March, 1978 ESL-P-804

ADVANCES AND OPEN PROBLEMS ON THE CONTROL OF LARGE SCALE SYSTEMS *

by

Michael Athans

ABSTRACT

The purpose of this paper is to present an overview of practical and theoretical issues associated with problems of analyzing and controlling complex large scale systems using decentralized control. The state of the art with respect to large scale systems and decentralized control will be briefly overviewed, with emphasis upon the methodology available and required, and some of the theoretical results that have appeared during the past five years. It is concluded that the currently available centralized design methodologies, associated both with classical servomechanism theory as well as modern control and estimation theory, has reached a certain limiting level for fundamental understanding of the complex design issues that are associated with the control of large scale systems. It is suggested that future relevant theoretical directions for research in large scale systems must contain novel and nontraditional philosophical approaches, methodologies, theories and algorithms.

*Plenary Paper; IFAC 1978, Helsinki, Finland. Research supported by ONR Contract N00014-76-C-0346 and ERDA Contract E-(49-18)-2087.

ADVANCES AND OPEN PROBLEMS ON THE CONTROL OF LARGE SCALE SYSTEMS*

Michael Athans

Director, Electronic Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U.S.A.

Abstract. The purpose of this paper is to present an overview of practical and theoretical issues associated with problems of analyzing and controlling complex large scale systems using decentralized control. The state of the art with respect to large scale systems and decentralized control will be briefly overviewed, with emphasis upon the methodology available and required, and some of the theoretical results that have appeared during the past five years. It is concluded that the currently available centralized design methodologies, associated both with classical servomechanism theory as well as modern control and estimation theory, has reached a certain limiting level for fundamental understanding of the complex design issues that are associated with the control of large scale systems. It is suggested that future relevant theoretical directions for research in large scale systems must contain novel and nontraditional philosophical approaches, methodologies, theories and algorithms.

Keywords.

WHAT ARE LARGE SCALE SYSTEMS?

The term "large scale systems" is a vague one and we do not have available today a well understood and universally accepted definition. I shall list below a typical and partial list of physical problems which fall into the category of large scale systems.

(a) Power networks, which include the interconnection of several different types of generators via transmission lines, serving different customers with distinct load characteristics.

(b) Urban traffic networks, which include the coordination of traffic signals in signalized arterials, and perhaps the coordination of signalized arterials with limited access freeway systems.

(c) Automated transportation networks, in which personal rapid transit vehicles, or group rapid transit vehicles travel through a network from different origin nodes to different destination nodes. These include not only super-automated futuristic transportation networks, but currently tested systems often described as "dial-a-ride".

(d) Digital communication networks, involving point-to-point, or broadcast transmission of digitized voice, or computer data, often characterized by packet-switched strategies rather than line-switched strategies.

(e) Flexible batch manufacturing networks, which produce a variety of similar but nonidentical products through a system of conveyor links and machines that perform specific tasks for the production of partial components and subsequent assembly.

(f) Complex industrial systems, such as chemical refineries, process control factories, steel mills, etc.

(g) Command and control systems, involving distributed sensors, data bases, decision nodes and communication networks.

One can continue the list of additional physical and socio-economic systems which can be characterized as large scale systems. Obviously, such systems are essential in the context of the economic well-being of modern technologically oriented societies, and contributing to the quality of life. However, if we take for granted the fact that we live in a world with limited resources, and in particular energy resources, it is important to realize that such large scale physical systems do not operate necessarily in the most reliable and economic fashion, and any improvements in their operation, reliability, and productivity, via optimization, will have significant economic

* Plenary Paper; IFAC 1978, Helsinki, Finland. Research supported by ONR Contract N00014-76-C-0346 and ERDA Contract E-(49-18)-2087.

impact on the society as a whole.

In a world of limited and dwindling resources we can see a greater need for coordination between such large scale systems even if such a coordination has be be carried out under conflicting and fuzzy objectives. The technological society and the associated economics generate a need for greater interconnection between such systems. The global economic system is a good example of this; the economic policies of one nation, the energy policies of another nation, and the agricultural policies of yet another nation, can have significant impact upon the economic welfare of several other nations. If we turn our attention to physical large scale systems, we can see several examples in which existing and operating large scale systems work in a relatively inefficient and unreliable way due to poor planning, lack of systematic decentralized and yet coordinated control, and failure in emergency situations. In the area of power systems, we see an increasing degree of interconnection, with subsequent ill-understood dynamic phenomena which can result in severe balckouts in a domino-like fashion. Transportation systems work extremely inefficiently; consider the usual traffic jams, the dubious effects of exclusive lanes, and the failure of deterministic scheduling algorithms to function effectively in a dynamic stochastic environment encountered in recent "dial-a-ride" demonstrations, because the control strategy did not take into account the reliability of resources, the location strategies for the vehicles, and the stochastic nature of customers' demands. In the case of complex data communication networks, such as the ARPANET, only about 30% of the network resources are used to transmit real information, while the remaining 70% are used to transmit protocol (control) information; in such data communication networks sudden changes in demand, and link or node failures, can set up dynamic instability. In batch manufacturing, which involves metal cutting by several interconnected machines, recent statistics in the United States show that the productivity is extremely low when viewed from the point of machine utilization; machines cut metal only about 3% of the time, while over 90% of the time the metal parts are either moving from machine to machine or simply waiting in queues. Each reader can add his own experiences with respect to the overall inefficiency, unreliability and low productivity of such complex systems.

It is important to realize that even a relatively small improvement in the efficiency of operation and productivity of such large scale systems, and any improvement in the inherent reliability of providing service, can have significant economic and social impacts because of the integrated effect over time of the quality, cost, and reliability of the services that are being delivered, and the large number of customers served by these complex systems.

