Fall 2020

Enabling flexibility through strategic management of complex engineering systems

Walter Louis Barnes II

Department: Engineering Management and Systems Engineering

ENABLING FLEXIBILITY THROUGH STRATEGIC MANAGEMENT OF

COMPLEX ENGINEERING SYSTEMS

by

WALTER LOUIS BARNES II

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ENGINEERING MANAGEMENT

2020

Approved by:

Dr. Ruwen Qin, Advisor
Dr. Steven M. Corns
Dr. Cihan H. Dagli
Dr. Ming C. Leu
Dr. Suzanna K. Long

ABSTRACT

Flexibility is a highly desired attribute of many systems operating in changing or uncertain conditions. It is a common theme in complex systems to identify where flexibility is generated within a system and how to model the processes needed to maintain and sustain flexibility. The key research question that is addressed is: how do we create a new definition of workforce flexibility within a human-technology-artificial intelligence environment?

Workforce flexibility is the management of organizational labor capacities and capabilities in operational environments using a broad and diffuse set of tools and approaches to mitigate system imbalances caused by uncertainties or changes. We establish a baseline reference for managers to use in choosing flexibility methods for specific applications and we determine the scope and effectiveness of these traditional flexibility methods.

The unique contributions of this research are: a) a new definition of workforce flexibility for a human-technology work environment versus traditional definitions; b) using a system of systems (SoS) approach to create and sustain that flexibility; and c) applying a coordinating strategy for optimal workforce flexibility within the human-technology framework. This dissertation research fills the gap of how we can model flexibility using SoS engineering to show where flexibility emerges and what strategies a manager can use to manage flexibility within this technology construct.

ACKNOWLEDGMENTS

I would like to acknowledge the guidance, support, and discipline I received from my great advisor, Dr. Ruwen Qin. She was a pillar of support during my doctoral journey and I have appreciated her tremendous support and drive in teaching me what it means to be a researcher, academician, and to focus on the prize ahead.

I wish to thank my beautiful wife, Christine Johnson-Barnes, and my wonderful family (daughters, Crystal and Camille; grandkids, Zjamyran, Zjaremya, Zjakai, and Zinnia) for their support and permission to pursue my doctoral dreams and realization of that dream. The long hours, some missed events, and valuable time were precious gifts given to ensure my success in this process.

I would like to especially thank my parents, Walter & Mary Barnes, for their unconditional love, guidance, and support that provided a solid foundation for the person I have become. Their tenacity, love of life and giving spirits were invaluable lifetime gifts that I have appreciated over the years. And to the countless people throughout my life who have nurtured, encouraged, and challenged me to be the best person I can in every respect, both professionally and personally. Without their collective shoulders, I would not be standing here today so I give thanks, gratitude, and the promise to pass on to the next generation my knowledge and experiences so that they may succeed as well.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Page

1. INTRODUCTION

1.1. AGILITY AND FLEXIBILITY - INTERDEPENDENCE

Within the domain of operations research and the management sciences (ORMS), work flexibility is an important and broad area of investigation with implications for managerial practice, system operations, and organizational profitability. It is noted that flexibility is an operational flexibility obtained through workforce management practices. The attainment of flexibility has included the use of technologies for assigning workers, facilitating teamwork, scheduling overtime, and flexible time among a range of approaches on a workforce basis.

In ORMS flexibility has a significant connection to agility as a key attribute, but distinct differences between agility and flexibility are observable in comparing the literature from these two areas. It is notable that both agility and flexibility help organizations that are facing uncertainty. However, if we posit an agility definition that states that agility is the successful exploration of competitive bases (speed, innovation, proactivity, quality and profitability) through integration of reconfigurable resources and best practices in a knowledge-rich environment to provide customer driven products and services in a fast and changing market driven environment, the difference is clearer [21]. Thus, the term agility is associated with predominantly external sources of largely unpredictable changes such as market or sector shifts that pose both risks and opportunities to organizations. By contrast, flexibility is associated with predominantly internal sources of uncertainty such as worker absence, common system fluctuations, and external sources of more predictable uncertainty such as the random arrivals of customers and the random, but stationary, demand for products. Flexibility and agility also differ along other lines such as uncertainty types, decision methods, decision time horizons, attributes, human behaviors, and solution generation methods. Thus, the ability to identify one or more best suited flexibility methods for specific applications is a critical step of flexibility practice.

We further note that agility and flexibility are often interdependent. To achieve the best work management outcomes, decisions in either of these areas may need to be made with the other in mind. For instance, flexibility may form sub-problems of agility. Conversely, work agility solutions may constrain some work flexibility methods. But, given the important differences between flexibility and agility, it is readily apparent that these two areas provide different complementary capabilities. Therefore, advances in one area may benefit the other. Most importantly, flexibility may form a foundation for developing agility. Therefore, research into and application of these concepts is necessary for the development of a theoretical framework of understanding and measuring agility and illustrating the practical application of flexibility concepts to specific cases advances these knowledge frontiers.

The traditional definition of workforce flexibility is the management of organizational labor capacities and capabilities in operational environments using a broad and diffuse set of tools and approaches to mitigate system imbalances caused by uncertainties or changes. As systems become more complex in their form and function, the problem of managing these newer, more complex systems to be flexible, i.e. to create the internal capability and capacity to address work uncertainties and external threats to work, becomes more acute. Flexible systems are those that can make changes easily to cope with changes or uncertainty. Traditional methods of managing and sustaining flexibility to deal with the uncertainties of work in the current environment of advanced technologies and communications are inadequate and do not take into account the myriad factors that are generated within multiple complex systems. When one looks at the differences between a single system and multiple systems, as summarized in Table 1 below, it is evident that a technology systems management methodology for managing these complex systems is required to generate the ability in the form of flexibility needed to address work challenges, such as balancing labor capacities, utilizing or borrowing skill sets, or generating capacity to address specific problems, among many other challenges.

Table 1.1 Differences Between Systems and Systems of Systems as They Apply to Systems Engineering.

Systems tend to...	**Multiple complex systems tend to...**
Have a clear set of stakeholders	Have multiple levels of stakeholders with mixed and possibly competing interests
Have clear objectives and purpose	Have multiple, and possibly contradictory, objectives and purpose
Have clear operational priorities, with escalation to resolve priorities	Have multiple, and sometimes different, operational priorities with no clear escalation routes
Have a single lifecycle	Have multiple lifecycles with elements being implemented asynchronously
Have clear ownership with the ability to move resources between elements	Have multiple owners making independent resourcing decisions

Without a methodology or framework to generate and manage flexibility in multiple systems, they will fail to meet the goals set out for them due to contradictory objectives and interests among systems and system owners. The work within this dissertation addresses the emerging needs for flexibility as a method of managing technology systems in the new era of modern industry. For example, within the energy and technology sectors, multiple microgrid systems and human technology-artificial intelligence systems need new forms of flexibility that this flexible systems management (FSM) methodology contributes to.

Distributed energy generation and smart grids are terms for domains that have emerged as key components for the conception of tomorrow's energy systems. In contrast to today's energy system, a smart grid is characterized as being a non-hierarchical, non-centric, undirected network for power distribution including a multitude of actors and a variety of energy sources. Not only will energy, in this system, be generated geographically distributed and from a variety of sources – it is also expected to be generated not only in specialized facilities but by users who appear on the energy market both as consumers and producers of energy. In transforming today's energy system to a smart grid, a variety of aspects become relevant.

The ability of the energy system to allow for energy generation to take place in a geographically distributed fashion is of course of prime importance. In this regard, several technologies become necessary, among them the storage of energy during times of high production and low energy demand. Also, management of these numerous sources becomes crucial in order for a smart grid to offer the supply stability of the classic power grid. Highly relevant in this regard is suitable forecasting of not only demand but also supply, which is dependent on exogenous factors such as the weather. Overall, the smart grid, being characterized as above, poses a control, or operational management, problem which requires the efficient processing of information from multiple information sources and thus an increased amount of information technology compared with today's energy system.

Within the future energy system, the energy market is thought to change accordingly. With a great number of both consumers and producers, the price is fulfilling an important signaling function and with demand of users being more responsive the energy market itself is becoming an important control mechanism for energy supply and demand the smart grid. A system of systems approach is expected by many to improve the competitive position of their company or organization by providing increased flexibility and also by increasing the range of services which can be offered to their company's or organization's users or customers. Also, usable capacity is thought to increase through the interconnection of existing systems into a greater system of systems, which enhances the competitive position of the company or organization through the process innovations a system of systems makes possible. Beyond process innovations, there are many who expect a system of systems approach to even allow entirely new business models to emerge.

Flexibility is a highly desired attribute of many systems operating in changing or uncertain conditions. This dissertation presents a study of enabling flexibility through a flexible systems management (FSM) framework and methodology. The sections show analyses of flexibility mechanisms of complex systems and, accordingly, identifies needs for flexibility that a flexible systems management approach based on system of systems (SoS) principles can meet. Following that, the research proposes a hierarchical network as a more flexible solution for applying the flexible systems management approach for complex or distributed large-scale systems. Then, decision problems for forming and evolving a network of systems are defined. A case that involves integrating distributed renewable energy sources with the main grid is presented to illustrate the implementation of the proposed methodology. Results from this study support the idea of acquiring and maintaining flexibility with the FSM method. This research also identifies research needs for advancing this particular use of FSM.

Coordinating the constituent systems of a system of systems (SoSs) in operations is an important task for functionalizing the SoS. The choice of a coordinating strategy needs to consider the autonomy, belonging, and connectivity levels of constituent systems. The diversity and emergence characteristics of SoSs are outcomes of coordination. This research synthesizes different strategies for coordinating constituent systems from the perspective of SoS characteristics and, therefore, derives the mechanism for choosing a coordinating strategy. Challenges are found in implementing the coordinating strategies for SoS. The section summarizes representative challenges facing system engineers. Methods for addressing these are proposed, including the multi-stage multi-scale coordination and smart coordination of operating mode switches. An island energy system composed of both diesel engine generators and microgrids with renewable energy sources is presented in this section as an example of a SoS network. Various aspects of coordinating constituent systems in this SoS network are presented to illustrate how the choice and execution of coordinating strategies functionalize SoS.

1.2. PROBLEM STATEMENT

This dissertation is motivated to analyze the flexibility mechanisms of systems of systems (SoSs) to enable flexibility for technological systems management with a system of system framework. The novelty in this research is the derivation of a managerial strategy for forming and evolving SoSs to provide needed flexibility, the derivation of a model to measure that flexibility in a modern manufacturing system consisting of complex constituent systems, and a new definition of flexibility in a human-technology-artificial intelligence manufacturing context. Specifically, the research is focused on changes or uncertainty that cannot be handled by a simple system in a cost-effective manner, but by systems of systems (SoSs). A SoS is a reconfigurable arrangement of independent and useful systems to deliver unique capabilities for a mission. A capability is the ability to execute a specified course of action. The key research question that is addressed is: how do we create a new definition of workforce flexibility within a human-technology-artificial intelligence environment?

The purpose of this dissertation is to provide an engineering systems management framework of how an SoS forms and evolves to provide more flexibility as a whole than from the component systems. The emergence, diversity, and autonomy resulting from the application of this framework is applied to the issues of renewable energy sources (RESs) that are geographically distributed and deliver power that is volatile and intermittent. The application of this framework to microgrids of renewable energy sources (RESs) addresses how a network of SoSs can be designed to integrate spatially distributed stochastic RESs with the main grid to deliver power in a complex system that is reliable and constant. Finally, the application of the same framework to intelligent systems and human-artificial intelligence (AI) interfaces that form a complex system highlights the research question of how we generate and sustain flexibility in these more complex systems that are at the frontiers of human-technology and machine learning influences that impact teams undergoing change from automation and how they function to meet to the new challenges of trends and technologies like Industry 4.0, Big Data analysis, the Internet of Things (IoT), cloud, remote sensing, and automation via robotics.

1.3. CHALLENGES IN THE SYSTEM OF SYSTEMS (SOS) APPROACH

From the management perspective, SoSs have multiple flexibility mechanisms. Decisions for forming SoSs and evolving them over time are complex, because the decisions are interdependent and across multiple time scales. From the technology perspective, the executions of SoS reconfiguration, system performance re-calibration, and SoS type change, require advanced control technologies. Addressing these challenges necessitates a seamless collaboration between systems engineers and domain experts.

The computational complexity in implementing multi-stage multi-scale coordination grows quickly as the size of a SoS, or SoS network, increases; how an outcome of coordination is reached and how it is evolving over time while using fully, or somewhat, decentralized coordination strategies for SoS are less understood. Finally, the use of smart technologies to improve coordination effectiveness requires more thorough analysis and exploration and is not evident in the literature or practice.

1.4. OVERVIEW OF DISSERTATION RESEARCH

An analysis of the literature associated with workforce flexibility was conducted to classify and determine the scope and effectiveness of each flexibility method. The analysis spanned the ORMS area and included flexible working time, floaters, cross-training, teamwork, and temporary workers. To facilitate managerial decision making, the key aspects of each method was summarized from the operations research and management science (ORMS) perspective. While flexibility is a potentially useful mitigator of uncertainty, specific methods are often designed to focus on one type or another. We notice that some flexibility methods may serve multiple purposes. Based on this analysis, a traditional view of flexibility was drawn as a baseline for researching flexibility in complex systems and as a basis for reforming that definition in an advanced technical environment.

A framework for enabling and evolving an SoS approach as applied to renewable energy sources and their associated power grids was created and applied to a case study to illustrate the creation of flexibility using a system of systems. The coordination of constituent systems is critical to preserving constituent systems' autonomy, belonging, and connectivity, and consequently generating diversity and emergence as desired SoS behaviors. However, not much formal work has been done to specify how to functionalize SoS through coordinating constituent systems. We are motivated to explore this topic and fully develop knowledge in this area of SoS.

Finally, an application of the SoS approach to a system comprised of a human, artificial intelligence, and an intelligent system (H-T system) was researched. It was concluded that the application of a SoS approach allowed for the recognition of where flexibility within this system is formed and the formulation of a limited ORMS model of the human-technology system provided a method to quantify flexibility and identify the emergence of flexibility within a SoS network of H-T systems. This research addressed this knowledge gap and created a model to establish a method to measure these systems in terms of the costs for attaining flexibility and the costs of the training needed for such a system to achieve its planning objectives. This portion of the dissertation research resulted in a modern definition of flexibility applied to the general context of complex intelligent systems with a measure of the flexibility of that system.

1.5. ANTICIPATED CONTRIBUTION

This research is anticipated to fill gaps identified in Section 2 regarding the summary of workforce tools available for practitioners to use to elicit flexibility-enabling behaviors within the traditional workforce that would enable teams to be able to address work uncertainties which would advance the goals and objectives of a work team in a dynamic environment.

The application of a flexible systems management (FSM) approach to renewable energy sources and their associated power grids is a unique, but practical application for smart grid design and implementation. Confronting this management problem enables energy providers to efficiently utilize their limited resources and improves timing for providing power to energy grids that provide for the needs of communities. Issues of coordination among the various systems are analyzed and addressed under a variety of approaches that allow for scalability in larger complex adaptive system applications without loss of flexibility in the individual systems, but a net gain in the flexibility of the overall system that is greater than the individual components comprising the system.

The traditional definition of workforce flexibility has been adequate to describe traditional work systems and environments. But, as systems become more complex, integrated, or hybridized with human-technology interfaces, the traditional definition of workforce flexibility is forced to adjust to the more complex systems and uncertainty in generating and sustaining the flexibility required to meet the challenges produced from these advances in work technologies. The application of the flexible systems management approach is a tool that allows us to generate flexibility through the emergence and diversity characteristics as well as the underlying system of systems characteristics of feedback loops, complexity, self-organization, and adaptability.

The final contributions of this research are: a) a new definition of workforce flexibility for a human-technology work environment versus the traditional flexibility definition; b) the creation of a flexible systems management framework based on system of systems (SoS) principles that creates and sustains flexibility for complex engineering systems; and c) application of a hierarchical coordinating strategy for optimal workforce flexibility within the human-technology framework. This dissertation research resulted in the creation of a flexible systems management methodology and a mathematical model that provides managers of complex engineering systems a method for determining where flexibility emerges and what strategies a manager can use to manage and sustain flexibility for human-technology systems.

2. AN ASSESSMENT OF WORKFORCE FLEXIBILITY METHODS IN OPERATIONS MANAGEMENT

2.1. WORKFORCE FLEXIBILITY OVERVIEW

Workforce flexibility is the management of organizational labor capacities and capabilities in operational environments using a broad and diffuse set of tools and approaches to mitigate system imbalances caused by uncertainties (e.g. worker absenteeism) or changes (e.g. seasonality). In this section, we review the literature associated with workforce flexibility in order to provide a reference for choosing flexibility methods for specific applications, and to better observe the connections and gaps in the literature with respect to future research needs. In this review, we span application contexts, research questions, and solution generation methodologies, for each flexibility method. Research opportunities for continuing and advancing the use of workforce flexibility are suggested.

Within the domain of operations research and the management sciences (ORMS), workforce flexibility is an important and broad area of investigation with implications for managerial practice, system operations, and organizational profitability. We note that workforce flexibility is an *operational* flexibility obtained through workforce management practices. Several researchers examined aspects of flexibility primarily from the perspective of workforce cross-training [1–5]. For instance, work by Hopp and Van Oyen [1] provided a comprehensive evaluation system for workforce flexibility obtained from cross-training and coordination. In addition to cross-training, workforce flexibility has included technologies for assigning workers, teamwork, floaters, overtime, and flexible time among a range of approaches. Given the breadth of workforce flexibility methods, a survey and evaluation of these may form a useful guide for practitioners to choose methods, or for researchers to note areas of opportunity.

Workforce flexibility has been widely researched for decades from various perspectives. Yet much remains to be done to advance the use of workforce flexibility. There had been a tendency for individual methods to be created for very specific conditions. Current systems where workforce flexibility is desirable are evolving based on new and emerging research questions on workforce flexibility.

We note that a stream of research attempted to differentiate workforce agility from workforce flexibility, yet they both help organizations that are facing uncertainty and change. For example, Yusuf et al. [6] stated that agility is the successful exploration of competitive bases (speed, innovation, proactivity, quality and profitability) through integration of reconfigurable resources and best practices in a knowledge rich environment to provide customer driven products and services in a fast and changing market driven environment. Qin and Nembhard [7] found that flexibility and agility also differ along other lines such as uncertainty types, decision methods, decision time horizons, attributes, human behaviors, and solution generation methods. Thus, the term *agility* is associated with predominantly external sources of largely unpredictable changes such as market or sector shifts that pose both risks and opportunities to organizations. By contrast, workforce *flexibility* is associated with predominantly internal sources of uncertainty such as worker absence, common system fluctuations, and external sources of more predictable uncertainty such as the random arrivals of customers and the random, but stationary, demand for products. We note that workforce agility and workforce flexibility are not mutually exclusive. To achieve the best workforce management outcomes, either flexibility or agility decisions may need to be made with the other in mind (e.g., [8]). Thus, workforce flexibility is an important characteristic of operations management. Workforce flexibility may form an important foundation for developing workforce agility [7].

In this section we review the research literature on workforce flexibility to outline the application settings, research questions, and solution generation methodologies, for each flexibility method. This review is written from the perspective lens of workforce flexibility including internal worker uncertainty and system fluctuations, as well as external sources of more predictable uncertainty such as random arrivals of customers and random (stationary) demand. The objective of this review is two-fold. First, we aim to provide a reference for choosing workforce flexibility methods for specific applications. Second, we identify future research needs by observing connections, overlaps, and gaps in the literature.

2.2. AN OVERVIEW OF WORKFORCE FLEXIBILITY METHODS IN THE LITERATURE

The criteria for this review arose from an initial list of the work- force flexibility methods identified through a literature search. To focus this review, we limited the literature search to the subject of Operations Research and Management Sciences (ORMS) and searched method names and variations of these names in article title, abstract, and key words. This began with relatively well-known terms including cross-training, teamwork, floaters, and flextime. The process was iterated in order to identify additional methods and naming variants during the review process. The authors' experience as well as recommendations from other experts in workforce engineering and management aided the identification of relevant workforce flexibility methods. We finally identified the following five workforce flexibility methods that have been both researched in ORMS and implemented in practice:

Flexible working time: relaxing standard shift lengths and workweek hours;

Floaters: Designating classes of workers to float to stations with greatest need;

Cross-training: Training workers at multiple skills;

Teamwork: Functional or additive collaboration;

Temporary labor: Contracting additional short-term labor.

These methods provide flexibilities in workforce capabilities, capacities, or both. Workforce flexibility methods differ in flexibility generation mechanisms and levels, as illustrated in Table 2.1. From the perspective of generation mechanisms, flexibility can be obtained by letting workers work longer or shorter hours, transfer to where they are in greater need, master multiple skills, and to collaborate and assist each other. Flexibility can also be obtained by using multiple labor sources. Table 2.1 shows that workforce flexibility may be generated at different levels, including the individual worker level, group level, and organizational level. Each workforce flexibility method can produce flexible capacities, and some including cross-training and teamwork can additionally produce flexible worker capabilities.

We remark that the classification of such a broad area of research is ultimately arbitrary. Nonetheless, we choose to classify the existing methods using terms commonly used in industry (namely, flextime, floaters, cross-training, teamwork, and temporary labor). While practitioners may associate these terms with very specific implementations, we use them as broad categories that should remain recognizable to practitioners and researchers alike. That is, we arrive at these choices as a synthesis of the literature in the area. Additionally, this classification focuses on the fundamental aspects of the methods, such that both the individual methods and the classes themselves may overlap. Many of these complexities are discussed in the corresponding sections for each method class.

Table 2.1 Classification of workforce flexibility methods by flexibility generation mechanisms and levels.

		Level		
		Individual	**Group**	**Organization**
	Time	Flexible working time		
	Transfer	Floaters		
Mechanism	Multi-functionality	Cross-training[a]		
	Collaboration		Teamwork[a]	
	Labor Sources			Temporary labors

[a] Also creates flexible worker capabilities.

The objectives of this assessment are to provide a guideline for practitioners to select workforce flexibility method(s) to address their needs, and for researchers to provide a justified list of open problems and gaps presented in current research. Thus, we want to classify the references and summarize key aspects for each workforce flexibility method from the ORMS perspective, including, problem types, research questions, solution generation methodologies, and limitations. We organize the remainder of the section as follows. Each of the following sections is a detailed review and discussion of the literature for each workforce flexibility method. Lastly, Section 2.7 summarizes these methods to facilitate method selections for applications, as well as research gaps and open problems we identified from the review.

2.3. FLEXIBLE WORKING TIME

Flexible working time is a method for creating flexibility in workforce capacity by allowing for varying labor hours, subject to constraints set by laws or agreements between employers and workers. Flexible working time includes a range of approaches such as overtime, flexible workdays, annualized working hours, and working time accounts, each of which is outlined in the subsections that follow.

2.3.1. Overtime. Overtime is time beyond workers' regular working time (e.g., more than 8 h per day). It has been modeled as an additional internal capacity that helps mitigate the need for frequent hiring and firing of workers. Workers are commonly compensated by additional pay for overtime work. Overtime is probably the simplest practice of flexible working time, yet it can be relatively expensive when overtime work is compensated at a higher rate than regular time work. Overtime is often used to address fluctuations in de- mand and uncertainty in production or service time [9–12], to meet time critical deadlines [12–15], and to help reduce the high inven- tory cost of spare parts [10,11,14]. Research on the use of overtime is primarily focused on two issues. The first issue involves deter- mining the amount of overtime considering the trade-off between overtime costs and benefits such as reduced tardiness, improved match between demand and capacity, and lowered inventory [10,11,13–15]. The second issue broadly involves the timing of the use of overtime [10,12].

Holloway and Nelson [13] formulated a model to minimize overtime subject to meeting due dates and operational precedence. They proposed a scheduling procedure that begins with a feasible solution to the optimization problem and then searches for improved solutions. The scheduling procedure for the static problem was extended to a less constrained dynamic problem. Others, such as Gelders and Kleindorfer [14,15] examined the trade-offs between overtime cost and other costs including tardiness, WIP, and other flow-time-related costs. They formulated an optimization model to minimize the total cost in the single-machine job shop and proposed an *ad hoc* Branch-and-Bound (BNB) algorithm. The use of overtime is often associated with rules or policies for workforce assignment and scheduling. Scudder [10] used simulation to comprehensively evaluate six overtime policies within a hypothetical repair shop that was subject to random arrivals of failed assemblies. Both proactive and reactive scheduling rules were considered. For example, a proactive rule may indicate action as soon as the lead time suggests a future stockout, and a reactive rule would simply wait for the actual stockout. These two rules were later reexamined by Scudder and Chua [11] to incorporate demand uncertainty.

A finite-capacity real-time scheduling problem was considered by Akkan [12], wherein overtime was an option for meeting due dates when it became infeasible to meet the deadline using an existing schedule. Akkan proposed an approach that would insert overtime work, without substantially changing the current regular work-order schedule, using one of two heuristics: a TwoPass heuristic and a Flowtime heuristic. The TwoPass heuristic (forward and backward) uses a backward pass insertion of a new overtime work order when a forward insertion is infeasible. The Flowtime heuristic uses a modified backward pass insertion that determines priorities for each operation based on the standard times of the operations and outperformed the TwoPass approach [12].

2.3.2. Flexible Workdays. Flexible workdays are a form of capacity flexibility obtained by adjusting the effective length of workdays from the nominal 40 hours per week, to something more flexible such as a range of hours per week [16]. Flexible workdays have different formats that may include a combination of overtime and compensatory time off [16,17], a shift of certain hours of a week to the preceding or following week [16], or a shift of the flexible hours of each day (above standard hours and up to the scheduled regular hours per day) to other days [18]. Effectively, flexible workdays allow for shifting workforce capacity from days of low load to days of heavier load. This can be used to mitigate the impact of demand fluctuation and processing time stochasticity [16,17,19], as well as other uncertainties such as the uncertain availability of material [18]. Rules or agreements between employers and workers are often established to protect workers and might include defining the upper bounds of each worker's hours in a day, a week, and sometimes a longer period (e.g., two weeks), and the maximum compensatory hours. These restrictions directly impact the time horizon of working-time planning. While adding useful flexibility, the use of flexible workdays has certain limitations. Workers' lifestyles may be impacted by their varying work schedules. To employers, capacity loss is a limitation of flexible workdays if overtime hours are compensated for by reducing regular hours. Flexible workdays are less expensive than pure overtime because lesser incentives are required. Overall, planning flexible workdays is more complex than planning for overtime workers. A major research question on flexible workdays is to determine the planning method for using it. Particularly, how many flexible hours or time off to schedule for each day.

Dynamic methods have been developed for the use of flexible workdays. For example, Yang, et al. [16] proposed three workday policies and evaluated each using simulation. The first policy compensates each hour in excess of a 40-hour week with one and a half hours of time off and the cumulative compensated hours must be within the range from zero to a positive upper bound. The second policy is similar, except that the lower bound of the cumulative compensated hours can be negative. The third policy shifts a certain number of hours from one week to the preceding or following week so that the total hours during the two weeks is 80 hours. Yang, et al. proposed an input/output control approach (i.e. checking the difference between the available backlog and target backlog) to dynamically adjust the working hours for each day. Their results suggested that these simple policies improved the performance of job shops by reducing the mean percentage of tardy jobs and the mean flow time. Yang, et al. [19] studied another three workday policies in a job-shop setting. Two policies are dynamic, reducing the workday for each machine (or the entire shop) if the queue of jobs falls below a lower bound and increasing the workday if the queue exceeds an upper bound. The third policy is the traditional fixed 8-hour daily schedule, which provides a benchmark to gauge other policies. Yang, et al. [17] further examined the interaction of flexible workdays and workforce cross-training in a simulation study. From this research they found cross-training to be more effective than flexible workdays in improving the performance of job shops. As more workers are cross-trained, flexible workdays become less valuable.

