Discrete Event Dynamic Systems: Theory and Applications 1, (1992): 271-287

Event Rates and Aggregation in Hierarchical Discrete Event Systems*

KEVIN M. PASSINO
Department of Electrical Engineering, The Ohio State University, 2015 Neil Ave., Columbus, OH 43210

PANOS J. ANTSAKLIS
Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556

Received May 7, 1990; Revised November 28, 1990; Accepted June 4, 1991
Editor: P. Varaiya

Abstract. A discrete event system (DES) is a dynamical system whose evolution in time develops as the result of the occurrence of physical events at possibly irregular time intervals. Although many DES's operation is asynchronous, others have dynamics which depend on a clock or some other complex timing schedule. Here we provide a formal representation of the advancement of time for logical DES via interpretations of time. We show that the interpretations of time along with a timing structure provide a framework to study principles of the advancement of time for hierarchical DES (HDES). In particular, it is shown that for a wide class of HDES the event rate is higher for DES at the lower levels of the hierarchy than at the higher levels of the hierarchy. Relationships between event rate and event aggregation are shown. We define a measure for event aggregation and show that there exists an inverse relationship between the amount of event aggregation and the event rate at any two successive levels in a class of HDES. Next, we study how to design the timing structure to ensure that there will be a decrease in the event rate (by some constant factor) between any two levels of a wide class of HDES. It is shown that if the communications between the various DES in the HDES satisfy a certain admissibility condition then there will be a decrease in the event rate. These results for HDES constitute the main results of this paper, since they provide the first mathematical characterization of the relationship between event aggregation and event rates of the HDES and show how to design the interconnections in a HDES to achieve event rate reduction. Several examples are provided to illustrate the results.

Key Words: Discrete Event Systems, Hierarchical Discrete Event Systems, Event Rate, Aggregation, Manufacturing Systems

1. Introduction

In our main results we show that the interpretations of time which characterize the advancement of time in DES (introduced in Section 2) along with a timing structure provide a framework to study principles of the advancement of time for hierarchical DES (HDES). In Theorem 3.1 it is shown that for a wide class of HDES the event rate is higher for DES at the lower levels of the hierarchy than at the higher levels of the hierarchy. Relationships between event rate and event aggregation are shown. We define a measure for event aggregation and show that a high amount of event aggregation will result in a much lower event rate at higher levels in a certain class of HDES while a low amount of event aggregation

*The authors gratefully acknowledge the partial support of the Jet Propulsion Laboratory. Please address all correspondence to K. Passino (email: passino@eagle.eng.ohio-state.edu).

will result in higher event rates (Theorem 3.2). Event rate reduction in HDES is often desirable, so the processors implementing the higher level controls are permitted adequate time before they must attend to the lower level systems. Next, we study how to design the timing structure to ensure that there will be a decrease in the event rate (by some constant factor) between any two levels of a wide class of HDES. It is shown that if the communications between the various DES in the HDES satisfy a certain admissibility condition then there will be a decrease in the event rate (Theorem 3.3). Hence, we show that one only needs to restrict the interconnections in the HDES to achieve event rate reduction. The results are illustrated with a conventional discrete event control system, a hybrid dynamical system, and a manufacturing system. Some of the results in this paper were originally reported in [Knight and Passino 1990; Passino 1989; Passino and Antsaklis 1988; 1991].

We focus on timing characteristics of single DES or HDES which have as components DES that can be accurately modeled with

$$P = (X, U, Y, \delta, \lambda, X_0), \tag{1}$$

where

X is the set of plant states x
U is the set of plant inputs u
Y is the set of plant outputs y
δ: $U \times X \rightarrow \mathsf{P}(X)$ is the plant state transition function ($\mathsf{P}(X)$ denotes the power set of X)
λ: $U \times X \rightarrow Y$ is the plant output function
$X_0 \subset X$ is the set of possible initial plant states

The plant state transition function (a partial, point to set function) specifies for each current input u and state x the set of possible next states $x' \in \delta(u, x)$. The output function specifies, for the current input u and state x, the current output symbol $y = \lambda(u, x)$. Formally, P is equivalent to a directed graph with node set X and edges $x \rightarrow x'$ labeled with u/y for each triple (u, x, x') such that $x' \in \delta(u, x)$ and $y = \lambda(u, x)$. The model P is similar to a standard automaton, but X, U, and Y are not required to be finite. A *run* of P is defined as a sequence of triples (u_0, x_0, y_0), (u_1, x_1, y_1), (u_2, x_2, y_2), ..., such that $x_0 \in X_0$, u_0 is the initial input, $x_{k+1} \in \delta(u_k, x_k)$, and $y_k = \lambda(u_k, x_k)$. Notice that since it is possible that $\delta(u_k, x_k) = \emptyset$ for all $u_k \in U$ at some $x_k \in X$, a run may have a finite length. We note that our results are not restricted solely to the use of the DES model (1). The above model was chosen so that the results here would directly apply to a wide class of systems that can be represented with "logical DES models" (e.g., general and extended Petri nets [Peterson 1981], finite automata, and other DES models [Ramadge and Wonham 1987; Zhong and Wonham 1990]). The development of results analogous to ours for hierarchical timed or performance models for DES is an important research direction.

2. Characterizing the Advancement of Time In DES

When a physical plant is modeled via (1), the meaning of the advancement of time must be defined. If Z is an arbitrary set, then Z^* denotes the set of all finite strings of elements for Z. If Z and Z' are arbitrary sets, then $Z^{Z'}$ denotes the set of all functions mapping Z' to Z. Let $\mathbf{N}$ denote the set of natural numbers. In order to discuss timing issues for P, an *index set* J and *index sequences*

$$\alpha \in J^* \cup J^{\mathbf{N}}$$

are utilized similar to the approach in [Sain 1981]. The index set J is thought of as a set of times. Let $\mathbf{R}^+$ denote the set of strictly positive real numbers and $\mathbf{R}_+ = \mathbf{R}^+ \cup \{0\}$, the set of nonnegative reals. Note that $\mathbf{N}$ or $\mathbf{R}$ could be candidates for the set J. For convenience, we assume that $J = \mathbf{R}_+$. The index sequences $\alpha \in 2\ J^* \cup J^{\mathbf{N}}$ are sequences of time instants that can be of finite or infinite length. For $\alpha \in J^* \cup J^{\mathbf{N}}$ let $|\alpha|$ denote the cardinality of the size of the string α. Note that either α: $\mathbf{N} \to J$ or α: $[0, a] \to J$, where $[0, a] \subset \mathbf{N}$, and $\alpha(k)$ simply denotes an element in J. An index sequence (function) $\alpha \in J^* \cup J^{\mathbf{N}}$ is said to be *admissible* if

(i) it is order preserving; i.e.,
  (a) if $\alpha \in J^{\mathbf{N}}$, then for all $k_1, k_2 \in \mathbf{N}$, $k_1 \leq k_2$ implies $\alpha(k_1) \leq \alpha(k_2)$;
  (b) if $\alpha \in J^*$, then for all $k_1, k_2 \in \mathbf{N}$ with $k_1, k_2 \in [0, |\alpha| - 1]$, $k_1 \leq k_2$ implies $\alpha(k_1) \leq \alpha(k_2)$, and
(ii) it is injective and if $\alpha \in J^{\mathbf{N}}$ then $\alpha(k) \to \infty$ as $k \to \infty$.