The inefficient and unreliable operation of large scale interconnected physical systems cannot be attributed solely to the ignorance of the designers. <u>There is a great lack of fundamental understanding and modeling of the underlying interactions, from a static and dynamic point of view and in the presence of stochastic perturbations; and the lack of understanding of how to generate coordinated, yet decentralized control strategies</u>. It is the author's opinion that in view of their basic training, systems engineers and scientists should have a major role toward future improvement in the efficiency, productivity, and reliability for such complex systems.

It is important to appreciate the attributes of such large scaled systems which require the development of a methodology for decentralized decision making and control. Most complex large scale systems are characterized by a network structure, and even at the static deterministic level they involve the solution of complex multicommodity flow problems. There are several degrees of freedom and control variables that can be manipulated to improve the performance of the overall system, only if one could understand how a coordinated approach to decentralized control is to be accomplished. Such large scale systems often possess hundreds or thousands of distributed sensors, with different accuracy and reliability characteristics. Many systems are geographically separated, although geographical separation does not necessarily imply a weakening of the dynamic interactions that one can expect. In addition to the geographical separation, such complex systems are often characterized by different time-scale phenomena; some interactions are relatively static, others have slow dynamic characteristics, while others have very fast dynamic characteristics. Yet there is dynamic and stochastic coupling between the fast and slow modes that one may encounter in such large scale systems. The cost and reliability of the required communication network has to be taken into account in the analysis and design of control systems for such large scale problems. Both common sense, and the recent technological advances in microprocessors and large scale memories give rise to issues of using distributed data bases, distributed computation, and hierarchical decomposition and decision making structures. In addition to the above, one must understand that even though the static problems may be understood, the dynamic and time varying interactions, as well as the effects of stochastic disturbances, have not been modeled in a great degree of detail.

Thus, for economic and reliability reasons, there is a trend towards decentralized decision making, distributed computation, and hierarchical control. However, these desirable goals of structuring a distributed information and decision framework for large scale systems, do not fall in the category of problems which have been traditionally treated with the available centralized design methodologies and procedures associated with either classical or modern control theory; and it is not clear how the available tools for design can be adapted toward the analysis and design of such complex large scale systems. In the remainder of this paper I shall briefly outline the status of existing methodology, and discuss some research areas for fundamental theoretical research which is motivated by, and is clearly relevant to, issues on how to analyze and design distributed information and decision systems for complex large scale problems.

STATUS OF AVAILABLE METHODOLOGY

Both classical servomechanism theory and modern control and estimation theory are centralized methodologies. Figure 1 shows, in an extremely simplified manner the basic problem of control. One has a complex dynamic system subject to stochastic disturbances and with noisy sensors. The objective is to design the feedback controller, which transforms the noisy sensor measurements into a set of real-time commanded controls to the actuators associated with the dynamic system, in such a way so that the overall behavior of the system is satisfactory. Classical servomechanism theory dealt primarily with linear time invariant dynamic systems, with single inputs and single outputs, and the problem was to design the feedback compensator using primarily frequency domain methods. From a philosophical point of view the recent design procedures suggested by Prof. Rosenbrock and Prof. MacFarlane [1] provide a partial vehicle for extending the frequency domain oriented design methodologies to the case of multivariable systems.

From a conceptual point of view, modern control theory provides a methodology, a set of computer-aided algorithms for the design of multivariable control systems, which takes into account uncertainty with respect to system disturbances and corruption of the measured variables by noise. Although, in principle, we do have a theory available for stochastic optimal control of nonlinear systems of arbitrary complexity [2], we do not have, and we will not have, a set of economic and reliable computer-aided design algorithms that help us design the controller in Figure 1 in a systematic manner.

As far as modern control theory is concerned, from the point of view of practical design, we can visualize it in the general block diagram shown in Figure 2. It is interesting to note that the practical use [3] of modern control theory leads to a two level decentralized structure. As shown in Figure 2, we still have a physical process subject to disturbances and once more the problem is to translate the noisy sensor measurements into the commanded controls to the actuators of the physical process, so that the system as a whole responds in a satisfactory manner. The implementation of the mapping of the actual data into the actual controls, is generally done in two levels. I would like to call one level the strategic (economic) level, and the other the tactical (stabilizing) level.

Let me describe the general methodology of the strategic (economic) level. Because of modeling and real time computational constraints, the models of the physical process tend to be either static, which completely ignores the dynamics of the physical process, or dynamic, subject to a great degree of aggregation. The function of the strategic system is to generate the ideal system response, for example generating the set points or trim points for key system variables, or define the desired optimal open-loop time trajectories associated with key state variables. In addition, the strategic level system generates the ideal control inputs to be applied to the physical process, in terms of set or trim points for the control variables, or time-varying open-loop preprogrammed controls.

At the strategic level, in order to generate the ideal open-loop controls and the ideal desired responses, one often use an objective function to be optimized. If the models employed are static, then the deterministic optimization problem is attacked through the theory and algorithms associated with nonlinear programming. If the underlying mathematical model is dynamic, then one uses the maximum principle [4], which leads to the solution of a two-point-boundary-value-problem. It is important to realize that the digital computer computations associated with the strategic level are off-line, and although they may be extremely complex, time consuming, and expensive they are not exercised in real time. The computations associated with either the solution of the nonlinear programming problem or the two-point-boundary-value-problem arising through the use of the maximum principle can be either centralized, or decomposition techniques such as the Frank-Wolfe method, and several others, can be used to ease the computational requirements.

In a parenthetic manner, it is important to stress that much of the current literature in large scale system theory, deals with static models and the methodologies of decomposition and coordination are in principle very similar to the decomposition methods in nonlinear programming problems [5]-[9].