Galbreth, et al. [18] then studied the role of workday flexibility under uncertain material availability. In their study, standard working hours per day were divided into regular hours and flexible hours. Overtime exceeding the standard working hours was allowed, yet flexible hours were charged with a premium lower than the premium for overtime. The study focused on determining how many flexible hours to use, as well as overtime. The value of workday flexibility was dependent of the overall staffing level.

Galbreth, et al. built an optimization model for minimizing total labor costs and penalties for production delays. Simulation was used to find the optimal levels of regular hours, flexible hours, and overtime hours for various staffing levels.

2.3.3. Annualized Hours. Annualized hours are a form of labor time flexibility that contracts labor to work for a certain number of hours per year and then allocates those hours throughout the year in order to better match random demand fluctuations as well as demand seasonality [20]. Annualized hours can reduce the overall costs associated with flexible labor (including hiring, firing, training, temporary labor, and overtime), inventory costs, and penalties unmet demand [21]. However, annualized hours lead to irregular shifts that tend to lessen working conditions and job satisfaction. To compensate, workers are either paid more or given a reduction in total annualized hours, which attenuates some of the cost reduction gains obtained by taking this approach. That is, efficient trade-offs between the costs and benefits of annualized hours need to be determined. Moreover, determining practical workforce schedules under annualized hours is in itself a complex problem [22–25]. For example, Corominas, et al. [25] found that the diversity of production systems leads to great variety in the problems that annualized hours entail. Corominas, et al. [21] provided a scheme for classifying annualized working-hour problems, which leads to thousands of cases. These problems have, thus far, been addressed in one of two ways, either using simple heuristics [22–24], or more elaborate optimization approaches such as Integer Programming (IP), Mixed Integer Linear Programming (MILP), and Stochastic Programming (SP) [25–35].

Hung [22] developed simple algorithms for determining the workforce size and scheduling workers under an annualized hours scenario. The workforce size is determined with consideration of the number of operating days per week, the number of workers on duty per day, the length of shift, the number of contract hours per worker, and the labor cost. After the workforce size is determined, workers are scheduled for each week based on the workforce capacity requirements for that week. Hung [23] further extended this problem by considering the use of multiple shifts. Azmat and Widmer [24] studied the planning and scheduling for a more practical situation where holiday weeks for workers are considered. Azmat and Widmer proposed a straightforward three-step approach that involves the determination of minimal workforce size, searching for a feasible solution with a minimal amount of overtime, and the allocation of workdays per week to each available worker. Azmat, et al. [26] replaced the heuristic method in the third step of [24] with four Mixed Integer Programs (MIPs) to solve the worker allocation problem. Objectives of the MIPs are to minimize the workload difference among workers and the use of overtime.

A variety of annualized hours problems have been studied using IP or MILP models. For example, Corominas, et al. [25] studied the problem of planning workers' annual hours for each week of a production system's year with multiple types of workers and tasks. The objective is to minimize both the labor related costs (such as the use of temporary workers, overtime, and mismatch between tasks and workers) and the difference of workloads among workers. Corominas, et al. [27] developed an IP model to plan annualized hours with weekly working hours belong to a finite set. Lusa, et al. [28] proposed two mathematical models for selecting the most appropriate set of weekly working hours for planning annualized hours, which consider operational factors such as demand and costs. Corominas, et al. [29] developed a MILP model for planning annualized hours of cross-trained workers. Lusa, et al. [30] further considered the efficiency differences among cross-trained workers in planning annualized hours. Corominas, et al. [31] developed MILP models for planning and scheduling annualized hours in a multi-product system where workers may take holidays at the same time, products are perishable, demand can be deferred, and temporary workers can be used. Hertz, et al. [32] built a MILP to plan annualized hours with multiple shifts.

Planning annualized hours under demand uncertainty was studied as well. Lusa, et al. [33] formulated it as a multistage stochastic optimization problem. The stochastic demand was represented by a scenario tree. When a node of the tree is reached, the number of working hours per period from that node to the end of that stage must be determined. The objective is to minimize expected mismatch between planned hours and realized requirement on capacity. Corominas and Pastor [34] proposed a method for re-planning working time under annualized hours when demand is uncertain. A MILP model was developed for the re-planning problem with an objective to minimize the cost of new plan as well as the change of worker schedules. Corominas and Pastor noted that generally a new plan cannot be executed if it does not take the previous plan into account. Consequently, a four-step process for re-planning working hours was proposed: (1) evaluation of the cost of the new plan; (2) evaluation of the impact of restrictions limiting the change; (3) calculation of the discrepancy between the new plan and old plan; and (4) evaluation of the cost for decreasing the discrepancy.

The trade-offs between the costs and benefits of annualized hours was studied by Corominas, et al. [35]. They developed a MILP model to help negotiate the best conditions for annualized hours systems. Their work was motivated by the fact that the flexibility offered by annualized hours is at a cost of working time reduction, for example. Corominas, et al. used numerical experiments to assess the trade-off between the benefits and costs of annualized hours.

2.4. FLOATERS

Floaters (also named floating workers, utility workers) are generalists who are dynamically allocated to jobs where they are needed, either in real time, or to fill in over daily or weekly periods. They naturally should be capable of performing a greater number of operations than the regular workforce who commonly have only minimal training on tasks other than their primary task set. Correspondingly, floaters tend to have either a higher pay grade, or receive additional pay when used in this role [43–45]. Floaters are commonly used in production and service systems that produce multiple products, or serve multiple classes of customers, at multiple stations/plants/departments organized as a chain or network. Workloads in the various operational units or locations may be unbalanced due to changes in workforce demand and supply, processing times, and operational conditions. The unbalance impacts the overall performance of the system. However, floaters' skill sets are not optimized at a system level, which is often the case for cross-trained workers. Yet, floaters share some similarity with cross-training and other work-sharing systems. They may be distinguished from other work sharing systems by intent. That is, they generally only cover a job/task when the real-time need arises. For example, floaters, generally do not have any pre-planned activities, rather they behave as dynamic agents or resources that are reassigned in an ad hoc real-time manner. The flexibility generation mechanism of floaters is simple. A small portion of the workforce is set to be available to float as generalists. The use of floaters potentially creates additional costs such as higher worker wages, set-up costs, set-up times, and longer working times (due to less task-specific experience) [45, 46]. The level of flexibility increases linearly with increased portion of floaters in the system, yet costs associated with the use of floaters grow very fast and can quickly outweigh the benefits. Therefore, the use of floaters is appropriate for systems that can use a modest amount of flexible capacity to accomplish significant benefits.

Current research on floaters is divided into two streams. One stream is focused on the optimal use of floaters such as the allocation of a system's floaters to maximize the system's performance or minimize the system's cost. This problem has often been modeled as a Markov Decision Process (MDP) with optimal policies derived through either Dynamic Programming (DP) or specialized heuristic methods. The second stream is dedicated to the integration of floaters with other flexible methods.

The use of floaters in clearing systems (no external arrival of jobs until originally scheduled jobs clear) has been well studied [46–49]. Particularly, dynamic allocation of floaters in clearing systems of two tandem queues have been modeled as DP problems. The optimal rules for dynamically allocating floaters are expressed as a boundary on the space of two job queues. Farrar [47] assumed a fixed server at each station, and a floater that may be switched off or dynamically allocated between the two stations. Floater movement was assumed to be at no cost with the objective of minimizing expected holding costs. Ahn, et al. [48] studied the case of two servers with external job arrivals, where the first server was dedicated to processing job type-1, the second server was primarily processing job type-2 and floated to job type-1 when necessary. Three optimal policies were derived for different conditions, which include: (i) an exhaustive policy for job type-2, (ii) a non-increasing boundary in the queue of job type-1, and (iii) a non-decreasing boundary in the queue of job type-1. Pandelis [49] extended these models and considered conditions of jobs leaving the system after leaving the first station, as well as less constrained cases where the floater can work on both stations. Pandelis [46] later considered the operating cost of using floaters, finding that it may be optimal to idle the floater when the operating costs are higher than holding costs.

General cases having more than two serial stations and external arrivals were studied in [45, 50]. Sennott, et al. [45] developed an MDP model for K-station serial production lines (K 2) with one specialist dedicated to each station and one floater who is able to perform all of the tasks on a line. Holding costs and set-up costs were modeled at each station, with the objective of minimizing the long run expected average cost per unit time. Since more than two stations were considered, the optimal solution to the MDP model was obtained numerically. Sennott, et al. found that the use of floaters is beneficial for fairly short lines or U-shaped work cells. Wu, et al. [50] studied a K-station serial system that faces possible failure of servers. To address 'the curse of dimensionality' for systems with many stations, they proposed two-pairing heuristics (i.e. looking at only two stations at a time) for solving the MDP model aimed at maximizing the long run average throughput.

The optimal use of floaters can be modeled as a deterministic problem in some applications. For example, Cevikcan and Durmusoglu [51] examined the use of floaters, which they termed utility workers, in a mixed-model assembly line that faces an uneven demand distribution for different models, as well as heterogeneous processing time of the models. Products are launched to the line at a fixed cycle time and no buffer is allowed between stations. Utility workers will help with stations where regular workers cannot finish their work within the cycle time (since some models take longer time than the cycle time to process). Cevikcan and Durmusoglu [51] has focused on task sequencing in this study and sequentially solved two MIP models. The first model minimizes the total utility time, and this solution becomes a constraint in the second model that minimizes the number of floater transfers. Ultimately, they introduced three heuristics to address this problem based on the number of floaters required to meet demand. Boysen, et al. [52] also modeled the use of floaters in mixed-model lines. They built a binary linear program for sequencing products aiming at minimizing the total number of overload situations by using utility workers appropriately.

The integration of floaters with other workforce flexibility methods was investigated also. For example, Wild and Schneewei [43] found that the mix of floaters and non-floaters, and the mix of floaters with other flexibility methods such as temporary workers and overtime, needs to be optimized for cost-effective system performance. For this they presented a hierarchical decision model spanning multiple time horizons. Francas, et al. [44] addressed a similar problem with a two-stage SP model, where the first stage involved a strategic investment decision in workforce capacity and machine capacity, and the second stage involved the operational decision on the optimal use of floaters and temporary workers.

The use of floaters has existed in production systems for a very long time, and along with it, a general natural understanding of the concept. This along with its power to provide real-time dynamic system balancing and flexibility are its greatest strengths.

Workers who are designated as floaters most typically are paid at a higher rate given their responsibilities, leading to higher overall costs. Also, relative to other approaches, floaters may have low task specific expertise. For instance, if a floater is needed on a production line for the first time, it is possible that the throughput rate decreases initially, while they learn to integrate themselves in the unfamiliar processes.

2.5. CROSS-TRAINING

Cross-training is one of the more broadly discussed workforce flexibility methods, particularly for complex systems where the delivery of products or services cannot be accomplished by finishing singular tasks. Example systems include production lines (either serial or parallel), job shops, manufacturing cells, call centers, health departments, and field service departments. Those systems may require flexibilities in workforce capacities and/or worker capabilities. Through cross-training, workers obtain and maintain skills to work on multiple tasks, so that they can be re-assigned where and when they are needed [1, 53]. However, mechanisms for cross-training are often complex. Surveys on cross-training re- search have included Nembhard [54] which summarized cross- training theory, practice, and challenges. Earlier, Treleven [55] and Hottenstein and Bowman [56] reviewed cross-training research in dual resource constrained (DRC) systems, wherein machines and the workforce constitute the two constraining quantities. Hopp and Van Oyen [1] built a framework for evaluating cross-training strategies and analyzed the cross-training literature in manufacturing and service operations. Aksin, et al. [57] reviewed the literature of cross-training for call center applications. Different from other review papers dedicated to cross-training, the discussion of cross-training in this paper is focused on summarizing the problem types, research questions, and solution generation methodologies for cross-training by reviewing relevant cross-training research.

Both multi-functionality and redundancy are important flexibility designs for complex systems of multiple tasks. Much as multi-functionality is the number of different tasks a worker can perform; redundancy is the number of workers able to perform a specific task. Systems with workforce multi-functionality and redundancy are robust to workforce supply and demand uncertainty and to varying distributions of workforce supply and demand among tasks. Such systems are also more capable of efficiently utilizing their workforce. Cross-training is a method that can build and maintain workforce multi-functionality and redundancy in an integrated fashion. Thus, it is useful for systems where considerable task heterogeneity is present and both workforce redundancy and multi-functionality are valuable. Sources of task heterogeneity can be either internal or external, wherein product and service complexity is a primary internal source of task heterogeneity. Complexity requires the delivery process to be designed as a set or sequence of heterogeneous tasks. External sources of heterogeneity include product mix or service mix. Different classes of customers need service agents with particular skills and knowledge; customized products are produced by workers capable of performing particular tasks.

Systems with high task heterogeneity are particularly vulnerable to workforce supply and demand uncertainty, as well as varying workforce supply and demand distributions among tasks. In such systems a worker's role cannot be naturally replaced by a random worker because specific skills or knowledge may be required. Uncertainty and/or variability in workforce demand are caused by, for example, product/model mix or service mix, random arrivals of jobs, stochastic time requirements, demand seasonality, short product life cycles, and market changes. Uncertainty and/or variability in workforce supply can also come from absenteeism and attrition. The flexibilities provided by cross-training can allow for better matching of workforce supply and demand.

Cross-training may bring a variety of benefits including lower labor costs, shorter lead times, higher quality, and increased production flexibility [1]. These benefits, however, often come along with costly side-effects. Cross-training is best used with careful consideration of the trade-offs between these costs and benefits. The cross-training costs modeled in the literature are summarized in Table 2.2.

Table 2.2 Costs and other side-effects of cross-training.

Cost Type	Literature
Reduced efficiency: Workers may have varying levels of efficiency on different tasks. In the practice of cross-training, workers are expected to be 100% efficient on the primary task they are responsible for and uually at lower efficiency on secondary tasks. The lower efficiency is commonly modeled as a cost associated with cross-training.	[64, 69, 72, 86-91, 93, 94, 101, 102, 106, 108]
Productivity/quality loss: When the productivity of workers or the quality of their work depends on experience on the same or similar tasks, the effect of learning and/or forgetting presents a loss due to frequent switches among different tasks. The effect of learning and/or forgetting may also cause the loss of quality associated with cross-training.	[60, 63, 73, 77-79, 82, 83, 98, 109, 110]
Transfer costs: Transfer (transition, switching) costs refer to costs associated with moving between tasks. For example, workers may spend time traveling to other departments, machines, or service sites. Workers may need to access new information, set up equipment, or adjust their status before they start a new task. In some circumstances, the administrative cost for worker transfer may be counted as well.	[65, 68, 84, 107]
Training costs: Training costs may include the time spent on direct training, investment in the infrastructure used by training, coaching costs, and costs for obtaining certifications.	[75, 77, 84, 86, 94, 97, 105, 107]
Additional wages: Additional employee compensation may be associated with cross-training. This is often skill-based compensation, where multi-skilled workers are paid more than those with single skills. Also, incentives may be given for undertaking additional on-the-job training.	[58, 75, 105-107]

In addition to the five types of costs, other unintended consequences of cross-training may include additional turnover since a multi-skilled worker may become more mobile, lower social identity, lower morale, and responsibility confusion. While these are discussed throughout the literature, they tend not to be modeled primarily due to the difficulty in quantifying these consequences. Moreover, costs for acquiring and maintaining cross-trained workers may become less affordable, or significantly dominate the benefits for systems in non-stationary environments where changes are significant, occurring too fast, or highly unpredictable.

Another important cross-training decision involves the choice among specific configurations, which addresses the question of who should be cross-trained on what tasks. That is, one may view this as the design of the worker-task skill matrix, which also must reflect levels of workforce multi-functionality and redundancy. There are four common representative configurations illustrated in Figure 2.1 and described below.

<u>No cross-training:</u> no task is performed by more than one worker (or a class of workers) and no worker performs more than one task, as Figure 2.1(a) illustrates. This configuration, which provides neither workforce redundancy nor multi-functionality, is often used as the benchmark case to which cross-training is compared.

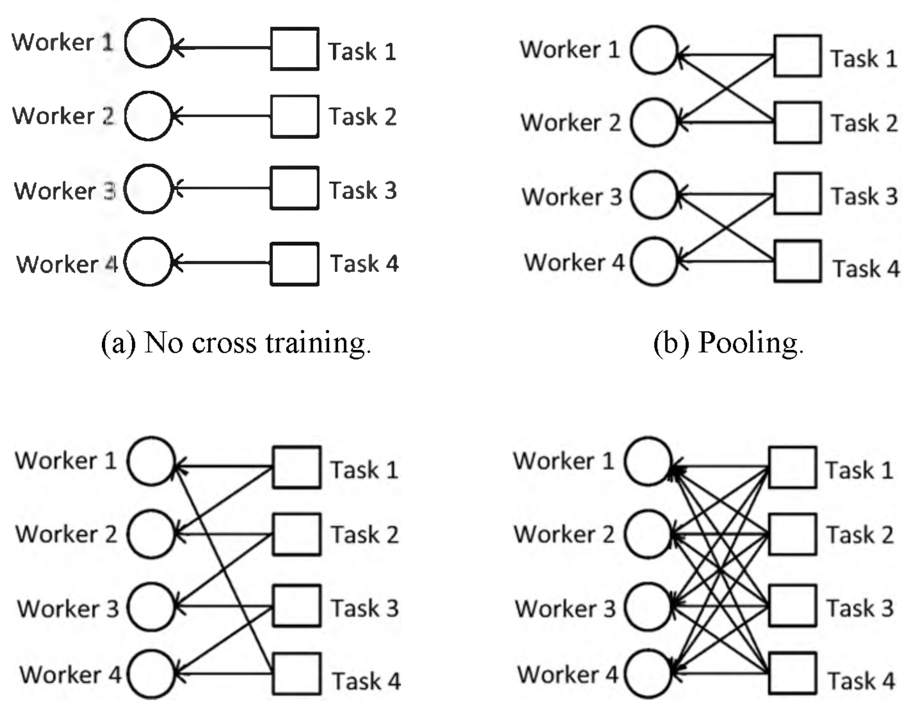
<u>Pooling:</u> pooling is a partial cross-training strategy suitable for systems where tasks can be pooled as a few larger sets because of task similarities. It is more cost-effective to cross-train workers to perform tasks within a pool, but not across pools. As illustrated in Figure 2.1(b), tasks 1 and 2 are pooled. Workers 1 and 2 are cross trained to perform these two tasks, but not tasks 3 and 4, which are

outside the pool. We note though, that it is an important combinatorial question to determine which tasks to pool. Aksin, et al. [111] discussed the design of pooling strategies in call centers. Tekin, et al. [104] developed pooling strategies for a call center via cross-training wherein pooling departments with highest service time coefficients of variation reduces the expected service delay more when mean service times are similar among departments. Easton [88] found that the correlation among multiple demand streams plays an important role in determining the practical effects of pooling. A corresponding broad research question is to determine how many workers to be pooled via cross-training. Robbins, et al. [105] analyzed a partial pooling strategy that considered a subset of the workforce to be pooled rather than a single homogeneous policy.

Chaining: (or skill-chaining) is another partial cross-training strategy wherein each worker can directly or indirectly aid in performing other tasks in the system. Chaining is suitable when there is a locative or functional proximity. For example, in Figure 2.1(c), worker 1 directly works on tasks 1 and 4, and worker 2 can perform tasks 1 and 2. In contrast to pooling, each worker may have a unique subset of skills with proximity. Detailed studies of chaining strategies for cross-training were presented in [53,59,69,76,85,86,89,95]. The chaining strategy is found to be an effective, robust configuration of cross-training if, for example, cross-training is costly, systems are of high variability or uncertainty, and WIP is required to be low.

Full cross-training: every worker is trained to perform all tasks, as Figure 2.1(d) depicts. Full cross-training provides the greatest level of flexibility. However, it may not be feasible or economical considering the growing costs and penalties associated with full flexibility. A straightforward constraint on this approach is to limit cross-training to a maximum number of tasks, particularly when the full task set is large.

Thus, pooling is a means of grouping tasks often by some measure of similarity, chaining groups by task proximity, and full training forms one large pool. We remark that there are a number of other specific partial cross-training configurations with irregular worker-task patterns. While we do not enumerate all of these, noting that other configurations tend to be combinations, extensions, or special cases of these archetypes.



(a) No cross training. (b) Pooling.

(c) Chaining. (d) Full cross training.

Figure 2.1 Representative configurations of cross-training.

The performance of cross-training configurations depends on a variety of factors. Choosing the most appropriate configuration is a critical step in designing a cross-training system. A broad class of literature evaluates and selects cross-training configurations through comparative studies. Generally, these studies simulate the system where cross-training is implemented. Also, numerical experiments are designed to compare the performance of candidate configurations in order to examine the manner in which performance is influenced by other factors such as environment, system settings, and workforce assignment policies (see [64,65,81,82,84,94–97,100,107]). A related stream of research uses constrained optimization models to identify an optimal cross-training configuration. IP is a commonly used model for these investigations (see [58, 62, 66, 77, 86, 87, 89, 90, 92, 93, 99, 101, 106]). In these IP models, the determination of optimal cross-training configurations is formulated with a variety of decision variables. For example, one may want to identify the number of workers to be trained on each task, and specifically who should be cross-trained on each task. Unfortunately, these IP problems are not solved easily in general, due to the combinatorics, and in some cases non-linearity (for mixed integer non-linear variants). Algorithms for solving the IP problems have been intensively investigated, including Branch-and-Bound (BNB), Genetic Algorithms (GA), linear approximation, and heuristics. Queueing is another tool considered by some for determining cross-training configurations when systems can be represented by standard queueing models (see [83,102–104]). Other methods are also used to facilitate the selection of cross-training configurations. Tiwari and Roy [79] proposed a fuzzy system-based method to determine the amount of cross-training. Iravani, et al. [97] developed a small world network structure for which the average shortest path length was used to evaluate cross-training configurations. Also, Kim and Nembhard [78] introduced association rule mining, a data mining technique, to conduct knowledge discovery of useful cross-training configurations.

The issue of how to effectively use cross-trained workers and to maintain their skills is another broad category of research. Assignment methods address the allocation of cross-trained workers to specific tasks that they are qualified to perform. For example, these methods refer to and determine worker-task combinations, as well as who should be assigned if more than one cross-trained worker is available. Hopp and Van Oyen [1] and Hopp, et al. [69] provided a classification of assignment and coordination policies.

Schedule oriented assignment methods are appropriate for deterministic systems for which workforce supply and demand during the scheduling time horizon are known. Workers are rotated among tasks based on a schedule that can be visualized by showing changes to the assignment of workers to tasks over time. The worker schedule timetable is derived based on the supply and demand forecast, as well as the worker-task matrix. A mathematical programming model is often used for this purpose, which minimizes the total cost or maximizes the total profit subject to a set of constraints. The most commonly used model is IP (e.g., [77,80,86,89,92,99,106]). In addition to IP, others have used Goal Programming (GP) (e.g., [91]), Integer Goal Programming (IGP) (e.g., [62]), and MIP (e.g., [67,88,101]).

Dynamic assignment policies are often used by systems facing uncertainties in workforce demands and supplies or other unpredictable changes. These policies dynamically determine ‘‘when’’ to transfer cross-trained workers and ‘‘where’’ they should be dispatched to. Moreover, if multiple workers are idle, the question of ‘‘who’’ should be assigned to the next incoming job must be determined. When to transfer a cross-trained worker from her current job to other jobs is often determined by either a centralized rule or a decentralized rule [60]. A centralized rule allows workers to transfer when they complete their current job whereas a decentralized rule allows workers to transfer only after they finish all jobs in their current machines, stations, or departments. Kher and Malhotra [60] stated that the centralized rule maximizes the utilization of workforce flexibility yet increases the transfers of workers. Their study showed that decentralized rules performed better than centralized because the frequent transfers may outweigh the benefit of workforce flexibility.

When cross-trained workers are available for transfer, the question of where they should be assigned must also be addressed. Generally, there are three categories of ‘‘where’’ assigning policies, which are based on the system status, worker types, and task types, respectively. System status-based assignment policies use the information of system status variables such as job queue lengths, queue gaps, workloads, workload gaps, and WIP, to determine what task an available cross-trained worker should perform. The general mechanism underlying these assignment methods is dynamically assigning cross-trained workers to places where they are not only capable of work but also the most in need. The most commonly used policies are to assign cross-trained workers to workstations with the longest queue (LQ), the heaviest workload, or the largest WIP (e.g., [69,76,95,97]). If workers are cross-trained to perform tasks on consecutive workstations, the maximum gap of queues, workloads, or WIP between two consecutive stations may also be the indication of where cross-trained workers may be most effective.

Job types or characteristics are also used to determine where available cross-trained workers should go. A commonly used policy is Fixed before Share (FbS) that assigns workers to their fixed or primary tasks (on which they are most efficient) if any available; otherwise, assigns them to cross-trained or shared tasks (on which they have reduced efficiency) [69,72,76]. Some studies prioritize jobs and, accordingly, assign workers to jobs with the highest priority. For example, Kher and Malhotra [60] set the highest priority on the job with the least slack time toward its due time. Parvin et al. [76] used a last-buffer first-served rule for workers who are cross-trained on shared zones. An available job at the most downstream station would have the highest priority. Agnihothri and Mishra [84] placed the higher priority on jobs, which were previously assigned to technicians who were not able to process and reassigned to these with the greater skills. Hopp, et al. [69] determined the priority of jobs according to long-run worker time efforts required. Tekin, et al. [104] considered a Non-Pre-Emptive Priority (NPP) policy that prioritizes jobs in the same pool and the available cross-trained workers choose jobs according to the priority level of the jobs.

Some literature uses the viewpoint of “routing tasks to workers” based on worker types rather than “assigning workers to tasks” for worker assignments. The Specialists before Generalists (SbG) policy assigns a job to a worker who is dedicated to this job type if a such worker is available; otherwise, the job is assigned to an available cross-trained worker. This policy is used in both service systems (e.g., [83,94,105]) and production systems (e.g., [64,95]). A possible reason for the SbG policy being more popular is that dedicated workers are usually more efficient than cross- trained workers. Fully utilizing dedicated workers first tends to maximize the system performance. Less attention has been given to the use of the Generalists before Specialists (GbS) policy. Yet GbS may be appropriate if the correct classification of customers can- not be done easily based on the information provided by customers themselves and when the services provided by dedicated workers are significantly more expensive than those provided by cross- trained workers. A common example of this dynamic policy is in the health delivery domain, wherein specialists tend to be more expensive than generalists. Furthermore, when multiple workers of a type (dedicated or cross-trained) are available, a ‘‘who’’ rule is needed for determining the worker who will take the next avail- able job arrival. The Longest Idle Time (LIT) policy is often used for this purpose (e.g., [62,64,81,95,103]).

Some dynamic policies are too complex to be presented as straightforward rules. These are usually derived based on an MDP (e.g., [68,70,75]), optimal control of queues (e.g., [71]), or fuzzy expert systems (e.g. [79]).

