# Distributed Fault Diagnosis With Overlapping Decompositions: An Adaptive Approximation Approach

Riccardo M. G. Ferrari, *Student Member, IEEE*, Thomas Parisini, *Senior Member, IEEE*, and Marios M. Polycarpou, *Fellow, IEEE* 

Abstract—This technical note deals with the problem of designing a distributed fault detection methodology for distributed (and possibly large-scale) nonlinear dynamical systems that are modelled as the interconnection of several subsystems. The subsystems are allowed to overlap, thus sharing some state components. For each subsystem, a Local Fault Detector is designed, based on the measured local state of the subsystem as well as the transmitted variables of neighboring states that define the subsystem interconnections. The local detection decision is made on the basis of the knowledge of the local subsystem dynamic model and of an adaptive approximation of the interconnection with neighboring subsystems. The use of a specially-designed consensus-based estimator is proposed in order to improve the detectability of faults affecting variables shared among different subsystems. Simulation results provide an evidence of the effectiveness of the proposed distributed fault detection scheme.

Index Terms—Adaptive estimation, distributed detection, fault diagnosis, large-scale systems, nonlinear systems.

### I. INTRODUCTION

The problem of automated fault diagnosis and accommodation is motivated by the need to develop more autonomous and intelligent systems that operate reliably in the presence of faults. In dynamical systems, faults are characterized by critical and unpredictable changes in the system dynamics, thus requiring the design of suitable fault diagnosis schemes [1]. Moreover, with current technological trends, several systems of practical interest are large-scale and/or physically distributed and thus the decomposition and spatial distribution of highly demanding computational tasks is of critical importance.

Recently there has been significant research activity in modeling, control and cooperation methodologies for distributed systems (see, for example, [2], and the references cited therein). This activity is motivated by several applications, especially in complex large-scale systems, such as traffic networks, environmental systems, communication networks, power grid networks, water distribution networks, etc. Such systems, although their dynamics and control objectives may appear to be completely different, have some important common characteristics: their dynamics are complex and spatially distributed, and, as a result, it is typically more convenient to decompose the system into smaller subsystems which can be controlled locally (or regionally). The study of controlling spatially distributed systems is not a new problem. As

Manuscript received June 26, 2007; revised May 15, 2008. Current version published April 08, 2009. This work was supported in part by the Italian Ministry for University and Research and by the Research Promotion Foundation of Cyprus. Recommended by Associate Editor M. Demetriou.

R. M. G. Ferrari and T. Parisini are with the Department of Electrical, Electronic and Computer Engineering, DEEI-University of Trieste, Trieste, Italy (e-mail: rferrari@units.it; parisini@units.it).

M. M. Polycarpou is with the KIOS Research Center for Intelligent Systems and Networks, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus (e-mail: mpolycar@ucy.ac.cy).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2008.2009591

far back as in the 1970s, researchers sought to develop so called "decentralized control" methods [3]. Since then, there have been many enhancements in the design and analysis of distributed control schemes. On the other hand, one area where there has been much less research activity is the problem of designing fault diagnosis schemes specifically for distributed systems.

Due to the complexity of the problem, in practice it is difficult to achieve robust fault diagnosis in large-scale distributed systems within a centralized architecture. As far as the literature is concerned, considerable efforts were aimed at developing distributed fault diagnosis methodologies in the context of discrete event systems (see, for instance, [4]–[8] and the references cited therein). On the other hand, very few works are available for discrete or continuous-time systems (for example, concerning large-scale networked control systems, see [9]).

In a previous work [10], the authors developed some preliminary results on a quantitative distributed fault detection scheme where a large-scale system was decomposed into a set of disjoint subsystems, and the physical interaction between neighboring subsystems was described by uncertain nonlinear functions. A network of Local Fault Detectors (LFD) was developed so that each LFD monitored a single subsystem by making use of the measurement of local variables, as well as the value of some interconnection variables communicated by neighboring LFDs. But apart from this exchange of measurements, the neighboring LFDs were not involved in the process of deciding whether a fault happened to a subsystem. In this note, based on the results recently presented in [11], the above distributed detection scheme is extended to allow cooperation between neighboring LFDs by using overlapping decompositions [12] of the initial large-scale system. In this way, more than one LFD may be monitoring a single shared variable and collectively decide on the presence of faults influencing it. This will be implemented by means of a specially designed consensus-based estimation scheme that may improve the detection capability of the LFDs with respect to the consensus-less, non overlapping case.

The note is organized as follows: in Section II, a problem formulation is developed for fault diagnosis of distributed dynamical systems. The design and analysis of a distributed fault detection architecture is presented in Section III, while simulation results for illustrating the methodology are given in Section IV. Finally, Section V provides some concluding remarks.

# II. PROBLEM FORMULATION

Let us consider a generic nonlinear system S (possibly large-scale) described as (see [13])

$$S: \dot{x} = f(x, u) + \beta(t - T_0)\phi(x, u) \tag{1}$$

where  $x \in \mathbb{R}^n$  and  $u \in \mathbb{R}^m$  denote the state and input vectors, respectively, and  $f: \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$  represents the nominal healthy dynamics. The term  $\beta(t-T_0)\phi(x,u)$  denotes the changes in the system dynamics due to the occurrence of a fault. More specifically, the vector  $\phi(x,u)$  represents the functional structure of the deviation in the state equation due to the fault and the function  $\beta(t-T_0)$  characterizes the time profile of the fault, where  $T_0$  is the unknown fault occurrence time. In this note, we only consider the case of abrupt (sudden) faults and, accordingly,  $\beta(\cdot)$  takes on the form of a step function, i.e.,  $\beta(t-T_0)=0$ , if  $t< T_0$  and  $\beta(t-T_0)=1$ , if  $t\geq T_0$ .