In the absence of the tactical stabilizing level in Figure 2, one is attempting to control the stochastic physical processes using an open-loop control strategy. The physical process will not respond in a satisfactory manner, due to the presence of stochastic disturbances and the modeling errors that have been made at the strategic level. Thus, a feedback loop is necessary which senses in real time the deviations of the actual sensor measurement from the desired trajectories and translates them into real time control corrections to the precomputed open-loop (static or dynamic) controls generated by the strategic level; in this manner one generates in real time the commanded controls to the process actuators. The design of the tactical (stabilizing) level has been the main subject of both classical and modern control theory. In the sequel some of the key issues associated with this multivariable tactical stabilizing level will be outlined.

The design of the feedback tactical controller has to rely once more on certain mathematical models. Both in classical and modern control design, these models are often linear and stochastic, obtained in general by static or dynamic linearization of the process dynamics about the desired controls and trajectories established by the strategic level. In addition, "fast" dynamics are often explicitly modeled at the tactical stabilizing level, although they may have been ignored at the strategic or economic level. One of the reasons for the use of linear models for the design of the feedback tactical level has to do with the fact that the feedback control must operate in real time, and real time computational constraints must be observed. It is very important to realize that the current design methodology hinges on a key assumption of centralized information, or classical information pattern, in the sense that the feedback controller has instantaneous access to all sensor measurements and transforms them simultaneously into real-time control corrections.

The function of the tactical (stabilizing) level is to provide, through the use of feedback, overall systems stability and good performance in the sense that the actual physical response of the system tracks, with small errors, the desired response of the system as defined by the strategic level. Although not often stressed explicitly, stability implies that the tactical level controller must be extremely reliable. It should not be forgotten that the use of feedback for improving system performance can be destabilizing in the case of certain sensor or actuator failures.

Modern control theory provides an appropriate computer-aided design methodology for practical designs of the multivariable tactical feedback controller shown in Figure 2. The methodology employed is the so-called Linear-Quadratic-Gaussian (LQG) one with constant gains, and appropriate gain scheduling logic, [10]-[12]. In this design methodology the models employed are linear, dynamic, and stochastic with additive white process and sensor noise. The performance index involves quadratic penalties on the deviations of key state variables from their desired values, and quadratic penalties on the magnitudes of the control signal corrections (necessary to ensure finite gain and finite bandwidth control systems). As is well known, the structure of the feedback controller involves the construction of a constant gain Kalman filter which estimates the deviations of the state variables from their desired values. These estimated state variable deviations are then multiplied by constant control gains and the result is the set of real time control corrections signals shown in Figure 2. The use of an optimization methodology for designing the feedback controller, especially the optimization of a quadratic performance index, should be viewed as a means to an end rather than the solution of an optimization problem for its own sake. The LQG methodology can be justified in terms of its usefulness to the control system designer in terms of general purpose design software, as well as with respect to its robustness properties as will be outlined below.

There are several important considerations that are associated with the design of the stabilizing feedback controller. The two most important ones, in the opinion of the author, are robustness and reliability. A very brief overview of these two key issues will be given below.

Robustness refers to the desirable property of the feedback system to tolerate significant modeling errors without resulting in errors in steady state behavior or dynamic instability. Modeling errors are inherent since the models used at the strategic (economic) level are, in general, inaccurate due to either their static nature or, if dynamic, due to the fact that aggregation has taken place, fast dynamics have been ignored, changes in significant parameters of the real system have not been modeled, and all stochastic effects have not been taken into account. Recent research at both the theoretical and applied level have shown that LQG based multivariable control system designs enjoy excellent robustness properties, and these robustness properties can be often expressed as multivariable extensions of the classical concepts of proportional-integral-derivative (PID) designs [13], as well as traditional measures of gain margins (gain increase and gain reduction) and phase margins [14]-[16]. These robustness properties are maximal if all state variables are measured in the absence of any errors. The robustness properties decrease whenever sampling at increasingly lower frequencies is employed, or noisy sensors are used that necessitate the introduction of Kalman filters for state reconstruction. It should be remarked that the study of robustness properties of feedback control systems in the full stochastic case and in the sampled-data case is still an active research area.

Another extremely important issue associated with the tactical feedback level is that of reliability. In order to insure both satisfactory performance and stability the feedback control system must be designed in such a way so that it will continue to operate, perhaps in a degraded mode, but without loss of stability in the presence of sensor failures, actuator failures, and, in the case of digital implementations, intermittent software failures. A systematic procedure and methodology for designing control systems with a guaranteed level of reliability with respect to their performance is lacking at this time. What has been accomplished during the past five years can be characterized by serious investigations on certain important aspects of the problem. One way of guaranteeing a great reliability is through triplex or quadruplex redundancy of sensors and actuators, and perhaps of the digital hardware that performs the operations. Although it is possible through hardware redundancy to increase the overall reliability of the system, it should be noted that this is very expensive. For this reason, in several applications it is very important to replace hardware redundancy with analytical redundancy, especially in the area of sensors. Quite often triply redundant sensors and actuators are employed for failure detection and isolation through simple majority voting algorithms. The use of centralized Kalman filtering methods, combined with dynamic hypotheses testing, can provide analytical redundancy [17], [18] and lead to a design of the control system which provides degraded operation in the case of multiple actuator and sensor failures, sometimes in the absence of control system reconfiguration, and sometimes requiring control system reconfiguration following the detection and isolation of the failure. As remarked before, we still do not have a systematic methodology that addresses the issue of reliable designs which takes into account the failure rates of key system components, redundancy levels of different sensors and actuators, and the architecture of the digital microprocessor-based controllers.

In summary, a systematic methodology for the design of multivariable stochastic and dynamic systems is available today, and multivariable design methods based either on frequency domain or time domain methods are better understood. It should be self-evident that large scale systems have a multiplicity of sensors and actuators, and their control can only benefit by improved understanding of the key issues associated with multivariable control and their additional desirable properties of robustness and reliability.