The strengths of cross-training for flexibility include the ability to address both demand and product mix uncertainties. It is also very well suited for systems with high task heterogeneity, where there is enhanced ability to match workers to their best suited tasks. However, since cross-training generally involves training on multiple tasks as a preplanned contingency for its potential future use, training costs tend to be high and grow disproportionately larger with higher cross-training levels. If the workforce has relatively low turnover, this fact is somewhat mitigated, yet turnover tends to inflate cross-training costs, generally. Nonetheless there are considerable trade-offs between the costs, and the potential benefits from the flexibility obtained. Optimizing the performance of cross training additionally requires considerable planning, perhaps including simulation and optimization modeling.

2.6. TEAMWORK

Teamwork has been studied from a variety of perspectives, including psychological and creative viewpoints. A team may be de- fined as a group of two or more people with specific or non-specific roles or functions, who interact dynamically, interdependently, and adaptively to target a common objective or mission [112]. Teamwork is often intended to accomplish more work than sets of workers working individually [1]. For instance, team members may have complementary skills, making the team more productive than individuals; teamwork may improve quality and reduce productivity loss, or improve motivation and reduce work fatigue. Besides these benefits, teamwork is also found to be a potential means toward facilitating flexibility. Yauch [113] specifically analyzed team attributes necessary to facilitate agile manufacturing, including multi-functionality, dynamism, cooperation (or collaboration), and virtualness. The first three of these are relevant to flexibility generation. Multi-functional teams create flexibility in workforce capacities or capabilities by combining the skills needed to enable high performance, decreasing the time needed for product design and development, and adding workforce capacity when and where it is most needed. Yauch stated that multi-functionality can be interpreted from two perspectives. First, it can be viewed as a set of individuals or team members with unique skills. The second perspective more broadly considers teaming up cross- trained workers. Dynamic teams are formed temporarily for special purposes to add flexibility and to enable rapid reconfiguration. Collaborative teams promote flexibility through higher group productivity, the generation of new ideas and insights, and higher levels of learning and skill development. In operations management research, collaboration and multi-functionality are the most commonly discussed teamwork attributes of flexibility. More often than not, these two attributes co-exist in many teams.

2.6.1. Collaborative Teams. In collaborative systems where at least a portion of tasks for a job can be performed simultaneously, teamwork may help increase the average task speed or labor productivity. A question regarding collaboration is perhaps the identification of circumstances where collaborative teams perform better than sets of individuals. Buzacott [114] used queueing models to analyze the performance of collaborative teams and found that the mean job completion time was shorter for teams than for individuals. However, the improvement depends on several important factors including the variability level of task processing times, and the utilization of servers. Van Oyen, et al. [4] also showed that collaborative teams are beneficial for systems with high variability. Under some circumstances, such as operational environments with low utilization, low variability, and a lack of balance, cooperative teams may not improve system performance unless collaborative efficiency is very high. Mandelbaum and Reiman [115] examined the effectiveness of pooling servers into teams. They found that the pooling of tasks and servers may reduce the steady-state average sojourn times for some circumstances such as under light-traffic and low variability in pooled tasks. Kim and Burton [116] examined project team performance using computational simulation studies and found that decentralized teams (decisions can be made by team members) help reduce the time and cost more than centralized teams (decisions are mainly made by project managers) when the task uncertainty is high (the extent to which information needed to complete a task is available when it starts and the likelihood that a task will fail and be reworked).

Sengupta and Jacobs [117] conducted computational simulation studies for comparing multi-functional collaborative teams in assembly cells with specialized workers in assembly lines. Sengupta and Jacobs found that assembly cells based on teamwork perform better than assembly lines when setup times and variances of processing times are high. This is due to the fact that team-oriented systems have a greater flexibility to absorb variations in processing times. When the total number of tasks increases, the team-oriented systems are superior to other systems although the efficiency of teamwork is low.

A second question regarding collaboration is how to achieve the maximum effectiveness. Andradóttir, et al. [118] studied dynamic policies of server assignments to maximize long-run average throughput for finite queueing systems. In their study, servers can collaborate on the same job with additive service rates. Collaboration between servers reduces the chance of idling servers and optimizes systems where all servers are generalists. The model was based on an MDP for a serial system with two stations and two servers. Then it was extended to the optimal policy of assigning collaborative servers for serial systems where the number of stations is equal to the number of servers. Van Oyen, et al. [4] developed an expediting policy for dynamically assigning cross- trained workers to collaborate. The attractiveness of this policy depends on the collaborative efficiency of workers. Collaborative efficiency was measured by differentiating the effective processing rate of collaborative teams from that of non-collaborative workers.

2.6.2. Multi-Functional Teams. Multi-functional (cross-trained, multi-skilled) teams may be a net benefit or loss with respect to productive output. An essential research question regarding multi-functionality is to what extent workers can work on multiple tasks in a nearly simultaneous fashion. Associated questions relate to optimal team sizes and configurations.

Molleman and Slomp [119] examined how multi-functionality improves team performance considering the uncertainty in labor supply caused by absenteeism. They measured team performance using shortage (the amount of work that cannot be done by the team), make span (the minimum time needed to perform all tasks), and production time (the cumulative time needed to perform all tasks). Using an LGP model to examine impacts on team performance, multi-functional teams were found to have some advantages, particularly those that are uniformly multi- skilled (e.g., each team member has skills for two tasks). Slomp and Molleman [61] examined the impact of cross-training on the team performance. The latter study indicated diminishing marginal returns from increasing the level of multi-functionality through cross-training. Sengupta and Jacobs [117] analyzed the loss of worker specialization of multi-functional teams and found that the value of flexibility produced by multi-functionality outweigh the losses from specialization when the variances of processing times are high.

Powell [120] modeled the multi-functional effect by assuming that the mean value of a worker's processing time increases faster than linearly with the number tasks performed, and the standard deviation of the processing time rises with the square root of the number of tasks. Noting that multi-functionality may be impacted by collaboration, Powell also modeled collaboration by assuming that as more workers join a team, the mean and standard deviation of job completion time for the collaborative team fall faster than for non-cooperative workers, initially. However, the marginal return from collaboration diminishes so that these measures eventually become greater than those of the non-collaboration case. In order to find preferred levels of multi-functionality and collaboration, Powell simulated line cycle times (LCT) and factory cycle times (FCT). Results showed that moderate levels of each are often appropriate. Inman and Blumenfeld [121] also reported on the determination of team size by considering teams having a specified leader and working on an assembly line where absenteeism was a source of uncertainty in the labor supply. Inman and Blumenfeld reported that the throughput of conforming production increases with team size initially (due to reduced overhead of team leaders) and then decreases (because the productivity suffers from reduced average support of the team leader to each of the workers on the serial line). As the team size increases, the negative effect of insufficient multi-skilled labor capability outweighs the benefit of cost saving. Further, the optimal team size decreases with the rate of absenteeism. Yet cross-training mitigates the negative impact of absenteeism on team size. The optimization of team size can be studied from a variety of perspectives. Though the question of optimal team size is outside the scope of this review and remains a generally open issue, the common consensus is that having a minimum level of multi-functionality achieves many of the benefits with minimal cost.

Optimal configurations for multi-functional teams is a problem studied in operations management also. Most of the studies focus on the multi-functionality created by assembling a set of individuals who each bring a unique skill to the team. For example, Bordoloi and Weatherby [122] developed an LP model to determine the optimal mix of staff in different categories for a hospital medical unit. The objective was to minimize the total cost subject to constraints of patient demand and staffing policies. Perron [123] developed an optimization model both planning jobs over time by assembling and re-assembling multi-functional teams for jobs and scheduling workers in each planned teamwork. Perron primarily focused on the development of models and solution approach, particularly for large-scale problems (i.e., 800 workers and 2000 jobs).

As an approach for obtaining flexibility, teamwork is particularly strong as a self-organizing approach. That is, management may set the teams, but the teams often address operational issues. This may as a secondary effect yield some level of cross-training, though some of this will likely be only observational, rather than procedural (hands-on). Thus, teams are particularly useful for handling high variability systems, and those with absenteeism. One common weakness of using teamwork, is that simply determining ideal team sizes, is in itself a difficult problem to address. For instance, larger teams, and teams with larger task loads may suffer from lowered efficiency due to a higher communication, and overhead load. Also, it is well understood that team dynamics often determine the success or failure of a team's performance, with multi-functional teams at somewhat of a disadvantage when there are simultaneous tasks to perform.

2.7. SUMMARY OF WORKFORCE FLEXIBILITY METHODS

In this section, literature was reviewed that was associated with workforce flexibility to classify and determine the scope and effectiveness of each flexibility method. The review spanned the ORMS area and included flexible working time, floaters, cross-training, teamwork, and temporary workers. The selection of workforce flexibility method(s) for a specific application is often a decision with multiple considerations. To facilitate such decision making, we summarize the key aspects of each method from the ORMS perspective in Table 2.3. These include problem types for which a method is appropriate, primary research questions in the use of the method, methodologies for addressing these research questions, and representative ORMS references of this method.

Table 2.3 A summary of workforce flexibility methods.

Flexible working time Relaxing standard shift lengths and workweek hours **Primary problem types** • Demand uncertainty • Seasonality **Research questions** • What is an optimal trade-off between flexible time costs and benefits? • How to plan working hours and schedule workers under flexible working time? **Models/methodologies** • Mathematical programming such as Integer Programming (IP) and Mixed Integer Linear Programming (MILP) • Stochastic programming (SP) and dynamic policies • Simulation-based optimization • Heuristic methods **Literature** [9–42]
Floaters Designating classes of workers to float to stations with greatest need **Primary problem types** • Product mix uncertainty • Unbalanced workloads in various operational units or locations **Research questions** • How to allocate floaters to operational units to optimize the system performance? • How to integrate floaters with other workforce flexibility methods? **Models/methodologies** • Queueing model based Markov Decision Process (MDP) solved with Dynamic Programming (DP) or heuristic methods • Mathematical programming models such as IP, primarily solved with heuristic methods • Mathematical programming models, as part of a hierarchical model or two-stage SP, solved directly or with heuristics **Literature** [43–52]

Table 2.3 A summary of workforce flexibility methods. (cont.)

Cross-training Training workers at multiple skills **Primary problem types** • High product mix uncertainty • Demand uncertainty • Workforce supply uncertainty **Research questions** • How to model the costs and side-effects of cross-training, and to achieve the best trade-off between the costs and benefits? • Who should be cross-trained on what tasks? • How to assign cross-trained workers to tasks to best utilize and maintain their multi-functionality? **Models/methodologies** • Optimization based methods including mathematical programming models (e.g., IP, Goal Programming (GP), Integer Goal Programming (IGP)), network, SP, DP, optimal control, MDP, and evolutionary fuzzy expert systems • Simulation based comparison studies • Data mining methods such as association rules **Literature** [1, 53–111]
Teamwork Functional or additive collaboration **Primary problem types** • Product mix uncertainty • Workforce supply uncertainty **Research questions** • Under what circumstances can collaboration improve performance at the team level? • How to coordinate team members to maximize the performance improvement by teams? • To what extent can workers perform multiple tasks simultaneously? • What is the optimal size or configuration of teams? **Models/methodologies** • Queueing modeling, control, and MDP • Comparative studies based on computational simulations • Mathematical programming models (LP, LGP, and IP) **Literature** [4, 61, 112–123]
Temporary labor Contracting additional short-term labor **Primary problem types** • Seasonality • Demand uncertainty • Uncertainty in the supply of permanent workers **Research questions** • How many temporary workers to hire to yield cost effective capacity? • How to optimize the mix of temporary workers, permanent workers, and/or other labor flexibility such as overtime or floaters? **Models/methodologies** • Single-period models maximizing expected profit or minimizing expected total cost, generally solved with DP • Single-period, two-stage SP models maximizing expected profit or minimizing expected total labor cost • Multi-period DP models minimizing expected labor and backlog costs **Literature** [43, 44, 124–130]

Among the wide range of possible problem types, we distinguish between problems arising from changes in demand volume or worker supply, and problems arising from changes in product mix. Examples of the former include seasonality, demand uncertainty, and worker absenteeism; and the latter concerns systems producing multiple products or providing multiple services. While flexibility is a potentially useful mitigator of both types of uncertainty, specific methods are often designed to focus on one type or the other. For instance, flexible working time and temporary labor are primarily used for dealing with changes in demand volume and worker supply. Floaters are often used to mitigate system imbalances caused by changes in product mix, for example. Yet we notice that some flexibility methods may serve multiple purposes. For example, cross-trained workers can be used or not.

3. A FRAMEWORK FOR FORMING FLEXIBLE SYSTEM OF SYSTEMS: MECHANISMS, NEEDS, ARCHITECTURE, AND DECISIONS

3.1. THE ROLE OF FLEXIBILITY IN A SYSTEM OF SYSTEMS

Flexibility is a desired attribute of various systems operating in changing or uncertain conditions [124], [125]. Therefore, it has been an important consideration of system design and widely implemented in practice. For example, a production line may be designed to be flexible in switching among product models, or accommodating product updates, to respond to changes that may be unknown early in the lifecycle. Today, systems are increasingly more complex or larger than they were used to be, to adapt to the growing and expanding social needs of human beings and rapid technological advancements. The complex architecture of those systems, and the growing importance of embedding flexibility to these systems, make the design and operations of flexible systems a research question for systems engineers.

What flexibility can be created through the design and operations of a system, and how, remain research questions. The current literature on flexibility is largely centered on specific application domains, such as manufacturing flexibility (e.g. [126], [127]), workforce flexibility (e.g. [128], [129]), and others. Despite many successful cases of creating flexibility in various domains, the literature has not been generalized enough to readily support the design and use of flexibility for any engineered systems that are growing in both types and complexity. Moreover, system performance is often characterized by multiple attributes, and flexibility is usually one attribute strongly interdependent of others. Flexibility induces increased or new interactions among systems or system components. Flexibility desired by a system or its elements usually does not naturally exist. The creation and use of the flexibility unavoidably affect other elements or need the collaboration of them. The actual contribution of flexibility to a system is a derivative in that the contribution depends on the evolution of underlying variables driving the needs for flexibility. All the features above make it important to calibrate the flexibility level during the design phase to ensure that the created flexibility is executable and will effectively produce the anticipated benefits later in operations.

This section is motivated to analyze flexibility mechanisms of systems of systems (SoSs) to propose the adoption of a framework, or hierarchical network, for creating flexibility. The novelty in this paper is the derivation of a strategy for forming and evolving SoSs to provide needed flexibility. Specifically, the study is focused on changes or uncertainty that cannot be handled by a simple system in a cost-effective manner, but by systems of systems (SoSs) [130]. A SoS is a reconfigurable arrangement of independent and useful systems to deliver unique capabilities for a mission [131]. A capability is the ability to execute a specified course of action. It is unlikely the central mission can be accomplished by an individual system. We remark that a SoS is not designed to be a simple collection of systems that each brings one of the required capabilities to the SoS [132], [133]. Five characteristics of SoS distinguished a SoS from a system [134], [135], which are autonomy, belonging, connectivity, diversity, and emergence. The remainder of this section is organized as follows: Section 3.2 briefly summarizes the relevant literature to acknowledge the status of current research. Then, Section 3.3 presents the proposed framework for enabling flexibility through designing and operating SoSs, followed by an illustrative case that demonstrates the rationale and feasibility of the proposed methodology in real-world applications. Important findings from this study and identified research needs are summarized in Section 3.5.

3.2. LITERATURE REVIEW

Generally speaking, flexible systems are those that can make changes easily to cope with changes or uncertainty. While it is a desirable characteristic, flexibility is an ambiguous concept. Within the domain of systems engineering, three streams of research efforts have particularly tried to address this issue to improve the communication and capability of designing and analyzing flexible systems among systems engineering practitioners and academics. The first stream of efforts is about defining flexibility (e.g., [124], [125], [136], [137], [138]). These studies all emphasized the critical aspects of flexibility including the existence of needs for flexibility, flexibility mechanisms, and effects of flexibility. The second stream has been focused on measuring and quantifying flexibility (e.g., [137], [139], [140], [141]). The degree to which changes can be made to a system's architecture is a way of quantifying flexibility [139], [142], [143]. Metrics for flexibility have been developed based on system's architecture and used to measure the flexibility of generic system architectures [143]. The third stream studied the interdependence of flexibility with other attributes of systems (e.g., [139], [144], [145]), and the impact of flexibility on system capabilities, performance, and others (e.g., [125], [136], [139]). All these efforts have built a foundation for the study in this paper.

A few research papers explicitly studied the flexibility of SoSs. Gorod et al. [146] examined the flexibility of a SoS as the flexibility of autonomy, flexibility of belongs, flexibility of connectivity, flexibility of emergency, and flexibility of diversity. They developed a concept of flexibility dynamic in their study. Recently, Dagli, et al. [147] led a series of research for developing flexible and intelligent learning architectures for SoSs.

Different than the literature, the work of this research is focused on analyzing the flexibility mechanisms of SoS and, accordingly, deriving a strategy for forming and evolving SoSs to provide needed flexibility.

3.3. THE SOS FRAMEWORK FOR FLEXIBILITY

Flexibility is valuable to stakeholders who not only face changes or uncertainty but are sensitive to these. With appropriate flexibility, the stakeholders are able to make changes quickly and easily to effectively meet their needs despite of the uncertainty. Among all identified needs for flexibility, some can be met by forming and evolving SoSs.

3.3.1. Flexibility Mechanisms of SoSs. To create flexibility through forming SoSs, we first need to determine flexibility mechanisms of SoSs. A SoS is an arrangement of independently operated and managed systems, which are integrated into a larger system that delivers unique capabilities. A SoS, as well as each of its constituent systems, consists of parts and relationship, and a whole of these is greater than the parts [131]. Flexibility can be acquired using one or a combination of the following flexibility mechanisms.

Flexibility of SoS type: There are four SoS types ([132], [148]), which are directed SoS, collaborative SoS, acknowledged SoS, and virtual SoS. A SoS may be designed to be able to switch from one SoS type to another, and it makes the SoS more adaptive to a wider range of operating conditions.

<u>Flexibility of SoS configuration:</u> A SoS is usually reconfigurable in terms of selecting constituent systems to participate in the SoS, as well as determining the collaboration among selected constituent systems, in a dynamic manner.

<u>Flexibility of constituent systems:</u> Constituent systems that are flexible in the capabilities to provide to the SoS, as well as in the performance of providing the capabilities, provide another degree of flexibility to the SoS.[149]

The three mechanisms above form a foundation for designing SoSs for the purpose of enabling flexibility.

3.3.2. The Needs for Flexibility. SoSs would generally meet the needs for flexibility that fall into the categories of changes or uncertainty in constituent systems and moving or ambiguous SoS objectives. Constituent systems' willingness to participate in the SoS, as well as their participating performances, may change over time for many reasons. New systems may emerge, becoming better choices for the SoS than existing constituent systems. Most of the time these changes are unpredictable. The specified outcomes, or objectives, of SoS may evolve for different reasons, such as changing requirements or behavior of stakeholders, dynamic operating environments for the SoS, and others.

All the needs for flexibility are caused by dynamics of various aspects, including operations, technology, market, environment, human behavior and perception, resources, and others. Facing the dynamics, either the optimum of a SoS is transient or SoS objectives are a moving target. A SoS can be quickly and easily revised, reconfigured, and re-calibrated, to always fulfill the SoS central mission.

3.3.3. Hierarchical Network: A More Flexible SoS Architecture. A single SoS may not be able to meet all the needs for flexibility. Hierarchical SoS network is a more flexible SoS architecture that is able to adapt to complex and widely distributed large-scale systems. We categorize hierarchical SoS networks into two major types, which are discussed in the following.

Network of SoSs: Forming a single SoS on a wide area might not be the best design. Instead, a network of multiple SoSs that are geographically distributed may be a better choice. These SoSs are homogeneous in that they are designed to fulfill the same central mission. They are also heterogeneous because each is specifically formed to serve a local group of stakeholders. These SoSs are locally optimal, but maybe not on the entire range of area. Connecting spatially distributed SoSs as a network allows them to collaborate with each other to move towards the global optimum, better serving all stakeholders over a wider area. Since its components are SoSs, this network inherits the properties of SoS.

Super SoS: Some constituent systems of a super SoS are heterogeneous SoSs. This architecture is appropriate when the bigger system needs heterogeneous capabilities that some are impossible to be delivered by simple systems but SoSs.

3.3.4. Decisions for Forming and Evolving a Flexible SoS. Flexibility provides choices to stakeholders facing changes or uncertainty. They dynamically make choices to respond to changed conditions reactively or potential changes proactively. To best use the flexibility they have, agents of SoSs make the following decisions.

SoS architecting: This involves selecting an SoS configuration to create or evolve to, which can be represented by a graph with N selected constituent systems or

SoSs and K edges among them, $G(N, K)$, in fulfilling the SoS mission. Choices of SoS architectures are discrete.

Operational planning and execution: If some or all of the selected constituent systems are flexible, the SoS agent needs to determine what capacities to request from the flexible constituent systems, the specifications of participation and collaboration for them. The deviation of real contributions from the specifications needs to be managed in a near real-time manner.

Collaboration approach: When needed, the approach to coordinating constituent systems in the SoS can be switched from one to another. A SoS with a dedicated agent coordinating constituent systems may be switched to one relying on peer collaboration.

The decisions discussed above are interdependent. For example, the SoS architecting uses the inputs from operational planning and execution; meanwhile, the result from the former constrains the latter.

3.4. A CASE STUDY: MICROGRIDS AS AN SOS NETWORK

Renewable energy sources (RESs) are often geographically distributed, volatile, and intermittent. Using a single RES or a conventional source usually does not meet requirements on energy generation and supply, such as affordability, reliability, sustainability, efficiency and others. The distributed nature of RESs suggests an evolutionary change of the central energy generation, transmission, and distribution. Distributed energy sources and the main grid are not competitors but complements. In this paper, we describe how a network of SoS can be designed to integrate spatially distributed stochastic RESs with the main grid.

3.4.1. A Microgrid as an SoS. Microgrids (MGs) have become an effective solution for utilizing distributed RESs. A MG is a localized group of power sources and loads, which can operate both in a stand-alone mode or a grid-connected mode of operations. Figure 3.1 illustrates a MG with RESs. This MG in a stand-alone mode includes heterogeneous energy generators such as photovoltaic (PV) systems, wind turbines (WT), and diesel engine (DE) generators; energy storage devices such as battery systems; and loads. When different generation sources and storage devices collaborate with each other, becoming a bigger system, the energy supply of the system is more reliable, safe, and cost-effective than that with a single energy source. This concept may seem counterintuitive due to the intermittency of RESs, but renewable energy actually can be effectively utilized when multiple generators are connected and supported by energy storage systems [150], [151]. As applied to renewable energy, the Law of Large Numbers dictates that the combined output of every PV, WT, and DE connected to the grid is far less volatile than the output of a single RES [150], [151]. The way they operate and collaborate can be controlled to evolve over time to better meet stochastic local loads. When connected with other MGs and the main grid, this MG can support or be supported by them.

A MG can be seen as a SoS [152], [153], [154]. A MG is an integration of heterogeneous and independently operated and managed systems (power generators, storage devices, and the main grid if under the grid-connected mode), for generating and supplying energy to meet load demand. The MG is flexible because many aspects of the

MG can be dynamically adjusted, including the connection and disconnection of existing or newly developed constituent systems, the way in which the selected constituent systems collaborate, and the operations of these systems. The integration of distributed generation and storage devices as a SoS provides greater ability to meet load demand than a simple system does.

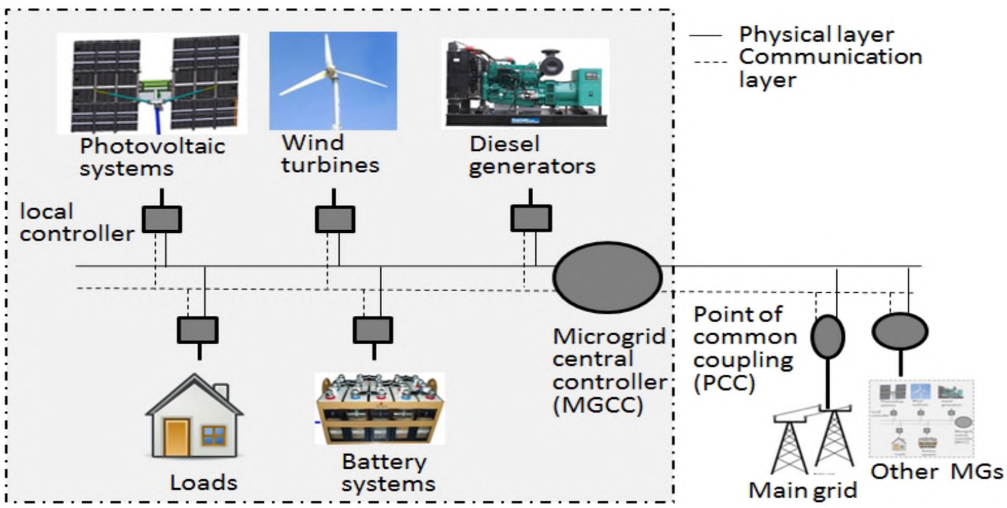


Figure 3.1 A microgrid with renewable energy sources.

Connectivity is an important characteristic of SoS. Figure 3.1 shows that constituent systems of the MG are connected as a two-layer network. At the physical level, it is a power network wherein the power flows from supplies (generation and/or storage devices) to demands (local loads, charging storage devices, and loads from the main grid and other MGs). At the communication level, it is an information network responsible for system monitoring, information collection, data exchange, and transferring control signals. Each generation or storage device is connected to the network by a device controller and, similarly, a load is connected to the network by a load controller. All these controllers are named local controllers (LCs) shown in Figure 3.1.

Coordinating constituent systems of a SoS is critical to the delivery of expected outcomes. The microgrid central controller (MGCC) of an MG is designed for this purpose, particularly under the stand-alone mode. It controls the MG in terms of assessing the operating status of MG, forecasting and planning power generations, dispatching power to loads, and managing load demand. MGCC communicates with the local controllers of the MG. When the MG is connecting to the main grid or other MGs, its MGCC also communicates with these external systems through a distribution management system (DMS). There are multiple control methods for coordinating elements of MGs [155]. The control method for an MG determines the SoS type of it.

Central hierarchical control - directed SoS: When a central hierarchical control method is used, the MG can be a directed SoS in that it is built and centrally managed during long-term operations to fulfill specific purposes. The constituent systems maintain an ability to operate independently, but their normal operational mode is subordinated to the central managed purpose. In a central hierarchical control, LCs follow and execute the orders of MGCC, but they may still have certain degree of autonomy or intelligence.

Decentralized hierarchical control - acknowledged SoS: When a decentralized hierarchical control method is used, the MG is more likely to be an acknowledged SoS. An acknowledged SoS has its objectives, independent management, and resources for the SoS; however, the component systems are also independently operated and managed in that they retain their independent objectives, sources, and development and sustainment approaches. Changes in the constituent systems are based on collaborations between the SoS and the systems. In a decentralized hierarchical control, LCs demonstrate a higher degree of autonomy and they optimize the control of the local devices. The MGCC attempts to influence the local optimization. The optimality of decisions by MCGG and LCs in this control method is sensitive to the system reliability and the communication speed, particularly in geographically distributed large-scale SoSs. Implementing the decentralized hierarchical control currently can still be technically challenging.