The model in (1) may be impractical for fault detection (FD), either because of its size, or because the system it represents is physically distributed, so that a centralized FD architecture is neither possible nor desirable. This problem can be overcome by considering  $\mathcal S$  as decomposed into N subsystems  $\mathcal S_i$ , each characterized by a *local* state vector

 $x_i \in \mathbb{R}^{n_i}$ , so that each  $S_i$  will be separately monitored. To this end, the dynamics of  $S_i$  can be modeled as

$$S_i: \dot{x}_i = \bar{f}_i(x, u) + \beta(t - T_0)\phi_i(x, u)$$

where the vectors  $\bar{f}_i$  and  $\phi_i$  are built upon the components of f and  $\phi$  that account for the dynamics of subsystem  $S_i$ . Generally, due to the distributed nature of the system, the nominal dynamics are unlikely to depend on the entire global vectors x and u, so that, as in [12],  $\bar{f}_i$  will be conveniently split into two parts

$$S_i: \dot{x}_i = f_i(x_i, u_i) + g_i(x_i, \bar{x}_i, u_i) + \beta(t - T_0)\phi_i(x, u)$$
 (2)

with  $f_i: \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \mapsto \mathbb{R}^{n_i}$  being the local nominal function,  $g_i: \mathbb{R}^{n_i} \times \mathbb{R}^{\bar{n}_i} \times \mathbb{R}^{m_i} \mapsto \mathbb{R}^{n_i}$  the interconnection function,  $u_i \in \mathbb{R}^{m_i}$ ,  $(m_i \leq m)$ , the local input, and  $\bar{x}_i \in \mathbb{R}^{\bar{n}_i}$ ,  $(\bar{n}_i \leq n - n_i)$ , the vector of interconnection state variables.

The vectors  $x_i, u_i$  and  $\bar{x}_i$  may be conveniently defined by relying on graph theory. The structure of  $\mathcal{S}$  will be described via a digraph  $\mathcal{G} \triangleq \{\mathcal{N}, \mathcal{E}\}$ , where  $\mathcal{N} \triangleq \{x^{(i)}, i=1,\ldots,n\} \cup \{u^{(j)}, j=1,\ldots,m\}$  is the set of nodes and  $\mathcal{E} \triangleq \{(e_i, e_j) : e_i \text{ acts on } e_j; e_i, e_j \in \mathcal{N}\}$  is the set of edges [12]. Here and in the following the notation  $x^{(i)}$  denotes the ith component of the generic vector x, while  $e_i$  denotes a generic node of the graph. The expression "acts on" is equivalent to "appears in the state equation of". Given  $n_a$  vectors  $a_1,\ldots,a_{n_a}$ , we define the vector  $a=\operatorname{col}(a_1,\ldots,a_{n_a}) \triangleq (a_1^\top,\ldots,a_{n_a}^\top)^\top$ . Each subsystem  $\mathcal{S}_i$  will be defined by introducing an extraction index set  $\mathcal{I}_i \in \mathbb{N}_+^{n_i}$  so that  $x_i \triangleq \operatorname{col}(x^{(j)}: j \in \mathcal{I}_i), u_i \triangleq \operatorname{col}(u^{(k)}: (u^{(k)}, x^{(j)}) \in \mathcal{E}, j \in \mathcal{I}_i, k=1,\ldots,m)$  and  $\bar{x}_i \triangleq \operatorname{col}(x^{(k)}: (x^{(k)}, x^{(j)}) \in \mathcal{E}, j \in \mathcal{I}_i, k=1,\ldots,n)$ . For every  $i=1,\ldots,N$ ,  $\mathcal{I}_i$  should fulfill the following conditions:

- 1)  $\mathcal{I}_i \neq \emptyset$ ;
- 2)  $\dim(\mathcal{I}_i) < n$ , that is  $\mathcal{S}_i \subset \mathcal{S}$ ;
- 3) the subdigraph of  $\mathcal{G}$  induced by  $x_i$  must be weakly connected, that is, each component of  $x_i$  must act on or must be acted on by at least another component of  $x_i$ .

The set of subsystems acting on  $\mathcal{S}_i$  through  $\bar{x}_i$  will be defined by means of the *neighbors index set*  $\mathcal{J}_i \triangleq \{h: \exists (x^{(k)}, x^{(j)}) \in \mathcal{E}, k \in \mathcal{I}_h, j \in \mathcal{I}_i\}$ . The structure of  $\mathcal{S}_i$  is described by the subdigraph  $\mathcal{G}_i \triangleq (\mathcal{N}_i, \mathcal{E}_i)$  induced on  $\mathcal{G}$  by the node set  $\mathcal{N}_i \triangleq \{x_i^{(j)}, j=1,\ldots,n_i\} \cup \{u_i^{(k)}, k=1,\ldots,m_i\}$ . Now, we have the following.

Definition 2.1: A decomposition of dimension N of the large-scale system S is a set  $\mathcal{D} \triangleq \{S_1, \ldots, S_N\}$  made of N subsystems, such that:

- 1)  $\bigcup_{i=1}^{N} \mathcal{I}_i = \{1, \dots, n\};$
- 2)  $\mathcal{I}_i \neq \mathcal{I}_j$ ,  $\forall i \neq j$ .

Point 1) in Definition 2.1 implies that the decomposition "covers" the whole original state vector x whereas Point 2) avoids "duplicate" subsystem definitions. It is worth noting that we do not require that  $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset, \forall i \neq j$ . Hence, overlapping decompositions are allowed, where the state vectors of any two subsystems may have common components. Overlapping decompositions [14] were found to be rather a useful tool when addressing large-scale systems. In particular, problems of stability, control and estimation [15], and fault diagnosis [16] for large-scale linear system were successfully solved by using overlapping decompositions.

As a result of overlaps, some components of the global state vector x will be assigned to more than a subsystem thus giving rise to the concepts of *shared state variable* and *overlap index set*.

Definition 2.2: A shared state variable  $x^{(s)}$  is a component of x such that  $s \in \mathcal{I}_i \cap \mathcal{I}_j$ , for some  $i, j \in \{1, \dots N\}, i \neq j$  and a given decomposition  $\mathcal{D}$  of dimension N.