DECENTRALIZED CONTROL: GENERAL ISSUES

The general structure of the decentralized control problem is illustrated in Figure 3. We are given a complex dynamic system, subject to external distrubances, which may be geographically distributed. We do have available a set of measurements as generated by distributed sensors and we can influence the response of the dynamic system by possibly distributed sets of actuators. At a conceptual level we can visualize a distributed set of controllers, possibly implemented by microprocessors or minicomputers, in such a way so that each controller receives the measurements only of a subset of the available sensors and in return generates commanded controls to the local actuator group. The key assumption is that each controller has available at each instance of time only a subset of the system measurements and can generate only a subset of the control commands; and that we do not allow instantaneous and error-free communication of any other measurements and control commands generated by the rest of the controllers. It is important to realize that if instantaneous and error-free communications were allowed, then the problem degenerates to a centralized solution, which is precisely what we wish to avoid [19].

There are many unresolved issues with respect to the situation depicted in Figure 3. To visualize some of the issues, let us imagine that the dynamic system which is to be controlled represents an urban traffic grid of one-way streets. Each local controller may represent the signal light at each intersection. The local control decisions would represent the timing and duration of the green, red and yellow cycles for each traffic signal, while the measurements may represent the queue lengths in the two local one-way links as measured by magnetic loop detectors. In this traffic example it is clear that some sort of coordination of the signals is necessary in order to achieve the objective of smooth traffic flow, without unecessary waste of time and energy consumption. For example, we would like to determine how adjacent signals may be dynamically synchronized, so as to minimize the conditions of stop-and-go traffic. It is clear that some coordination of the strategies, and communication of the queue lengths, is necessary between adjacent traffic signals. This would represent the dashed communication links with question mark shown in Figure 3. In addition, one could postulate a coordinator whose job would be to coordinate the decentralized decisions of groups of traffic signals. For example, we may postulate a coordinator that may control the traffic in, say, one dozen neighboring intersections. The situation can get much more complicated if we have a large number of traffic signals, so that there are several levels of coordination leading into a hierarchical traffic control system.

As another example, suppose that the dynamic system in Figure 3 represents an aircraft. One can postulate several levels of control for aircraft. For example, one of the controllers may be responsible for maintaining the position and average velocity of the aircraft, viewed as a point mass, so that the commands of the air traffic control system are obeyed. Another of the controllers may be responsible for the control of the slow fugoid mode of the aircraft, which represents a slow oscillatory mode, linking the velocity and the flight path angle of the aircraft. A third controller may be responsible for the control of the so-called short period mode, whose time constant is of the order of a few seconds, and which represents the interchange of energy between pitch rate and the angle of attack of the aircraft. A fourth controller may be responsible for the coordination of the aerodynamics surfaces to minimize the effect of the bending modes of the aircraft, with the objective of decreasing the fatigue induced in the structure. Yet, a fifth controller may be responsible for the control of the aerodynamic surfaces to avoid flutter, and hence enlarge the operating envelope of the aircraft. All these physical phenomena, namely trajectory control, the control of the fugoid mode, the control of the short period mode, the control of the bending modes and the control of flutter, are characterized by time constants that are widely different. Thus the complex dynamic system, in this case an aircraft, is not necessarily a large scale system which is distributed in space, but rather a dynamic system which has certain characteristics that can be identified with respect to different time constants of dynamic phenomena that occur; it makes engineering sense to postulate decentralized controllers that have the objective for appropriately controlling these different dynamic phenomena.

There are several issues that have to be addressed in reference to Figure 3. First of all, how do we decide the total number of individual controllers that we should employ? Once we have decided on the total number of controllers, perhaps based upon common engineering sense and the real time computational capability of each controller, then how do we decide which measurements are to be available in each controller and how do we decide which controls are to be generated by each controller? For coordination purposes, should the measurements that are available on each controller contain a subset of the measurements that are available to another controller? (In the traffic example cited above, this would correspond to the situation that each traffic signal microprocessor not only has available the measurements of the downstream queues, but also of the upstream queues that are controlled by the adjacent traffic lights.) The next question is: what information should be transmitted between controllers? Should the information be transmitted only to local neighbors or should the information be transmitted further along the system? A further question arises with respect to the objectives for control that are to be specified for each local controller. How do we translate, in general, an overall objective into subobjectives so that, in some sense, each local controller is doing the right thing?

For real time control it may be necessary, as shown in Figure 3, to have a coordinator. Presumably, the coordinator should receive some real time information from each local controller, and transmit a certain amount of coordinating information to each controller. How do we find out how to aggregate the information to be transmitted from each controller to the coordinator, and what type of information should each coordinator transmit back to the controller? In the real system the communication channels between each local controller, as well as to the coordinator, may be subject to errors, loss of communication, and communication delays. How then does this reliability of the communication system and the delays that it introduces, affect the performance of the overall system? It is important to realize that these questions do not arise in the available centralized control methodology, because we have assumed that the tactical stabilizing level has access to all the sensor measurements instantaneously and generates all control decisions instantaneously.

The general questions posed in the preceding paragraph are extremely important for any large scale system for which we must deploy a decentralized information and decision system. Since each controller must operate on the basis of some mathematical model, and perhaps with respect to some local objective function, the problem becomes even more complicated because one has to decide upon the appropriate local model available for control by each local controller, which certainly should not be the model of the entire system. The problem of reliability is an extremely important one. One would like to minimize the communication links between the controllers and the coordinator, since communication links are subject to failures. On the other hand, loss of communication of key variables may lead to performance degradation. Another important reliability issue deals with the failure of the local data processing system. Suppose that the local system fails; how is this detected by the adjacent controllers and perhaps the coordinator? How can we guarantee that the failure of a particular local controller will not lead to an unacceptable overall system performance degradation or even instability? This leads back to the question on how we decide upon the number of controllers and the nature of the tasks to be performend by the controllers in the first place.