Decentralized control – collaborative or virtual SoS: When a decentralized control method is chosen, the MG is a collaborative or virtual SoS whose constituent systems interact more or less voluntarily to fulfill agreed upon central purposes. Compared to the decentralized hierarchical control method above, this distributed control lacks a dedicated central controller like the MGCC to coordinate local devices. LCs are responsible for optimizing the operations of distributed devices. LCs have no or limited communicate with each other and operate mainly based on local measurements.

A MG may operate in various conditions so that a single coordination method may not always be the best. A change in the operating condition or the central mission may require a switch of the MG operating mode. Consequently, the coordination method, and therefore the SoS type, of the MG may be changed too. For example, when a MG switches from the grid-connected operating mode to the stand-alone mode, the control method of the MG may be changed temporarily from a central control to a decentralized control. This triggers a change in the SoS type. To obtain the flexibility of using multiple coordination methods, the communication network for exchanging information and transmitting control signals needs to be well designed to adapt to any coordination methods.

3.4.2. Connected MGs and the Main Grid as an SoS Network. While an MG can be independently managed and operated to coordinate the RESs and loads locally, it can be connected to other MGs and/or the main grid to exchange energy. The full value of RESs requires the grid connection [106]. Connecting multiple MGs and the main grid as a SoS network adds additional flexibility, which helps achieve greater performance than the additive outcomes of unconnected individual energy systems. That is, an additional utility is added by forming a SoS network.

This section uses an example in [157] to illustrate the SoS network, which is composed of two MGs (MG_A and MG_B) and the main grid (MnG). Figure 3.2 illustrates the eight configurations of the SoS network. The network can be seen as a graph with three nodes. The eight configurations are different from one to another in terms of the number edges connecting nodes. The flexibility at the SoS network level lies in the possibility of intendedly switching from one configuration to another, for adapting to different operating conditions. There are 24 possible intended switches because only one edge can be added or removed at one time. There are unintended switches too, mainly occurring during unplanned outages of the main grid. For example, configuration II may be switched to configuration VII due to an outage of the main grid.

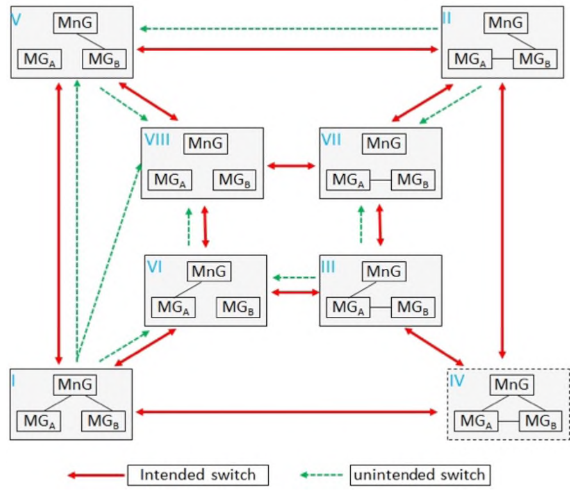


Figure 3.2 Configurations of a SoS network.

Generally speaking, a SoS network with N different MGs and the main grid has $2^{(N+1)N/2}$ configurations in total. Each configuration can be intendedly switched to one of another $(N+1)N/2$ configurations if only one edge can be changed at one time. Therefore, in total there are $2^{(N+1)N/2}$ $(N+1)N/2$ intended switches between configurations.

Similarly, there are different methods for coordinating constituent SoSs in the network, and so the type of the SoS network may be revised if the control method is changed. A MG network in configuration II and coordinated by a central hierarchical control method may be switched to configuration VII unintendedly due to the outage of the main grid. The control method may be switched to the distributed control temporarily, and so the SoS type of the network becomes a collaborative SoS. The configuration switch and the control method switch need to be coordinated carefully to avoid negative impacts [157].

3.4.3. Evolution of MGs and MG Network. A SoS or SoS network usually evolve over time, driven by the dynamics discussed in Section 3.3.2 and determined by the decisions described in Section 3.3.4. In the following we describe these decisions in this case study, which usually fall into the category of energy management.

Generation planning of individual MGs: Given the forecasts of local loads and RESs of an MG *N* periods of time into the future, as well as energy exchange requests from other MGs and the main grid, the MG agent plans the power generation of each source, the charge or discharge amount of each storage device, and the energy exchange commitments to other MGs and the main grid.

Power dispatch of individual MGs: The actual loads and RESs are measured every period of time. Given the measurements (e.g., wind speed, solar irradiance), as well as the energy exchange commitments to other MGs and the main grid, the SoS agent adjusts the power outputs of individual generation sources and the charge/discharge amount of storage devices to optimize the power flows and stabilize system voltages.

Planning and coordination of power exchanges: The power exchange, either between different MGs or between a MG and the main grid, aims to generate additional utility at the SoS network level by mitigating the unbalanced generation and load demand on a wide area.

If communication capability is provided, the three types of decisions can be made in an integrated manner in that they are interdependent. The first two decisions, which are coordinated by the MGCC and LCs of each MG, are more closely coupled. They can be modeled as a two-stage stochastic programming problem, a robust optimization problem, or a rolling optimization problem. Generally speaking, the third decision is a game-theoretic problem and coordinated by the DMS. The way in which the third decision is integrated with the first two depends on the type of the SoS network.

3.5. SUMMARY

In this section, we proposed a methodology for acquiring and maintaining flexibility for distributed large-scale or complex systems in changing or uncertain conditions through forming and evolving SoSs. Findings from the preliminary study of this topic positively support the proposed methodology.

Challenges are present in the implementation of the proposed methodology. From the management perspective, SoSs have multiple flexibility mechanisms. Decisions for forming SoSs and evolving them over time are complex; in that the decisions are interdependent and across multiple time scales. From the technology perspective, the executions of SoS reconfiguration, system performance re-calibration, and SoS type change, require advanced control technologies. Addressing these challenges needs a seamless collaboration between systems engineers and domain experts.

Therefore, in the next section we address the implementation problem of coordinating constituent systems within a SoS to preserve the creation of flexibility made possible by the SoS approach. The choice of strategy, barriers to applying a coordination strategy, and methods to address removing those barriers are discussed, evaluated, and applied to a SoS involving another distributed energy system on an island.

4. FUNCTIONALIZING SYSTEM OF SYSTEMS USING COORDINATION OF CONSTITUENT SYSTEMS

4.1. CHARACTERISTICS OF A SYSTEM OF SYSTEMS

A system of systems (SoS) is an arrangement of independently managed and operated systems. A SoS has five characteristics that differentiate it from a system, which are autonomy, belonging, connectivity, diversity, and emergence [158] as shown in Section 3. Therefore, SoSs provide unique capabilities for meeting special needs for intelligence [159], [160], flexibility [161], and synergy [162], which may not be accomplished by a single system.

The five characteristics have been widely adopted as the foundation for forming, analyzing, and modeling SoS. Dimario, et al. [163] stated that the characteristics of SoS play important roles in the SoS mechanism design and social function, concluding that the SoS formation is a result of balancing multiple objectives in a satisficing environment. Similarly, Gorod, et al. [164] proposed to use both dynamic and static doctrines to manage SoS in ever changing dynamic environments, wherein the five characteristics of SoS represent the dynamic doctrine. Baldwin and Sauser [165] and Baldwin et al. [166] modeled autonomy and connectivity using set theory, and formulated belonging using game theory. Agent-based model simulation was then used to implement the model and gain better understanding of SoS formation. Preservation of the five characteristics of a SoS as it is evolving in a dynamic environment is clearly critical to functionalizing the SoS; however, not much work has been dedicated to this need.

Coordinating constituent systems of a SoS is an important task for operating SoS and so preserving the characteristics of SoS. Coordination of constituent systems must fully consider their levels of autonomy, belonging, and connectivity.

Autonomy, to a certain degree, is currently being designed into many systems for various reasons. For example, geographically distributed systems are designed to be able to accomplish their goals without depending on the command from a central controller. Coordinating autonomous systems is not about taking charge of their functions. Instead, the coordination should respect the autonomous nature of constituent systems and still allow them to be operated and managed independently. Coordination of constituent systems takes into account the autonomous level of each individual constituent system and accordingly chooses an appropriate method to influence the management and operations of constituent systems. If the autonomous level of a constituent system can be varied, the coordination may also involve persuading constituent systems to adjust their autonomous level when needed.

Belonging can be interpreted as choices. Systems can choose to belong or not belong. From this perspective, the coordination of constituent systems should consider their belonging level, and if possibly, identify and provide constituent systems favorable ways of SoS participation, which are likely to be accepted by constituent systems due to the benefit of choosing to belong. Likewise, the SoS chooses to allow systems to belong. From the perspective of SoS, the coordination of constituent systems involves choosing participating systems and designing their participating specifications including both requirements and compensations.

Connections of constituent systems in a SoS are dynamic. An important aspect of coordinating constituent systems is to determine connections and disconnections of constituent systems, as well as execute the decisions. The decision and execution must take into account the connectivity of constituent systems, the anticipated benefits from connections, possible negative effects, and technical specifications of connection and disconnection.

The three characteristics discussed above are usually intertwined in coordination of constituent systems. For example, an opportunity for a system to choose to belong may require it to adjust its level of autonomy and connect with other constituent systems.

Coordination of constituent systems plays a key role in producing the diversity and emergence characteristics of SoS.

A SoS is designed and formed to provide multiple capabilities and be capable of responding to largely uncertain conditions. Yet, the realization of diversity heavily relies on the coordination of constituent systems from perspectives of reconfiguration, dynamic assignment, and others.

Emergence is a result of coordinating constituent systems to collaborate or cooperate. That is, the SoS is able to provide a unique function, behave in a special manner, and generate a level of utility that cannot be accomplished by a single system or a group of systems without coordination. The desired constituent system behaviors are the result of voluntary and collaborative interactions without central direction [167].

We can conclude that coordination of constituent systems is critical to preserving constituent systems' autonomy, belonging, and connectivity, and consequently generating diversity and emergence as desired SoS behaviors. However, not much formal work has been done to specify how to functionalize SoS through coordinating constituent systems. We are motivated to explore this topic and develop knowledge on it. In the remainder of the section, we first analyze coordinating strategies in operating SoS in Section 4.2. Then, in Section 4.3, representative challenges in implementing the coordinating strategies are summarized, followed by an overview of proposed approaches to addressing the challenges. Section 4.5 presents a case study that involves coordinating heteronomous generators, multiple storage systems, and distributed loads on an island. The section is concluded with a brief summary of identified future work for promoting effective coordination of constituent systems.

4.2. COORDINATING STRATEGIES FOR SOS

There are different strategies for coordinating constituent systems of a SoS. A pure centralized strategy for SoS means that a central controller (CC) that coordinates constituent systems at the SoS level exists and all constituent systems subordinate to the SoS. A pure decentralized strategy for SoS means neither a CC nor a pre-specified rule or agreement for constituent systems exists. A SoS can also choose a coordinating strategy that is a mix of centralized and decentralized coordination. For a system engineer, choosing an appropriate strategy is important. Our classification of coordinating strategies in Figure 4.1 takes multiple aspects into account, including the autonomy, belonging and connectivity levels of constituent systems; resulting diversity and emergence of SoS; and other issues such as system reliability. The classification, although sharing similarity with that composed of directed, acknowledged, collaborative, and virtual SoSs [167], [168], is more from the perspective of designing or choosing a strategy with consideration of the SoS characteristics.

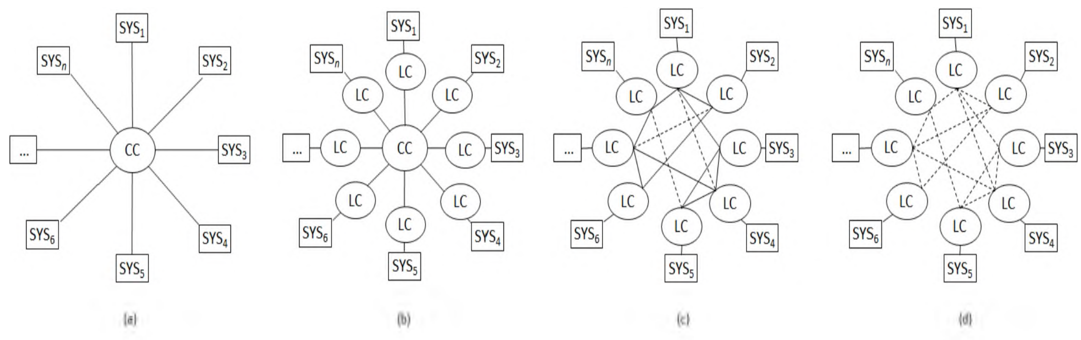


Figure 4.1 Coordination strategies for SoS.

4.2.1. Centralized Coordination. In Figure 4.1(a), centralized coordination can be used when goals of constituent systems are highly consistent with those of the SoS. Constituent systems are willing and able to follow the command of a CC in the general operating mode, but they can also be independently operated and managed. Although constituent systems have their own controllers for management and operations (named local controllers), their control mechanism is not the necessary information for the CC due to the high cooperative level and response capability. The presence of a CC and the high cooperative level of constituent systems make the centralized control an effective strategy to enhance the characteristics of diversity and emergence. Yet, the global optimal outcome for the SoS highly relies on the speed and reliability of communication between constituent systems and the SoS.

4.2.2. Hierarchical Coordination. Hierarchical coordination shown in Figure 4.1(b) has a CC just like the centralized strategy. Yet, constituent systems have local controllers (LCs) that are primarily designed at their own interests and for fulfilling their own purposes. The CC should not assume that constituent systems are always willing and able to follow its command; instead, the CC needs to influence the belonging and autonomy levels of constituent systems, and thus, leverage the output of LCs. Therefore, besides the need for reliable and fast communication between SoS and individual constituent systems, the use of hierarchical strategy also requires some knowledge about the control mechanism of each individual LC, for example, through learning. LCs use not only local measurements, but other information sent from the CC. The hierarchical coordination strategy has a greater chance than centralized control to see deviations between the realized outcome of LCs and the anticipated outcome for various reasons; for example, incomplete information about LCs, lead time of knowledge acquisition due to learning, unanticipated changes of constituent systems, and others.

4.2.3. Peer Coordination. Peer coordination is partially decentralized as seen in Figure 4.1(c). A constituent system communicates with only a subset of constituent systems due to various reasons such as geographical obstacle, communication band limit, and others. Therefore, the communication network is not a full connected network. In Figure 4.1(c), links between the LCs of any two systems represents connectivity in the communication network. A solid link indicates the two systems are connected, whereas a dashed link indicates they are not connected at that moment. Unlike centralized and hierarchical coordination, the peer coordinating strategy does not have a CC; instead, the fulfillment of the purpose of SoS relies on voluntary collaboration among constituent systems. Therefore, the use of peer coordinating strategy indicates firstly that the information on variables with common interest is shared among constituent systems;

secondly, a mechanism must exist, or it can be created, which motivates constituent systems to adjust their autonomy and belonging levels. Under such circumstances they will converge to a consensus or equilibrium on the variables of common interests.

4.2.4. Decentralized Coordination. A CC does not exist in the decentralized coordinating strategy as diagrammed in Figure 4.1(d). LCs use only local measurements as the input. Therefore, communication between constituent systems is not needed. The decentralized coordinating strategy is highly tolerant to changes and failures of constituent systems. However, because constituent systems are high in the autonomous level and low in connectivity and belonging levels, it is more difficult to attain and maintain desired emergent capabilities of SoS, as well as to correct unintended SoS behaviors.

4.3. CHALLENGES TO IMPLEMENTATION

Selecting the most appropriate coordinating strategy is about balancing the limitations and advantages of candidate strategies to determine the best fit. Yet this still does not guarantee intended SoS behaviors can always be obtained because every strategy has certain implementation difficulties.

4.3.1. Imperfect Capability of Central Controller and Local Controllers. Centralized and hierarchical coordinating strategies use a CC that is designed to be able to monitor and communicate with all constituent systems. They aim at achieving an optimal solution at the SoS level. Therefore, the CC makes the SoS more capable of engendering the diversity and emergence characteristics. In the perfect situation where all constituent systems operate as they have been planned or advised by the CC, or as they have committed, the role of CC is maximized. However, constituent systems may not be able to deliver the performance as planned or promised for various reasons, such as system failures, disturbance of external environments, performance uncertainty, and other sudden and intended changes that constituent systems have. To make the CC resilient to these factors is a challenge of implementing the centralized and hierarchical coordinating strategies.

When the CC is missing, like in the peer and decentralized coordinating strategies, the diversity and emergence characteristics of SoS become sensitive to the speed at which LCs can converge or reach an agreement on the SoS participation. The speed is largely influenced by constituent systems' autonomy, belonging, and connectivity levels. Without a CC or connections with all other constituent systems, LCs, even highly cooperative ones, may converge to an agreement that is just a local optimum at the SoS level. At the SoS level, emergent behaviors are versatile and difficult to predict. Addressing limitations of the coordination strategies that have no CC is a need.

4.3.2. Operating Mode Changes. A SoS may operate under different modes. The selection of a coordinating strategy for a SoS is usually made according to its regular operating mode. Yet the regular operating mode may temporarily switch to another mode, either intentionally or unintentionally. If the coordinating strategy designed for the regular operating mode is not applicable to other operating modes, a portfolio of coordinating strategies is needed to handle different operating modes. The seamless switch between coordinating strategies during an operating model change is critical to the reliable operation of SoS.

It should be noted that the challenges discussed above become more notorious in large-scale SoS due to induced high computational requirements, unpredictability, and complexity.

4.4. METHODS TO APPROACH THE CHALLENGES

A relevant method of meeting the challenge of having imperfectly capable central and local controllers is to apply multi-stage, multi-scale coordination. Also, a way to address the issue of operating mode changes is to select the most appropriate coordinating strategy to balance the possibility of operating mode switching being intentional or unintentional. The relative advantages of candidate strategies will determine the best fit. These two approaches are not guarantees of intended SoS behaviors, but can be used to mitigate certain implementation difficulties.

4.4.1. Multi-stage Multi-Scale Coordination. The coordination of constituent systems when a CC is present is an analogy to stochastic control considering the coordination outcome is partially random, i.e. the realized outcome may not reach, or it may deviate away from the expected outcome and the gap between them is unpredictable. The coordination of constituent systems when a CC is missing is analogous to simultaneous games. Therefore, coordinating constituent systems over multiple stages allows for dynamically adjusting the coordination for the next stage according to the observed outcome from the current stage. But different than regular stochastic control or repeated simultaneous games, coordination of constituent systems at different stages may be at a different scale and, thus, can be based on different models.

Take the two-stage coordination as an example. The first stage coordination can be focused on coordinating constituent systems in moving towards the target outcome. Once the realized outcome is observed, the second stage coordination will be focused on coordinating the constituent systems to minimize the impact of the gap between the target outcome and the realized outcome from the first stage coordination. The first stage coordination should consider two aspects: (i) possible outcomes of a chosen course of action and the chances those outcomes will occur; and (ii) the effectiveness of the second stage coordination in moving from the first stage coordination outcome towards the target outcome. Therefore, taking the ability and outcome of second stage coordination as an input, the first stage coordination chooses an action that can achieve the best expected overall coordination outcome over the two stages. The second stage coordination is for addressing the gap between the target outcome and the outcome from the first stage coordination. On one hand, this gap is in a much smaller scale than that before the first stage coordination is taken; and on the other hand, this small gap needs to be addressed in a much shorter time period than that for the first stage coordination.

4.4.2. Coordination of Operating Mode Switches. Operating mode switches can be intended or unintended. Intended switches can be relatively easily coordinated in that it is more like a deterministic scheduling problem that pre-specifies actions are planned over a timeline to meet all requirements on a mode switch. Unintended operating mode switches are more difficult to coordinate than intended ones. The ability of coordinating unintended operating mode switches is critical to the control of any negative effects that may be caused. This ability is built up by a smart system that integrates the following capabilities: 1) Sensors are deployed in the SoS and may be networked to: (i) monitor the

SoS to increase the preparedness for, and responsiveness to, operating mode switches; (ii) detect mode switches in real-time manner; and (iii) check the effectiveness of coordination to provide feedback; 2) Models and algorithms are developed to: (i) determine the operation specifications for constituent systems under each operating mode; (ii) capture the coordination mechanism and capability of controllers; (iii) derive the optimal coordination actions to take; and 3) Execution tools that execute the coordination commands of controllers.

4.5. CASE STUDY: SOS NETWORK OF AN ISLAND ENERGY SYSTEM

This section uses an energy system installed on an island and originally presented in [169] and [170] as a case to further discuss the implementation of coordination strategies. The energy system is composed of renewable energy source (RES) based microgrids (MGs) and diesel engine (DE) generators. Figure 4.2 shows its configuration. Microgrid A (named MG_A) is composed of wind turbines (WT) (500kW) and energy storage systems (one ultra-capacitor (UC) in 500kWx15s + lithium iron phosphate batteries in 500kWx2h), which is mainly focused on meeting the load from feeder 1. Microgrid B (named MG_B) has the same WT and energy storage systems as in MG_A. Besides these, MG_B also contains photovoltaics panels (PV) (660kWp+175kWp), and electric vehicle (EV) charging stations. The load from feeder 2 is the local load for MG_B. Five DE generators in a total capacity of 1,700 kW is named MnG in this case.

MG_A and MG_B are two SoSs, and MnG is a system. They form a SoS network as Figure 4.2 shows. Any two of them can be connected by a fast-tracking switch (FTS) to provide reliable and quality power supply to the entire island. Specifically, each individual MG is primarily focused on using RES to meet the local load but is willing to exchange energy with the other MG to achieve a more globally optimal power supply to unbalanced distributed loads. MnG can be connected to either or both MGs to mitigate the limitation of intermittent and volatile RESs. Communication of the SoS network is based on the IEC61850 standard that uses an MMS protocol for between SoSs communication and a GOOSE protocol for within SoS communication. The standard adopts a bi-layer communication network that separates the transmittal of control signals and monitored information of SoS states.

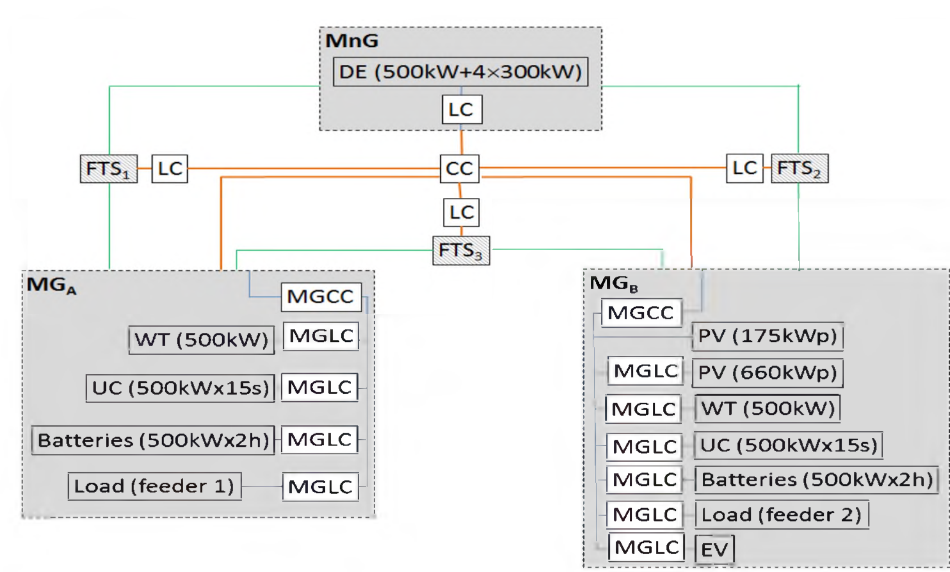


Figure 4.2 Configuration of the SoS network with dual microgrids and diesel engine generators.

Figure 4.2 shows that each MG has a CC named MGCC; controllable distributed devices in each MGs have their local controller named MGLC. The MnG has a LC, and so each of the three FTS. MG_A, MG_B, MnG, and the three FTSs for connecting them are coordinated by a CC, as Figure 4.2 shows. In the following, we discuss various aspects of coordination in this SoS network.

4.5.1. Hierarchical SoS Coordinating Strategy. Each MG implements a hierarchical coordinating strategy to ensure that the requirements on economic operations and system stability are met. Specifically, given the forecasted loads, RESs, and energy storage capability, the coordination schedules renewable generation, demand adjustment, and energy storage to achieve the economic objective of operations. Given the power shortage forecast and participation factors of distributed devices, it allocates the system power shortage to each controllable unit to ensure the power balance of the MG.

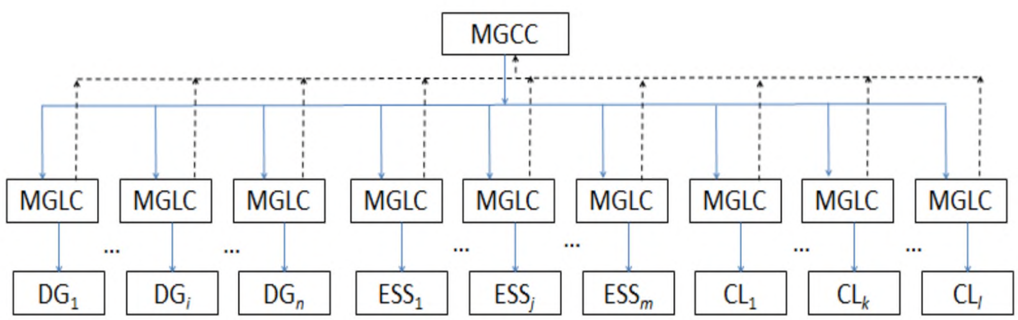


Figure 4.3 The schematic diagram of the hierarchical coordinating strategy for SoS.

Figure 4.3 shows the relationship between the MGCC and MGLCs of a MG. The MGCC collects and processes global and local information, including frequency, voltage, capacity, and load demand; determines power reference values for MGLCs of individual distributed devices and send control commands to the MGLCs. MGLCs control their local devices including distributed generators (DGs), the energy storage systems (ESSs), and controllable load (CLs), according to the control commands received from MGCC.

Meanwhile, MGLCs provide MGCC with the operating status of local devices as the coordination feedback.

4.5.2. Coordinating Heterogeneous Constituent Systems. The energy management of this island MG aims to strategically realize economic, environment-friendly, and reliable operations. These objectives are met through optimally coordinating distributed generators and ESSs, and load demands. Therefore, from the perspective of SoS, energy management involves coordinating constituent systems in the SoS.

This island has a fishing industry and a famous tourist attraction. The total load on this island is classified into three categories based on the controllability of the load:

Important load: requirements on the quantity and time period of power supply must be met. Examples include the electricity for fishery production.

Shiftable load: the time period of supply can be adjusted, but not the quantity. For example, the electricity consumed by electric vehicle charging stations on the island.

Adjustable load: both the quantity and timer period of supply can be adjusted. The electricity consumed by air-conditioners falls into this category.

The partial controllability of load demand expands the regulation ranges of distributed power supplies and the ESSs, easing the stability control of the MG. Moreover, the ESSs help mitigate the limitation of renewable generation, increasing the penetration rate of RESs and further improving the system stability. The MGs here use two types of ESSs: ultra-capacitors and lithium iron phosphate batteries. The former has fast response speed and high-power output, but low in energy density. The latter has longer service lives and lower energy losses. They can help each other to achieve better energy storage performance.

Given the ESSs and the partial ability to control load, the energy management of this island implements a two-stage coordination of load, sources and storage, which are in two different time scales.

First stage coordination: Based on the forecasts of load and renewable generations N time periods into the future, the first stage coordination schedules the load, the power output for distributed generators, and charging/discharging amount of the ESSs, on a time horizon with N periods.