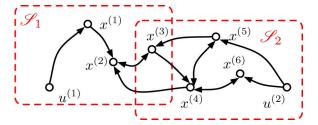


Fig. 1. Example of decomposition of a large-scale system  $\mathcal S$  into two overlapping subsystems  $\mathcal S_1$  and  $\mathcal S_2$  such that:  $x_1 = [x^{(1)}x^{(2)}x^{(3)}]^{\top}$  and  $x_2 = [x^{(3)}x^{(4)}x^{(5)}x^{(6)}]^{\top}$  are the local states,  $u_1 = u^{(1)}$  and  $u_2 = u^{(2)}$  the local inputs,  $\bar x_1 = [x^{(4)}x^{(5)}]^{\top}$  and  $\bar x_2 = x^{(2)}$  the interconnection variables, and  $x^{(3)} \equiv x_1^{(3)} \equiv x_2^{(1)}$  is a shared variable with  $\mathcal O_3 = \{1,2\}$ .

Definition 2.3: The overlap index set of subsystems sharing a variable  $x^{(s)}$  is the set  $\mathcal{O}_s \triangleq \{i : s \in \mathcal{I}_i\}$ , whose dimension is  $d_s \triangleq \dim(\mathcal{O}_s)$ .

In the following, the notation  $x_i^{(s_i)}$ , with  $x_i^{(s_i)} \equiv x^{(s)}$ , will be used to denote that the sth state component of the original large-scale system has been shared and became the  $s_i$ th of the ith subsystem,  $i \in \mathcal{O}_s$ . To gain some more insight into the afore-described decomposition approach, consider the simple example depicted in Fig. 1, where a specific decomposition of a system  $\mathcal{S}$  into two overlapping subsystems  $\mathcal{S}_1$  and  $\mathcal{S}_2$  is considered.

The following assumptions are now needed.

Assumption 1: For each  $S_i$ , i = 1, ..., N, the state variables  $x_i(t)$  and control variables  $u_i(t)$  remain bounded before and after the occurrence of a fault, i.e., there exist some stability regions  $\mathcal{R}_i = \mathcal{R}_i^x \times \mathcal{R}_i^u \subset \mathbb{R}^{n_i} \times \mathbb{R}^{m_i}$ , such that  $(x_i(t), u_i(t)) \in \mathcal{R}_i^x \times \mathcal{R}_i^u$ ,  $\forall i = 1, ..., N, \forall t \geq 0$ .

Clearly, as a consequence of Assumption 1, for each subsystem  $\mathcal{S}_i$ ,  $i=1,\ldots,N$ , it is possible to define some stability regions  $\mathcal{R}_i^{\bar{x}}$  for the interconnecting variables  $\bar{x}_i$ . The first reason for introducing such a boundedness assumption is a formal one in order to make the problem of detecting faults well-posed. Moreover, from an application point of view, Assumption 1 does not turn out to be very restrictive as the difficult issue generally is the early detection of faults characterized by a relatively small magnitude. Indeed, since no fault accommodation is considered in this note, the feedback controller acting on the system  $\mathcal S$  must be such that the measurable signals x(t) and u(t) remain bounded for all  $t \geq 0$ . However, it is important to note that the proposed Distributed Fault Detection (DFD) design is not dependent on the structure of the controller.

Assumption 2: The decomposition  $\mathcal{D}$  is given a priori and is such that, for each  $\mathcal{S}_i$ , the local nominal function  $f_i$  is perfectly known, whereas the interconnection term  $g_i$  is an uncertain function of  $x_i$ ,  $\bar{x}_i$  and  $u_i$ . For each  $k = 1, \ldots, n_i$ , the kth component of  $g_i$  is bounded by some known functional, i.e.,

$$\left|g_i^{(k)}(x_i,\bar{x}_i,u_i)\right| \leq \bar{g}_i^{(k)}(x_i,\bar{x}_i,u_i), \forall (x_i,\bar{x}_i,u_i) \in \mathcal{R}_i^x \times \mathcal{R}_i^{\bar{x}} \times \mathcal{R}_i^u$$

where the bounding function  $\bar{g}_i^{(k)}(x_i, \bar{x}_i, u_i) \geq 0$  is known, integrable, and bounded for all  $(x_i, \bar{x}_i, u_i)$  in some compact region of interest  $\bar{\mathcal{R}} \supseteq \mathcal{R}_i^x \times \mathcal{R}_i^{\bar{x}} \times \mathcal{R}_i^u$ .

Assumption 2 captures situations where each  $S_i$  corresponds to a known physical subsystem or a component, interacting through uncertain physical links as part of a complex large-scale system or to attain a higher goal (several application contexts can be found where such modeling approach turns out to be useful—see, for example, [17]). This uncertainty will be overcome in the following sections by employing an adaptive approximator  $\hat{g}_i$  in lieu of  $g_i$ .

Remark 2.1: It is worth noting that the determination of non-conservative bounding functions  $\bar{g}_i^{(\bar{k})}(x_i, \bar{x}_i, u_i) \geq 0$  may turn out to be rather difficult in practice and has to be carried out by exploiting prior knowledge by plant technicians and extensive off-line simulation trials.

#### III. DISTRIBUTED FAULT DETECTION ARCHITECTURE

In this section, the proposed DFD scheme will be described. In general, the DFD architecture is made of N communicating Local Fault Detectors (LFDs)  $\mathcal{L}_i$ , which are devoted to monitor the global system S, described by an overlapping decomposition into N subsystems. In the next subsections, the design of the LFDs will be addressed according to the fault detection methodology presented in [1]. First, we consider the system under nominal (healthy) mode of behavior and, subsequently, under faulty conditions.