What we are lacking is a systematic and scientific methodology for deciding upon the decentralized information and decision structure [5], [20]. In my opinion, we should not be looking for "optimal" information and decision structures, because the truly "optimal" ones are the fully centralized structures. What we would like to have is a theory that combines the overall system dynamics, the availability of measurements and their accuracy, the nature of controls, the overall system objective function, the reliability of the communication interfaces, computational delays, communication delays, so that we can at least answer the question of whether or not one structure is clearly preferable to another; we may have to accept the fact that certain distributed information and decision structures may not be directly comparable. To put it another way, we need a theory that would isolate a subset of distributed information and decision structures that are perferable to other ones, while satisfying certain gross criteria of goodness related to performance reliability. Such a theory is lacking, and in the opinion of the author, such a theory will not be the outcome of extensions of current centralized methodologies (such as nonlinear programming, stochastic dynamic programming, and the maximum principle, to mention just a few).

Thus, for real time stochastic decentralized control we must understand how important each and every bit of information is for stability and performance. We must understand the following issues: from whom should this bit of information originate, to whom should it be transmitted, with what maximum delay, why it should be transmitted, and what are the consequences of not obtaining this bit of information either accurately or in a timely fashion? It is obvious that we have no theoretical basis for answering these questions at the present time.

DECENTRALIZED CONTROL USING CENTRALIZED THEORIES

Whenever the dynamics of a complex system exhibit a certain structure, which is captured in the mathematical model of the system as well as in the performance objectives, then a combination of engineering common sense and mathematical analysis can indeed lead to a decentralized structure which can be analyzed and designed using the available centralized methodology, theory and algorithms. In this section, two classes of such problems will be discussed, because they are quite important for several physical applications. These two classes of systems are characterized by either weak dynamic coupling of otherwise separate systems, or by time-scale spearation phenomena involving fast and slow dynamics interacting in the same overall system.

First, I shall discuss the case of stochastic control of linear systems that are characterized by fast and slow modes [21]. As indicated before, such systems arise in several engineering applications; for example, in the longitudinal dynamics of the aircraft as well as in power systems applications. In reference to Figure 4, suppose that we have a linear stochastic dynamic system that is characterized by fast and slow modes. We assume that we have a set of noisy sensors, some of them which measure variables related to the fast dynamics, while others measure variables associated with the slow dynamics. The objective is to find the real time control inputs to the system for the purpose of regulation. The hierarchical and decomposed structure of the dynamic compensator for such systems is illustrated in Figure 4, [22]. Effectively, the generation of the overall control inputs split into the problem of generating "slow" controls and "fast" controls. The "slow" controls are generated by driving a Kalman filter from the "slow" sensor measurements so that estimates of the "slow" state variables are obtained, which in turn get multiplied by control gains to generate the "slow" controls. The theory tells us that the design of the "slow" Kalman filter should not only take into account in a natural way the process noise and the sensor noise, but in addition, the effects of the fast dynamics; these are modeled as additional white process noise as far as the "slow" Kalman filter is concerned. On the other hand, the "fast" Kalman filter is influenced by the estimates of the "slow" states generated by the "slow" Kalman filter. In the limit, as the separation between the "fast" and "slow" modes becomes infinite, the intuitive effect of the corrections generated by the "slow" Kalman filter look like bias corrections to the "fast" Kalman filter. Once more, one generates the "fast" controls by multiplying the estimates of the "fast" state variables by the appropriate control gains. This structure arises naturally from the mathematical formulation of the problem, and is not the outcome of an ad-hoc decomposition. This then is a case in which the mathematical properties of the system lead to a decomposition; these are indeed reflected in the limiting optimal hierarchical structure of the compensator, which has benefits with respect to the decomposition and reduction of both the off-line and the on-line computations. It is important to stress that one could have decomposed a priori the compensator structure into a "fast" compensator and a "slow" compensator; if one did this, one would not get the solution shown in Figure 4, because the bias corrections from the "slow" Kalman filter to the "fast" Kalman filter would not have been obtained; in addition, the effects of the "fast" dynamics in the calculation of the gains in the "slow" Kalman filter also would not have been obtained.

The structure shown in Figure 4 is an optimal one, only if the time scale separation of the fast and slow modes is infinite. It is important to understand the effects of the relative ratio of the fast to slow dynamics, before this decomposed compensator structure causes robustness and stability problems. This problem is currently under investigation, and it is not fully understood. At the present time, neither is the performance degradation that one can expect as the slow states become faster and the faster states become slower; what is needed is a systematic study in the definition of the quadratic performance index used for optimization, in which the existence of the fast and slow modes is explicitly reflected in the weights employed in the performance index.

These types of designs can be readily extended for systems that employ a whole hierarchy of slow and fast states. The mathematical tool is stochastic dynamic programming together with singular perturbation theory, both of which are centralized theoretical methodologies.

Another situation [20] that can be handled with the available centralized methodology is the one illustrated in Figure 5. Suppose that one has two distinct linear dynamic systems with weak dynamic coupling. The scalars w_1, w_2 denote the degree of dynamic coupling between the two systems. The systems become decoupled if both w_1 and w_2 go to zero. Under the assumption of no coupling between the two systems, one can design the individual dynamic compensators for each system using either modern or classical control techniques. This would fix the structure and the dimension of the necessary dynamic compensators. For example, if the LQG methodology is employed, then one would fix the dynamic compensator order for system 1 on the basis of the dimension of the Kalman filter or Luenberger observer, and similarly for system 2. Both common sense and mathematical analysis indicate that the system would perform adequately for weak dynamic coupling. One can justify this from state space arguments, or using the diagonal dominance method of Rosenbrock, or vector Lyapunov functions [23].

