Next Generation Factory Layouts: Research Challenges and Recent Progress

Saïf Benjaafar • Sunderesh S. Heragu • Shahrukh A. Irani
Graduate Program in Industrial Engineering, Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota 55455
Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, Troy, New York 12180
Department of Industrial and Systems Engineering, Ohio State University, Columbus, Ohio 43210
saif@ie.umn.edu • herags@rpi.edu • irani.4@osu.edu

This paper was refereed.

Recent trends in industry suggest that existing layout configurations do not meet the needs of multiproduct enterprises and that there is a need for a new generation of factory layouts that are flexible, modular, and easy to reconfigure. Although most of the academic literature on layout design is based on a deterministic paradigm that assumes production requirements are known far in advance or change very little over time, a growing body of research focuses on designing layouts for dynamic and uncertain environments. An example is the research being carried out by the newly formed Consortium on Next Generation Factory Layouts (NGFL). The consortium, which involves multiple universities and several companies, is developing alternative layouts, new performance metrics, and new methods for designing flexible and reconfigurable factories.


There is an emerging consensus that existing layout configurations do not meet the needs of multiproduct enterprises and there is a need for a new generation of factory layouts that are more flexible, modular, and easy to reconfigure (Askin et al. 1997, Benjaafar and Sheikhzadeh 2000, Irani and Huang 2000, Kochhar and Heragu 1999, Montreuil 1999, National Research Council 1998, Yang and Peters 1998). With increased flexibility, modularity, and reconfigurability, factories could avoid redesigning their layouts each time their production requirements changed. Creating new layouts can be expensive and disruptive, especially when factories must shut down. Because factories that operate in volatile environments or introduce new products regularly cannot afford frequent disruptions, plant managers often prefer to live with the inefficiencies of existing layouts rather than suffer through costly redesigns, which may quickly become obsolete. In our work with over 20 companies in the last five years, we have encountered mounting frustration with existing layout choices, particularly in companies that offer many products with variable demand and short life cycles. These companies value layouts that retain their usefulness over many product mixes or can easily be reconfigured. Equally important are layouts that permit shorter lead times, lower inventories, and a greater degree of product customization.

Conventional layouts, such as product, process, and cellular layouts, do not meet these needs. They are typically designed for a specific product mix and production volumes that are assumed to continue for a sufficiently long period (usually, three to five years). The evaluation criterion used in most layout design procedures—long-term material-handling efficiency—fails to capture the priorities of the flexible factory (for example, scope is more important than scale, responsiveness is more important than cost, and reconfigurability is more important than efficiency). Consequently, layout...
performance deteriorates as product volumes, mix, or routings fluctuate (Afentakis et al. 1990, Braglia et al. forthcoming, Lahmar and Benjaafar 2002, Norman and Smith 2001, Palekar et al. 1992). A static measure of material-handling efficiency also fails to capture the impact of layout configuration on aspects of operational performance, such as work-in-process accumulation, queue times at processing departments, and throughput rates. Consequently, layouts that improve material handling often cause inefficiencies elsewhere in the form of long lead times or large in-process inventories (Benjaafar 2002).

When product variety is high or production volumes are small, a functional layout, with all resources of the same type in one location, is often thought to provide the greatest flexibility (Figure 1). However, a functional layout is notorious for its material-handling inefficiency and scheduling complexity, which can lead to long lead times, large work-in-process inventories, and inefficient material handling (Flynn and Jacobs 1986, Shafer and Chames 1988, Montreuil 1999, Sarper and Greene 1993). While grouping resources based on function provides some economies of scale and simplicity in allocating workloads, it makes the layout susceptible to manufacturing inefficiencies when there are changes in product mix or routings. Such changes often require a costly redesign of the plant layout or the material-handling system (Yang and Peters 1998, Lahmar and Benjaafar 2002).

An alternative to a functional layout is a cellular configuration, in which the factory is partitioned into cells (Figure 1), each dedicated to a family of products with similar processing requirements (Heragu 1994). Although cellular factories can simplify work flow and reduce material handling, they are generally designed to produce a specific set of products whose demand levels are assumed to be stable and product life cycles sufficiently long. In fact, cells are usually dedicated to single product families with little allowance for inter-cell flows. Cellular factories are inefficient when demand for existing products fluctuates or new products are introduced often (Benjaafar 1995, Askin et al. 1997, Irani et al. 1993, Suresh and Meredith 1994, Wemmerlöv and Hyer 1989, Wemmerlöv and Johnson 2000, Heragu et al. 2000). Some authors have proposed alternative cellular structures to overcome these problems, such as overlapping cells (Irani et al. 1993), cells with machine sharing (Benjaafar 1995, Suresh and

---

**Figure 1**: In a functional layout, resources of the same type are placed in the same location, while in a cellular layout, resources are partitioned into cells, each dedicated to a family of products.
Meridith (1994), and fractal cells (Montreuil 1999, Venkatadri et al. 1997). Although an improvement, these alternatives remain bounded by their cellular structure.