### A. Normal (Healthy) Operating Conditions

The local fault detection algorithm is based on a nonlinear adaptive estimator based on the subsystem model (2) and for the ith LFD it takes

$$\dot{\hat{x}}_{i}^{(s_{i})} = -\lambda \left[ \sum_{j \in \mathcal{O}_{s}} \left( \hat{x}_{i}^{(s_{i})} - \hat{x}_{j}^{(s_{j})} \right) + d_{s} \left( \hat{x}_{i}^{(s_{i})} - x_{i}^{(s_{i})} \right) \right] 
+ \frac{1}{d_{s}} \sum_{j \in \mathcal{O}_{s}} \left[ f_{j}^{(s_{j})}(x_{j}, u_{j}) + \hat{g}_{j}^{(s_{j})}(x_{j}, \bar{x}_{j}, u_{j}, \hat{\vartheta}_{j}) \right]$$
(3)

for each  $s_i = 1, \ldots, n_i$ , where

- $\hat{x}_i^{(s_i)}$  denotes the estimate of the local state component  $x_i^{(s_i)}$ ;  $x_i^{(s_i)}$  corresponds to the sth component of the global state vector, that is  $x_i^{(s_i)} \equiv x^{(s)}$ ;
- $\mathcal{O}_s$  is the index set of the  $d_s$  LFDs sharing the variable  $x^{(s)}$ , as specified in definition 2.3;
- $\hat{g}_{j}(\cdot)$  is an adaptive approximator to be described later;
- $\hat{\vartheta}_i$  is the vector of the adaptive approximator parameters;
- $-\lambda < 0$  represents the value of the estimator poles.

As the entire state  $x_i$  is assumed to be measurable, it must be stressed that the estimate  $\hat{x}_i^{(s_i)}$  is not used for estimation, but will be employed in the fault detection process for residual error generation and for adaptive approximation. A consensus mechanism is embedded in the estimator for shared components of the local state of  $S_i$ , allowing LFDs in  $\mathcal{O}_s$  to share their knowledge about the local and the approximated interconnection part of the model. Consensus and agreement problems were extensively treated in the computer science literature concerning distributed fault diagnosis of synchronous and asynchronous systems [18], and recently in the framework of average-consensus on static and dynamic quantities by sensor networks [19].

It is worth noting that, in order to implement (3), the LFD  $\mathcal{L}_i$  does not need the information about the expressions of  $f_j^{(s_j)}$  and of  $\hat{g}_j^{(s_j)}$ ; instead, it suffices that  $\mathcal{L}_j$ ,  $j \in \mathcal{O}_s$ , computes the term  $f_j^{(s_j)} + \hat{g}_j^{(s_j)}$  and communicates it to other LFDs in  $\mathcal{O}_s$  alongside its actual state estimate  $\hat{x}_i^{(s_j)}$ . Furthermore, each  $\mathcal{L}_j$ , with  $j \in \mathcal{J}_i$ , must communicate to  $\mathcal{L}_i$  its values of the local state components needed to populate the interconnection state vector  $\bar{x}_i$ .

Clearly, for non-shared state components the overlap index set is a singleton and (3) simplifies to an estimator without consensus as fol-

Since it is assumed that, for each  $S_i$ , the interconnection function  $q_i$  is uncertain (or unknown), a key point in the proposed scheme is that each LFD will adaptively learn the uncertain function  $g_i$  using a linearly

parameterized adaptive approximator  $\hat{g}_i(x_i, \bar{x}_i, u_i, \hat{\theta}_i) : \mathbb{R}^{n_i} \times \mathbb{R}^{\bar{n}_i}$  $\mathbb{R}^{m_i} \times \mathbb{R}^{p_i} \mapsto \mathbb{R}^{n_i}$  of the form

$$\hat{g}_{i}^{(k)}(x_{i}, \bar{x}_{i}, u_{i}, \hat{\vartheta}_{i}) = \sum_{l=1}^{r_{i}} c_{i}^{(k)} \varphi_{i}^{l}(x_{i}, \bar{x}_{i}, u_{i})$$

where  $\varphi_i^l(\cdot)$  are given parameterized basis functions,  $c_i^{l(k)} \in \mathbb{R}$  are the parameters to be determined, i.e.,  $\hat{\vartheta}_i \in \mathbb{R}^{p_i}, \hat{\vartheta}_i \stackrel{\Delta}{=} \operatorname{col}(c_i^{(k)}: l =$  $1, \ldots, r_i, k = 1, \ldots, n_i$ ). Here the term adaptive approximator [20] may represent any linear-in-the-parameters, but otherwise nonlinear multivariable approximation model, such as neural networks, fuzzy logic networks, polynomials, spline functions, wavelet networks, etc. By introducing the gradient matrix  $Z_i \stackrel{\Delta}{=} \partial \hat{g}_i(x_i, \bar{x}_i, u_i, \hat{\theta}_i) / \partial \hat{\theta}_i$  w.r.t. the adjustable parameter vector [21], the approximator output can be written as  $\hat{q}_i(x_i, \bar{x}_i, u_i, \hat{\vartheta}_i) = Z_i \hat{\vartheta}_i$ .

Using adaptive parameter estimation techniques, the learning law for the parameter vector takes on the form

$$\dot{\hat{\vartheta}}_i \stackrel{\Delta}{=} \mathcal{P}\left(\Gamma_i Z_i^{\top} \epsilon_i\right)$$

where  $\mathcal{P}$  is a *projection operator* [13] that restricts  $\hat{\vartheta}_i$  to a pre-defined compact and convex set  $\Theta_i \subset \mathbb{R}^{p_i}$ ,  $\Gamma_i \in \mathbb{R}^{p_i \times p_i}$  is a symmetric and positive definite learning rate matrix and  $\epsilon_i(t) \stackrel{\Delta}{=} x_i(t) - \hat{x}_i(t)$  is the estimation error, which plays a double role: it provides a measure of the residual error for fault detection purposes and it also provides the error measure used for adaptively learning the unknown interconnection term  $g_i$ .