Following [Sain 1981], the state of the plant $x \in X$ is associated with the index $\alpha(k)$ for some $\alpha \in J^* \cup J^{\mathbf{N}}$ and is denoted $x(\alpha(k))$, meaning "the state at time $\alpha(k)$." Similarly, iputs $u \in U$ and outputs $y \in Y$ are associated with that same index and denoted $u(\alpha(k))$ and $y(\alpha(k))$, respectively. The transition to a state in the set $\delta(u, x)$ can be thought of as leading to the next state, with "next" quantified with the index sequence α as $\alpha(k + 1)$. With this, the transition function is given as $x(\alpha(k + 1)) \in \delta(u(\alpha(k)), x(\alpha(k)))$, which is often abbreviated $x_{k+1} \in \delta(u_k, x_k)$. Similarly, the output is often denoted by $y_k = \lambda(u_k, x_k)$ for $k \in \mathbf{N}$. Each run of $P(u_0, x_0, y_0), (u_1, x_1, y_1), \ldots$ has an associated index sequence $\alpha \in J^* \cup J^{\mathbf{N}}$, $\alpha = \alpha(0), \alpha(1), \ldots$, specifying the time instants at which the triples are defined. Notice that if for some $x_k \in X$ and all $u_k \in U$, $\delta(u_k, x_k) = \emptyset$, then $\alpha(k + 1)$ is undefined and in this case α has finite length (in this situation we say that P is "deadlocked").

A DES often *activates* or *triggers* other DES to act. For instance, in the case where P represents a plant, P may trigger a controller to generate an input to P. In this case, the trigger often represents certain changes that occur in the plant. For instance, *events* can be used as the trigger. Similar to [Ramadge and Wonham 1987], we let $E \subset X \times X$ denote the set of *events* e, where

$$E = \{(x, x') \in X \times X : x' \in \delta(u, x)\}. \tag{3}$$

An event $e = (x, x')$ is said to *occur* if the state transition from x to $x' \in \delta(u, x)$ takes place. For convenience we shall assume that the event occurs (is defined) at the time instant $\alpha(k + 1)$ where the next state is defined. Due to the injective part of the admissibility requirement for α the variables x, u, and y are defined at time instants which are distinct from one another. By condition (ii) of the admissibility requirement when state transitions occur it is guaranteed that time will advance (although it may be a very small amount) and if an infinite number of events occur this will take an infinite amount of time. The other important implication is that using the definition of events E in (3) it is automatically assumed that events occur at distinct times; i.e., simultaneous events are not allowed because the index sequences are required to be admissible. Suppose for a moment that condition (ii) of the admissibility requirement is omitted, so that for $\alpha \in J^* \cup J^{\mathbf{N}}$ it is possible that $\alpha(k + 1) = \alpha(k)$ for any $k \in \mathbf{N}$ such that $\alpha(k)$ and $\alpha(k + 1)$ are defined. This will allow events to occur simultaneously at a particular time instant. In fact, for $\alpha \in J^{\mathbf{N}}$ it will allow even an infinite number of events to occur at one time instant, resulting in the possibility that time will not advance. Normally, to treat simultaneous events, only a finite number of events are allowed to occur at a single time instant; hence, other events representing the case that "several events occur at once" can often be defined. So the problem of dealing with simultaneous events is often transformed to the case where only a single event occurs at each time instant, so that time is guaranteed to advance and admissibility can be assumed (e.g., this can be done for Petri nets [Peterson 1981]).

The pair $I = (A, J)$, where J is an index set and $A \subset J^* \cup J^{\mathbf{N}}$, will be referred to as an *interpretation of time* since it specifies the meaning of the advances in time for the occurrence of state transitions; i.e., it specifies the time instants where the variables of the DES P are defined. In general, a system P is said to have a particular interpretation of time $I = (A, J)$ as long as the time instants associated with the elements of the runs of P are elements of J and the index sequences associated with the runs of P are elements of A. The *admissible interpretation of time* will be denoted $I_{\text{ad}} = (A_{\text{ad}}, J_{\text{ad}})$, where J_{ad} is an index set and

$$A_{\text{ad}} = \{\alpha \in J_{\text{ad}}^* \cup J_{\text{ad}}^{\mathbf{N}} : \alpha \text{ is admissible}\}. \tag{4}$$

Most often we can choose $J_{\text{ad}} = J = \mathbf{R}_+$ and this is what we will assume here. It is common to discuss the timing characteristics of DES relative to a clock. By a *clock* we mean a device which has a fixed interval $T \in \mathbf{R}^+$ between *ticks* and which does not stop ticking (if there is deadlock, the clock keeps ticking but no events occur). Next we provide definitions for several standard interpretations of time used in DES studies:

DEFINITION 2.1. The *asynchronous interpretation of time* is $I_a = (A_a, J_a)$, where $J_a = \mathbf{R}_+$ and

$$A_a = \{\alpha \in A_{\text{ad}} : \alpha(0) = 0\}.$$

According to convention $J_a = J_{\text{ad}} = \mathbf{R}_+$, with the time instant of zero corresponding to the case where no state transitions have occurred. Here I_a represents the situation where

the plant P is asynchronous (out of sync, not synchronous) with a clock. For I_a the time instants at which the plant variables are defined are at nonuniform (irregular) distances from one another along the time line $\mathbf{R}_+$. Notice that $A_a \subset A_{ad}$ if $J_{ad} = J_a$, so that asynchronous interpretations of time are admissible interpretations of time, but not necessarily vice versa.

DEFINITION 2.2. The *partially asynchronous interpretation of time* is $I_{pa} = (A_{\gamma\beta}, J_{pa})$, with $J_{pa} = \mathbf{R}_+$ and $A_{\gamma\beta} = \{\alpha \in A_a : \alpha(k) + \gamma \leq \alpha(k+1) \leq \alpha(k) + \beta\}$ for $\gamma, \beta \in \mathbf{R}^+$, where $\beta \geq \gamma$.

Here I_{pa} represents the case where we know that the time instant where the next state is defined is constrained to occur at least γ, and no more than β time units later. Notice that $A_{\gamma\beta} \subset A_a$ if $J_{pa} = J_a$; that is, partially asynchronous interpretations of time are asynchronous interpretations of time, but not necessarily vice versa.

DEFINITION 2.3. The *general synchronous interpretation of time* is $I_s = (A_T, J_s)$ with $J_s = \mathbf{R}_+$ and $A_T = \{\alpha \in A_a : \alpha(k+1) = \alpha(k) + nT$, where $n \in \mathbf{N}-\{0\}\}$ with $T \in \mathbf{R}^+$.

For the general synchronous interpretation of time, the time instants at which the plant variables x, u, y are defined are at distances nT, for $n \in \mathbf{N}-\{0\}$, from one another along the time line $\mathbf{R}_+$. Notice that, in general, a state transition may not occur between any two particular ticks of the clock (since $n > 0$), and that after each state transition occurs another may not eventually occur. When $n = 1$, we shall refer to I_s simply as the *synchronous interpretation of time*. For the synchronous interpretation of time it is not necessarily the case that $|A_T| = 1$ since any finite length index sequence may be possible. Notice that if $n = 1$, $A_T \subset A_{\gamma\beta}$ provided that $\gamma \leq T \leq \beta$ so that the synchronous interpretation of time is a partially asynchronous interpretation of time, but not necessarily vice versa.