Layout design procedures, whether for functional or cellular layouts, have been largely based on a deterministic paradigm. Such design parameters as product mix, product demands, and product routings are assumed to be known with certainty (Meller and Gau 1996b, Norman and Smith 2001, Benjaafar 2000, Kochhar and Heragu 1999). The design criterion is often a static measure of material-handling efficiency (a total adjacency score, total material-handling cost, or a combination of both), which does not capture the need for flexibility and reconfigurability (Benjaafar 2000, 2002; Kochhar and Heragu 1999; Braglia et al. forthcoming). In fact, the relationship between layout flexibility and layout performance is poorly understood and analytical models for its evaluation are lacking. The structural properties of layouts that affect their flexibility are also not well understood (Pullington and Webster 1987, Gupta 1986, Sethi and Sethi 1990, Tompkins 1980, Webster and Tyberghein 1980). Current design criteria do not capture the effect of layout on such performance measures as congestion, cycle time, and throughput rate. They also ignore the impact of such operational parameters as setup, batching, and loading and unloading at work centers. More important, they measure only average performance and in so doing cannot guarantee effectiveness under all operating scenarios. Clearly, we need a new class of layouts, new evaluation criteria, and new design models and solution procedures.

Literature Review

Facility layout has been formally studied as an academic area of research since the early 1950s. Balakrishnan and Cheng (1998), Meller and Gau (1996a), and Kusiak and Heragu (1987) survey their vast literature. We focus on papers that are pertinent to the design of layouts in dynamic environments.

Several authors have addressed the design of layouts in settings where product mix and demand volume vary from period to period. In these settings, it may be possible to reconfigure the layout when the changes are sufficiently large, although there may be associated re-layout costs. Assuming demand information for each period is available at the initial design stage, the objective is to identify a layout for each period such that both the material handling and re-layout costs are minimized over the planning horizon. This problem is often called the dynamic facility-layout problem. Hicks and Cowan (1976) incorporated the costs of relocating departments in analyzing a single period problem. Rosenblatt (1986) was first to develop a formal model and an optimal solution procedure for determining optimal layouts for multiple periods. His model considers material-handling cost as well as the cost of relocating departments from one period to the next. Since then, a number of researchers, including Batta (1987), Urban (1992, 1998), and Balakrishnan (1993), improved on Rosenblatt's solution procedure. Others, such as Conway and Venkatramanan (1994), Kochhar and Heragu (1999), Lacksenon and Enscoe (1993), Urban (1993), and Kaku and Mazzola (1997), proposed heuristics. Balakrishnan et al. (1992), Afentakis et al. (1990), and Kouvelis and Kiran (1991) studied variations of the basic dynamic layout problem. Montreuil and Venkatadri (1991) assumed that a goal for the last period is provided by the designer and developed a model that uses this goal layout as an input and provides intermediate layouts for the intermediate planning periods. A limitation of this approach is that the relative positions of departments are fixed over all the planning periods, with only their sizes and shapes being allowed to vary. Balakrishnan and Cheng (1998) provide a comprehensive review of papers on the dynamic facility-layout problem.

In environments where changes in product mix and demand volumes are frequent or where re-layout costs are high, a plant manager may prefer a layout that is robust under multiple production demand scenarios, for example, optimistic, pessimistic, and most likely. Although the layout may not be optimal for any of the scenarios, it is robust in the sense that it is suitable under each. Rosenblatt and Lee (1987) introduced the concept of robustness in analyzing single period layouts. It was further elaborated by Rosenblatt and Kropp (1992). The work of Rosenblatt and Lee and Rosenblatt and Kropp builds on that of Shore and Tompkins (1980), who were first to consider the design of layouts under uncertainty.
present heuristic strategies for developing robust layouts for multiple planning periods. Palekar et al. (1992) consider uncertainties explicitly in determining plant layout. They formulate a stochastic dynamic layout problem assuming the following are known: (1) material flows between departments for several periods, and (2) the probability of transitioning from one flow matrix to another. They solve the model using dynamic programming for small problems and heuristics for large ones. Kochhar and Heragu (1999) describe an algorithm for single- and multiple-period-dynamic-layout problems that considers layout changeover costs.