In general, the approximated interconnection term  $\hat{g}_i$  cannot be expected to perfectly match the true term  $g_i$ . This can be formalized by introducing an *optimal weight vector*  $\hat{\vartheta}_i^*$  within the compact convex set

$$\hat{\vartheta}_i^* \stackrel{\triangle}{=} \arg \min_{\hat{\vartheta}_i \in \Theta_i} \max_{\hat{\mathcal{R}}} \left\| g_i(x_i, \bar{x}_i, u_i) - \hat{g}_i(x_i, \bar{x}_i, u_i, \hat{\vartheta}_i) \right\| \tag{4}$$

and the corresponding minimum functional approximation error

$$\nu_i(t) \stackrel{\Delta}{=} g_i\left(x_i(t), \bar{x}_i(t), u_i(t)\right) - \hat{g}_i\left(x_i(t), \bar{x}_i(t), u_i(t), \hat{\vartheta}_i^*\right). \tag{5}$$

By introducing the *parameter estimation error*  $\hat{\vartheta}_i \stackrel{\Delta}{=} \hat{\vartheta}_i - \hat{\vartheta}_i^*$ , the dynamics of the generic estimation error component for  $t < T_0$  can

$$\dot{\epsilon}_{i}^{(s_{i})} = \frac{1}{d_{s}} \sum_{j \in \mathcal{O}_{s}} \left( -z_{j,s_{j}}^{\top} \tilde{\vartheta}_{j} + \nu_{j}^{(s_{j})} \right) - \lambda \left[ \sum_{j \in \mathcal{O}_{s}} \left( \epsilon_{i}^{(s_{i})} - \epsilon_{j}^{(s_{j})} \right) + d_{s} \epsilon_{i}^{(s_{i})} \right]$$

where the notation  $z_{j,s_j}^{\top}$  stands for the  $s_j$ th row of the gradient matrix  $Z_i$ . The solution of the above equation can be written as

$$\epsilon_i^{(s_i)}(t) = \frac{1}{d_s} \left[ \int_0^t e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left( -z_{j,s_j}^\top \tilde{\vartheta}_j + \nu_j^{(s_j)} \right) d\tau + e^{-2\lambda d_s t} \sum_{j \in \mathcal{O}_s} \left( e^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \epsilon_j^{(s_j)}(0) \right]$$

where  $\delta_{ij}$  is the Kronecker delta function defined as  $\delta_{ij} = 1$  if i = jand  $\delta_{ij} = 0$  otherwise. Again, this expression can be simplified in the case of a non-shared state component as follows:

$$\epsilon_i^{(k)}(t) = \int_0^t e^{-\lambda(t-\tau)} \left( -z_{i,k}^\top \tilde{\vartheta}_k + \nu_i^{(k)} \right) d\tau + e^{-\lambda t} \epsilon_i^{(k)}(0).$$

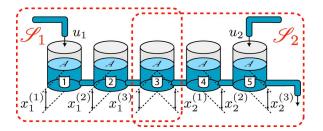


Fig. 2. Structure of the five-tanks system under consideration.

This solution shows that, because of the parameter estimation error  $\tilde{\vartheta}_i$  and of the MFAE, the estimation error will be nonzero even in the absence of a fault. By applying the triangle inequality, it can be shown that the absolute value of  $\epsilon_i$  can be upper-bounded as follows:

$$\left| \epsilon_i^{(s_i)}(t) \right| \leq \bar{\epsilon}_i^{(s_i)}(t)$$

$$\stackrel{\Delta}{=} \frac{1}{d_s} \left[ \int_0^t e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left( \kappa_j(\tau) \left\| z_{j,s_j}^\top \right\| + \bar{\nu}_j^{(s_j)} \right) d\tau \right]$$

$$+ e^{-2\lambda d_s t} \sum_{j \in \mathcal{O}_s} \left( e^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \left| \epsilon_j^{(s_j)}(0) \right|$$
(6)

where  $\kappa_i(t) \geq \|\tilde{\vartheta}_i\|$  depends on the geometric properties of the set  $\Theta_i$ . For instance, letting the parameter set  $\Theta_i$  be a hyper-sphere centered in the origin and with radius equal to  $M_i$ , we have  $\kappa_i(t) \triangleq M_i + \|\hat{\vartheta}_i\|$ . Moreover

$$\left|\nu_i^{(k)}(t)\right| \le \bar{\nu}_i^{(k)}(t)$$

where

$$\bar{\nu}_i^{(k)}(t) \triangleq \bar{g}_i^{(k)}\left(x_i(t), \bar{x}_i(t), u_i(t)\right) + K_{\hat{g}} \left\|\operatorname{col}\left(x_i(t), \bar{x}_i(t), u_i(t)\right)\right\|$$

and  $K_{\bar{g}}$  denotes the Lipschitz constant of the adaptive approximator on the compact set  $\bar{\mathcal{R}}$  introduced in Assumption 2. The bound described by (6) represents an adaptive threshold on the state estimation error that can be easily implemented by linear filtering techniques [1]. The bound  $\bar{\epsilon}_i(t)$  will be exploited in the next section in the fault detection context.

## B. Faulty Operating Condition

Now, a methodology for detection of faults by the DFD scheme proposed in this note, will be described. The occurrence at time  $T_0$  of a fault affecting the state of a subsystem  $\mathcal{S}_i$  will result in an additional term in the dynamics of the estimation error

$$\begin{split} \dot{\epsilon}_i^{(s_i)} &= -\lambda \left[ \sum_{j \in \mathcal{O}_s} \left( \epsilon_i^{(s_i)} - \epsilon_j^{(s_j)} \right) + d_s \epsilon_i^{(s_i)} \right] \\ &\quad + \frac{1}{d_s} \sum_{i \in \mathcal{O}} \left( -z_{j,s_j}^{\top} \tilde{\vartheta}_j + \nu_j^{(s_j)} \right) + \phi_i^{(s_i)}(x,u). \end{split}$$

Due to the fault function  $\phi_i^{(s_i)}(x,u)$ , the inequality (6) may no longer hold and the fault may be detected at a finite-time  $T_d \geq T_0$  by the ith LFD, according to the the following decision logic.

Definition 3.1: A fault occurring at time  $T_0$  is detectable in finite-time if there exists a  $\bar{T} < \infty$  and a  $k \in \{1,\ldots,n_i\}$  such that  $|\epsilon_i^{(k)}(\bar{T})| > \bar{\epsilon}_i^{(k)}(\bar{T})$ . The fault detection time  $T_d$  is defined as  $T_d \stackrel{\Delta}{=} \inf \bigcup_{k=1}^{n_i} \{t \geq T_0 : |\epsilon_i^{(k)}(t)| > \bar{\epsilon}_i^{(k)}(t)\}$ .