Note that it is not, in general, required that the timing characteristics of the plant be defined relative to a clock although they are often treated as such. In general, in a manner similar to that with the clock, the plant may be in sync (out of sync) with changes in other systems. Also notice that for asynchronous time if $\alpha_a \in A_a$ and the current time is $\alpha_a(k)$, then the next time is $\alpha_a(k+1) = \alpha_a(k) + r$, where $r \in \mathbf{R}^+$. On the other hand, for general synchronous time if $\alpha_s \in A_T$ the current time is $\alpha_s(k)$, then the next time is $\alpha_s(k+1) = \alpha_s(k) + r'$, where $r' \in R_1$ and $R_1 = \{n\mathrm{T} : n \in \mathbf{N}-\{0\}\}$ for a given $T \in \mathbf{R}^+$. Since R_1 is equinumerous with $\mathbf{N}-\{0\}$, a proper subset of $\mathbf{R}^+$, it is the case that card $(\mathbf{R}^+)$ > card (R_1). Hence, no matter what the time interval T, the number of possible "next" time instants is always greater if an asynchronous interpretation of time is used rather than a synchronous one. This helps to clarify the intuitions we have about the relationships between the synchronous and asynchronous interpretations of time. It is clear that synchronous time cannot be used if the underlying system can only be accurately represented with an asynchronous intepretation. However, it is possible that the synchronous interpretation of time with T very small may result in an accurate model for some asynchronous systems. This will depend on the particular plant to be modeled and the design objectives to be studied.

3. Timing Characteristics of Hierarchical DES

The formation of a control theory for HDES is just beginning [Zhong and Wonham 1988, 1989, 1990], even though such systems occur quite frequently. Some principles of the evolution of time in hierarchical systems have been postulated but not fully investigated [Albus et al. 1981; Antsaklis et al. 1989; Mesarovic et al. 1970; Passino and Antsaklis 1988; Saridis 1983; Valavanis 1986]. As in [Gershwin 1989] what these researchers have recognized is that "systems usually operate at higher rates at the lower levels in a hierarchical system." We shall verify this intuition for a wide class of HDES here.

3.1. A Hierarchical DES Model

We shall focus on HDES that have as components two types of DES, G_j, $1 \leq j \leq m$, and P_i, $1 \leq i \leq n$, all defined via (1) except with different timing characteristics. We think of the P_i as modeling the physical system, and hence its timing characteristics are given by the interpretation of time for the respective portions of the system modeled by each P_i. We think of the G_j as modeling controllers and hence as having timing characteristics that are influenced by the physical system and the other component controllers in the HDES. We introduce what we call a *timing structure*, which will define how the various components of the HDES influence (are influenced by) the timing characteristics of other components of the HDES. The definition of the timing structure is based on the interpretations of time defined in Section 2 and what will be called *input* and *output triggers*. Each P_i, $1 \leq i \leq n$, in the HDES has timing characteristics that are simply specified via their own interpretation of time denoted with $I_{pi} = (A_{pi}, J_{pi})$. Roughly speaking, each G_j, $1 \leq j \leq m$, has timing characteristics that depend on other P_i and G_ℓ via the timing structure as we now discuss in more detail.

Let E_{pi} denote the set of events for P_i, and E_{gj}, the set of events for G_j both defined in a similar manner to the events E for P in (3). Let C_{pi} (C_{gj}) denote the set of *communications* that can be transmitted from P_i (G_j) via the timing structure to other G_ℓ. The *output triggers* for the P_i (resp. for G_j) are defined via

$$\phi_i : E_{pi}^* \to C_{pi}, \quad 1 \leq i \leq n \quad (\text{resp., } \psi_j : E_{gj}^* \to C_{gj}, \quad 1 \leq j \leq m) \tag{5}$$

(or restrictions of these maps). The output triggers define how the P_i and G_j connect to other components of the HDES to influence their timing characteristics. If $\phi_i(\mathbf{e}) = c$ or $\psi_j(\mathbf{e}') = c'$, then c and c' are communications that are said to *occur* due to the occurrence of event string $\mathbf{e}$ or $\mathbf{e}'$ ($\mathbf{e}$ triggers communication c). We will have occasion below to utilize a null communication $\emptyset$ which cannot cause an event occurrence in any other DES in the HDES. We use the standard notation for *concatenation*; e.g., if $\mathbf{e}, \mathbf{e}' \in E_{pi}^*$, then $\mathbf{ee}'$ denotes the concatenation of $\mathbf{e}$ and $\mathbf{e}'$. The time instant at which the communication c (c') occurs is the same time instant that $a \in E_{pi}$ ($a' \in E_{gj}$) occurs where $\phi_i(\mathbf{e}a) = c$ ($\psi_j(\mathbf{e}'a') = c'$). The *input triggers* for the G_j are defined by the τ_j maps for j, $1 \leq j \leq m$, where

$$\tau_j : C_{p1} \times \cdots \times C_{pn} \times C_{g1} \times \cdots \times C_{gj-1} \times C_{gj+1} \times \cdots \times C_{gm} \to \{0, 1\}, \quad (6)$$

and $\tau_j(\cdot) = 1$ $(= 0)$ indicates that an event $e_{gj}(\alpha(k + 1)) \in E_{gj}$, where $e_{gj}(\alpha(k + 1)) = (x_{gj}(\alpha(k)), x_{gj}(\alpha(k + 1)))$ is forced (not) to occur in G_j. The τ_j indicate which P_i and G_ℓ communicate with G_j via the timing structure; hence τ_j describes how various components of the HDES cause events in $G_j(e_{gj}(\alpha(k + 1)))$ to occur at time instant $\alpha(k + 1)$. Equation (6) indicates the *form* for the τ_j maps; the absence of a C_{pi} or $C_{g\ell}$ from the cross product in the domain of τ_j indicates that P_i or G_ℓ does not communicate with G_j via τ_j. It is assumed that the τ_j maps form a "tree structured" timing structure as we describe next.

Let each DES component P_i, $1 \leq i \leq n$, or G_j, $1 \leq j \leq m$, of the HDES represent a *node* (e.g., denoted with boxes as in Figure 1) of a directed graph $\mathcal{S}$ and let the τ_j define the *arcs* (e.g., denoted with shaded arcs in Figure 1) that connect the P_i and G_ℓ to G_j in the following manner: Let $i1$, $i2$, $\ell 1$, $\ell 2 \in \mathbf{N}$. If there exists $i(\ell)$ such that $\tau_j : C_{pi1} \times \cdots \times C_{pi} \times \cdots \times C_{pi2} \times C_{g\ell 1} \times \cdots \times C_{g\ell} \times \cdots \times C_{g\ell 2} \to \{0, 1\}$, then there exists an arc pointing from P_i *to* G_j (G_ℓ to G_j). In this paper we assume that the HDES has a *tree structured* timing structure; i.e., we assume that $\mathcal{S}$ has no *closed cycles.* In this way we eliminate the possibility that some G_j can directly force its own events to occur via the timing structure. Although this limits the manner in which the various P_i and G_j can influence the timing characteristics of other G_ℓ, it does not restrict the manner in which the inputs and outputs of the various P_i and G_j are connected. Notice that the P_i, $1 \leq i \leq n$, are the "leaves" of the tree structured timing structure.