Second stage coordination: During each individual period, after the realized wind speed, solar irradiance, and load are measured, the second stage coordination adjusts the power outputs of distributed generators and the ESSs to optimize the system voltages and power flows.

If the second stage coordination is able to make adjustments during any of the N periods, a new plan of first-stage coordination will not be created until the current plan is fully executed. Yet another scenario may happen before the current plan can be done. As time is moving forward, the actual operations may largely deviate away from the original plan due to the accumulation of prediction errors. Consequently, in one of the N periods, the second stage coordination may fail in making the adjustments. If that happened, a new plan for the next N periods will be created by the first stage coordination.

4.5.3. Operating Mode Switches. To ensure the reliability of power supply, a large island energy system with RES usually is equipped with two kinds of main power sources. For this purpose, the island maintains a system of DE generators, MnG, besides using renewable energy. MnG can be connected to either of the two MGs or both. Consequently, a MG can operate on either of the following two modes: the energy-saving operating mode with DEs serving as the main power source, and the green operating mode with ESSs serving as the main power source.

In the energy-saving mode, DEs in MnG provide the reference for MG power regulation and all distributed devices of MGs are under the active reactive power control (P/Q) control. The green operating mode may be a more complex situation, depending on what control method is used for the main power source. The MGs here use master-slave control to maintain the safe and stable operation. The use of master-slave control makes the operating mode switches more challenging; that is, an operating mode switch requires a switch between two voltage control methods – P/Q control and voltage-frequency (V/f) control. Since the control method switch and the operating mode switch are not completely synchronized, a current spike may occur. Therefore, the time sequence of operating mode switch needs to be accurately determined to reduce the chance of transient voltage fluctuations or a power failure.

> <ins>Intended ES2G switch:</ins> An intended switch from energy-saving operating mode to green operating mode occurs when RESs are abundant and sufficient energy is stored in ESSs. Once the CC of the SoS network determines that a MG should operate in the green operating mode, the MGCC of the MG will first verify if the current operating mode is the energy-saving mode. Then, the master ESS is chosen and its MGLC determines the voltage and frequency of the system based on V/f control. MGLCs of slave ESSs and RESs follow the power regulation for them. Once power balancing (between FTS and MG) is completed, the MGCC

will ask the LC of FTS to disconnect the MG with the MnG, switching the system to the green operating mode.

Intended G2ES switch: An intended switch from the green operating model to the energy-saving operating mode occurs when RESs are insufficient. Once the CC of the SoS determines that a MG should operate in the energy-saving operating mode, the MGCC of the MG will first verify if the current operating mode is the green mode. Based on the present active and reactive power of the master ESS, the MGCC sets the initial power value in P/Q control and send the synchronization command to the master MGLC and the LC of FTS that will connect the MG with the MnG. The LC of FTS put the switch on and after a pre-specified time delay the master MGLC switches from V/f control to P/Q control.

Unintended ES2G switch: An unintended switch to green operating mode occurs when an outage of MnG happens. By disconnecting the FTS and switching the control mode of the master ESS from P/Q to V/f, the supply to important load can still be ensured. The response time of the LC for FTS is different than that for the master ESS; therefore, an accurate control of the time sequence of these two actions would help. During the unintended ES2G switch, the MGCCs may not meet the time requirement on switch. Therefore, load shedding and fast control of distributed local devices are necessary.

If the peer-to-peer control is used for the main power source in MGs, the voltage control method remains unchanged when switching between the two operating modes. The possible current spike caused by the switch can be mitigated by a pre-synchronous control. The use of peer-to-peer control, particularly in large MG network, effectively reduces the complexity of mode switching strategy and increasing the success rate of switch. However, the peer-to-peer control process is ‘unsupervised’ compared to the master-slave control.

4.6. SUMMARY AND FUTURE RESEARCH

Motivated by the importance of coordinating the constituent systems of a SoS in operations, this section analyzed the mechanism of choosing coordinating strategies for the SoS, discussed challenges in the implementation, and proposed methods for addressing these. The section further examined the proposed work in a case study that involves coordinating heterogeneous power generators, multiple parallel energy storage systems, and distributed loads on an island. The work of this section has confirmed the importance of considering SoS characteristics in choosing coordinating strategies. It has also showed that SoS characteristics are affected by the effectiveness of coordinating constituent systems.

From the study in this section we found that the computational complexity in implementing multi-stage multi-scale coordination grows quickly as the size of a SoS, or SoS network, increases; how an outcome of coordination is reached and how it is evolving over time in use of fully or somewhat decentralized coordination strategies for SoS are less understood; the use of smart technologies to improve the coordination effectiveness requires more thorough analysis and exploration. Addressing these technical needs, as well as others to be identified, will largely improve the effectiveness of coordinating constituent systems and so the capabilities of SoS.

Now that a framework has been established to form and evolve SoS in uncertain environments and a strategy has been developed for choosing the proper SoS coordination method, we can apply these combined techniques to an area that has not seen the application of SoS to create and sustain flexibility. The application of SoS to intelligent systems with a human-AI technology interface is evaluated in Section 5.

5. A NEW DEFINITION OF WORKFORCE FLEXIBILITY USING AN SOS APPROACH WITHIN A HUMAN-INTELLIGENT SYSTEM-ARTIFICIAL INTELLIGENCE SYSTEM

5.1. A NEW CONCEPTUALIZATION OF FLEXIBILITY IN WORK

The National Science Foundation (NSF) has recognized that the landscape of jobs and work is changing at unprecedented speed, enabled by advances in computer and engineering technologies such as artificial intelligence and robotics, deeper understanding of societal and environmental change, advances in the learning sciences, pervasive, intelligent, and autonomous systems, and new conceptions of work and workplaces [171]. This technological and scientific revolution presents an opportunity in the creation of new industries and occupations, enhanced productivity and quality of work life, and the potential for more people to participate in the workforce, ultimately yielding sustained innovation and global leadership. This new environment creates some uncertainties in the areas of work and how we define concepts like workforce flexibility. Even with the advent of technologies such as the use of artificial intelligence (AI), intelligent systems, machine automation and learning, or additive manufacturing, they still require a Human-Technology (H-T) or Human-AI hybrid interface. Therefore, flexibility among such systems is required to meet the new challenges and operational characteristics of these environments.

Again, according to the NSF, the future of work at the Human-Technology (H-T) frontier is a conceptualization of work, and by extension workforce flexibility, in the future that will be enabled or improved by advances in intelligent technology and their synergistic integration with human skill to achieve broad participation in the workforce and improve the social, economic, and environmental well-being of society [171]. Technology should be integrated with learning sciences, research on education and workforce training, and social, behavioral, and economic science perspectives to advance the science of the human-technology team. Potential results should contribute to fundamental advances in the science and technology of future workforce development and education, work environments, and positive workforce flexibility outcomes for workers. This research broadly speaking is oriented toward the future of work at the human-technology frontier and is not overly couched in current technology or work practices. Figure 5.1 indicates the stages of the industrial revolution as systems become more complex over time and morphs into sociotechnical or cyber physical systems. The newest evolution of this is called Industry 4.0.

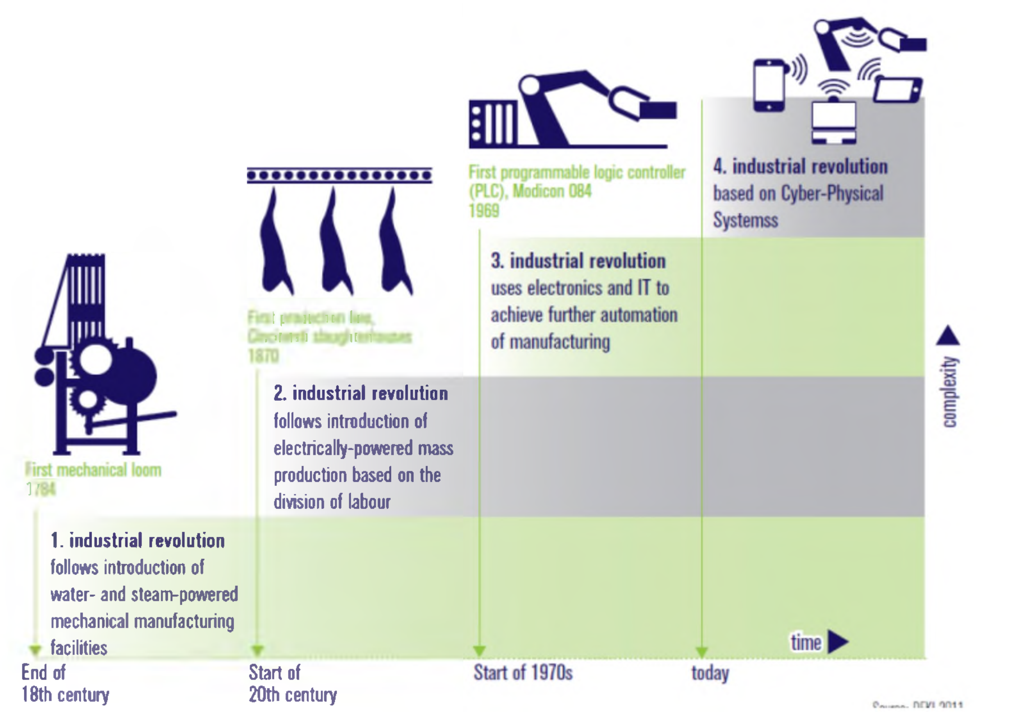


Figure 5.1 Timeline of industrial revolution cycles from Industry 1.0 to 4.0.

The key research question to ask is: how do we define workforce flexibility within a human-technology interfacial environment? This section proposes: a) a new definition of workforce flexibility within the human-technology interface of a system; b) using a system of systems approach to create and sustain that flexibility; and c) applying a coordinating strategy for optimal workforce flexibility within the human-technology interface of such a system.

Workforce flexibility is the management of organizational labor capacities and capabilities in operational environments using a broad and diffuse set of tools and approaches to mitigate system imbalances caused by uncertainties (e.g., worker absenteeism) and/or changes (e.g., seasonality). In Section 2, we reviewed the literature associated with workforce flexibility in order to provide a reference for choosing flexibility methods for specific applications, and to better observe the connections and gaps in the literature with respect to future research needs. In that review, we spanned application contexts, research questions, and solution generation methodologies, for each flexibility method. Five flexibility methods that have been researched in ORMS and implemented into practice are: flexible working time, floaters, cross-training, teamwork, and temporary labor. These methods were classified using terms that were assumed to be recognizable by practitioners and academics as a common basis of discussion.

This research provided a detailed review and discussion of the literature for each workforce flexibility method. It also formed the foundation in flexibility knowledge necessary to advance research into ways to create and sustain flexibility in complex systems. As discussed in Sections 2.6 and 2.7, workforce flexibility methods may be used in complex systems where less predictable uncertainties and rapid changes are more common. But current methods may not be effective in addressing these new challenges. Methods that help overcome limitations of current flexibility methods and application approaches should be researched to improve scalability and robustness of flexibility in complex systems.

In the new context of Industry 4.0 and the use of technologies such as machine learning and Big Data analysis, the labor capacities and capabilities are complex systems that can be full automation or hybrid systems that combine human capabilities (e.g. planning, review, and innovative thought) with technology that makes labor more efficient, e.g. robotics, cloud computing, remote sensing, etc. Some of these technologies mitigate the uncertainties in the work environment such as labor or overtime costs and efficient working time for the project. Traditional flexibility methods like having floaters, cross-training, and temporary labor are nearly zeroed out with complex systems and automation, but the human component is still vital to creating the flexibility needed with the interaction of humans and technology. This interaction is fundamentally changed with how people interface with technology.

5.2. HUMAN-TECHNOLOGY INTERFACE AS A SYSTEM

G.E. Wang of Stanford University's Institute for Human-Centered Artificial Intelligence (HAI) had an excellent discussion having humans-in-the-loop of designing AI. The key reason is that full automation may not yield the optimal solution to solving problems, but also that automation does not equal complete removal of human involvement in a task [172]. As stated in the article, the human-in-the-loop approach changes the framing of the problem from an automation only to a human-technology design problem. Design problems like these are amenable to "interactive machine learning in which intelligent system are designed to augment or enhance the human, serving as a tool to be wielded through human interaction" [172]. Fails and Olsen (2003) were the first to introduce the term interactive machine learning in the human-computer interaction community, characterizing it with rapid train-feedback-correct cycles, where users iteratively provide corrective feedback to a learner after viewing its output.[173] They demonstrated this process with their Crayons system, which allowed users with no machine-learning background to train pixel classifiers by iteratively marking pixels as foreground or background through brushstrokes on an image. After each user interaction, the system responded with an updated image segmentation for further review and corrective input. Fails and Olsen's work on Crayons demonstrated that users modify their behavior based on a learner's outputs, which is an underlying premise for much of the following research on interactive machine learning. Figure 5.2 [174] shows that in machine learning, people iteratively supply information to a learning system and then observe and interpret the outputs of the system to inform subsequent iterations. In interactive machine learning, these iterations are more focused, frequent, and incremental than traditional machine learning. The tighter interaction between users and learning systems in interactive machine learning necessitates an increased focus on studying the user's involvement in the process [174].

The benefits highlighted in Wang's article of: 1) transparency to understand the system, 2) incorporation of human judgment in the decision loop, 3) reducing the need to build "perfect" algorithms, and 4) enabling of more powerful systems to achieve "functional excellence" align with the objectives of a system that is capable of achieving its goals and having the ability to improve over time as the system receives feedback from its environment as well as from a human perspective.

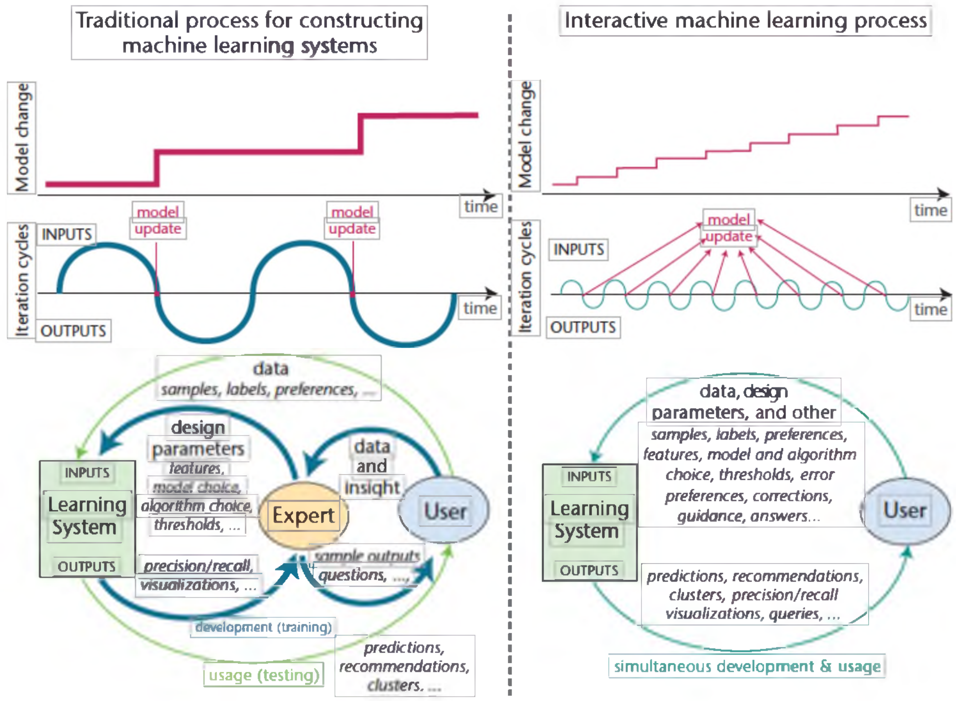


Figure 5.2 Comparison of traditional vs. interactive machine learning processes.

While the focus of this section is not on interactive machine learning *per se*, since it is a constituent part of a complex system it is necessary to give a description of what it is and its role as a *deus ex machina* for the human-AI interface.

Artificial intelligence and its progeny have not always been this way. From the outset, there were two schools of thought regarding how understandable, or explainable, AI ought to be. Many thought it made the most sense to build machines that reasoned according to rules and logic, making their inner workings transparent to anyone who cared to examine some code. Others felt that intelligence would more easily emerge if machines took inspiration from biology and learned by observing and experiencing. This meant turning computer programming on its head. Instead of a programmer writing the commands to solve a problem, the program generates its own algorithm based on example data and a desired output. The machine-learning techniques that would later evolve into today's most powerful AI systems followed the latter path: the machine essentially programs itself.

At first this approach was of limited practical use, and in the 1960s and '70s it remained largely confined to the fringes of the field. Then the computerization of many industries and the emergence of large data sets renewed interest. That inspired the development of more powerful machine-learning techniques, especially new versions of one known as the artificial neural network. By the 1990s, neural networks could automatically digitize handwritten characters.

But it was not until the start of this decade, after several clever tweaks and refinements, that very large, or "deep" neural networks demonstrated dramatic improvements in automated perception. Deep learning is responsible for today's explosion of AI. It has given computers extraordinary powers, like the ability to recognize spoken words almost as well as a person could, a skill too complex to code into the machine by hand. Deep learning has transformed computer vision and dramatically improved machine translation. It is now being used to guide all sorts of key decisions in medicine, finance, manufacturing and beyond. A network's reasoning is embedded in the behavior of thousands of simulated neurons, arranged into dozens or even hundreds of intricately interconnected layers. The neurons in the first layer each receive an input, like the intensity of a pixel in an image, and then perform a calculation before outputting a new signal. These outputs are fed, in a complex web, to the neurons in the next layer, and so on, until an overall output is produced. Plus, there is a process known as backpropagation that tweaks the calculations of individual neurons in a way that lets the network learn to produce a desired output.

The many layers in a deep network enable it to recognize things at different levels of abstraction. In a system designed to recognize dogs, for instance, the lower layers recognize simple things like outlines or color; higher layers recognize more complex stuff like fur or eyes; and the topmost layer identifies it all as a dog. The same approach can be applied, roughly speaking, to other inputs that lead a machine to teach itself: the sounds that make up words in speech, the letters and words that create sentences in text, or the steering-wheel movements required for driving. It is the interplay of calculations inside a deep neural network that is crucial to higher-level pattern recognition and complex decision-making, but those calculations are a quagmire of mathematical functions and variables.

The objective of this research is to provide a flexible systems management methodology (FSM) for practitioners that allows them to generate and sustain flexibility as a core ability to meet and address any system disparities, lack of capacity, and coordinating system capabilities to meet production needs and customer requirements. It is also to highlight the future of manufacturing and how workers can collaborate with robotics and AI as the model of the future for manufacturing. The gap this research addresses in this new area is a model and measure the benefits using SoS engineering. Therefore, the need to research and develop these flexibility methods for complex systems is established here and further research into an effective method(s) of generating flexibility are shown herein as well.

5.2.1. Definition of the Human-Technology System – Pyramidal Model of Manufacturing. First, we need to identify what this future manufacturing system looks like in terms of its principal components or operators. A system is a group of interacting or interrelated entities that form a unified whole. A system is described by its spatial and temporal boundaries, surrounded and influenced by its environment, described by its structure and purpose and expressed in its functioning [175]. Figure 5.3 is a pyramidal diagram of the components in a generic high technology manufacturing setting.

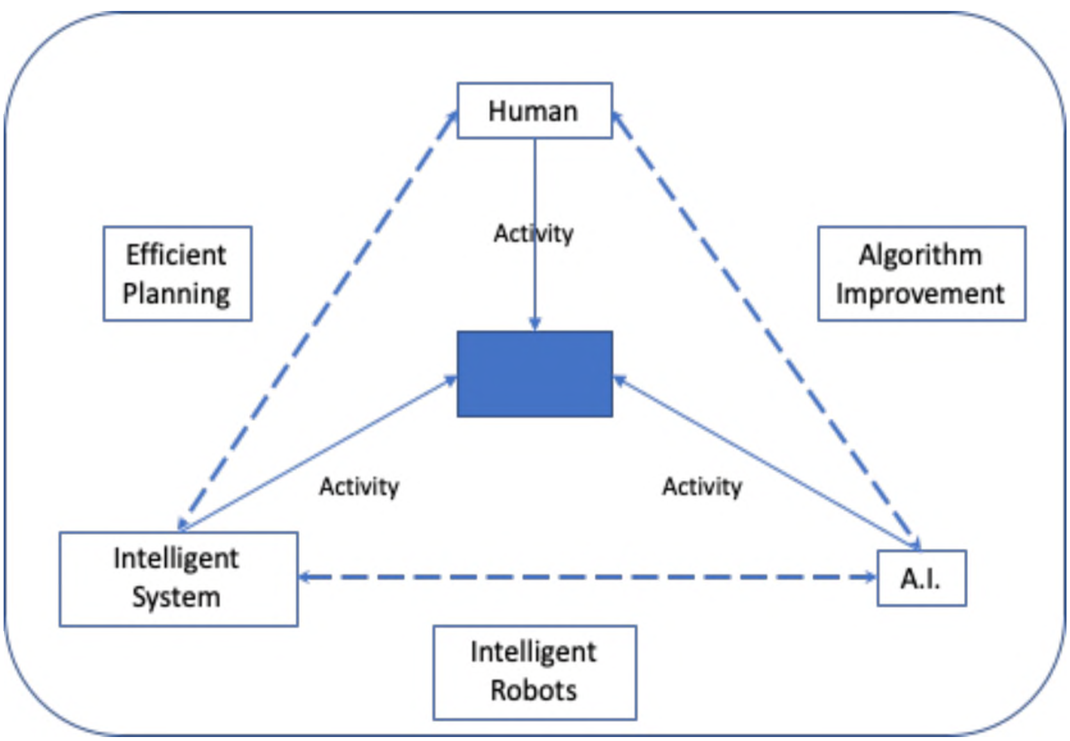


Figure 5.3 A model of a human-AI interface with an intelligent system.

This human-technology (H-T) unit is composed of a person (H) who interfaces with the AI to contribute creativity, knowledge, and any corrections to the algorithm of the AI which interfaces with the intelligent system to perform the work required to achieve objectives. But the human also interfaces with the intelligent system to provide the plan, objectives, and overall goals for the system to achieve. The interface occurs in both directions such that the human is aware of the intelligent system's progress and can make adjustments to increase efficiencies and minimize inefficiencies in achieving objectives or helping to match skill levels and capabilities to current or new tasks or exogenous conditions. The human in this system works with assessing, improving, and modifying the algorithm that drives the artificial intelligence that runs the robots that work with or as an intelligent system. The human factor is important in that it is the vehicle through which innovative thought and creativity are inserted into the system to allow system adaptation to external influences or factors. We still preserve the benefits of automation with the selective inclusion of human participation in this system. Also, there is an inherent level of control that comes from including a person in this system. The human can receive feedback from the intelligent system or the AI itself to guide any choices in modifications or changes to the AI algorithm or even adjustments to the functioning of the intelligent system. Technology has enabled interfaces that facilitate this interaction, whether it is through cellular technology or distance-enabled communications via satellites, telepresence, graphical user interfaces (GUIs), or other hybrid means of contemporary communications. In this vision of future manufacturing, intelligent systems are designed to augment or enhance the human, serving as a tool to be wielded through human interaction.

5.2.2. Artificial Intelligence Progression. AI initially started as being a pure sub-branch of computer science that aimed at making computers and machines intelligent; where intelligence is restricted in the short term to mean reasoning, knowledge representation, planning, learning, natural language processing, vision and perception. In the long-term view, the ambition is to achieve AGI (Artificial General Intelligence), or Strong AI as it is sometimes referred to, where the idea of intelligence involves a much more complex problematization of an amalgamation of various scientific disciplines such as mathematics, psychology, engineering etc. For the purposes of this research, we disregard AGI as the “singularity” (which is a precondition for AGI) is not anywhere near the horizon of achievability based on the consensus among those in the global AI community.

Alan Turing, considered by many as the father of modern computer science, published in 1950 his popular Turing test that consisted of a machine that can make conversation that is indistinguishable from a conversation with a human being [176]. If the machine passed the test, it would be labelled as “Intelligent”, per Turing. He dreamed of the day when humanity would make its Last and Final Invention. Since then, advancements in computer science coupled with the revolution in technology pertaining to higher processing power via pursuit of Moore’s Law, has made it somewhat possible for these purely theoretical musings to take some tangible shape and form.

AI has made incredible progress in the past few years. The AI of today do specific tasks such as driving a car, booking meetings or even talking on your behalf on an audio call. All these enhancements were brought forth by AI’s subsets and techniques. The following Venn diagrams in Figures 5.4 and 5.5 [177] depict the layers of AI as it stands today.

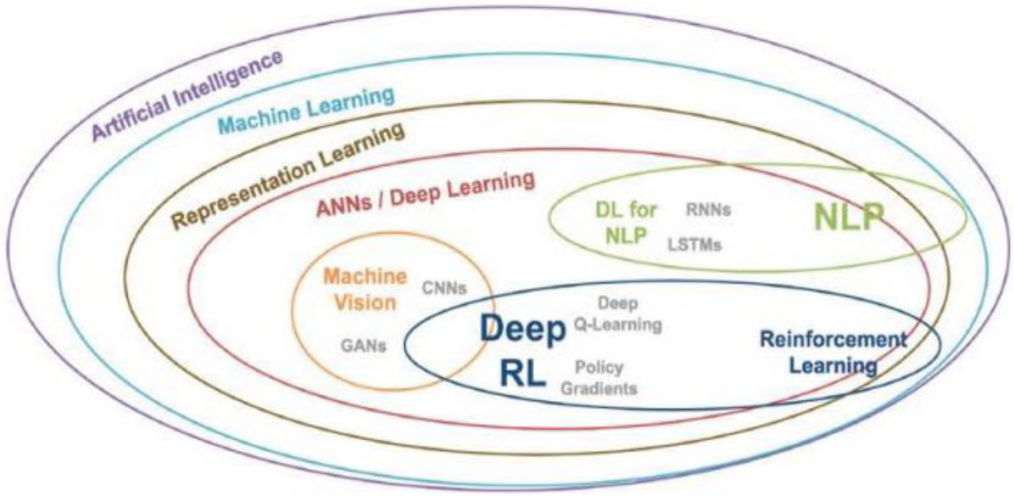


Figure 5.4 Venn diagram representing AI and its subsets.

Artificial Intelligence (AI): AI can be defined as a machine that inputs data from the real world, processes it and makes specific decisions as a result in order to achieve a goal. Today's applications of AI include driving cars, chatbot, image/voice recognition, etc.

Machine learning (ML): ML is a subset of AI which focuses on developing software, mostly algorithms that can learn to accomplish tasks by themselves without a developer explicitly telling it how to. For ML to properly work it needs clean and relevant data.

Representation Learning (RL): RL is a branch of ML which goes deeper than Traditional ML which needs more human intervention. RL models take in huge amounts of data and learn representations also called features by themselves.

ANNs: Artificial neural networks are the most popular RL technique. It was inspired by the human brain. It is a collection of artificial neurons that are arranged in such a way that they can send and receive information among them in order to produce the desired output.

Deep learning: Deep learning is also an RL technique. It is made of five or more layers of artificial neurons. A single input layer takes the data, three or few hidden layers that processes the data and learn new features and a single output layer to show results.

Machine vision: It is a branch of deep learning that focuses on object recognition. It is used for self-driving cars algorithms, image recognition and any AI that needs to at some point to recognize objects.