It must be acknowledged that the adaptive threshold  $\bar{\epsilon}_i(t)$  defined by (6) is designed to avoid false positives, since a certain level of estimation error is always present due to the uncertainty in the learning process. Of course, this may lead to certain faults being undetectable if they cause an estimation error small enough to be indistinguishable from the estimation error due to the uncertainty. This intuitive point will be formalized in Theorem 3.1. First, analogously to (4) and (5), the following quantities are defined:

$$\hat{\vartheta}_{i}^{*f} \triangleq \arg\min_{\hat{\vartheta}_{i} \in \Theta_{i}} \max_{\hat{\mathcal{R}}} \left\| g_{i}(x_{i}, \bar{x}_{i}, u_{i}) + \phi_{i}(x, u) - \hat{g}_{i}(x_{i}, \bar{x}_{i}, u_{i}, \hat{\vartheta}_{i}) \right\|$$

$$\mu_{i}(t) \triangleq g_{i}(x_{i}, \bar{x}_{i}, u_{i}) + \phi_{i}(x, u) - \hat{g}_{i}\left(x_{i}, \bar{x}_{i}, u_{i}, \hat{\vartheta}_{i}^{*f}\right)$$

as well as the mismatch function

$$\xi_{i}^{(k)}(t) \stackrel{\Delta}{=} z_{i,(k)}^{\top} \left( \hat{\vartheta}_{i}^{*f} - \hat{\vartheta}_{i}^{*} \right)^{(k)} + \mu_{i}^{(k)} - \nu_{i}^{(k)}.$$

Now, we can state the following result.

Theorem 3.1 (Fault Detectability): Given a variable  $x^{(s)}$  with an overlap set  $\mathcal{O}_s$ , suppose that for some time-interval  $[t_1, t_2]$  the corresponding components of the mismatch functions  $\xi_j(t), j \in \mathcal{O}_s$ , fulfill the following inequality for at least the ith LFD,  $i \in \mathcal{O}_s$ :

$$\int_{t_1}^{t_2} e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left| \left( \xi_j^{(s_j)} \right) \right| d\tau$$

$$\geq 2 \int_{0}^{t_2} e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left( \kappa_j(\tau) \left\| z_{j,s_j}^{\mathsf{T}} \right\| + \bar{\nu}_j^{(s_j)} \right) d\tau$$

$$+ 2 e^{-2\lambda d_s t} \sum_{j \in \mathcal{O}_s} \left( e^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \left| \epsilon_j^{(s_j)}(0) \right|.$$

Then, a fault will be detected at time-instant  $t=t_2$  by the *i*th LFD, that is  $|\epsilon_i^{(s_i)}(t_2)| > \overline{\epsilon}_i^{(s_i)}(t_2)$ . Moreover,  $t_1$  is an upper bound on the fault occurrence time  $T_0$ .

*Proof:* Following the proof of Theorem 3.2 in [10], the error dynamics at the generic time instant t can be written as follows:

$$\dot{\boldsymbol{\epsilon}}_{i}^{(s_{i})} = \begin{cases} -\lambda \left( \sum_{j \in \mathcal{O}_{\mathcal{S}}} \left( \boldsymbol{\epsilon}_{i}^{(s_{i})} - \boldsymbol{\epsilon}_{j}^{(s_{j})} \right) + d_{s} \boldsymbol{\epsilon}_{i}^{(s_{i})} \right) \\ + \frac{1}{d_{s}} \sum_{j \in \mathcal{O}_{s}} \left( -z_{j,s_{j}}^{\top} \tilde{\boldsymbol{\vartheta}}_{j} + \boldsymbol{\nu}_{j}^{(s_{j})}(t) \right) & \text{if } t < T_{0} \\ -\lambda \left( \sum_{j \in \mathcal{O}_{s}} \left( \boldsymbol{\epsilon}_{i}^{(s_{i})} - \boldsymbol{\epsilon}_{j}^{(s_{j})} \right) + d_{s} \boldsymbol{\epsilon}_{i}^{(s_{i})} \right) \\ + \frac{1}{d_{s}} \sum_{j \in \mathcal{O}_{s}} \left( -z_{j,s_{j}}^{\top} \tilde{\boldsymbol{\vartheta}}_{j}^{f} + \boldsymbol{\mu}_{j}^{(s_{j})}(t) \right) & \text{if } t \geq T_{0} \end{cases}$$

where  $\hat{\vartheta}_j^f \stackrel{\Delta}{=} \hat{\vartheta}_j - \hat{\vartheta}_i^{*f}$ . By using the mismatch function, the solution to the error dynamics is

$$\begin{split} \epsilon_i^{(s_i)}(t) &= \frac{1}{d_s} \int_0^t \mathrm{e}^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left( -z_{j,s_j}^\top \tilde{\vartheta}_j + \nu_j^{(s_j)} \right) \mathrm{d}\tau \\ &+ \frac{1}{d_s} \int_{T_0}^t \mathrm{e}^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \xi_j^{(s_j)} \mathrm{d}\tau \\ &+ \frac{1}{d_s} \mathrm{e}^{-2\lambda d_s t} \sum_{j \in \mathcal{O}} \left( \mathrm{e}^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \epsilon_j^{(s_j)}(0) \end{split}$$

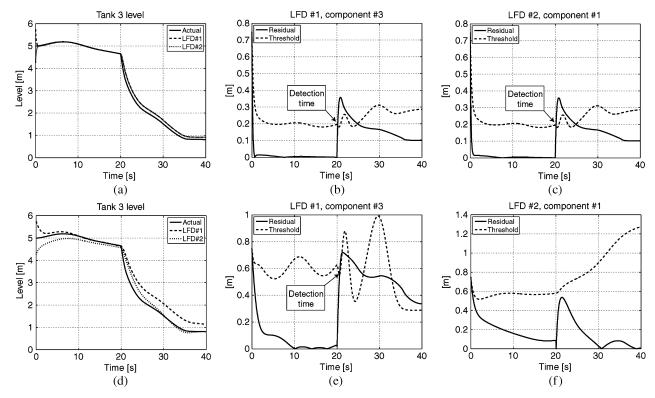


Fig. 3. Time-behaviors of signals related to tanks no. 3 when a leakage is introduced at time 20 s, with (a,b,c) and without (d, e, f) consensus.