Yang and Peters (1998) present a method for developing flexible layouts. Flexible layouts are based on the notion that layouts neither remain static for multiple planning periods nor change during every period. Instead, a layout may remain static for a block of periods, at the end of which the production has changed so much that a new layout is necessary. The layout designer must decide how and when to change the layout. Assuming the flow matrices and their probability of occurrence are known for multiple periods, the designer first determines the block of periods for which a layout is to remain static. He or she then solves the layout problem for each block of periods and combines the results to produce a layout plan for multiple periods. Montreuil and LaForge (1992) also assume that future production scenarios and their probability of occurrence are known and propose another method for developing multiple-period layouts. Like Montreuil and Venkatadri’s (1991) approach, a limitation of this method is that the relative positions of departments are fixed for all periods and only their sizes and shapes can vary.

To address the limitations that come from fixed department locations, several authors proposed that functional departments should be duplicated and strategically distributed throughout the plant. Duplication would not necessarily mean acquiring additional capacity but could be achieved simply by disaggregating existing departments, which may consist of several identical machines, into smaller ones. Montreuil et al. (1991) suggested a maximally distributed, or holographic, layout in which functional departments are fully disaggregated into individual machines, which are then placed as far from each other as possible to maximize coverage. Benjaafar and Sheikhzadeh (2000) showed that, while some disaggregation and distribution are desirable, full disaggregation and distribution are rarely justified. In fact, the benefits of disaggregation and distribution diminish with most of the benefits achieved with only a few duplicates of each department. Benjaafar and Sheikhzadeh also showed that, even in the absence of reliable information about product volumes and routings, the simple fact of having duplicates placed throughout the plant can significantly improve layout robustness. Drolet (1989) showed how distributed layouts can be used to form virtual cells temporarily dedicated to particular job orders. Lahmar and Benjaafar (2002) extend Benjaafar and Sheikhzadeh’s (2000) approach to problems with multiple periods and consider relocation costs. Askin (1999) used simulation to compare the holographic and fractal layouts proposed respectively by Montreuil et al. (1991) and Venkatadri et al. (1997).

Several of these approaches have the shortcoming of assuming known production data for future periods. Even authors who associate a probability of occurrence with each production scenario implicitly assume that the production resources (type and quantity) remain fixed. In today’s environment, drastic production changes take place frequently. Manufacturing equipment is regularly decommissioned and new equipment deployed. Plant managers often know about changes in product mix and demand volumes only slightly before a new production cycle starts. It seems reasonable for plant managers and facility designers not to look beyond the next period and instead develop layouts that can be reconfigured quickly and without much cost to suit the upcoming period’s production requirements. Heragu and Kochhar (1994) discussed this idea and argued that advances in materials and manufacturing processes, such as lighter composite materials with excellent vibration absorption properties and laser cutting, point towards lighter machine tools that will allow companies to reconfigure machines easily and frequently. Kochhar and Heragu (1999) present a genetic algorithm to solve the associated dynamic layout problem.

**Layout Classification**

In view of the above discussion, we can classify approaches to design of factory layouts for dynamic...
environments into two categories: (1) layouts that are robust for multiple production periods or scenarios, and (2) layouts that are flexible or modular enough to be reconfigured with minimal effort to meet changed production requirements. The first approach assumes that either the production data for multiple periods is available at the initial design stage itself so that the designer can identify a layout that is robust (and causes minimal materials handling inefficiency overall) over the multiple periods; or the designer can develop a layout with inherent features (for example, duplication of key resources at strategic locations within the plant) that will ensure reasonable material-handling efficiency through the various production periods. The first assumption suffers from the fact that production data must be available at the outset, which is unlikely in a dynamic environment. Designing features that allow future flexibility is more promising. However, the research in this area remains limited.

The second approach assumes that layouts would be reconfigured after each period and should be designed to minimize reconfiguration cost while guaranteeing reasonable material-flow efficiency in each period. To carry out this balancing, designers would need knowledge of production for all future periods. An alternative is to design reconfigurable features into the layout so that re-layout costs are always minimal. As with flexible layouts, research on reconfigurable layouts is still limited (Heragu and Zijm 2000).

Methods for designing layouts for dynamic environments could also be classified based on the design criteria used to evaluate alternate layouts. Most models in the literature, including those that deal with dynamic environments, rely on measures of expected material-handling efficiency—a weighted sum of travel distances incurred by the material-handling system—in evaluating candidate layouts (Meller and Gau 1996b). Some authors, such as Rosenblatt and Lee (1987) and Kouvelis et al. (1992), have used a robustness criterion based, not on mean performance, but on a layout’s ability to guarantee performance for each period or under each scenario. Others have used a combined mean and variance criterion to minimize fluctuation in performance, for example, Norman and Smith (2001). A few authors have considered operational performance as an evaluation criterion. This includes Fu and Kaku (1997), who argued that the conventional measure of average travel distances is indeed a good predictor of operational performance, as measured, for example, by expected work in process. Benjaafar (2002) showed that this is not always the case. Layouts designed using operational performance as a criterion can sometimes be very different from those that minimize average material-handling effort.