NLP: Natural language processing is a machine learning technique that is used to teach the machine to recognize characters and language. Deep learning for NLP is a much efficient technique that allows AI to interact via natural languages (spoken or typed). The following diagram shows that NLP sits at the intersection of the computer science, AI and linguistics fields.

Deep reinforcement learning: DRL is a reinforcement learning technique that involves artificial neural networks. Reinforcement learning is good at taking the appropriate decision among many options. DRL is better when it comes to processing a huge variety of data coming from an external environment.

While AI is still in the early majority phase of adoption, Kathleen Walch, in her 2019 publication “The Seven Patterns of AI” [178], revealed that probably many of the companies developing AI solutions are also the ones applying and experimenting with AI approaches to their project management. We believe that leading AI start-ups like those listed from CB Insight's 2020 list of top 100 AI start-ups are the best place to look for emerging practices in this field [179]. Within a system as outlined in Figure 5.3, the appropriate AI would be chosen to fit the application, but with the consideration of including the human with the opportunity to interface with the AI to add any creativity, guidance, or contribution to the AI's learning such that the system addressing issues, focuses on core goals, increases its skill level and capabilities and allows the AI to communicate with the intelligent system to achieve system goals and objectives. This interaction between these components is essential to the generation of flexibility within the system.

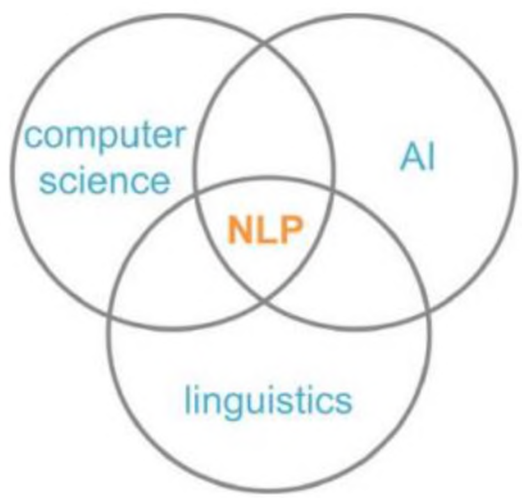


Figure 5.5 Venn diagram of natural language processing.

5.2.3. Intelligent Systems – A Critical Component. In the preceding paragraphs, the discussion of AI concerned the intelligence, or "brains", of the overall pyramidal system and its potential for learning and acquiring skills and capabilities that are used to teach the machine of the system and to carry the kernels of innovation and creativity imparted by the human component to the AI. Intelligent systems (or agents) are technologically advanced machines that perceive and respond to the world around them by taking actions that maximize their chances of success. By this definition, simple programs that solve specific problems are “intelligent systems”, as are human beings and organizations of human beings, such as firms. The intelligent system paradigm defines AI research as the study of intelligent systems [180]. This is a generalization of some earlier definitions of AI: it goes beyond studying human intelligence; it studies all kinds of intelligence. Intelligent systems can take many forms, from automated vacuums such as the Roomba to facial recognition programs to Amazon's personalized shopping suggestions. The field of intelligent systems also focuses on how these systems interact with human users in changing and dynamic physical and social environments. Early robots possessed little autonomy in making decisions: they assumed a predictable world and performed the same action(s) repeatedly under the same conditions. Today, a robot is considered to be an autonomous system that can sense the environment and can act in a physical world in order to achieve some goals. An intelligent agent is a system that perceives its environment and takes actions which maximize its chances of success.

5.2.4. Challenges in Intelligent Systems. Research in intelligent systems faces numerous challenges, many of which relate to representing a dynamic physical world computationally.

Uncertainty: Physical sensors/effectors provide limited, noisy and inaccurate information/action. Therefore, any actions the system takes may be incorrect both due to noise in the sensors and due to the limitations in executing those actions.

Dynamic world: The physical world changes continuously, requiring that decisions be made at fast time scales to accommodate for the changes in the environment.

Time-consuming computation: Searching for the optimal path to a goal requires an extensive search through a very large state space, which is computationally expensive. The drawback of spending too much time on computation is that the world may change in the meantime, thus rendering the computed plan obsolete.

Mapping: A lot of information is lost in the transformation from the 3D world to the 2D world. Computer vision must deal with challenges including changes in perspective, lighting and scale; background clutter or motion; and grouping items with intra/inter-class variation.

5.3. HUMAN-TECHNOLOGY UNITS AS A SYSTEM OF SYSTEMS (SOS) – CHARACTERISTICS

A system of systems (SoS) is an arrangement of independently managed and operated systems. A SoS has five characteristics that differentiate it from a system as shown in Figure 5.6. SoSs provide unique capabilities for meeting special needs for intelligence [181-182], flexibility [183], and synergy [184], which may not be accomplished by a single system. Since we have defined above that the components in Figure 9 are a system, one can see that if multiple systems interface with other similar or identical H-T units of H-T-intelligent system interfaces then one can see that teams of these H-T units can be seen as a type of system of systems. Sauser and Boardman defined the five characteristics of a SoS as: Autonomy, Belonging, Connectivity, diversity, and emergence [185]. Each of these attributes are intrinsic to the H-T system. It is apparent that using human-technology combinations that utilize AI across multiple locations is no longer limited solely by geography but have begun to operate in collaborative manufacturing networks that possess SoS attributes.

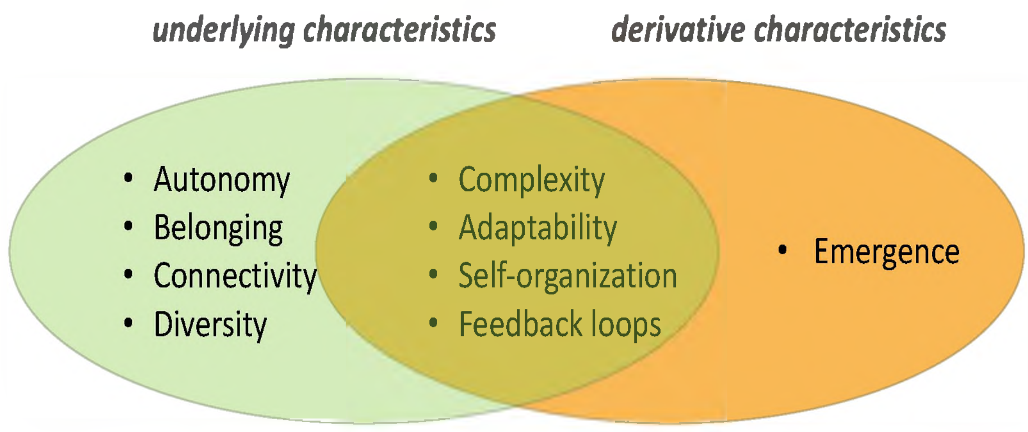


Figure 5.6 Underlying & derivative characteristics of a System of Systems (SoS).

A description of how those attributes relate to the H-T network is described below.

Autonomy: An autonomous system is "situated within and a part of an environment that senses its local environment and acts upon that environment in pursuit of its own agenda" [186]. Based on the foregoing discussion, an intelligent system matches this definition of autonomy. The H-T unit, which incorporates the intelligent system, functions as an independent system that can choose how it acts to achieve its objectives or pursues its own agenda.

Belonging: Constituent systems within an SoS choose to be part of the larger system because of their needs, beliefs or fulfillment [187]. The inherent collaborative nature of the human-AI-intelligent system is a defining characteristic. H-T units enter the network of their own accord and set the terms of their involvement with other H-T units upon entering. Aligning with units of similar purpose and form facilitates this belonging to the over SoS network.

Connectivity: System-of-systems feature interoperability and a communication capability between the constituents of the SoS so that social functionality is enabled [188]. This interoperability is essential for operations with integrated H-T units. These systems use information and communications technology to transmit and share information across the network of units, therefore distributing the total production load, sharing of algorithm knowledge, intelligent unit configuration(s), and interfacing among humans to stimulate creativity and connectivity among multiple units.

Diversity: Another attribute of system-of-systems is that they feature visible heterogeneity. That is, they include "distinct or unlike elements or qualities in a group" [185]. Whether inter-unit or intra-unit, each of the collaborative manufacturing networks is an amalgam of such heterogeneous entities. Each H-T unit has distinctive capabilities and competencies and participates in the network so that it can obtain access to those complementary capabilities and competencies that it does not possess.

Emergence: The final core concept in Sauser and Boardman's model is that system-of-systems exhibit emergent attributes, including unexpected structures and behaviors [185]. These multi-H-T unit collaborative manufacturing networks are each expected to be transitory in nature as they dissolve following the delivery of the customer's or manufacturing planning's requirements. However, there is the potential for collaborations to endure and to take on new forms beyond the completion of the initial network's objectives. The emergence factor is the key characteristic that is observed and measured as the representation of flexibility. The new structures and behaviors shown by a SoS network of H-T units embodies the response to exogenous factors in the environment that applies external force(s) that threaten SoS goals and objectives.

5.4. DEFINING FLEXIBILITY WITHIN THE H-T SOS - EMERGENCE

The emergence factor is the key derivative characteristic for the SoS configuration that is observed and measured as the representation of flexibility, i.e. flexibility equals emergence. The new structures and behaviors shown by a SoS network of H-T systems embodies the response to exogenous factors in the environment that applies external force(s) that threaten SoS goals and objectives. The SoS on whole being an intelligent system monitors and responds in this way to the environment surrounding it.

Figure 5.6 illustrates the overlay of the SoS characteristics of complexity, adaptability, self-organization, and feedback loops. In starting with defining the H-T unit as a system of components that are inherently complex, the complexity of the system is multiplied as the components work together as their own singular system. If, as shown in Figure 5.7, we have a network of systems to create an SoS network the connections become intertwined and the interoperability of the systems become even more complex as the H-T units communicate plans, objectives, capacities, capabilities (i.e. skill levels), and potentially share resources. This SoS network facilitates the creation and sustainment of the SoS utilizing feedback loops as represented by the bi-directional arrows in Figure 5.7. These loops carry information that increases the underlying and derivative characteristics that drive these H-T systems into a network of systems. The feedback loops also carry old, new, and shared learning of skills to accomplish goals and plans more effectively than the H-T units do individually.

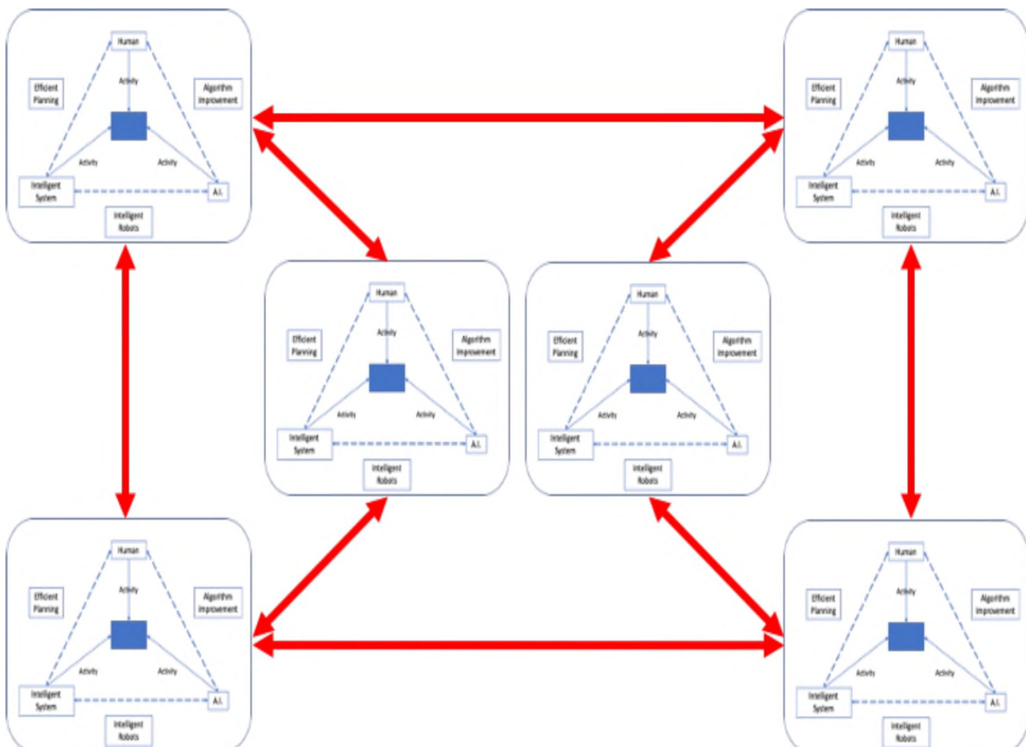


Figure 5.7 Configuration of an SoS network of Human-Technology (H-T) units interfacing with other H-T units as a system of systems connected via communication links at multiple nodes.

Figure 5.8 shows a fully operable SoS network that chooses to communicate or share information via feedback loops or to belong to each other. The adaptability of the

SoS comes in the form of the formation of those structures and connections either as needed for additional capability or capacity or to shift the load for producing materials or products that fulfill the production plan requirements that meet system goals and objectives. The self-organization characteristic is revealed in terms of which connections each H-T unit chooses to establish and collapse as a result of meeting objectives, acquiring skills to complete those tasks and objectives, or to increase capacity on an ad hoc basis as indicated by the dashed lines in Figure 5.8.

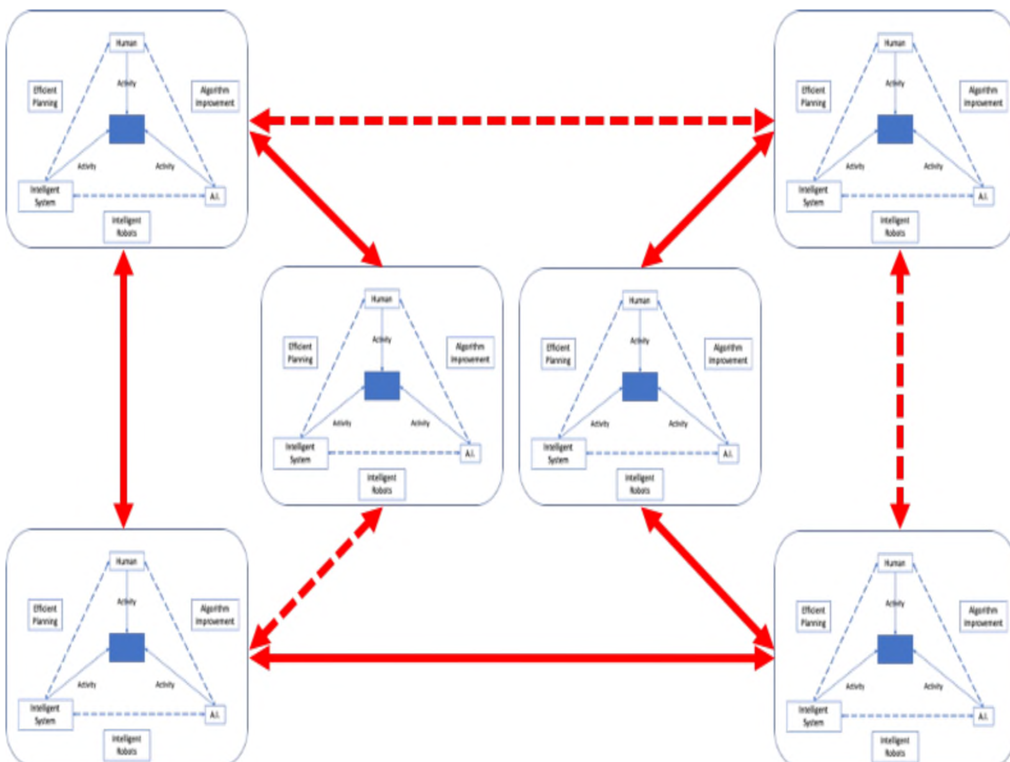


Figure 5.8 Illustration of a SoS with partial connectivity & unavailable H-T systems.

The dynamic and static manners in which these connections form and dissolve defines where flexibility will emerge. Flexibility is the ability of the H-T interface to work with the intelligent system to achieve SoS objectives in accordance with a production plan to meet the requirements of the customer in the most efficient and least costly manner.

5.5. LIMITED ORMS MODEL

A primary objective of this research was to develop a mathematical model which can be used to help managers decide on optimal tactical plans for training or retraining of a new or existing workforce according to the skill levels demanded by a forecasted production schedule for a defined planning horizon. Instead of humans only as the workforce, we have our H-T systems. The objective function for this model reflects the desire of a manager. Three such goals were identified:

1. To minimize the total cost of machine learning via algorithm training,
2. To minimize the total time of machine learning via algorithm training, and
3. To maximize the flexibility of the workforce.

Since there is a trade-off between training cost and H-T system flexibility, a multi-objective model was also developed to consider these two opposing goals simultaneously. Constraints in this model were developed to represent restrictions on overtime, total number of production hours and rates required to meet the master schedule and budget.

The analogy of a skill acquired by a worker is seen in the H-T system wherein the human helps to refine the learning algorithm, which is designated by a level number, k.

An intelligent robot is said to be skilled at a certain level if it has gone through a certain number of hours of algorithm efficiency levels of training or possesses a defined set of capabilities. Thus, skill gained from the H-T interface is defined as a discrete variable with a finite number of states. The higher the level number, generally the more complex is the required training. When an intelligent robot has received training at the highest level of an intelligent system, then it may be considered an expert operator of that intelligent system which interfaces with the AI.

This approach implies that all intelligent robots, or machines, learn at the same rate, and the cost or time associated with that learning represents an average value s. Since the actual learning rate is not generally known or would be difficult and time consuming to obtain, this approach is reasonable and practical. However, if true learning rates are known for each of the individual robotic systems, these can be taken explicitly into account in the formulations with the appropriate cost or time coefficients.

All intelligent robot systems need not have the same number of levels. The number of levels a particular type of machine can have depends on several factors, some of which are:

- whether the organization is AI-constrained or machine-constrained,
- complexity and sophistication of the intelligent system or robot,
- number and types of process plans available,
- for large organizations where there are several manufacturing cells with H-T systems, the need for intercellular H-T system transfer,
- strategic and tactical plans of the organization,
- algorithm efficiency training budget.

A small number of skill levels for a complex machine will reduce the flexibility of the overall system. On the other hand, too many skill levels could be costly for the organization. A compromise must be made between meeting the organizational needs and staying within the budget constraints.

A process plan and planning horizon are needed to determine the quantity of work, skill levels required, and the time constraints within which the production must be completed. Each product may have several alternate process plans, but one must be selected for use in the formulations.

The planning horizon may include two time spans.

1. the time required to train the H-T system with the requisite skills for the production in a particular process plan, i.e., the training horizon, and
2. the time required to produce all products (or perform all tasks) contained in the process plan, i.e., the production horizon.

If the planning horizon is particularly short or the skill levels demanded are atypical for a specific planning horizon, it may be short-sighted to train only for these skill levels. It may be desirable then either to extend the planning horizon or to include additional requirements to account for long-term skill level demand not required in the current planning horizon.

There are three major constraint sets applicable to this model's formulation.

1. Each intelligent robot is available for a specified maximum amount of time during the time horizon.
2. Each intelligent system is available for a specified maximum amount of time during the time horizon.

3. Production requirements for the production horizon must be fulfilled.

Constraint (1.1) ensures that no intelligent robot is allowed to work more than R time units within a shift. One constraint is required for each of the I robots. Constraint (1.2) ensures that no robot produces more than is possible for one intelligent robot during one shift and forces the 0/1 variable to 1 when a robot must be trained. Constraint (1.3) ensures that no intelligent system is used more than R time units within a shift. One constraint is required for each of the J intelligent system types. Constraint (1.4) ensures that the production requirements are met. One constraint is required per intelligent system/skill-level combination

$$\sum_{i\in I}\sum_{j\in J,}\sum_{k\in K_j} T_{pjk}\, Y_{pijk} \leq R \qquad \text{for each } i\in I, \qquad (1.1)$$

$$Y_{pijk} \leq U_{pjk}\, X_{ijk} \qquad \text{for each } p\in P,\ j\in J,\ i\in I,\ k\in Kj, \qquad (1.2)$$

$$\sum_{i\in I}\sum_{j\in J,}\sum_{k\in K_j} T_{pjk}\, Y_{pjk} \leq RM_j \qquad \text{for each } j\in J, \qquad (1.3)$$

$$\Sigma \quad Y_{pijk} = \frac{Np}{HQ} \qquad \text{for each } j\in J,\ i\in I \qquad (1.4)$$

where, X_{ijk} =1, if robot i is trained on intelligent system i at algorithm skill level k, and X_{ijk} = 0, otherwise. Y_{pijk} is the number of H-T units per shift of product p that intelligent robot i processes on intelligent system j at algorithm skill level k (note: $Ypijk$ is a continuous variable and may not be an integer), $Cijk$ is the cost of training intelligent robot i on intelligent system j at algorithm level k, $Cijk$ = 0, if already trained, I is the index set of all intelligent robots in the system, J is the index set of all intelligent systems in the system, K_j is the index set of all algorithm skill levels of intelligent system j, P is the index set of all products, N_p is the total number of product p required in the production horizon, H is the number of days in the production horizon, Q is the number of shifts per day, R is the duration of each shift, T_{pjk} is the processing time per unit for product p on intelligent system j at algorithm skill level k, M_j is the number of intelligent systems of type j, and U_{pjk} is the maximum number of units of product p that can be produced in one shift by one intelligent robot on intelligent system j at level algorithm skill level k (calculated as R/T_{pjk}).

Solution infeasibility would indicate that one or more of the following resources would have to be increased:

- number of intelligent systems,
- number of intelligent robots,
- production horizon.

If production is constrained by one or more of the resources, then the following options for increasing one or more of these resources may be considered:

a) acquire additional intelligent systems,
b) operate critical intelligent systems on additional shifts or around the clock,
c) install additional intelligent robots,
d) assign or exchange intelligent robots with the required skills or capabilities in place of those that don't possess those skills or capabilities,
e) examine alternate process plans.

This objective model may be applied in a situation where management wants to achieve workforce flexibility in a cost-effective manner. The objective of this model is to minimize simultaneously the cost of training and the cost of attaining flexibility of the workforce for a future production schedule subject to the standard constraints. The model is given below:

Minimize

$$W_{ct} \sum_{i \in I} \sum_{j \in J,\, k \in K_j} \sum C_{ijk} X_{ijk} + W_{cf} \sum_{i \in I} \sum_{j \in J,\, k \in K_j} \sum F_{ijk} X_{ijk} \tag{1.5}$$

subject to constraint equations (1.1), (1.2), (1.3) and (1.4), where

F_{ijk} = cost (penalty) of attaining flexibility if H-T *i* is trained on intelligent system *j* at level *k*.

W_{ct} = weight assigned to the cost of training.

W_{cf} = weight assigned to the cost of flexibility.

The cost of attaining flexibility is quantified, and the multiple objectives are represented by a composite objective function. Different weights may be assigned to these two objectives to reflect priorities of the decision maker. The ratio of these two weights would indicate the relative desirability of each objective. Several approaches have been used to evaluate flexibility based on machines, processes, products, routings, volumes, expansions, operations, and productions. For example, one can refer to Kumar [190], Brill and Mandelbaum [191], Primrose and Leonard [192], Son and Park [193], Gupta and Goyal [194], and Stewart, et al. [195]. Chryssolouris and Lee [196] give a procedure for determining the costs of attaining flexibility. In their approach, the measure of flexibility accounts for the penalty for change and the probability of change, i.e.,

STC = Sensitivity to Change,

= Penalty * Probability.

In general terms:

$$\text{STC} = \sum_{s=1}^{S} \text{Pn}(Xs)\ \text{Pr}(Xs), \tag{1.6}$$

where

S = number of potential state changes.

s = state transition index.

X_S = random variable for the potential state change *s*.

$\text{Pn}(X_s)$ = penalty of potential state change *s*.

$\text{Pr}(X_s)$ = probability of the potential state changes.

STC is actually a measure of inflexibility. The lower the STC, the higher the flexibility. Thus, flexibility is inversely related to STC. If change can be implemented without penalty, the system has maximum flexibility, and STC is zero. On the other hand, if change results in a large penalty, the system is very inflexible, and the STC value is large.

Applying this approach to the current problem, if an H-T unit has to be trained on an intelligent system or skill level combination, our objective is to minimize this cost so that the flexibility attained per dollar spent will be maximized.

An event related to the H-T “worker” that might affect production may be the system refusing a project or being unavailable for work. The costs (penalties) that accrue to a manufacturer when a H-T “worker” is unavailable or does not have the skill level for the job are the cost of training, retraining, or acquiring another H-T system to fill the position and the cost of production loss, if any. These same cost (penalty) components may apply if a H-T system (or its components) is unavailable or non-functioning. Thus, in general terms, the cost of attaining H-T system flexibility (SF) may be determined as follows.

$SF = \mathrm{PEN_L}*\mathrm{PROB_L}+\mathrm{PEN_A}*\mathrm{PROB_A}$, (1.7)

where

$\mathrm{PEN_L}$ = penalty to the manufacturer if a H-T unit leaves the job.

$\mathrm{PROB_L}$ = probability that a H-T unit will leave the job.

$\mathrm{PEN_A}$ = penalty per day to the manufacturer if a H-T unit is unavailable.

$\mathrm{PROB_A}$ = probability that a H-T unit will be unavailable in any one day during the production horizon.

The above-mentioned penalties can be estimated as follows:

$\mathrm{PEN_L}$ = cost of hiring, training, or retraining a H-T unit to fill the position, and

$\mathrm{PEN_A}$ = production loss and/or cost of an additional H-T work unit.

The values of $\mathrm{PEN_L}$ and $\mathrm{PEN_A}$ will vary from system to system as the cost of installing new robots or intelligent systems and production loss both depend on H-T unit and algorithm efficiency level combination(s) of the H-T unit who leaves the SoS or is unavailable for the work. Therefore, $\mathrm{PEN_L}$ and $\mathrm{PEN_A}$ should be defined in terms of each H-T system and/or algorithm-level combination. Thus,

$\mathrm{PEN_{ujk}}$ = cost of hiring, training, or retraining a H-T unit to fill an open position i on intelligent system j at level k.

$\mathrm{PEN_{Aijk}}$ = production loss and/or cost of additional H-T unit per day due to H-T unit i being absent on intelligent system j at algorithm skill level k.

To estimate the probabilities, for example, let

$\mathrm{PROB_L}$ = (number of intelligent robots and H-T units who are unavailable for the job during last 365 available working days) ÷ (total number of intelligent robots and H-T units working during that period), (1.8)

$PROB_A$ = (number of unavailable intelligent robots and H-T unit days during last 365 working days) ÷ (total number of intelligent robots and H-T unit days available during that period). (1.9)

Here,

total unavailable intelligent robot-days = $\sum_{i=1}^{I} A_i$, (1.10)

where A_i = number of days that intelligent robot i was unavailable in the last 365 days.

The values of $PROB_L$ are assumed to be the same for all intelligent robots or H-T units. Thus, the cost or penalty (STC) of attaining flexibility if intelligent robot *i* is trained on intelligent system *j* at level *k* is given by

$SF_{ijk} = PEN_{Lijk}*PROB_L + PEN_{Aijk}*PROB_A$ (1.11)

H-T system flexibility is achieved by training robots on more intelligent system-algorithm efficiency level combinations than required under the present production plan. H-T system flexibility is achieved by first quantifying the cost of intelligent robot flexibility and then by minimizing the combined costs (cost of training plus the cost of intelligent robot flexibility).