Intuitively, the ϕ_i and ψ_j specify what each DES will communicate (C_{pi} and C_{gj}) to the other DES in the HDES. The τ_j define *communication channels* (the arcs and *paths* in $\mathcal{S}$) and where the communications are distributed in the HDES. Next, we define the time instants at which events occur when they are forced to do so by other DES components of the HDES via the timing structure.

Whereas the interpretation of time is always specified for the P_i, $1 \leq i \leq n$, the interpretations of time for the G_j are specified in terms of the other G_ℓ and the P_i via the timing structure. Let $\alpha_{cpi}(k + 1)$ and $\alpha_{cg\ell}(k' + 1)$ denote the time instants at which communications $c_{pi} \in C_{pi}$ and $c_{g\ell} \in C_{g\ell}$ occur, respectively. Suppose that at some time instant $\alpha_{gj}(k + 1)$, $\tau_j(\cdot) = 1$ so that $e_{gj}(\alpha_{gj}(k + 1)) \in E_{gj}$ occurs. This time instant at which $e_{gj}(\alpha_{gj}(k + 1)$ occurs is given by

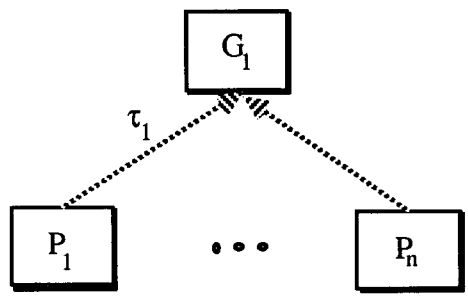

Figure 1. Hierarchical DES with single-branch.

$$\alpha_{gj}(k+1) = \max\{\alpha_{cpi}(k+1), \alpha_{cg\ell}(k'+1) : \exists \text{ an arc in } \mathcal{S} \text{ from } P_i \text{ or } G_\ell \text{ to } G_j\} \quad (7)$$

and corresponds to the time instant at which the last communication accessible to G_j occurred and caused $\tau_j(\cdot) = 1$. Each time a communication occurs which forces $\tau_j(\cdot) = 1$, an event occurs in G_j; hence the "1" represents a *pulse* sent to G_j via τ_j which forces an event to occur. Hence, if $\tau_j(\cdot)$ is set equal to 1 at some time instant, an event in G_j must occur at that time instant (unless G_j is deadlocked); if every communication in a sequence of communications all cause $\tau_j(\cdot) = 1$, then there is one event occurrence in G_j for each communication in the sequence. The interpretation of time for any G_j is found by executing all possible runs (in all possible orders) of the P_i, $1 \leq i \leq n$, and G_ℓ for which there exists a path in $\mathcal{S}$ from P_i or G_ℓ to G_j. Then via (7), the time instants and hence index sequences and interpretations of time for the G_j are specified. We shall study HDES where there is at least one P_i and the intepretations of time for the G_j can be uniquely defined in terms of the P_i.

Note that although we consider only tree structured timing structures we place no restrictions on the manner in which the DES inputs U and outputs Y are connected. This allows our results to apply to a large class of HDES with a wide variety of input/output connecting structures. Tree structured timing structures allow us to study properties of what has been called a "time scale hierarchy" [Antsaklis et al. 1989]. In this hierarchy a DES component is "higher in the hierarchy" than another DES component if its timing characteristics can be influenced by the other DES (i.e., there exists a path in $\mathcal{S}$ from one to the other).

3.2. Lower Event Rates at Higher Levels in the HDES

To analyze the timing characteristics of HDES we study one fundamental component (shown in Figure 1) of the HDES defined above. Even though we consider only P_i at the lower level, it requires only a simple modification to consider a mix of P_i and G_j at the lower level, and all of our results below are still valid. Moreover, our results easily generalize to the fully interconnected HDES by repeated application of the derived relationships which pertain to the two levels in Figure 1.

Let the admissible interpretation of time for P_i be $I_{pi} = (A_{pi}, J_{pi})$, $1 \leq i \leq n$, with $J_{pi} = \mathsf{R}_+$, and for G_1 let it be $I_{g1} = (A_{g1}, J_{g1})$. For any possible run made by the P_i with an index sequence $\alpha_{pi} \in A_{pi}$, the corresponding run in G_1 has index sequence denoted by $\alpha_{g1} \in A_{g1}$.

DEFINITION 3.1. The *event occurrence rate* (*event rate*) in P_i or G_j is the number of events that occur in the time interval $T_u = (r_1, r_2]$, where $(r_1, r_2] \subset \mathsf{R}^+$, and it will be denoted $\#(P_i, T_u)$ and $\#(G_j, T_u)$, respectively.

Notice that if P_i has a synchronous interpretation of time with $T \in \mathsf{R}^+$ and we choose T_u such that $|r_2 - r_1| = T$, then $\#(P_i, T_u) = 1$; i.e., there is one event occurrence in the time interval T_u no matter what the values of r_1 and r_2 are. If P_i has an asynchronous

interpretation of time, then no matter how T_u is chosen it is possible that $\#(P_i, T_u) = 0$, since we cannot guarantee that an event will occur in the given time interval T_u. In fact, we do not know how many events will occur in T_u. It would appear that our definition of event rate is too restrictive. This is, however, not the case, since the focus here is on *comparing* the event rates of different DES components in the HDES, and this comparison is made relative to T_u, an interval of the real time line.

THEOREM 3.1.

$$\sum_{i=1}^{n} \#(P_i, T_u) \geq \#(G_1, T_u) \geq 0 \text{ for all } T_u.$$

Proof. Suppose that an event $e \in E_{pi}$ occurs at time $\alpha_{pi}(k)$ in some P_i, $1 \leq i \leq n$, resulting in $\tau_1(\cdot, \cdot, \ldots, \phi_i(\mathbf{e}e), \ldots, \cdot, \cdot) = 1$. If another event $e' \in E_{p\ell}$ occurs in some P_ℓ at time $\alpha_{p\ell}(k')$ resulting in $\tau_1(\cdot, \cdot, \ldots, \phi_\ell(\mathbf{e}'e'), \ldots, \cdot, \cdot) = 0$, then the index sequence α_{g1} will contain $\alpha_{pi}(k)$ but not $\alpha_{g\ell}(k')$. If $[\alpha]$ denotes the set of elements that make up α, then $[\alpha_{g1}] \subset \bigcup_i [\alpha_{pi}]$, so clearly for any T_u the relationship holds (even if there is deadlock in any P_i or simultaneous events occurring in any number of P_i).

Theorem 3.1 states the intuitively clear fact that the timing structure can *mask* events and hence remove the time instants at which events occur in higher levels of the hierarchy. This means that the event rate is lower in DES at the higher levels of the HDES and higher in lower levels of the HDES no matter what the interpretations of time are for the P_i, $1 \leq i \leq n$.

Remark 3.1. Repeated application of Theorem 3.1 to the *multi-level hierarchy* in Figure 2 results in $\#(P_1, T_u) \geq \#(G_1, T_u) \geq \cdots \geq \#(G_m, T_u) \geq 0$ for all T_u. This result supports the studies in [Gershwin 1989], where the author assumes that the event rates can

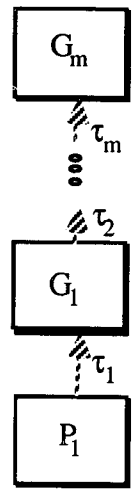

Figure 2. Multi-Level hierarchical DES with m + 1 levels.

be split into "spectra" according to the level in the hierarchy. It also shows that in the more general case, e.g., for P_1 with any admissible interpretation of time, the event rates in DES at the higher levels are also less than or equal to the event rates at the lower levels.

