engineering Research Council of Canada.

References

[1] A. M.Bonvik, "Performance analysis of manufacturing systems", *Massachusetts institute of technology, Department of Electrical Engineering and Computer Science*, 1996.
[2] A. M. Bonvik, Y. Dallery, S.B. Gershwin, "Approximate analysis of production systems operated by a CONWIP/finite buffer hybrid control policy", *International Journal of Production Research*, Vol 30 (13), Sept. 2000, pp 2845-2869.
[3] S.Y. Chiang, C.T.Kuo, and S.M.Meerkov, "DT- bottlenecks in serial production lines: Theory and application", *IEEE Transactions Robotics and Automation*, Vol 16, Oct. 2000, pp 567-580.
[4] Y.Dallery and S.B.Gershwin, "Manufacturing flow line systems : A review of models and analytical results", *Queueing Syst.*, Vol 12, 1992, pp 3-94.
[5] Y.Dallery and H.Le Bihan, "An improved decomposition method for the analysis of production lines with unreliable machines and finite buffers", *International Journal of Production Research*, Vol 37 (5), 1999, pp 1093-1117.
[6] S.B.Gershwin, "An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking", *Operations Research*, Vol 35 (2), 1987, pp 291-305.
[7] R. Levantesi, A. Matta and T. Tolio, "Performance evaluation of continuous production lines with machines having different processing times and multiple failure modes", *Performance Evaluation*, Vol 51, 2003, issn 0166-5316, pp 247-268.
[8] J. Sadr, R. P.Malhamé, "Decomposition/aggregation-based dynamic programming optimization of partially homogeneous unreliable transfer lines", *IEEE Transactions Automatic Control*, Vol 49 (1), 2004, pp 68-81.
[9] Y. Frein, C.Commault, Y. Dallery, "Modeling and analysis of closed-loop production lines with unreliable machines and finite buffers", *IIE Transaction*, Vol 28, 1996, pp 545-554.
[10] J.Kimemia, S. B. Gershwin, "An algorithm for the computer control of a flexible manufacturig system", *IIE Trans.* , Dec. 1983, pp 353-362.
[11] R. Malhamé, "A jump-driven markovian electric load model", *Advances in Applied Probability*, Vol 22 (3), Sept. 1990, pp 564-586.
[12] S. El-Férik, R.P. Malhamé, "Padé approximants for the transient optimization of hedging control policies in manufacturing", *IEEE Transactions Automatic Control*, Vol 42 (4), Apr. 1997, pp 440-457.
[13] J. Sadr, R.P.Malhamé, "Unreliable transfer lines: decomposition/ aggregation and optimisation", *Annals of Operations Research*, Vol 125, Jan. 2004, pp 167-190.
[14] F.Z. Mhada, R.P. Malhamé, "Aggregation based approximate performance analysis of CONWIP disciplines in unreliable partially homogeneous transfer lines", *Cahiers du GERAD* , G-2008-21, 2008. http://www.gerad.ca/fichiers/cahiers/G-2008-21.pdf
[15] Stanley B. Gershwin, "Manufacturing systems engineering", Englewood Cliffs, NJ : PTR Prentice Hall, 1994