so that the application of the triangle inequality leads to

$$\left| \epsilon_i^{(s_i)}(t) \right| \ge \frac{1}{d_s} \left| \int_{T_0}^t e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \xi_j^{(s_j)} d\tau \right|$$

$$- \frac{1}{d_s} \int_0^t e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left| -z_{j,s_j}^\top \tilde{\vartheta}_j + \nu_j^{(s_j)} \right| d\tau$$

$$- \frac{1}{d_s} e^{-2\lambda d_s t} \sum_{j \in \mathcal{O}_s} \left( e^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \left| \epsilon_j^{(s_j)}(0) \right|.$$

A sufficient condition for the previous inequality to hold is

$$\begin{aligned} \left| \epsilon_i^{(s_i)}(t) \right| &\geq \frac{1}{d_s} \int_{T_0}^t e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left| \xi_j^{(s_j)} \right| d\tau \\ &- \frac{1}{d_s} \int_0^t e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left| -z_{j,s_j}^\top \tilde{\vartheta}_j + \nu_j^{(s_j)} \right| d\tau \\ &- \frac{1}{d_s} e^{-2\lambda d_s t} \sum_{j \in \mathcal{O}_s} \left( e^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \left| \epsilon_j^{(s_j)}(0) \right| \end{aligned}$$

and, if  $|\epsilon_i^{(s_i)}(t_2)| > \bar{\epsilon}_i^{(s_i)}(t_2)$  for some  $t_2 > T_0$ , then a fault is detected. This translates into the following inequality:

$$\int_{T_0}^{t_2} e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left| \xi_j^{(s_j)} \right| d\tau$$

$$\geq \int_0^{t_2} e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left( \kappa_j(\tau) \left\| z_{j,s_j}^\top \right\| + \bar{\nu}_j^{(s_j)} \right) d\tau$$

$$+ \int_0^{t_2} e^{-\lambda d_s(t-\tau)} \sum_{j \in \mathcal{O}_s} \left| -z_{j,s_j}^\top \tilde{\vartheta}_j + \nu_j^{(s_j)} \right| d\tau$$

$$+ 2e^{-2\lambda d_s t} \sum_{j \in \mathcal{O}_s} \left( e^{\lambda d_s t} - 1 + d_s \delta_{ij} \right) \left| \epsilon_j^{(s_j)}(0) \right|$$

which implies the thesis when  $t_1 \ge T_0$ , thus proving the theorem.

Remark 3.1: It is worth noting that in a decentralized fault-diagnosis system where no information is exchanged between the LFDs, a detection decision may be difficult to reach in presence of a low mismatch  $\xi_j^{(s_j)}$  and/or high uncertainties  $\kappa_j(\tau)\|z_{j,s_j}^\top\|+\bar{\nu}_j^{(s_j)}$ . On the other hand, we expect a consensus mechanism like the one proposed in the technical note to be of benefit in such a scenario. However, due to the generality of the framework considered here (we do not make any assumption on the structural/geometric properties of the faults with respect to the structure of the distributed plant, and we do not assume persistency of excitation) proving that the proposed consensus-based methodology in general performs better than a decentralized one turns out to be difficult and is beyond the scope of the present note.

### IV. SIMULATION RESULTS

Now, a simple example to illustrate the effectiveness of the proposed FD scheme will be presented. It is based on the well-known three-tank problem, extended to encompass a five-tank string and two LFDs (see Fig. 2). The two LFDs monitor three tanks each, while sharing the third tank. Clearly, here the local nominal functions  $f_1$  and  $f_2$  describe the flows through the pipes linking tanks assigned to the same LFD, while the interaction terms  $g_1$  and  $g_2$  are due to the flow between tanks 3 and 4 and between tanks 2 and 3 (for details about the dynamical equations of a multi-tank system the reader is referred for example to [1]). All the tanks are cylinders with a cross-section A=1 m², whilst every pipe has a cross-section  $A_p=0.1$  m² and unitary outflow coefficient. The tank levels are denoted by  $x_1^{(i)}$  and  $x_2^{(i)}$ , with i=1,2,3, and are limited between 0 and 10 m. The scalars  $0 \le u_i \le 1$  m³/s, i=1,2, correspond to the inflows supplied by two pumps.

correspond to the inflows supplied by two pumps. The interconnection variables being  $\bar{x}_1 = x_2^{(2)}$  and  $\bar{x}_2 = x_1^{(2)}$ ,  $g_1(x_1, \bar{x}_1, u_1)$  and  $g_2(x_2, \bar{x}_2, u_2)$  should be 5-inputs, 3-outputs functions. Because of the topology of this specific example, both  $g_1$  and  $g_2$  have only one non-zero output component and depend only on  $(x_1^{(2)}, x_2^{(1)})$  and  $(x_2^{(2)}, x_1^{(3)})$  respectively. Therefore, the adaptive approximators  $\hat{g}_1$  and  $\hat{g}_2$  were realized with two 2-inputs, 1-output

radial basis neural networks. The network  $\hat{g}_1$  is implemented with 49 basis functions, while the network  $\hat{g}_2$  is made of 4 basis functions only. In both cases the basis functions are equally spaced over the square  $[0,10]^2$ ; the learning rate matrices are  $\Gamma_i=\mathrm{diag}(0.75)$  and the estimator constants are  $\lambda_i=1.5$ . After suitable offline simulations the parameter domains  $\Theta_1$  and  $\Theta_2$  are chosen to be hyperspheres with radii equal to 0.75 and 1.5, respectively. The non-zero bounds on the approximation error are set to  $\bar{v}_1^{(3)}=0.025$  and  $\bar{v}_2^{(1)}=0.2$ . Finally, the inflows are  $u_1=0.2\cdot\cos(0.3t)+0.3$  and  $u_2=0.25\cdot\cos(0.5t)+0.3$ ; the nominal tank initial levels were 8, 6.5, 5, 3.5 and 3 m, while the estimated ones are 15% higher and 15% lower, respectively for the first and the second LFD.