Example 3.1. (*Conventional Discrete Event Control System*) Consider the controlled DES shown in Figure 3. We have $\phi_1 : E_{p1}^* \rightarrow C_{p1}$ and $\tau_1 : C_{p1} \rightarrow \{0, 1\}$, and for the standard control configuration it is most often assumed that for all $\mathbf{e} \in E_{p1}^*$ such that $\mathbf{e} = \mathbf{e}'e$ ($e \in E_{p1}$), $\tau_1(\phi_1(\mathbf{e}'e)) = 1$, so that each time an event occurs in P_1, G_1 is forced to act by having an event in G_1 occur (it is normally assumed that one always exists). Clearly, then if $I_{p1} = (A_{p1}, J_{p1})$ is the interpretation of time for P_1 and $I_{g1} = (A_{g1}, J_{g1})$ for G_1 where $J_{g1} = J_{p1}$, then $A_{g1} = A_{p1}$. The interpretaion of time for the plant and controller are the same. In this way we think of specifying the interpretation of time for G_1 by I_{p1} and P_1 via τ_1 and ϕ_1. Via Theorem 3.1, for general ϕ_1 and τ_1 we see that we can expect fewer events to occur in G_1 than in P_1, since P_1 may not communicate the occurrence of an event or G_1 may not recognize the communication.

Example 3.2. (*Hybrid Dynamical System*) Let $n = 1$ in Figure 1 and for $P_1 = (X_p, U_p, Y_p, \delta_p, \lambda_p, X_{0p})$ let $X_p = \mathbf{R}^{n_p}$, $U_p = \mathbf{R}^{m_p}$, $Y_p = \mathbf{R}^{r_p}$, and I_{p1} be an admissible interpretation of time, so that P_1 is a model for a nonlinear discrete time system. For instance, P_1 could represent a zero-order-hold followed by a nonlinear continuous time system and a sampler. We think of G_1 as our DES model. For instance, G_1 can model any system that can be represented by a General or Extended Petri net [Peterson 1981], and P_1 and G_1 constitute a hybrid dynamical system. Theorem 3.1 shows that if the event rate for P_1 is the number of state transitions per unit time, then the event rate in the higher level DES model G_1 will be lower no matter how the communications are defined.

Remark 3.1 and Examples 3.1 and 3.2 illustrate the generality of Theorem 3.1; the result applies to hierarchical DES currently being studied (in additional to the work in [Gershwin 1989] the result also applies to the work in [Zhong and Wonham 1988, 1989, 1990]),the standard discrete event control systems, and to hybrid dynamical systems. Example 3.3 (a manufacturing system) in Section 3.4 is used to further illustrate the use of Theorem 3.1.

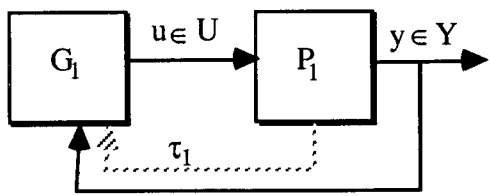

Figure 3. Discrete event control system.

3.3. *Event Rates and Aggregation in a Class of HDES*

Next we study how aggregation affects the event rates in HDES. Again we shall focus on the HDES component shown in Figure 1, but note that the result easily generalizes to a fully interconnected HDES. Let $E_{ai}^* \subset E_{pi}^*$ and $\phi_i' : E_{ai}^* \to C_{pi}$ for all i, $1 \leq i \leq n$, denote restrictions of ϕ_i. We use ϕ_i' maps for aggregation rather than the ϕ_i; if $\mathbf{e} \in E_{pi}^*$, $\mathbf{e} \notin E_{ai}^*$, then ϕ_i is said to *ignore* $\mathbf{e}$ (rather than *mask* $\mathbf{e}$). Let $B_j \subset \mathbf{N} - \{0\}$ for j, $1 \leq j \leq n$.

DEFINITION 3.2. $\{P_j = (X^j, U^j, Y^j, \delta^j, \lambda^j, X_{0j}), \phi'_j : 1 \leq j \leq n\}$ satisfies the $(\pi^1, \pi^2, \ldots, \pi^n)$-*event aggregation property* if for each j, $1 \leq j \leq n$,

(i) there exists a family of sets $X_{ij} \subset X^j$, $i \in B_j$ such that
  (a) $X_{ij} \cap X_{kj} = \emptyset$ for all $i \neq k$, and $X_{0j} \cap X_{ij} = \emptyset$ for $i \in B_j$;
  (b) if P_j first enters a state $x \in X_{ij}$ for some $i \in B_j$, it will take (for all possible runs) at least $\pi^j > 0$ state transitions before the state of P_j, say x', is such that $x' \notin X_{ij}$;

(ii) $\phi'_j : E_{aj}^* \to C_{pj}$, where $E_{aj}^* = \{\mathbf{e} \in E_{pj}^* : \mathbf{e} = \mathbf{e}'e, \mathbf{e}' \in E_{aj}^*$ and $e = (x, x')$ with $x \in X_{ij}$, $x' \notin X_{ij}$ for some $i \in B_j\}$.

THEOREM 3.2. If $\{P_j = (X^j, U^j, Y^j, \delta^j, \lambda^j, X_{0j}), \phi'_j : 1 \leq j \leq n\}$ satisfies the $(\pi_1, \pi_2, \ldots, \pi^n)$-event aggregation property and $T_u = (r_1, r_2]$ and $|r_2 - r_1|$ is sufficiently large, then

$$\sum_{j=1}^{n} \left\{ \frac{\#(P_j, T_u)}{\pi^j} + 1 \right\} \geq \#(G_1, T_u). \tag{8}$$

Proof: To prove the Theorem it must be shown that there exists a time interval T_u such that (8) holds under the stated assumptions. It is first shown how to construct such a T_u. Assume that a run is made in each of the P_j, $1 \leq j \leq n$ and that the corresponding index sequences are $\alpha_{pj} \in A_{pj}$; the corresponding run in G_1 has index sequence $\alpha_{g1} \in A_{g1}$. Let $\alpha_{g1}(k')$ and $\alpha_{g1}(k' + 1)$ be two elements on some $\alpha_{g1} \in A_{g1}$ such that $|\alpha_{g1}(k' + 1) - \alpha_{g1}(k')|$ is an upper bound on the time between sequential events (that actually occurred) in G_1 for all possible runs in the P_j. Let $T_u = (r_1, r_2]$ such that $|r_2 - r_1| > |\alpha_{g1}(k' + 1) - \alpha_{g1}(k')|$. Next, it is shown that this choice of T_u results in the satisfaction of (8). Assume that $\ell > 0$ events occur in G_1 in T_u, where $\ell = \ell_1 + \ell_2 + \cdots + \ell_n$ and ℓ_j is the number of events in G_1 that occurred due to a communication from P_j (if simultaneous events occur in P_j and P_k, $j \neq k$, causing an event occurrence in G_1, then the event occurrence can be attributed to either P_j or P_k and counted only once in ℓ). For each ℓ_j, at least $(\ell_j - 1)\pi^j$ events have occurred in P_j, $1 \leq j \leq n$. Therefore $\#(P_j, T_u) \geq (\ell_j - 1)\pi^j$ so that

$$\sum_{j=1}^{n} \left\{ \frac{\#(P_j, T_u)}{\pi^j} + 1 \right\} \geq \sum_{j=1}^{n} \ell_j = \ell = \#(G_1, T_u). \tag{9}$$

Clearly if any P_j deadlocks it can be the case that $\ell_j = 0$, but relationship (8) still holds.