The interesting question is: what happens as the dynamic coupling between the two systems increases? One can retain the structure of the dynamic compensators, in the sense that each dynamic compensator operates using the same sensor measurements as in the decoupled case and generates the control; however, the parameters of the dynamic compensators are globally optimized off-line by taking into account the degree of dynamic coupling. This would give adequate performance up to a point, and one may obtain an unstable overall system as the dynamic coupling increases. In this case a problem that has not been resolved is how to utilize a subset of the measurements of system 2 to be communicated into the dynamic compensator for systeml, and vice-versa, as illustrated by the dashed lines in Figure 5.

In general this type of a situation can be extended to an arbitrary set of dynamic systems which are weakly coupled. A combination of engineering common sense and mathematical analysis <u>can fix the structure of each decentralized compensator, but one can optimize the parameters of the compensator in a global manner</u>. Some studies [20] indicate that the gains of the decentralized compensators are indeed influenced by not only the degree of dynamic coupling, but, in addition, the level of uncertainty, quantified by covariance matrices, associated with the stochastic part of the problem [24]. It is important to note that in such situations the <u>off-line</u> computational requirements, which are necessary to optimize the free parameters of the decentralized compensators in a global way, are far more severe than the necessary off-line calculations in the centralized case. On the other hand, this increase in off-line computation is accompanied by a decrease in the amount of the real-time computation for the system as a whole.

The system shown in Figure 6 represents a combination of the two cases discussed in this section. Each system is composed of a slow sybsystem and a fast subsystem, with weak dynamic coupling between the slow modes of each slow subsystem, as well as weak dynamic coupling of the fast modes of each fast subsystem. It is felt that a detailed analysis and optimization of this class of structures will give valuable insight into both the decomposition and global optimization of the free parameters of the resultant compensators. The results are not available as yet, but the issues are under investigation.

In summary, if with a combination of engineering common sense and mathematical analysis one can fix the structure of the decentralized dynamic compensators for linear systems, then one can optimize their parameters in a global manner. The optimal values of the decentralized compensator parameters will naturally take into account the existing dynamic interactions, as well as the relative separation of time scales that may exist in a complex system. The centralized methodologies will give us a better design that pure ad-hoc decomposition. In this sense the available centralized results are useful in decentralized situations.

DECENTRALIZED CONTROL: HARD PROBLEMS

In this section we shall briefly discuss the problems associated with the design of the decentralized information and decision system, as illustrated in Figure 3, for which the issue of fixing the number and the structure of the decentralized controllers and their interfaces has not been fixed a priori. Considerable attention has been given to the problem using the techniques of stochastic dynamic programming, in order to obtain some idea of the nature of the optimal decentralized solution. The results can be best described as a "can of worms" and "pitfalls", which illustrate the need for fresh and innovative theoretical approaches to these classes of problems [14], [20].

The existence of a non-classical information pattern, in other words that each local controller does not have access to the measurements and decisions made from the other controllers, results in two main difficulties. One of them is referred to in the literature as signaling strategies, while the other is referred to as the second guessing phenomenon.

Signaling strategies, [5], are by no means fully understood. Intuitively speaking, whenever a communication channel is denied between the distributed controllers, the centralized theory would like to communicate the available information by whatever means possible. This can be exhibited by extreme nonlinear coding phenomena of the information over the available communication channels between the distributed controllers [25], or, in the absence of any communication interfaces, by modulating the control actions in such a way so that the dynamic system itself acts as the communication medium, and these signals are decoded, in an as yet ill understood way, by the other controllers. This is an extremely unrealistic phenomenon, because in most real large scale system problems the dynamics are extremely uncertain. It is perhaps the fault of the theory that some model of the dynamic system is postulated, and that the sources of uncertainty tend to be additive rather than multiplicative in nature. At any rate, it makes no engineering sense to try to communicate missing bits of information using the system dynamics as the communications medium. Nonetheless, this is the nature of the solution that one obtains if one uses the available centralized methodology. Even for relativ^ly trivial examples, the optimal strategies are extremely complex, very nonlinear, corresponding to non-convex optimization problems. This implies that even if the problem were important from a mathematical point of view, any hope for numerical solutions remains a dream.

The second guessing phenomenon is primarily reflected on the dimensionality of the dynamic decentralized compensators. This problem has been studied in the context of decentralized LQG problems in which a Kalman-like filter is used for estimating the state variables. A particular decentralized Kalman filter has to generate an estimate of the local state variables; this estimate is not only influenced by the local control decisions, but also by the controls of the other decentralized controllers. Therefore, the local Kalman filter has to attempt to estimate the controls applied by the other controllers, which in turn depend on the best estimate of the state variables of the particular Kalman filter, and so on and so on. This implies that the Kalman filters increase in dimensionality and local computation. Once more the second guessing phenomena should be eliminated from a well designed information and decision decentralized structure, but it is the consequence of using the available centralized methodology.

It is not clear at this time how to tell the mathematics to avoid these signaling strategies or these second guessing phenomena. One way for avoiding these issues is to fix the structure and dimensionality of the dynamic compensators and optimize their parameters as discussed in the previous section. On the other hand, unless there is a good understanding of any time-scale separations and weak dynamic coupling, the issue of how to select the structure of the decentralized system is an open question. Thus, we desperately need innovative approaches on how to select a good structure, and in my mind this theory will have to take into account delays in information transmission, as well as the reliability of the communication subsystem, and of the control system as a whole.