5.6. COORDINATING STRATEGIES FOR MANAGING FLEXIBILITY FOR THE H-T SOS

Multi-faceted systems of the future entail complex logic and reasoning with many levels of reasoning in intricate arrangements. The organization of these systems involves a web of connections and demonstrates self-driven adaptability. They are designed for autonomy and may exhibit emergent behavior that can be visualized. The challenge in complex systems design is to design an organized complexity that will allow a system to achieve its goals. The system-of-systems (SoS) approach is developed to handle this huge uncertainty in socio-technical systems.

Per Dahmann, et al., four categories of SoS are described in the literature namely: directed, acknowledged, collaborative and virtual [197]. These four types of SoS vary based on their degree of managerial control over the participating systems and their structural complexity. The spectrum of SoS ranges from directed SoS that represents complicated systems to virtual SoS that are complex systems. Figure 5.9 is a summary illustration of the four types of coordinating strategies for an SoS.

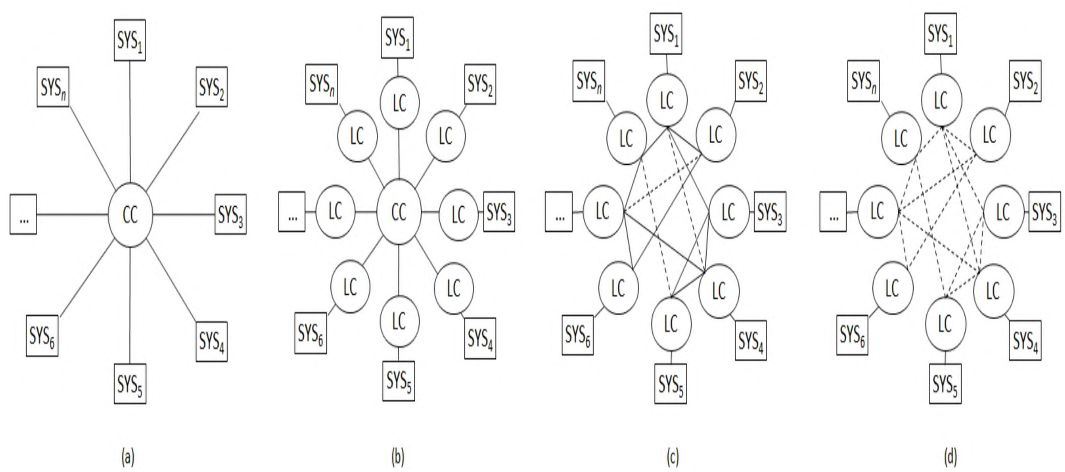


Figure 5.9 Coordination strategies for SoS.

5.6.1. Centralized Coordination (Directed SoS). The strategy in Figure 5.9(a) is used when the goals of constituent systems are highly consistent with the SoS. Systems are willing and able to follow the central controller (CC) in general operating mode. Local controllers are subordinated due to high cooperation levels and response capability with the central controller. In this strategy the local AI and human for the H-T unit is subordinated to a central AI that administers the production plan within the planning horizon. It directs the other H-T systems to meet the needs of the overall SoS according to each system's overall capability and skill level. This is an effective strategy for enhancing the diversity and emergence SoS characteristics which translate into high flexibility for the SoS. The global optimal outcome for the SoS heavily relies on the speed & reliability of communication between constituent systems and the SoS. Given the communication loops of the H-T systems this communication is highly efficient. As noted, the goals of each of the H-T systems must be consistent with the SoS, otherwise a degradation in the emergence characteristic is expected and flexibility is not globally optimized.

5.6.2. Hierarchical Coordination (Acknowledged SoS). The strategy shown in Figure 5.9(b) is similar to the centralized strategy with a CC. Constituent system have local controllers (LCs) designed for their own interests and purposes. These LCs in the context of H-T systems may take the form of a local AI or even a human that interfaces with their LC to evaluate goals and planning requirements as a member of the SoS. The CC cannot assume the LCs are always available or willing to follow its commands. The CC has to influence the belonging and autonomy levels of each system and leverage the output of the LCs. The use of this strategy requires knowledge of the control mechanism of each LC, e.g. via learning. This learning process is more easily facilitated in a H-T system context due to the ease of connection and conversely, disconnection, which in itself can be a learning on how to influence the LC. This configuration has a better chance than centralized coordination of seeing deviations between realized outcomes of LCs vs. the anticipated outcome (variance analysis).

5.6.3. Peer Coordination (Collaborative SoS). This coordination strategy is partially decentralized, i.e. no CC, as illustrated in Figure 5.9(c). Constituent subsystems communicate with subsets of each other for various reasons, e.g. geography, communication band limit. Links between LCs of any 2 systems equals connectivity. A solid link indicates connection, but a dashed link indicates no connection at the moment. Fulfillment of the purpose of the SoS relies on voluntary collaboration among constituent systems. For H-T systems this collaboration will require agreement within the component systems of the H-T unit before connecting and belonging to another system. Therefore, use of this strategy shows that information on variables with common interest is shared among systems. A mechanism must exist, or be created, that makes other systems adjust their autonomy and belonging. With this, they will converge to consensus on common interest variables such as learned skills and sharing H-T system capacities to achieve common objectives.

5.6.4. Decentralized Coordination (Virtual SoS). As shown in Figure 5.9(d), a CC does not exist in this strategy. LCs use only local measurements as input. Therefore, communication between constituent systems is not needed. This strategy is highly tolerant to changes and failures in constituent systems due to avoidances in belonging and connectivity. But, since these systems are extremely autonomous and low in connectivity and belonging levels, it is difficult to attain & maintain the emergent capabilities of the SoS and therefore flexibility within H-T systems. The unintended SoS behaviors that evolve from applying this strategy are difficult to correct here due to literally lack of coordination among systems.

5.6.5. Optimal Coordination Strategy for H-T SoSs. An acknowledged SoS lies in between this continuous spectrum SoS types. This particular SoS is the focal point of our coordination strategy. The acknowledged SoS and directed SoS share some similarities such as both have SoS objectives, management, funding and authority. Nevertheless, unlike directed SoS, acknowledged SoS systems are not subordinated to SoS. However, acknowledged SoS systems retain their own management, funding and authority in parallel with the SoS. Collaborative SoS are similar to acknowledged SoS systems in the fact that systems voluntarily work together to address shared or common interests.

Therefore, based on the research and case studies in prior sections it is determined that a hierarchical coordination strategy (acknowledged SoS) is the optimal coordination strategy for a SoS network of H-T systems.

5.6.6. New Definition of Flexibility in Modern Manufacturing. Given the development of the H-T systems model for modern manufacturing, the mathematical model of flexibility and its associated costs and dependence on constituent systems and their capacities and capabilities, we can state that flexibility within the H-T systems construct is the ability and capacity of the H-AI interface to learn and work with its intelligent system under a hierarchical coordination strategy to achieve human-technology system objectives in line with a production plan to meet the requirements of the customer in the most efficient and least costly manner using inherent and learned AI skill levels and machine capacities to diminish system disparities caused by uncertainties or deviations.

6. CONCLUSIONS

6.1. SUMMARY

In Section 2 an analysis of the literature associated with workforce flexibility was conducted to classify and determine the scope and effectiveness of each flexibility method. The analysis spanned the ORMS area and included flexible working time, floaters, cross-training, teamwork, and temporary workers. The selection of workforce flexibility method(s) for a specific application is often a decision with multiple considerations. To facilitate such decision making, the key aspects of each method was summarized from the ORMS perspective.

Among the wide range of possible problem types, a distinction was made between problems arising from changes in demand volume or worker supply, and problems arising from changes in product mix. While flexibility is a potentially useful mitigator of both types of uncertainty, specific methods are often designed to focus on one type or the other. For instance, flexible working time and temporary labor are primarily used for dealing with changes in demand volume and worker supply. Yet, we notice that some flexibility methods may serve multiple purposes. Based on this analysis a traditional view of flexibility was drawn as a baseline for researching flexibility in complex systems and as a basis for reforming that definition.

Next, a methodology for acquiring and maintaining flexibility for distributed large-scale or complex systems in changing or uncertain conditions through forming and evolving SoSs was proposed. Findings from the study of this topic positively support the proposed methodology. Challenges are present in the implementation of the proposed methodology. From the management perspective, SoSs have multiple flexibility mechanisms. Decisions for forming SoSs and evolving them over time are complex; in that the decisions are interdependent and across multiple time scales. Addressing these challenges needs a seamless collaboration between systems engineers and domain experts.

Therefore, this research also addressed the implementation problem of coordinating constituent systems within a SoS to preserve the creation of flexibility made possible by the SoS approach. The choice of strategy, barriers to applying a coordination strategy, and methods to address removing those barriers were discussed, evaluated, and applied to a SoS involving another distributed energy system on an island.

Motivated by the importance of coordinating the constituent systems of a SoS in operations, this portion of the research analyzed the mechanism of choosing coordinating strategies for the SoS, discussed challenges in the implementation, and proposed methods for addressing these. Section 4 further examined the proposed work in a case study that involved coordinating heterogeneous power generators, multiple parallel energy storage systems, and distributed loads on an island. The work of this section confirmed the importance of considering SoS characteristics in choosing coordinating strategies. It also showed that SoS characteristics are affected by the effectiveness of coordinating constituent systems.

Since an analysis of the mechanisms of how to form and evolve systems of systems in uncertain environments was established within this dissertation research and a management coordination strategy was developed for choosing the proper SoS coordination method, the application of a flexible systems management methodology consisting of these principles and techniques were applied to an area that had not seen the application of this methodology to create and sustain flexibility. The application of flexible systems management to a system comprised of a human, artificial intelligence, and an intelligent system (H-T system) was evaluated in Section 5. It was concluded the FSM methodology applied to this modern manufacturing system allowed for the recognition of where flexibility within this system is formed and the formulation of an ORMS-based model of the human-technology system provided a method to quantify flexibility and identify the emergence of flexibility within a network of H-T systems. This research addressed this knowledge gap and the novel model formulation establishes a new method to measure these systems in terms of the costs for attaining flexibility and the costs of the training needed for such a system to achieve its planning objectives. This model and FSM methodology are unique tools that engineering management practitioners can use for generating and sustaining flexibility in complex engineering systems. As a result, we have a modern definition of flexibility applied to the general context of complex intelligent systems with a measure of the flexibility of that system. This modern definition states that flexibility is the ability and capacity of the H-AI interface to work with its intelligent system under a hierarchical coordination strategy to achieve human-technology system objectives in line with a production plan to meet the requirements of the customer in the most efficient and least costly manner using inherent and learned AI skill levels and machine capacities to diminish system disparities caused by uncertainties or deviations.

6.2. CONTRIBUTIONS

The first contribution of this research resulted in a baseline reference for engineering management practitioners to use in choosing flexibility methods for specific applications and to determine the scope and effectiveness of traditional flexibility methods. The analysis spanned the organizational research and management science (ORMS) area and included flexible working time, floaters, cross-training, teamwork, and temporary workers. This analysis is based under this traditional flexibility definition.

A second contribution of this flexibility research analyzed flexibility mechanisms of SoSs and, accordingly, identified needs for flexibility that SoSs can meet. The research showed that a hierarchical network is a more flexible SoS design for complex or distributed large-scale systems. A case that involved integrating distributed renewable energy sources with the main grid was presented to illustrate the implementation of the methodology. Another related result of this research is that the coordination of the constituent systems of a system of systems (SoS) in operations is an important task for functionalizing the SoS. The choice of a coordinating strategy needs to consider the autonomy, belonging, and connectivity levels of constituent systems. The diversity and emergence characteristics of SoS are outcomes of this coordination which is evidence of flexibility generation and sustainability. A recent paper by Zhao, et al. is based on the flexibility mechanisms and coordination strategies elucidated based on the SoS architecture in Sections 3 and 4.[198] These scholars and others in power systems are discussing their visions for system architecting which is closely related to the work completed in Sections 3 and 4.

In the new context of Industry 4.0 and the use of technologies such as machine learning, the labor capacities and capabilities are complex systems that can be full automation or hybrid systems that combine human capabilities (e.g. planning, review, and innovative thought) with technology that makes labor more efficient, e.g. robotics, cloud computing, remote sensing, etc. The final contributions of this research are: a) a new definition of flexibility for a human-technology work environment versus the traditional flexibility definition; b) the creation of a flexible systems management framework based on system of systems (SoS) principles that creates and sustains flexibility for complex engineering systems; and c) application of a hierarchical coordinating strategy for optimal workforce flexibility within the human-technology framework. This dissertation research resulted in the creation of a flexible systems management methodology and a mathematical model that provides managers of complex engineering systems a method for determining where flexibility emerges and what strategies a manager can use to manage and sustain flexibility for human-technology systems.

Finally, the work developed in Section 5 has been an important contribution to the system component of an awarded NSF study concerning collaborative research for a project called "Assistive Intelligence for Cooperative Robot and Inspector Survey of Infrastructure Systems.[199] This project is intended to transform inspection with a new integrated bridge inspection capability, a multi-university team will develop and implement a cooperative robot-inspector system with assistive intelligence (AI) in order to make future bridge inspections significantly faster, cheaper, safer, and more consistent. A robotic platform equipped with infrared cameras and a central processing unit with intelligent algorithms will operate in both flying and crawling modes, travel in proximity to various parts/elements of a bridge and collect high-fidelity images of the entire bridge. This NSF study's system is closely related to the H-T system described and developed in Section 5 and is influencing change in this industry.

6.3. FUTURE RESEARCH

From the research conducted in Section 4 it was found that the computational complexity in implementing multi-stage, multi-scale coordination grows quickly as the size of a system of systems network increases; how an outcome of coordination is reached and how it is evolving over time in use of fully or somewhat decentralized coordination strategies for flexible systems management are less understood; the use of smart technologies to improve the coordination effectiveness requires more thorough analysis and exploration. Addressing these technical needs, as well as others to be identified, will largely improve the effectiveness of coordinating constituent systems and in doing so the capabilities of the FSM approach. Given the mathematical model on flexibility in Section 5, an exploration of the multiple ways to solve the model for multiple networks of H-T systems or the optimization of the model for such systems is a suggested pathway for extending this research.

APPENDIX

COPYRIGHT INFORMATION

Copyright policy of Elsevier publisher is given below:

1. COPYRIGHT

Describes the rights related to the publication and distribution of research. It governs how authors (as well as their employers or funders), publishers and the wider general public can use, publish and distribute articles or books.

2. JOURNAL AUTHOR RIGHTS

In order for Elsevier to publish and disseminate research articles, we need publishing rights. This is determined by a publishing agreement between the author and Elsevier. This agreement deals with the transfer or license of the copyright to Elsevier and authors retain significant rights to use and share their own published articles. Elsevier supports the need for authors to share, disseminate and maximize the impact of their research and these rights, in Elsevier proprietary journals* are defined below:

For subscription articles
Authors transfer copyright to the publisher as part of a journal publishing agreement, but have the right to: • Share their article for Personal Use, Internal Institutional Use and Scholarly Sharing purposes, with a DOI link to the version of record on ScienceDirect (and with the Creative Commons CC-BY-NC-ND license for author manuscript versions) • Retain patent, trademark, and other intellectual property rights (including research data). • Proper attribution and credit for the published work.

For open access articles
Authors sign an exclusive license agreement, where authors have copyright but license exclusive rights in their article to the publisher**. In this case authors have the right to: • Share their article in the same ways permitted to third parties under the relevant user license (together with Personal Use rights) so long as it contains a CrossMark logo, the end user license, and a DOI link to the version of record on ScienceDirect. • Retain patent, trademark and other intellectual property rights (including research data). • Proper attribution and credit for the published work.

*Please note that society or third party owned journals may have different publishing agreements. Please see the journal's guide for authors for journal specific copyright information.

**This includes the right for the publisher to make and authorize commercial use, please see "Rights granted to Elsevier" for more details.

3. RIGHTS GRANTED TO ELSEVIER

For both subscription and open access articles, published in proprietary titles, Elsevier is granted the following rights:

• The exclusive right to publish and distribute an article, and to grant rights to others, including for commercial purposes.

• For open access articles, Elsevier will apply the relevant third party user license where Elsevier publishes the article on its online platforms.

• The right to provide the article in all forms and media so the article can be used on the latest technology even after publication.

• The authority to enforce the rights in the article, on behalf of an author, against third parties, for example in the case of plagiarism or copyright infringement.

Dissertation Section 2:

Reprinted from *Surveys in Operations Research and Management Science*, Volume 20 (**1**), Qin, R., Nembhard, D.A., and Barnes II, W.L., "Workforce flexibility in operations management.", pp. 19-33, 2015, with permission from Elsevier.

Dissertation Section 3:

Reprinted from *Procedia Computer Science*, Volume 114, Barnes II, W.L., Qin, R., and Dagli, C.H. "Enabling flexibility through forming an evolving system of systems", pp. 19-27, 2017, with permission from Elsevier.

Dissertation Section 4:

Reprinted from *Procedia Computer Science*, Volume 114, Barnes II, W.L., Calvin, M., Linville, K., Song, Y., Xu, R., Zhao, B. and Qin, R., "Coordination of Constituent Systems for Functionalizing Systems of Systems: An Exploration", pp. 375-383, 2017, with permission from Elsevier.

BIBLIOGRAPHY

[1] W.J. Hopp, M.P. Van Oyen, Agile workforce evaluation: a framework for cross-training and coordination, IIE Trans. 36 (2004) 919–940.

[2] E.S. Gel, Stochastic models of workforce agility in production systems (Ph.D. thesis), Northwestern University, 1999.

[3] E. Tekin, Performance opportunity for workforce agility (Ph.D. thesis), Northwestern University, 2003.

[4] M.P. Van Oyen, E.S. Gel, W.J. Hopp, Performance opportunity for workforce agility in collaborative and noncollaborative work systems, IIE Trans. 33 (2001) 761–777.

[5] S.M.R. Iravani, V. Krishnamurthy, Workforce agility in repair and maintenance environments, Manuf. Serv. Oper. Manage. 9 (2007) 168–184.

[6] Y.Y. Yusuf, M. Sarhadi, A. Gunasekaran, Agile manufacturing: The drivers, concepts and attributes, Int. J. Prod. Econ. 62 (1999) 33–43.

[7] R. Qin, D.A. Nembhard, Workforce agility in operations management, Technical Report, Missouri University of Science and Technology, Pennsylvania State University, 2014.

[8] A. Martel, A. Al-Nuaimi, Tactical manpower planning via programming under uncertainty, Oper. Res. Q. 24 (1973) 571–585.

[9] J.C. Goodwin Jr., D. Elvers, J.S. Goodwin, Overtime usage in a job shop environment, Omega 6 (1978) 493–500.

[10] G.D. Scudder, An evaluation of overtime policies for a repair shop, J. Oper. Manage. 6 (1985) 87–98.

[11] G.D. Scudder, R. Chua, Determining overtime policies for a repair shop, Omega 15 (1987) 197–206.

[12] C. Akkan, Overtime scheduling: an application in finite-capacity real-time scheduling, J. Oper. Res. Soc. 47 (1996) 1137–1149.

[13] C.A. Holloway, R.T. Nelson, Job shop scheduling with due dates and overtime capability, Manage. Sci. 21 (1974) 68–78.

[14] L. Gelders, P.R. Kleindorfer, Coordinating aggregate and detailed scheduling decisions in the one-machine job shop: Part I., Oper. Res. 22 (1974) 46–60.

[15] L. Gelders, P.R. Kleindorfer, Coordinating aggregate and detailed scheduling decisions in the one-machine job shop: Part II. Computation and structure, Oper. Res. 23 (1975) 312–324.

[16] K.-K. Yang, S. Webster, R.A. Ruben, An evaluation of flexible workday policies in job shops*, Decis. Sci. 33 (2002) 223–250.

[17] K.-K. Yang, S. Webster, R.A. Ruben, An evaluation of worker cross training and flexible workdays in job shops, IIE Trans. 39 (2007) 735–746.

[18] M.R. Galbreth, P.R. Philipoom, M.K. Malhotra, Planning with uncertain materials availability: The value of workday flexibility, Oper. Manage. Res. 5 (2012) 91–100.

[19] K.-K. Yang, S. Webster, R.A. Ruben, Exploratory analysis of flexible workday policies in a job shop, Int. J. Prod. Res. 41 (2003) 1721–1737.

[20] R. Hung, Annualized hours and aggregate planning, Prod. Inventory Manage. J. 38 (1997) 48–50.

[21] A. Corominas, A. Lusa, R. Pastor, Characteristics and classification of the annualised working hours planning problems, Int. J. Serv. Technol. Manage. 5 (2004) 435–447.

[22] R. Hung, Scheduling a workforce under annualized hours, Int. J. Prod. Res. 37 (1999) 2419–2427.

[23] R. Hung, A multiple-shift workforce scheduling model under annualized hours, Naval Res. Logist. (NRL) 46 (1999) 726–736.

[24] C.S. Azmat, M. Widmer, A case study of single shift planning and scheduling under annualized hours: A simple three-step approach, European J. Oper. Res. 153 (2004) 148–175.

[25] A. Corominas, A. Lusa, R. Pastor, Using MILP to plan annualised working hours, J. Oper. Res. Soc. (2002) 1101–1108.

[26] C.S. Azmat, T. Hürlimann, M. Widmer, Mixed integer programming to schedule a single-shift workforce under annualized hours, Ann. Oper. Res. 128 (2004) 199–215.

[27] A. Corominas, A. Lusa, R. Pastor, Planning annualised hours with a finite set of weekly working hours and joint holidays, Ann. Oper. Res. 128 (2004) 217–233.

[28] A. Lusa, R. Pastor, A. Corominas, Determining the most appropriate set of weekly working hours for planning annualised working time, Int. J. Prod. Econ. 111 (2008) 697–706.

[29] A. Corominas, A. Lusa, R. Pastor, Planning annualised hours with a finite set of weekly working hours and cross-trained workers, European J. Oper. Res. 176 (2007) 230–239.

[30] A. Lusa, A. Corominas, R. Pastor, An exact procedure for planning holidays and working time under annualized hours considering cross-trained workers with different efficiencies, Int. J. Prod. Res. 46 (2008) 2123–2142.

[31] A. Corominas, A. Lusa, R. Pastor, Planning production and working time within an annualised hours scheme framework, Ann. Oper. Res. 155 (2007) 5–23.

[32] A. Hertz, N. Lahrichi, M. Widmer, A flexible MILP model for multiple-shift workforce planning under annualized hours, European J. Oper. Res. 200 (2010) 860–873.

[33] A. Lusa, A. Corominas, N. Muñoz, A multistage scenario optimisation procedure to plan annualised working hours under demand uncertainty, Int. J. Prod. Econ. 113 (2008) 957–968.

[34] A. Corominas, R. Pastor, Replanning working time under annualised working hours, Int. J. Prod. Res. 48 (2010) 1493–1515.

[35] A. Corominas, A. Lusa, R. Pastor, Using a MILP model to establish a framework for an annualised hours agreement, European J. Oper. Res. 177 (2007) 1495–1506.

[36] A. Lusa, A. Corominas, J. Olivella, R. Pastor, Production planning under a working time accounts scheme, Int. J. Prod. Res. 47 (2009) 3435–3451.

[37] A. Herzog-Stein, I. Zapf, Navigating the great recession the impact of working-time accounts in Germany, Ind. Labor Relat. Rev. 67 (2014) 891–925.

[38] A. Lusa, R. Pastor, Planning working time accounts under demand uncer- tainty, Comput. Oper. Res. 38 (2011) 517–524.

[39] A. Corominas, J. Olivella, R. Pastor, Capacity planning with working time accounts in services, J. Oper. Res. Soc. 61 (2010) 321–331.

[40] T. Sillekens, A. Koberstein, L. Suhl, Aggregate production planning in the automotive industry with special consideration of workforce flexibility, Int. J. Prod. Res. 49 (2011) 5055–5078.

[41] A. Corominas, A. Lusa, J. Olivella, A detailed workforce planning model including non-linear dependence of capacity on the size of the staff and cash management, European J. Oper. Res. 216 (2012) 445–458.

[42] A. Corominas, A. Lusa, R. Pastor, Human resource management using working time accounts with expiry of hours, J. Ind. Manag. Optim. 5 (2009) 569–584.

[43] B. Wild, C. Schneewei, Manpower capacity planning—a hierarchical approach, Int. J. Prod. Econ. 30–31 (1993) 95–106.

[44] D. Francas, N. Löhndorf, S. Minner, Machine and labor flexibility in manufacturing networks, Int. J. Prod. Econ. 131 (2011) 165–174.

[45] L.I. Sennott, M.P. Van Oyen, S.M.R. Iravani, Optimal dynamic assignment of a flexible worker on an open production line with specialists, European J. Oper. Res. 170 (2006) 541–566.

[46] D.G. Pandelis, Optimal control of flexible servers in two tandem queues with operating costs, Probab. Engrg. Inform. Sci. 22 (2008) 107–131.

[47] T.M. Farrar, Optimal use of an extra server in a two station tandem queueing network, IEEE Trans. Automat. Control 38 (1993) 1296–1299.

[48] H.-S. Ahn, I. Duenyas, R.Q. Zhang, Optimal control of a flexible server, Adv. Appl. Probab. 36 (2004) 139–170.

[49] D.G. Pandelis, Optimal use of excess capacity in two interconnected queues, Math. Methods Oper. Res. 65 (2007) 179–192.

[50] C.-H. Wu, D.G. Down, M.E. Lewis, Heuristics for allocation of reconfigurable resources in a serial line with reliability considerations, IIE Trans. 40 (2008), 595–611.

[51] E. Cevikcan, M.B. Durmusoglu, Minimising utility work and utility worker transfers for a mixed-model assembly line, Int. J. Prod. Res. 49 (2011) 7293–7314.

[52] N. Boysen, M. Kiel, A. Scholl, Sequencing mixed-model assembly lines to minimise the number of work overload situations, Int. J. Prod. Res. 49 (2011) 4735–4760.

[53] S.M. Iravani, M.P. Van Oyen, K.T. Sims, Structural flexibility: A new perspective on the design of manufacturing and service operations, Manage. Sci. 51 (2005) 151–166.

[54] D.A. Nembhard, Workforce Cross Training, CRC Press, Boca Raton, FL, 2007.

[55] M. Treleven, A review of the dual resource constrained system research, IIE Trans. 21 (1989) 279–287.

[56] M.P. Hottenstein, S.A. Bowman, Cross-training and worker flexibility: A review of DRC system research, J. High Technol. Manage. Res. 9 (1998) 157–174.

[57] O.Z. Aksin, F. Karaesmen, E.L. Örmeci, Workforce cross-training in call centers from an operations management perspective, in: D.A. Nembhard (Ed.), Workforce Cross Training, CRC Press, Boca Raton, FL, 2007, pp. 211–239.

[58] A.C. Ebeling, C.-Y. Lee, Cross-training effectiveness and profitability, Int. J. Prod. Res. 32 (1994) 2843–2859.

[59] R.R. Inman, W.C. Jordan, D.E. Blumenfeld, Chained cross-training of assembly line workers, Int. J. Prod. Res. 42 (2004) 1899–1910.

[60] H.V. Kher, M.K. Malhotra, Acquiring and operationalizing worker flexibility in dual resource constrained job shops with worker transfer delays and learning losses, Omega 22 (1994) 521–533.

[61] J. Slomp, E. Molleman, Cross-training policies and team performance, Int. J. Prod. Res. 40 (2002) 1193–1219.