COROLLARY 3.1. If $\{P_j = (X^j, U^j, Y^j, \delta^j, \lambda^j, X_{0j}), \phi'_j : 1 \leq j \leq n\}$ satisfies the $(\pi^1, \pi^2, \ldots, \pi^n)$-event aggregation property and $T_u = (r_1, r_2]$ with $r_1 = 0$, then for all $r_2 > 0$,

$$\sum_{j=1}^{n} \left\{ \frac{\#(P_j, T_u)}{\pi^j} \right\} \geq \#(G_1, T_u). \tag{10}$$

Remark 3.2. For the multi-level HDES in Figure 2, if the (π^j)-event aggregation property holds for each successive level and $T_u = (r_1, r_2]$ with $r_1 = 0$, then for all $r_2 > 0$, and for j, $1 < j \leq m$,

$$\frac{\#(G_j, T_u)}{\pi^j} \geq \#(G_{j+1}, T_u). \tag{11}$$

The ϕ_j, ψ_j, and τ_j can be viewed as maps that cause event aggregation; consequently, Theorem 3.2, Corollary 3.1, and Remark 3.2 provide a relationship between event aggregation and event rates for one class of HDES. If there is a high measure of aggregation at level j (large π^j), then there will be far fewer events occurring at level $j + 1$. This illustrates that there is an inverse relationship between event aggregation and event rate between two levels of a HDES. In general, hierarchical systems researchers have observed a similar inverse relationship between "time scale density" ("time granularity") and "model abstractness" [Antsaklis et al. 1989; Saridis 1983]. The above results provide the first mathematical validation of these researchers' intuition about relationships between event aggregation and event rates for a class of HDES. Example 3.3 in Section 3.4 is used to illustrate the use of Theorem 3.2.

3.4. HDES Timing Structure Design for Event Rate Reduction

Theorem 3.2 and its use above provides a characterization of how event rates are affected by aggregation for one class of HDES. In this section we study the program of how to *design* the timing structure to ensure that there is a decrease (by a some constant factor) in the event rate between any two levels of a wide class of HDES. A reduction in event rate is often desirable so that the processors implementing the higher level controls are permitted adequate time before they must attend to the lower level systems (e.g., take actions based on the occurrence of an event string).

Let

$$S_{pi} \subset E^*_{pi} \quad (S_{gj} \subset E^*_{gj}), \tag{12}$$

and let the communications be defined by

$$C_{pi} = \mathsf{P}(S_{pi}) \quad (C_{gj} = \mathsf{P}(S_{gj})). \tag{13}$$

The notation $\mathbf{e} \in \mathbf{e}'$ will be used to denote the fact that $\mathbf{e}$ is a substring of $\mathbf{e}'$ (of course, we are abusing the notation here since $\mathbf{e}'$ is not a set). Consider the case where the output trigger is defined so that $\mathbf{s} \in \phi_i(\mathbf{e})$ if $\mathbf{s} \in \mathbf{e}$ and $\mathbf{s} \in S_{pi}$ ($\mathbf{s}' \in \phi_i(\mathbf{e})$ if $\mathbf{s}' \in \mathbf{e}$ and $\mathbf{s}' \in S_{gj}$). This output trigger initiates a communication the first time an event string occurs and $|\phi_i(\mathbf{e}')| \geq |\phi_i(\mathbf{e})|$ if $|\mathbf{e}'| \geq |\mathbf{e}|$. Hence, if the same event string occurs twice (or more) in some run this fact cannot be reported by this output trigger. Similar problems can exist if we define the output trigger so that $\mathbf{s} \in \phi_i(\mathbf{e})$ if $\mathbf{e} = \mathbf{e}_1\mathbf{e}_2$, $p \geq |\mathbf{e}_2| > 0$, and $\mathbf{s} \in \mathbf{e}_2$. This output trigger does, however, have the interesting property that it will "forget" about event strings in the past (depending on the choice for p). Here, we shall define the output trigger so that

$$\mathbf{s} \in \phi_i(\mathbf{e}) \quad \text{if } \mathbf{e} = \mathbf{e}'\mathbf{s} \text{ and } \mathbf{s} \in S_{pi} \tag{14}$$

($\mathbf{s}' \in \psi_j(\mathbf{e})$ if $\mathbf{e} = \mathbf{e}''\mathbf{s}'$ and $\mathbf{s}' \in S_{gj}$). By definition, if $\phi_i(\mathbf{e}) = \emptyset$ ($\psi_j(\mathbf{e}) = \emptyset$), a "null communication" occurs which cannot directly cause an event occurrence in any other DES in the HDES (hence $\tau_j(\emptyset, \emptyset, \cdots, \emptyset) = 0$ for all j). These assumptions about C_{pi} (C_{gj}) and ϕ_i (ψ_j) in (12)–(14) are only mildly restrictive, since it is possible that there can be a different communication representing each possible set of finite event strings that have just occurred. Moreover, there will be no particular assumptions about the τ_j maps, and the definition for the output triggers via (14) and communications via (12)–(13) is quite practical since each component DES is allowed to communicate the fact that sequences of events have just occurred; other DES in the HDES can then act based on such behavior.

The design of the timing structure entails choosing the proper S_{pi} and, hence, the communications C_{pi} that can occur between the various DES in the HDES. It is shown that by restricting the choice of what communications are allowed, one can achieve a decrease in the event rate at the higher levels of the HDES. In this way we achieve event rate reduction by restricting the manner in which the DES communicate and not by making particular assumptions about the dynamics of each component DES (as was done for Theorem 3.2). As in Section 3.2 and 3.3, we shall focus only on the HDES of Figure 1, and the results easily generalize to fully interconnected tree-structured HDES. First, we introduce a fundamental property of communications within the HDES:

DEFINITION 3.3. The set S_{pi} is said to be γ_i-*admissible* if for all $\mathbf{s}, \mathbf{s}' \in S_{pi}$ such that $\mathbf{s} = \mathbf{ab}$, $\mathbf{s}' = \mathbf{cd}$, and $\mathbf{b} = \mathbf{c}$ with $|\mathbf{b}| = |\mathbf{c}| \geq 0$ it is the case that $|\mathbf{d}| \geq \gamma_i > 0$.

A similar definition can be given for the S_{gj}. Clearly, there may not exist $S_{pi} \subset E_{pi}^*$ such that S_{pi} is γ_i-admissible for some given γ_i; but there always exists some $\gamma_i > 0$ such that S_{pi} is γ_i-admissible. Hence, for some DES one may be able to achieve more event rate reduction than for others and γ_i-admissibility characterizes this property. Intuitively, if the behavior of some DES P_i is such that it generates event strings which do not frequently cause communications to other DES than γ_i is large. Next, we provide several examples of $S_{pi} \subset E_{pi}^*$ that are γ_i-admissible:

1. Assume that for all $\mathbf{s} \in S_{pi}$, $|\mathbf{s}| \geq \gamma_i$. If for all $\mathbf{s} \in S_{pi}$ and all $e \in \mathbf{s}$ where $\mathbf{e} \in E_{pi}$, there does not exist $\mathbf{s}' \in S_{pi}$, $\mathbf{s}' \neq \mathbf{s}$, such that $e \in \mathbf{s}'$ then S_{pi} is γ_i-admissible.