CONCLUSIONS

It is the author's opinion that future theoretical developments in the area of decentralized control for large scale systems have to be directed into the problem of defining superior decentralized structures versus inferior decentralized structures, accepting the fact that groups of structures may not be directly comparable to each other In addition to the usual performance criteria, the mathematical statements of the problem have to explicitly include penalties for communication channels, they must take into account the reliability of the communication interfaces, as well as issues associated with model aggregation.

On the basis of the existing theory and results, one can deduce a set of rules of thumb which are important for arriving at the decentralized information and decision structure.

With respect to model simplification, the aggregation of the model should not be carried in a vacuum. The aggregation should depend very strongly on the performance index, and sensors that are available for control. One aggregated model may be well suited for prediction of forecasting, and be very badly suited for control, and vice-versa. Detailed understanding of the structure of the overall model, the existence of fast and slow dynamics, and the isolation of weak dynamic couplings will only help to arrive at a reasonable aggregated model for both decentralized estimation and control.

Communication links between the decentralized controllers, and perhaps the coordinator should be modeled appropriately. No formulation should admit instantaneous and noise-free communication of any information, because this can lead to the signaling problem and to a certain extent the second guessing phenomenon. Delays in communication should be explicitly modeled, as well as the failure probabilities and repair probabilities of the communication links, as an integral part of any optimization problem.

In order to avoid the problem of signaling using the dynamic system itself as the communication channel, appropriate care must be exercised in the statement of the dynamics of the physical system. It appears that uncertainty in the system dynamics modeled by means of additive noise is not sufficient to eliminate the signaling problem. Consideration should be given in modeling to parameters of the dynamic system using multiplicative uncertainty, for example by modeling certain parameters as multiplicative white noise.

With respect to the second guessing phenomenon, an explicit limitation on the dimensionality of each dynamic compensator should be an integral part of the problem formulation, with a penalty for the increase of dimensionality and real-time computation built as an explicit penalty term in the objective function to be optimized.

Finally, one should re-examine in more detail the static or strategic issues, together with the dynamic tactical issues. Static optimization is overly deterministic, and it appears that certain stochastic elements should be taken into account in the design of the strategic controller. Otherwise, if all stochastic effects are delegated to the decentralized tactical level, the lack of real-time information interchanges may not lead to the best system configuration.

In spite of all these conceptual difficulties, the development of methodologies and theories for decentralized control are extremely challenging, and the payoffs are very important. What is needed is a combined program of looking at physical large scale systems together with theoretical analyses and methodology development. It is important to isolate the generic issues associated with decentralized control. And finally, one word of caution; many of the nasty issues associated with decentralized control can be fully appreciated only in a stochastic framework. There is no doubt that understanding decentralization of deterministic systems is important, but a deterministic theory of decomposition and decentralization does not take into account all of the real issues that are associated with communication channels, their reliability, their characteristics, and their cost. These are all reflected in the stochastic aspects of the problem.

REFERENCES

1. A.G.J. MacFarlane, "Relationships Between Recent Development in Linear Control Theory and Classical Design Techniques," Parts 1-5, Measurement and Control (ASME), 8, 1975.

2. K. J. Astrom, Introduction to Stochastic Control Theory, N.Y. Academic Press, 1970.

3. M. Athans, "The Role and Use of the Stochastic LQG Problem in Control System Design," IEEE Trans. on Auto. Control, AC-16, N° 6, 1971.

4. Pontryagin, et al, The Mathematical Theory of Optimal Processes, N.Y., Wiley Interscience, 1962.

5. M. Athans (ed.), "Special Issue on Large Scale Systems and Decentralized Control," IEEE Trans. on Auto. Control, AC-23, N° 2, April, 1978.

6. M. D. Mesarovic, et al, Theory of Hierarchical Multilevel Systems, N. Y., Academic Press, 1970.

7. W. Findeisen, "A Survey of Problems in Hierarchical Control," Workshop on Multilevel Control, Institute of Automatic Control, Technical University of Warsaw, Poland, 1975.

8. D. A. Wisner (ed.), Optimization Methods for Large Scale Systems with Applications, N. Y., McGraw-Hill, 1971.

9. G. Guardarbrasi and A. Loratelli (eds.), "Proc. IFAC Symposium on Large Scale Systems Theory and Applications," Udine, Italy, 1976.

10. B.D.O. Anderson and J. B. Moore, Linear Optimal Control, Englewood Cliffs, N.J., Prentice-Hall, 1971.

11. M. Athans (ed.), "Special Issue on the LQG Problem," IEEE Trans. on Auto. Control, AC-16, N° 6, 1971.

12. H. Kwackernaak and R. Sivan, Optimal Linear Control Systems, N. Y., Wiley, 1974.

13. E. J. Davison, "The Robust Decentralized Control of a General Servomechanism Problem," IEEE Trans. on Auto. Control, AC-21, N° 1, 1976.

14. M. G. Safonov and M. Athans, "Gain and Phase Margins for Multiloop LQG Regu-

lators," IEEE Trans. on Auto. Control, AC-22, N° 2, 1977.

15. M. G. Safonov, "Robustness and Stability Aspects of Stochastic Multivariable Feedback Design," Ph.D. Thesis, M.I.T., Sept., 1977.

16. P. K. Wong, et al, "Structural Reliability and Robustness Properties of Optimal Linear-Quadratic Multivariable Regulators," Proc. IFAC 1978, Helsinki, Finland.

17. A. S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems," Automatica, 12, 1976.

18. J. C. Deckert, et al, "F-8 DFBW Sensor Failure Identification Using Analytical Redundancy," IEEE Trans. on Auto. Control, AC-22, N° 5, 1977.