Fig. 3 shows the results of a simulation where at  $T_0=20~\rm s$  an abrupt leakage with cross section  $A_l=0.15~\rm m^2$  was introduced in tank 3, first when a consensus-filter is employed and then when it is not. In this respect, it can be observed that the LFD based on the network with fewer neurons (hence with more limited approximation capabilities) does not reach a detection decision in the absence of the consensus mechanism whereas a decision is reached in the presence of consensus using the information provided by the other LFD based on a networks with a much larger number of basis functions. The much better performance when the consensus mechanism is used is also due to the fact that the consensus equation dampens the difference between the estimates and the true values and also the difference among the two estimates. This can be seen very clearly by comparing the initial transient behaviors in Fig. 3(a) and (d).

#### V. CONCLUSION

In this note, a problem formulation and a distributed fault diagnosis architecture for large-scale dynamical systems was presented. The proposed scheme relies on overlapping decompositions of the system into sets of interconnected simpler subsystems. Each subsystem is monitored by a local fault diagnoser, which is able to detect the presence of faults in the corresponding subsystem based on its own measurements and information from neighboring subsystems. An adaptive approximation scheme is developed for learning the functional uncertainty in the interconnection between neighboring subsystems. As overlapping decompositions lead to some state components being shared between two or more subsystems, in the proposed scheme a specially designed consensus-based estimation scheme was implemented in order to allow the diagnoser to reach a common decision about faults affecting such variables and a detectability result was proved. Finally simulation results were provided to illustrate the effectiveness of the proposed scheme.

Future research efforts will be devoted to fully characterize the performances of the proposed consensus-based method as compared to decentralized schemes, to more practical detectability conditions and to extend the technique to the case where delays and disturbances affect the information exchanged among the local fault detectors.

#### REFERENCES

- X. Zhang, M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Trans. Automat. Control*, vol. 47, no. 4, pp. 576–593, Apr. 2002.
- [2] C. Langbort, R. Chandra, and R. D'Andrea, "Distributed control design for systems interconnected over an arbitrary graph," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1502–1519, Sep. 2004.
- [3] N. Sandell, P. Varaiya, M. Athans, and M. Safonov, "Survey of decentralized control methods for large scale systems," *IEEE Trans. Automat. Control*, vol. AC-23, no. 2, pp. 108–128, Apr. 1978.
- [4] P. Baroni, G. Lamperti, P. Pogliano, and M. Zanella, "Diagnosis of large active systems," *Artif. Intell.*, vol. 110, pp. 135–189, 1999.

- [5] I. Rish, M. Brodie, S. Ma, N. Odintsova, A. Beygelzimer, G. Grabarnik, and K. Hernandez, "Adaptive diagnosis in distributed systems," *IEEE Trans. Neural Networks*, vol. 16, no. 10, pp. 1088–1109, Sep. 2005.
- [6] Y. Wu and C. N. Hadjicostis, "Distributed non-concurrent fault identification in discrete event systems," in *Proc. CESA'03*, 2003, [CD ROM].
- [7] E. Athanasopoulou and C. Hadjicostis, "Probabilistic approaches to fault detection in networked discrete event systems," *IEEE Trans. Neural Networks*, vol. 16, no. 5, pp. 1042–1052, Sep. 2005.
- [8] M. Daigle, X. Koutsoukos, and G. Biswas, "Distributed diagnosis in formations of mobile robots," *IEEE Trans. Robotics*, vol. 23, no. 2, pp. 353–369, Apr. 2007.
- [9] R. J. Patton, C. Kambhampati, A. Casavola, P. Zhang, S. Ding, and D. Sauter, "A generic strategy for fault-tolerance in control systems distributed over a network," *Eur. J. Control*, vol. 13, no. 2–3, pp. 280–296, 2007.
- [10] R. Ferrari, T. Parisini, and M. Polycarpou, "A fault detection scheme for distributed nonlinear uncertain systems," in *Proc. Int. Symp. Intell. Control*, Munich, Germany, Oct. 2006, pp. 2742–2747.
- [11] R. Ferrari, T. Parisini, and M. Polycarpou, "Distributed fault diagnosis with overlapping decompositions and consensus filters," in *Proc. Amer. Control Conf.*, New York, Jul. 2007, pp. 693–698.
- [12] D. Siljak, Large-Scale Dynamic Systems. New York: North-Holland, 1978
- [13] M. Polycarpou and A. Helmicki, "Automated fault detection and accommodation: A learning systems approach," *IEEE Trans. Syst., Man, Cybern.*, vol. 25, no. 11, pp. 1447–1458, Nov. 1995.
- [14] M. Ikeda and D. Šiljak, "Overlapping decompositions, expansions, and contractions of dynamic systems," *Large Scale Syst.*, vol. 1, pp. 29–38, 1980.
- [15] M. Hodžič and D. Šiljak, "Decentralized estimation and control with overlapping information sets," *IEEE Trans. Automat. Control*, vol. AC-31, no. 1, pp. 83–86, Jan. 1986.
- [16] M. Singh, D. Li, Y. Chen, M. Hassan, and Q. Pan, "New approach to failure detection in large-scale systems," *Proc. Inst. Elect. Eng. D*, vol. 130, pp. 243–249, 1983.
- [17] M. Polycarpou, J. Uber, Z. Wang, F. Shang, and M. Brdys, "Feedback control of water quality," *IEEE Control Syst. Mag.*, vol. 22, no. 3, pp. 68–87, Jun. 2002.
- [18] T. Chandra and S. Toueg, "Unreliable failure detectors for reliable distributed systems," J. ACM, vol. 43, no. 2, pp. 225–267, Mar. 1996.
- [19] R. Olfati-Saber and J. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proc. 44th IEEE CDC ECC 2005*, Seville, Spain, Dec. 12–15, 2005, pp. 6698–6703.
- [20] J. Farrell and M. Polycarpou, Adaptive Approximation Based Control: Unifying Neural, Fuzzy, and Traditional Adaptive Approximation Approaches. Hoboken, NJ: Wiley, 2006.
   [21] A. Vemuri and M. Polycarpou, "On-line approximation methods for
- [21] A. Vemuri and M. Polycarpou, "On-line approximation methods for robust fault detection," in *Proc. 13th IFAC World Congress*, Jul. 1996, vol. K, pp. 319–324.