2. If for all $\mathbf{s} \in S_{pi}$ there exists $e \in E_{pi}$ and $\mathbf{s}_2$ such that $\mathbf{s} = e\mathbf{s}_2$ and $|\mathbf{s}_2| \geq \gamma_i - 1$, and there does not exist $\mathbf{s}' \in S_{pi}$, $\mathbf{s}' \neq \mathbf{s}$, such that $e \in \mathbf{s}'$, then S_{pi} is γ_i-admissible (similarly for $\mathbf{s} = \mathbf{s}_2 e$). And more generally:
3. If for all $\mathbf{s} \in S_{pi}$ there exists $\mathbf{s}_2$, $\mathbf{s}_2$ such that $\mathbf{s} = \mathbf{s}_1\mathbf{s}_2$, $|\mathbf{s}_1| \geq 1$, and $|\mathbf{s}_2| \geq \gamma_i - 1$, and there does not exist $\mathbf{s}' \in S_{pi}$, $\mathbf{s}' \neq \mathbf{s}$, such that $\mathbf{s}_1 \in \mathbf{s}'$, then S_{pi} is γ_i-admissible.

It is important to note that for a given S_{pi} that might be chosen in the design of a timing structure it is not difficult to test whether S_{pi} is γ_i-admissible for some γ_i (of course, this may be computationally intensive).

THEOREM 3.2. If S_{pi} is γ_i-admissible for all i, $1 \leq i \leq n$, and the output triggers are given by (14), then

$$\sum_{i=1}^{n} \left\{ \frac{\#(P_i, T_u)}{\gamma^i} \right\} \geq \#(G_1, T_u) \quad \text{for all } T_u. \tag{15}$$

Proof. Consider the case of $n = 1$; the case for all n follows immediately in a manner similar to that in the proof of Theorem 3.2. Clearly the relationship holds when no events have occurred in G_1. Assume that an event occurs in G_1, i.e., that $\mathbf{e} \in E_{p1}^*$ is a sequence of events in P_1 such that $\tau_1(\phi_1(\mathbf{e})) = 1$. It must be the case then that there exists $\mathbf{s} \in \phi_1(\mathbf{e})$ such that $\mathbf{s}' \in S_{p1}$. Let $\mathbf{e}'$ be an event string such that $\mathbf{e}\mathbf{e}' \in E_{p1}^*$. By γ_1-admissibility and assuming that an output trigger defined via (14) is utilized, there does not exist $\mathbf{s}' \in S_{pi}$ such that $\mathbf{s}' \in \phi_1(\mathbf{e}\mathbf{e}')$ unless $|\mathbf{e}'| \geq \gamma_1$. Therefore, $\tau_1(\phi_1(\mathbf{e}\mathbf{e}')) = 0$ so long as $|\mathbf{e}'| < \gamma_1$, and, hence, $\#(P_1, T_u) \geq \gamma_1 \#(G_1, T_u)$ and the relationship (15) holds.

Theorem 3.3 shows that if each S_{pi}, for i, $1 \leq i \leq n$, is γ_i-admissible then there results a special type of aggregation between two levels of a HDES so that event rate reduction can be obtained. It is important that this aggregation is achieved via conditions on the communications and not on the structure of the P_i, $1 \leq i \leq n$. Of course, for a given set of lower level P_i one may be able to achieve lower event rates than for another set of P_i; Theorem 3.3 shows how to design the communications in the timing structure to achieve event rate reduction for a given set of P_i.

Remark 3.3. For a conventional discrete event control system as in Example 3.1, Theorem 3.3 shows the conditions under which G_1 will be guaranteed to have to generate a new control action only after at least γ_i events have occurred in P_1. For the hybrid dynamical system of Example 3.2, γ_i-admissibility places restrictions on how the higher level G_1 can observe state trajectories in P_1 to ensure that G_1 will act more slowly that P_1.

Example 3.3. (*Manufacturing System*) A simple manufacturing system will be used to illustrate the results from Sections 3.2, 3.3, and 3.4. We consider a manufacturing system which consists of a machine that can process parts of two types, one at a time. The machine outputs each type part into a particular output bin and the machine can be idle. Let $X = \{\text{MI}, \text{M}_1, \text{M}_2, \text{OUT}_1, \text{OUT}_2\}$ be the set of states, where MI means "machine idle", M_i

means the machine is busy processing part type i, and OUT_i means that the machine outputs part type i. Let $U = \{u_1, u_2\}$, where u_i means input part type i into the machine. Let $Y = \{y_b, y_{1d}, y_{2d}\}$, where y_b indicates that the machine is busy processing a part of either type, and y_{id} indicates that the machine is done processing a part of type i. The transition function δ and the output function λ for the manufacturing system are specified via the bottom of Figure 4. We let $X_0 = \{\text{MI}\}$ and consider $P_1 = (X, U, Y, \delta, \lambda, X_0)$ to be our plant.

There is a higher level mechanism which forces the alternate processing of one part of type 1 and then two parts of type 2. This device is pictured in the top of Figure 4 and will be referred to as G_1. We have $G_1 = (X_g, U_g, Y_g, \delta_g, \lambda_g, X_{0g})$ and $X_g = \{x_1, x_2, x_3\}$, $U_g = Y - \{y_b\}$, $Y_g = U$, and $X_{0g} = \{x_1\}$. Also, initially the input to the plant is u_1. Notice that G_1 completely ignores output y_b, as it is unimportant in coordinating the alternation of processing. (Hence, the inputs and outputs are not connected in a conventional manner where in G_1, $\delta_g(u, x)$ must be defined for all u.)

Suppose we let $\tau_1 : C_{p1} \to \{0, 1\}$, $C_{p1} = E_{p1}$, and $\phi_1(\mathbf{e}e) = e$ for all $\mathbf{e}e \in E^*_{p1}$ such that $e = (\text{OUT}_i, \text{MI})$ for $i = 1, 2$ (otherwise, $\phi_1(\mathbf{e}e) = \emptyset$ so that no event occurs in G_1). If the manufacturing system operates asynchronously (synchronously) then the coordination mechanism will operate asynchronously (with a general synchronous interpretation of time). Via Theorem 3.1, $\#(P_1, T_u) \geq \}(G_1, T_u)$ for all T_u. In particular we see that since event strings ending with (MI, M_i) and $(\text{M}_i, \text{OUT}_i)$ for $i = 1, 2$ are masked, a greater number of events will occur in the plant P_1 (lower level system) than in the controller G_1 (higher level system). Hence, for this simple manufacturing system the event rate at the higher level is lower no matter what the interpretation of time in P_1 is. Also,