19. M. Athans, "Survey of Decentralized Control Methods," Annals of Economic and Social Measurement, 4, 1975.

20. N. R. Sandell, Jr., et al, "Survey of Decentralized Control Methods for Large Scale Systems," IEEE Trans. on Auto. Control, AC-23, N° 2, 1978.

21. P. V. Kokotovic and R. A. Yackel, "Singular Perturbation of Linear Regulators: Basic Theorems," IEEE Trans. on Auto. Control, AC-17, N° 1, 1972.

22. D. Teneketzis and N. R. Sandell, Jr., "Linear Regulator Design for Stochastic Systems by a Multiple Time Scales Method," IEEE Trans. on Auto. Control, AC-22, N° 4, 1977.

23. A. N. Michel and R. K. Miller, "Qualitative Analysis of Large Scale Dynamical Systems," N.Y., Academic Press, 1977.

24. C. Y. Chong and M. Athans, "On the Stochastic Control of Linear Systems with Different Information Sets," IEEE Trans. on Auto. Control, AC-16, N° 5, 1971.

25. H.S. Witsenhausen, "A Counterexample in Stochastic Optimal Control," SIAM J. Control, 6, 1968.

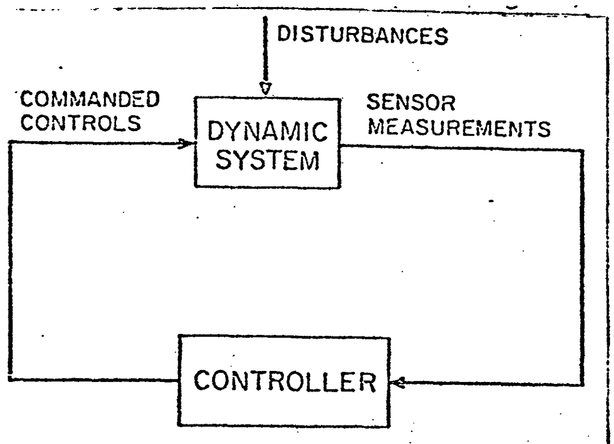

Figure 1

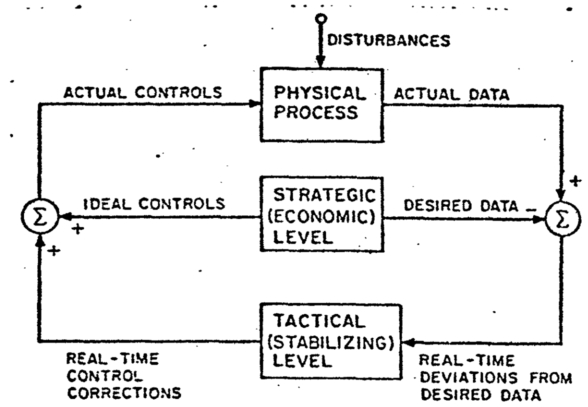

Figure 2

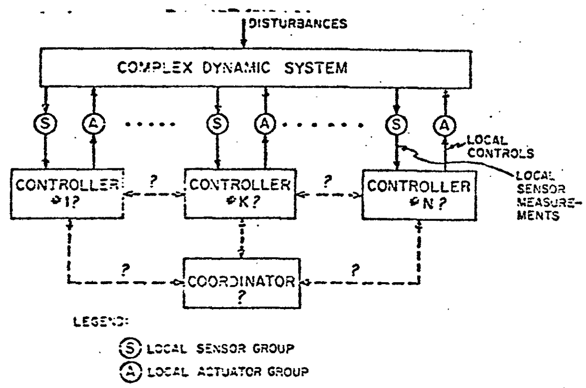

Figure 3

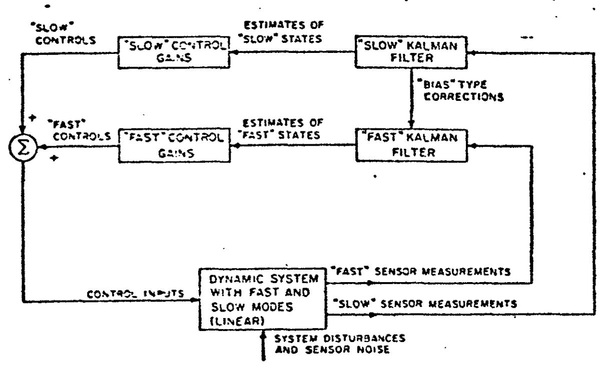

Figure 4

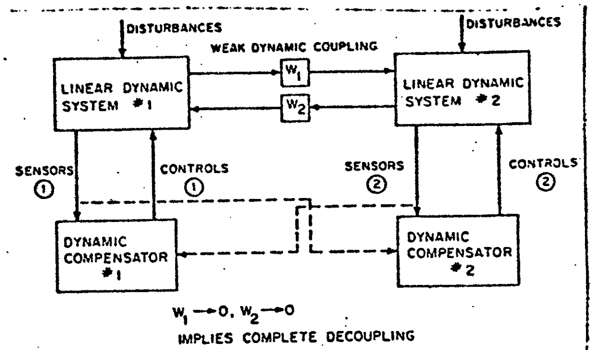

Figure 5

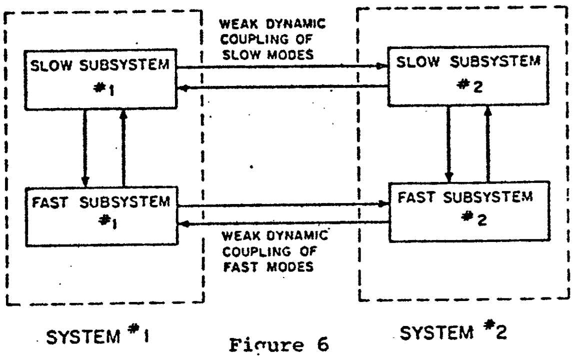

Figure 6