[62] J.A.C. Bokhorst, J. Slomp, E. Molleman, Development and evaluation of cross-training policies for manufacturing teams, IIE Trans. 36 (2004) 969–984.

[63] D.A. Nembhard, H.B. Nembhard, R. Qin, A real options model for workforce cross-training, Eng. Econ. 50 (2005) 95–116.

[64] K.K. Yang, A comparison of cross-training policies in different job shops, Int. J. Prod. Res. 45 (2007) 1279–1295.

[65] P.S. Park, The examination of worker cross-training in a dual resource constrained job shop, European J. Oper. Res. 52 (1991) 291–299.

[66] B.D. Stewart, D.B. Webster, S. Ahmad, J.O. Matson, Mathematical models for developing a flexible workforce, Int. J. Prod. Econ. 36 (1994) 243–254.

[67] J.F. Bard, Staff scheduling in high volume service facilities with downgrading, IIE Trans. 36 (2004) 985–997.

[68] O. Berman, R.C. Larson, A queueing control model for retail services having back room operations and cross-trained workers, Comput. Oper. Res. 31 (2004) 201–222.

[69] W.J. Hopp, E. Tekin, M.P. Van Oyen, Benefits of skill chaining in serial production lines with cross-trained workers, Manage. Sci. 50 (2004) 83–98.

[70] O. Berman, K.P. Sapna-Isotupa, Optimal control of servers in front and back rooms with correlated work, IIE Trans. 37 (2005) 167–173.

[71] O. Berman, J. Wang, K.P. Sapna-Isotupa, Optimal management of cross-trained workers in services with negligible switching costs, European J. Oper. Res. 167 (2005) 349–369.

[72] E.S. Gel, W.J. Hopp, M.P. Van Oyen, Hierarchical cross-training in work-in-process-constrained systems, IIE Trans. 39 (2007) 125–143.

[73] D.A. Nembhard, B.A. Norman, Cross training in production systems with human learning and forgetting, in: D. Nembhard (Ed.), Workforce Cross Training, CRC Press, 2007, pp. 111–130.

[74] O. Berman, E. Ianovsky, Optimal management of cross-trained workers, using Markov decision approach, Int. J. Oper. Res. 3 (2008) 154–182.

[75] M.E. Mayorga, K.M. Taaffe, R. Arumugam, Allocating flexible servers in serial systems with switching costs, Ann. Oper. Res. 172 (2009) 231–242.

[76] H. Parvin, M.P. Van Oyen, D.G. Pandelis, D.P. Williams, J. Lee, Fixed task zone chaining: worker coordination and zone design for inexpensive cross-training in serial conwip lines, IIE Trans. 44 (2012) 894–914.

[77] N. Azizi, M. Liang, An integrated approach to worker assignment, workforce flexibility acquisition, and task rotation, J. Oper. Res. Soc. 64 (2013) 260–275.

[78] S. Kim, D.A. Nembhard, Rule mining for scheduling cross training with a heterogeneous workforce, Int. J. Prod. Res. 51 (2013) 2281–2300.

[79] M.K. Tiwari, D. Roy, Application of an evolutionary fuzzy system for the estimation of workforce deployment and cross-training in an assembly environment, Int. J. Prod. Res. 40 (2002) 4651–4674.

[80] R.L. Daniels, J.B. Mazzola, D. Shi, Flow shop scheduling with partial resource flexibility, Manage. Sci. 50 (2004) 658–669.

[81] J.A.C. Bokhorst, The impact of the amount of work in process on the use of cross-training, Int. J. Prod. Res. 49 (2011) 3171–3190.

[82] D.A. Nembhard, K. Prichanont, Cross training in serial production with process characteristics and operational factors, IEEE Trans. Eng. Manage. 54 (2007) 565–575.

[83] E.J. Pinker, R.A. Shumsky, The efficiency-quality trade-off of cross-trained workers, Manuf. Serv. Oper. Manage. 2 (2000) 32–48.

[84] S.R. Agnihothri, A.K. Mishra, Cross-training decisions in field services with three job types and server–job mismatch*, Decis. Sci. 35 (2004) 239–257.

[85] R.R. Inman, D.E. Blumenfeld, A. Ko, Cross-training hospital nurses to reduce staffing costs, Health Care Manage. Rev. 30 (2005) 116–125.

[86] J. Slomp, J.A.C. Bokhorst, E. Molleman, Cross-training in a cellular manufacturing environment, Comput. Ind. Eng. 48 (2005) 609–624.

[87] M.J. Brusco, An exact algorithm for a workforce allocation problem with application to an analysis of cross-training policies, IIE Trans. 40 (2008) 495–508.

[88] F.F. Easton, Cross-training performance in flexible labor scheduling environments, IIE Trans. 43 (2011) 589–603.

[89] M.J. Brusco, T.R. Johns, Staffing a multiskilled workforce with varying levels of productivity: An analysis of cross-training policies*, Decis. Sci. 29 (1998) 499–515.

[90] G.M. Campbell, Cross-utilization of workers whose capabilities differ, Manage. Sci. 45 (1999) 722–732.

[91] L.L.X. Li, B.E. King, A healthcare staff decision model considering the effects of staff cross-training, Health Care Manage. Sci. 2 (1999) 53–61.

[92] G. Vairaktarakis, J.K. Winch, Worker cross-training in paced assembly lines, Manuf. Serv. Oper. Manage. 1 (1999) 112–131.

[93] G.M. Campbell, M. Diaby, Development and evaluation of an assignment heuristic for allocating cross-trained workers, European J. Oper. Res. 138 (2002) 9–20.

[94] S.R. Agnihothri, A.K. Mishra, D.E. Simmons, Workforce cross-training decisions in field service systems with two job types, J. Oper. Res. Soc. 54 (2003) 410–418.

[95] W.C. Jordan, R.R. Inman, D.E. Blumenfeld, Chained cross-training of workers for robust performance, IIE Trans. 36 (2004) 953–967.

[96] U. Kula, I. Duenyas, S.M.R. Iravani, Estimating job waiting times in production systems with cross-trained setup crews, IIE Trans. 36 (2004) 999–1010.

[97] S.M.R. Iravani, B. Kolfal, M.P. Van Oyen, Call-center labor cross-training: it's a small world after all, Manage. Sci. 53 (2007) 1102–1112.

[98] S. Sayın, S. Karabatı, Assigning cross-trained workers to departments: A two-stage optimization model to maximize utility and skill improvement, European J. Oper. Res. 176 (2007) 1643–1658.

[99] J.K. Winch, X. Cai, G.L. Vairaktarakis, Cyclic job scheduling in paced assembly lines with cross-trained workers, Int. J. Prod. Res. 45 (2007) 803–828.

[100] J. Chung, K.P. White, Cross-trained versus specialized agents in an inbound call centre: a simulation-based methodology for trade-off analysis, J. Simul. 2 (2008) 162–169.

[101] J.W. Fowler, P. Wirojanagud, E.S. Gel, Heuristics for workforce planning with worker differences, European J. Oper. Res. 190 (2008) 724–740.

[102] M. Ahghari, B. Balciogˆlu, Benefits of cross-training in a skill-based routing contact center with priority queues and impatient customers, IIE Trans. 41 (2009) 524–536.

[103] J.A.C. Bokhorst, G.J.C. Gaalman, Cross-training workers in dual resource constrained systems with heterogeneous processing times, Int. J. Prod. Res. 47 (2009) 6333–6356.

[104] E. Tekin, W.J. Hopp, M.P. Van Oyen, Pooling strategies for call center agent cross-training, IIE Trans. 41 (2009) 546–561.

[105] T.R. Robbins, D.J. Medeiros, T.P. Harrison, Cross training in call centers with uncertain arrivals and global service level agreements, Int. J. Oper. Quant. Manage. 16 (2010) 307–329.

[106] G.M. Campbell, A two-stage stochastic program for scheduling and allocating cross-trained workers, J. Oper. Res. Soc. 62 (2011) 1038–1047.

[107] P.J. Colen, M.R. Lambrecht, Cross-training policies in field services, Int. J. Prod. Econ. 138 (2012) 76–88.

[108] K.A. Marentette, A.W. Johnson, L. Mills, A measure of cross-training benefit versus job skill specialization, Comput. Ind. Eng. 57 (2009) 937–940.

[109] S.M. Shafer, D.A. Nembhard, M.V. Uzumeri, The effects of worker learning, forgetting, and heterogeneity on assembly line productivity, Manage. Sci. 47 (2001) 1639–1653.

[110] D.A. Nembhard, N. Osothsilp, Task complexity effects on between-individual learning/forgetting variability, Int. J. Ind. Ergon. 29 (2002) 297–306.

[111] Z. Aksin, M. Armony, V. Mehrotra, The modern call center: A multi- disciplinary perspective on operations management research, Prod. Oper. Manage. 16 (2007) 665–688.

[112] E. Salas, T.L. Dickinson, S.A. Converse, S.I. Tannenbaum, R.W. Swezey, E. Salas, Toward an understanding of team performance and training, in: Teams: their Training and Performance, Ablex Publishing, 1992, pp. 3–29.

[113] C.A. Yauch, Team-based work and work system balance in the context of agile manufacturing, Appl. Ergon. 38 (2007) 19–27.

[114] J.A. Buzacott, Commonalities in reengineered business processes: models and issues, Manage. Sci. 42 (1996) 768–782.

[115] A. Mandelbaum, M.I. Reiman, On pooling in queueing networks, Manage. Sci. 44 (1998) 971–981.

[116] J. Kim, R.M. Burton, The effect of task uncertainty and decentralization on project team performance, Comput. Math. Organ. Theory 8 (2002) 365–384.

[117] K. Sengupta, F. Jacobs, Impact of work teams: a comparison study of assembly cells and assembly line for a variety of operating environments, Int. J. Prod. Res. 42 (2004) 4173–4193.

[118] S. Andradóttir, H. Ayhan, D.G. Down, Server assignment policies for maximizing the steady-state throughput of finite queueing systems, Manage. Sci. 47 (2001) 1421–1439.

[119] E. Molleman, J. Slomp, Functional flexibility and team performance, Int. J. Prod. Res. 37 (1999) 1837–1858.

[120] S.G. Powell, Specialization, teamwork, and production efficiency, Int. J. Prod. Econ. 67 (2000) 205–218.

[121] R.R. Inman, D.E. Blumenfeld, Assembly line team sizing with absenteeism, Int. J. Prod. Res. 48 (2010) 6537–6558.

[122] S.K. Bordoloi, E.J. Weatherby, Managerial implications of calculating optimum nurse staffing in medical units, Health Care Manage. Rev. 24 (1999) 35–44.

[123] L. Perron, Planning and scheduling teams of skilled workers, J. Intell. Manuf. 21 (2010) 155–164.

[124] Ryan, E.T., Jacques D.R., and Colombi J.M. (2013) "An ontological framework for clarifying flexibility-related terminology via literature survey. *Systems Engineering* **16** (**1**): 99-110.

[125] Saleh, J.H., Mark, G., and Jordan N.C. (2009) "Flexibility: a multi-disciplinary literature review and a research agenda for designing flexible engineering systems." *Journal of Engineering Design* **20** (**3**): 307-23.

[126] Gupta, Y.P. and Goyal, S. (1989) "Flexibility of manufacturing systems: concepts and measurements." *European Journal of Operational Research* **43** (**2**): 119-35.

[127] Gerwin, D. (1993) "Manufacturing flexibility: a strategic perspective." *Management Science* **39** (**4**): 395-410.

[128] Hopp, W.J. and Oyen, M.P. (2004) "Agile workforce evaluation: a framework for cross-training and coordination." *IIE Transactions* **36** (**10**): 919-40.

[129] Qin, R., Nembhard, D.A., and Barnes II, W.L. (2015) "Workforce flexibility in operations management." *Surveys in Operations Research and Management Science* **20** (**1**): 19-33.

[130] Ackoff, R.L. (1971) "Towards a system of systems concepts." *Management Science* **17** (**11**): 661-71.

[131] OUSD AT&L, (2008) *Systems Engineering Guide for Systems of Systems*, Pentagon, Washington, DC.

[132] Maier, M.W. (1996) "Architecting principles for systems-of-systems." In *INCOSE International Symposium* **6** (**1**): 565-573.

[133] Jamshidi, M. (ed.) (2008) *Systems of Systems Engineering: Principles and Applications*. CRC press.

[134] Boardman, J. and Sauser, B. (2006) "System of systems-the meaning of of." In *System of Systems Engineering, IEEE/SMC International Conference on* 2006 Apr 24 (pp. 6-10). IEEE.

[135] Baldwin, W.C. and Sauser, B. (2009) "Modeling the characteristics of system of systems." In *System of Systems Engineering, SoSE 2009. IEEE International Conference on* 2009 May 30 (pp. 1-6). IEEE.

[136] Ross, A.M., Rhodes, D.H., and Hastings, D.E. (2008) "Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value." *Systems Engineering* **11** (**3**): 246-62.

[137] Nilchiani, R. and Hastings, D.E. (2007) "Measuring the Value of flexibility in space systems: A six-element framework." *Systems Engineering* **10 (1)**: 26-44.

[138] Suh, N.P. (1998) "Axiomatic design theory for systems." *Research in Engineering Design* **10 (4)**: 189-209.

[139] Shah, N.B., Viscito, L., Wilds, J., Ross, A.M., and Hastings, D.E. (2008) "Quantifying flexibility for architecting changeable systems." In *6th Conference on Systems Engineering Research*, Los Angeles, CA 2008 Apr.

[140] Cardin, M.A., Neufville, R., and Geltner, D.M. (2015) "Design catalogs: A systematic Approach to design and value flexibility in engineering systems." *Systems Engineering* **18 (5)**: 453-71.

[141] Konur, D., Farhangi, H., and Dagli, C.H. (2016) "A multi-objective system of systems architecting problem with inflexible and flexible systems: formulation and solution methods." *OR Spectrum*. **38 (4)**: 967-1006.

[142] Moses, J. (2002) "The anatomy of large scale systems." In *ESD Internal Symposium*. Citeseer.

[143] Moses, J. (2010) "Flexibility and its relation to complexity and architecture." In *Complex Systems Design & Management,* pp. 197-206. Springer Berlin Heidelberg.

[144] Broniatowski, D.A. and Moses, J. (2016) "Measuring flexibility, descriptive complexity, and rework potential in generic System Architectures." *Systems Engineering* **19 (3)**: 207-21.

[145] Eckert, C., Clarkson, P.J., and Zanker, W. (2004) "Change and customisation in complex engineering domains." *Research in Engineering Design* **15 (1)**: 1-21.

[146] Gorod, A., Gandhi, S.J., Sauser, B., and Boardman, J. (2008) "Flexibility of system of systems." *Global Journal of Flexible Systems Management* **9 (4)**: 21.

[147] Dagli, C.H., Enke, D., Kilicay-Ergin, N., Konur, D., Qin, R., Gosavi, A., Wang, R., Pape, L., Agarwal, S., and Gottapu, R.D. *Technical Report SERC-2015-RT-021-4: Flexible and intelligent learning architectures for SoS (FILA-SoS).* Systems Engineering Research Center, Tech. Rep., 2015, volumes 1-16.

[148] Dahmann, J.S. and Baldwin, K.J. "Understanding the current state of US defense systems of systems and the implications for systems engineering." In *Systems Conference, 2008 2nd Annual IEEE* 2008 Apr 7 (pp. 1-7). IEEE.

[149] Konur, D., Farhangi, H., Dagli, C.H. "On the Flexibility of Systems in System of Systems Architecting." *Procedia Computer Science 36*, *Proceedings of the 2014 Complex Adaptive Systems Conference*, pp. 65-71.

[150] Fares, R. "Renewable Energy Intermittency Explained: Challenges, Solutions, and Opportunities." *Scientific American*, March 11, 2015, blogs.scientificamerican.com/plugged-in/renewable-energy-intermittency-explained-challenges-solutions-and-opportunities/.

[151] Solanki, A., Montoya, L.F., Fu, Q., Nasiri, A., Bhavaraju, V., Abdallah, T., and Yu, D. "Managing intermittent renewables in a microgrid." In *Power & Energy Society (PES) Conference on Innovative Smart Grid Technologies (ISGT), 2012 3rd Annual IEEE* 2012 Jan 16-20. IEEE.

[152] Pérez, J., Díaz, J., Garbajosa, J., Yagüe, A., Gonzalez, E., Lopez-Perea, M. "Large-scale smart grids as system of systems." In *Proceedings of the First International Workshop on Software Engineering for Systems-of-Systems* 2013 Jul 2 (pp. 38-42). ACM.

[153] Mahmoud, M.S., Rahman, M.S., Fouad, M.S. (2015) "Review of microgrid architectures - a system of systems perspective." *IET Renewable Power Generation* **9** (**8**): 1064-78.

[154] Thanos, A.E., Moore, D.E., Shi, X., Celik, N. (2015) "System of systems modeling and simulation for microgrids using DDDAMS." In Rainey, L.B. and Tolk, A. (eds) *Modeling and Simulation Support for System of Systems Engineering Applications*. New Jersey, Wiley; pp. 337-360.

[155] Zamora, R. and Srivastava, A.K. (2010) "Controls for microgrids with storage: Review, challenges, and research needs." *Renewable and Sustainable Energy Reviews* **14** (**7**): 2009-18.

[156] Howard, M.W. VII. *The Integrated Grid: Realizing the Full Value of Central and Distributed Energy Resources*. The ICER Chronicle Edition 2, July 2014: 1:32.

[157] Zhao, B., Yang, Y., Zhang, X., Xue, M., Li, P., Xu, C., and Zhou, D. (2016) "Implementation of a dual-microgrid system with flexible configurations and hierarchical control in China." *Renewable and Sustainable Energy Reviews* **65**: 113-23.

[158] Boardman, J. and Sauser, B. "System of systems – the meaning of 'of'." In *Proceedings of the 2006 IEEE/SMC International Conference on System of Systems Engineering*, Los Angeles, CA, pp. 118-123. IEEE (April 2006).

[159] Fox-Penner, P.S. (2014) Smart power anniversary edition: Climate change, the smart grid, and the future of electric utilities. Island Press.

[160] Lucke, D.L., Constantinescu, C., and Westkämper, E. (2008) “Smart factory - a step towards the next generation of manufacturing.” *Manufacturing systems and technologies for the new frontier* pp. 115-118.

[161] Barnes II, W.L., Qin, R., and Dagli, C.H. (2017) “Enabling flexibility through forming an evolving system of systems.” *2017 Complex Adaptive Systems Conference*, Chicago, IL, October 2017.

[162] Henshaw, M. (2016) “System of systems, cyber-physical systems, the internet of things… whatever next.” *Insight* **19** (**3**): 51-54.

[163] DiMario, M.J., Boardman, J., and Sauser, B. (2009) “System of systems collaborative formation.” *IEEE Systems Journal* **3** (**3**): 360-368.

[164] Gorod, A., DiMario, M.J., Sauser, B., and Boardman, J. (2009) “‘Satisficing’ system of systems using dynamic and static doctrines.” *International Journal of System of Systems Engineering* **1** (**3**): 347-366.

[165] Baldwin, W.C. and Sauser, B. (2009) “Modeling the characteristics of system of systems.” In: *IEEE International Conference on System of Systems Engineering*, Albuquerque, NM, pp. 1-6. IEEE.

[166] Baldwin, W.C., Ben-Zvi, T., and Sauser, B. (2009) “Formation of collaborative system of systems through belonging choice mechanisms.” *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* **42**: 793-801.

[167] Maier, M.W. (1998) “Architecting principles for systems-of-systems.” *Systems Engineering* **1** (**4**): 267-284.

[168] Dahmann, J.S. and Baldwin, K.J. (2008) “Understanding the current state of US defense systems of systems and the implications for systems engineering.” In: *2nd Annual IEEE Systems Conference*, Montreal, Canada, pp. 1-7. IEEE.

[169] Zhao, B., Yang, Y., Zhang, X., Xue, M., Li, P., Xu, C., and Zhou, D. (2016) “Implementation of a dual-microgrid system with flexible configurations and hierarchical control in China.” *Renewable and Sustainable Energy Review* **65**: 113-123.

[170] Zhao, B., Chen, J., Zhang, X., and Qin, R. (2017) “Three representative island microgrids in the East China Sea: Key technology and experiences.” Submitted to *Renewable and Sustainable Energy Review*.

[171] National Science Foundation, Directorate of Engineering (2019, November). U.S. Census Bureau. Bureau of Economic Analysis. (2008, August). *Future of Work at the Human-Technology Frontier: Core Research (FW-HTF)* (Program Solicitation NSF 19-541). Retrieved from https://www.nsf.gov/pubs/2019/nsf19541/nsf19541.htm.

[172] Wang, G.E. "Humans in the Loop: The Design of Interactive AI Systems". *Human-Centered Artificial Intelligence, Stanford University*. October 20, 2019. Retrieved from https://hai.stanford.edu/blog/humans-loop-design-interactive-ai-systems.

[173] Fails, J. A., and Olsen Jr, D.R. (2003). Interactive Machine Learning. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, 39–45. New York: Association for Computing Machinery.

[174] Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*, 35(4), 105-120.

[175] Backlund, A. (2000), "The definition of system", *Kybernetes*, Vol. 29 No. 4, pp. 444-451. https://doi.org/10.1108/03684920010322055

[176] Harnad, Stevan. (2008). The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence. In: Epstein, R., Roberts, G. and Beber, G. (eds.) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Evolving Consciousness.* Springer, pp. 23-66.

[177] Krohn, J., Bassens, A., Beyleveld, G. (2019). Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence. (n.p.): Pearson Education, pp. 86 and 53, respectively.

[178] Walch, K. "The Seven Pattterns of AI". *Cognilytica Research and Insight*. April 4, 2019. Retrieved from https://www.forbes.com/sites/cognitiveworld/2019/09/17/the-seven-patterns-of-ai/#53594df12d01.

[179] CB Insights, Research Briefs. "AI 100: The Artificial Intelligence Startups Redefining Industries". *CB Insights*. March 3, 2020. Retrieved from https://www.cbinsights.com/research/artificial-intelligence-top-startups/.

[180] Russell, S. J. and Norvig, P. (2003). Artificial Intelligence: A Modern Approach (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall, p. 32.

[181] Saleh, J.H., Mark, G., and Jordan N.C. (2009). Flexibility: a multi-disciplinary literature review and a research agenda for designing flexible engineering systems. *Journal of Engineering Design, 20*, 3, 307-23.

[182] Gupta, Y.P. and Goyal, S. (1989). Flexibility of manufacturing systems: concepts and measurements. *European Journal of Operational Research, 43,* 2, 119-35.

[183] Gerwin, D. (1993). Manufacturing flexibility: a strategic perspective. *Management Science*, *39*, 4, 395-410.

[184] Hopp, W.J. and Oyen, M.P. (2004). Agile workforce evaluation: a framework for cross-training and coordination. *IIE Transactions, 36*, 10, 919-40.

[185] Sauser, B. and Boardman, J. (2008). Taking Hold of System of Systems Management. *Engineering Management Journal*, *20,* 4, 3-9.

[186] Garrett, R.K., Anderson, S., Baron, N.T. and Moreland, J.D. (2010). Managing the Interstitials, a System of Systems Framework Suited for the Ballistic Missile Defense System. *Systems Engineering*, *14,* 1, 87-109.

[187] Gorod, A., Sauser, B., and Boardman, J. (2008). System-of- Systems Engineering Management: A Review of Modern History and a Path Forward. *IEEE Systems Journal*, *2,* 4, 484-499.

[188] Dimario, M. J., Boardman, J.T., and Sauser, B.J. (2009). System of Systems Collaborative Formation. *IEEE Systems Journal*, *3*, 3, 360-368.

[189] Qin, R. (2018). Missouri University of Science & Technology, Engineering Management & Systems Engineering Department, Systems Analytics Laboratory.

[190] Kumar, V. (1987). Entropic measures of manufacturing flexibility. *Int. J. Prod. Res.*, 25(7), 957-966.

[191] Brill, P.H and Mandelbaum, M. (1989). On measures of flexibility in manufacturing systems, *Int. J. Prod. Res.*, 27(5), 747-756.

[192] Primrose, P.L. and Leonard, R. (1984). Conditions under which flexible manufacturing is financially viable. In *Proceedings 3rd International Conference on Flexible Manufacturing Systems*. Boeblingen, Germany, September 11-13, IFS Publications Ltd., Bedford, U.K. 121-132.

[193] Son, Y.K. and Park, C.S. (1990). Quantifying opportunity costs associated with adding manufacturing flexibility. *Int. J. Prod. Res.*, 28(7), 1183-1194.

[194] Gupta, Y.P. and Goyal, S. (1989). Flexibility of manufacturing systems: Concepts and measurements. *Eur. J. Oper. Res.*, 43(2), 119-135.

[195] Stewart, B.D., Wester, D.B, Ahmad, S., and Matson, J.O. (1994). Mathematical models for developing a flexible workforce. *Int. J. Production Economics*, 36, 243-54.

[196] Chryssolouris, G. and Lee, M. (1992). An assessment of flexibility in manufacturing systems. *Manufac. Rev.*, 5(2), 105-116.

[197] Dahmann, J., Rebovich, G., Lowry, R., Lane, J., & Baldwin, K. (2011). An Implementers' View of Systems Engineering for Systems of Systems. In *Systems Conference (SysCon), 2011 IEEE International*, 212-217. IEEE.

[198] Zhao, B., Wang, X., Lin, D., Calvin, M., Morgan, J., Qin, R., and Wang, C. (2018). Energy Management of Multiple Microgrids Based on a System of Systems Architecture. *IEEE Transactions on Power Systems*, *33*, 6, 6410-21.

[199] National Science Foundation, Directorate of Engineering, Division of Electrical, Communications, and Cyber Systems (2020, August). *Assistive Intelligence for Cooperative Robot and Inspector Survey of Infrastructure Systems (AI-CRISIS) (FW-HTF-RM)* (Award Number 2026357). Retrieved from https://www.nsf.gov/awardsearch/showAward?AWD_ID=2026357&HistoricalAwards=false.

VITA

Walter L. Barnes II was born in New Albany, Mississippi. He received his Bachelor of Science in Chemical Engineering in 1992 from Florida A&M University (FAMU). He worked for the Procter & Gamble Company for five summers. After undergraduate school, he worked at the 3M Company as a Manufacturing Process Engineer. While working at 3M, he earned a Master of Science degree in Engineering Management in 1997 from the University of Missouri-Rolla. After several years at 3M, Walter completed the Master of Business Administration (MBA) degree, a Juris Doctor (law) degree, and a Certificate in Intellectual Property & E-Commerce in 2002 and 2005, respectively, at the University of Missouri-Columbia. During this time, he worked at Arthur Andersen LLP and the Monsanto Company on summer internships. Walter was hired full-time as a Manager, Strategic Analysis at the Monsanto Company working in the Finance organization in 2005. He worked in the Corporate Strategy and Technology Finance groups. In 2013, Walter joined Brewer Science, Inc. and became the Director of Product Characterization & Engineering in the Product Development group working on chemical production of materials used in the photolithography process. In 2014, he began his doctoral studies in Dr. Ruwen Qin's Systems Analytics Laboratory group at Missouri University of Science & Technology focusing on the strategic management of complex engineering systems. He published two conference papers and a journal article and was 2nd Runner-Up for Best Theoretical Paper at the 7th Annual Complex Adaptive Systems Conference in 2017. He earned the Doctor of Philosophy degree in Engineering Management from Missouri S&T in December 2020.