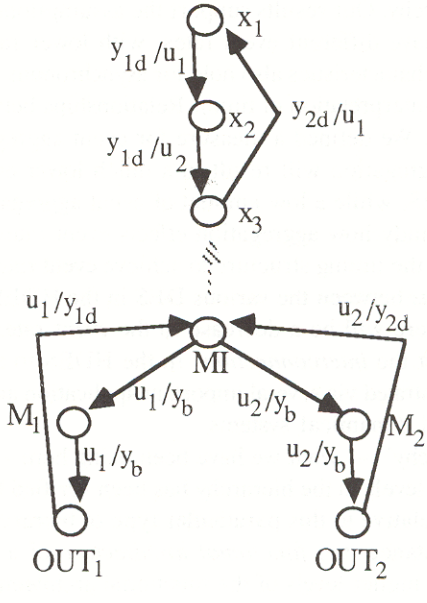

Figure 4. Model of a manufacturing system.

notice that if we choose $B_1 = \{1, 2\}$, and $X_{i1} = \{M_i, OUT_i\}$ for $i \in B$, then $E_{a1} = E_{p1}$, $E^*_{a1} \subset E^*_{p1}$ (given via Definition 3.3), and if we use the input trigger $\phi'_1 : E^*_{a1} \to \{0, 1\}$, a restriction of ϕ_1, and initial state as defined above, the conditions of Theorem 3.2 are satisfied with $\pi^1 = 2$, so $\#(P_1, T_u)/2 + 1 \geq \#(G_1, T_u)$ for the proper T_u and all τ_1. Hence, via Theorem 3.2 we find an inverse relationship between even aggregation and event rate for this simple manufacturing system.

Next, we show how Theorem 3.3 applies to this manufacturing system. Suppose that the same models for the manufacturing system and its controller are used but with a different interconnecting timing structure. In particular, using the approach of this section to specify the timing structure, we let $S_{p1} = \{MI, M_1)(M_1, OUT_1)(OUT_1, MI), (MI, M_2)(M_2, OUT_2)(OUT_2, MI)\}$, $C_{pi} = P(S_{pi})$, $\tau_1 : C_{p1} \to \{0, 1\}$, and ϕ_1 be defined as in (14). Notice that S_{p1} is 3-admissible (case (1) above) so that $\#(P_1, T_u)/3 \geq \#(G_1, T_u)$ for all T_u. We see that Theorem 3.3 can be used to produce a tighter bound on the number of events that occur in G_1; hence, the design of the timing structure via the Theorem 3.3 results in the guarantee of an even lower event rate in G_1.

4. Conclusions

We have provided a mathematical representation of the advancement of time in DES via index sets, index sequences, and interpretations of time. We discussed how deadlock and simultaneous events are characterized in our framework. In our main results we showed that the interpretations of time along with a timing structure provide a framework to study principles of the advancement of time for HDES. It was shown that for a wide class of HDES the *event rate* is higher for DES at the lower levels of the hierarchy than at the higher levels of the hierarchy. Our results support the assumptions in [Gershwin 1989] that the levels of an HDES have different event rates, with lower rates at higher levels, and show how similar timing characteristics also hold for asynchronous systems (and other system that have any admissible interpretation of time). Relationships between event rate and *event aggregation* were shown. We defined a measure for event aggregation and showed that a high amount of event aggregation will result in a much lower event rate at higher levels in a certain class of HDES, while a low amount of event aggregation will result in higher event rates. In order to study how aggregation effects event rates in more general HDES we studied how to design the timing structure to achieve event rate reduction. It was shown that if the *communications* between the various DES in the HDES satisfy a certain admissibility condition then there will be a decrease in the event rate. Hence, we showed that one only needs to restrict the *interconnections* in the HDES to achieve event rate reduction. The results were illustrated via several important application areas, including a manufacturing system and hybrid dynamical systems.

In a "time scale hierarchy" — what we have been using here — the inituition that event rates are higher for lower levels in the hierarchy has been verified here for a class of HDES, but the results here are relative to this paraticular type of hierarchy. If one defines a hierarchy relative to, for instance, the *functional architecture* of a system [Antsaklis et al. 1989], then clearly at the higher levels of the functional architecture there may be systems that are operating at higher rates than at the lower levels of the functional architecture.

It may take a rearrangement of the system components to place the system in a time scale hierarchy so that our results apply.

Acknowledgment

The authors would like to thank the reviewers for their helpful comments.

References

J.S. Albus, A.J. Barbera, and R.N. Nagel, "Theory and practice of hierarchical control," in *Proc. 23rd IEEE COMPCON*, 1981, pp. 19-39.

P.J. Antsaklis, K.M. Passino, and S.J. Wang, "Towards intelligent autonomous control systems: architecture and fundamental issues," *J. Intell. Robotic Syst.*, Vol. 1, pp. 315-342, 1989.

S.B. Gershwin, "Hierarchical flow control: a framework for scheduling and planning discrete events in manufacturing systems," *Proc. IEEE*, Vol. 77, No. 1, pp. 195-209, 1989.

J.F. Knight and K.M. Passino, "Decidability for a temporal logic used in discrete event system analysis," *Int. J. Control*, Vol. 52, No. 6, pp. 1489-1506, 1990.

M. Mesarovic, D. Macko, and Y. Takahara, *Theory of Hierarchical Multilevel Systems*, Academic Press: New York, 1970.

K.M. Passino, *Analysis and Synthesis of Discrete Event Regulator Systems*, Ph.D. Dissertation, Dept. of Electrical Eng., University of Notre Dame, Notre Dame, IN, 1989.

K.M. Passino and P.J. Antsaklis, "Fault detection and identification in an intelligent restructurable controller," *J. Intell. Robotic Syst.*, Vol. 1, pp. 145-161, 1988.

K.M. Passino and P.J. Antsaklis, "Relationships between event rates and aggregation in hierarchical discrete event systems," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Univ. of Illinois, Champaign-Urbana, pp. 475-484, Oct. 1990. (See also K.M. Passino and P.J. Antsaklis, "Timing characteristics of discrete event systems," Control Systems Technical Note #68, Univ. of Notre Dame, Dept. of Electrical Engineering, 1989.)

K.M. Passino and P.J. Antsaklis, "Timing characteristics of hierarchical discrete event systems," in *Proc. Amer. Control Conf.*, Boston, MA, pp. 2917-2922, 1991.

J.L. Peterson, *Petri Net Theory and the Modeling of Systems*, Prentice Hall: Englewood Cliffs, NJ, 1981.

P.J. Ramadge, W. M. Wonham, "Supervisory control of a class of discrete event processes," *SIAM J. Control Optim.*, Vol. 25, No. 1, 1987.

M.K. Sain, *Introduction of Algebraic System Theory*, Academic Press: New York, 1981.

G.N. Saridis, "Intelligent robot control," *IEEE Trans. Automat. Control*, Vol. AC-28, pp. 547-556, 1983.

K.P. Valavanis, *A Mathematical Formulation for the Analytical Design of Intelligent Machines*, Ph.D. Dissertation, Dept. of Electrical and Computer Eng., Rensselaer Polytechnic Inst., Troy, NY, 1986.

H. Zhong and W.M. Wonham, "On the hierarchical control of discrete-event system," in *Proc. 1988 Conf. of Inf. Sciences and Systems*, Princeton, NJ, 1988.

H. Zhong and W.M. Wonham, "Hierarchical control of discrete event systems: computation and examples," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Univ. of Illinois, Urbana, 1989.

H. Zhong and W.M. Wonham, "On the consistency of hierarchical supervision in discrete-event systems," *IEEE Trans. Automat. Control*, Vol. 35, No. 10, pp. 1125-1134, 1990.