From a practical point of view, depending upon the degree of uncertainty in the production mix and volume data for future periods and the cost of revising the layout, facility designers can choose among four types of layouts (Table 1). A dynamic layout is useful when uncertainties in the production data are low and cost of re-layout is modest. A robust layout is preferable when uncertainty in production data is low, but re-layout costs are high. A distributed layout is desirable when uncertainty and re-layout cost are both high, while a reconfigurable layout is more appropriate when re-layout costs are low but uncertainty is high.

**Emerging Trends in Industry**

Several important trends are emerging in industry that could transform the layout design problem or even eliminate it. We focus on five of these trends to highlight the interaction between new business practices, new technologies, and layout design.

**Contract Manufacturing**

In many industries, outside suppliers are increasingly doing most of the manufacturing and assembly for original equipment manufacturers (OEMs) (Gibson 2000, McHale 1999). Along with just-in-time deliveries, outsourcing has led to firms reconfiguring their final

<table>
<thead>
<tr>
<th>Cost of re-layout</th>
<th>Uncertainty of Future Production Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Dynamic layout</td>
</tr>
<tr>
<td>High</td>
<td>Robust layout</td>
</tr>
<tr>
<td></td>
<td>Reconfigurable layout</td>
</tr>
<tr>
<td></td>
<td>Distributed layout</td>
</tr>
</tbody>
</table>

*Table 1: The choice of a layout type depends on the uncertainty with respect to future production requirements and the cost of re-layout.*
assembly facilities to accommodate closer coupling between suppliers and OEMs. For example, many automobile manufacturers allow suppliers to deliver components directly to points of use on their assembly lines. They have designed multiple loading docks and multiple inventory drop-off points throughout their factories. The new Cadillac plant in Lansing, Michigan, for example, is T-shaped to maximize supplier access to the factory floor. Some automobile manufacturers, such as Volkswagen (VW), are allowing suppliers to carry out some or all of the manufacturing and assembly on site. The new VW truck plant in Resende, Brazil is a showcase for this modular plant concept. To support modular plants, designers are using spine layouts (Figure 2), with the product moving along a main artery, or spine, through the plant. Linked to the spine are mini-assembly lines owned by the suppliers, each attaching its own module to the moving product. The hybrid layout has features of a flow line and multiple, autonomous cells. The configuration allows the plant to add and remove suppliers without changing the main layout. It also gracefully accommodates the growth and contraction of supplier operations. Trotter, Inc., a manufacturer of exercise treadmills, has used similar ideas in its plant (Assembly Magazine 1995). Other companies have chosen to colocate suppliers in a single large complex. The GM Gravatai plant in Brazil, for example, houses a final assembly plant and 16 supplier plants, including plants owned by Delphi, Lear, and Goodyear, which deliver preassembled modules to GM’s line workers (Wheatley 2000). The 17 plants are within walking distance and are connected through a shared material-handling system of forklift trucks and conveyors. Facility planners had to choose layouts that make material handling efficient not only in each individual plant but throughout the complex.

The picture that emerges from the above cases is of layouts with fixed cores and variable peripheries. The challenge for facility planners is then to develop a layout and a material-handling system to permit high efficiency at the core and flexibility and reconfigurability at the periphery. The design metrics should certainly be different depending on the area of the plant, but the design tools should also support a variety of layout types within the same facility. The modular layouts we discuss later address in part the challenges of constructing such hybrid layouts.

Delayed Product Differentiation

Increased product variety and the need for mass customization has led many companies to delay product differentiation (Feitzinger and Lee 1997, Lee and Tang 1997, Gupta and Benjaafar 2002), postponing the point in the manufacturing process when products are assigned individual features. Companies do this, for example, by building a platform common to all products and differentiating it by assigning to it certain productspecific features and components only after actual demand becomes known. They create hybrid facilities consisting of flow-line-like components where they build the common platforms and job shop-like components where they customize the products. If final products are easily grouped into families, the job-shop structure could be replaced by cells, each dedicated to one of the product families (Figure 3). Taken to the extreme, delayed differentiation can eliminate the problem of designing layouts altogether. For example, if customization takes place at the point of sale or in distribution warehouses, as is increasingly the case for computers (Lee and Tang 1997), the factory becomes a single high-volume, low-variety production line. Hewlett-Packard has implemented such a strategy by carrying out the localization steps for its computers and printers in its overseas distribution centers (for example, its distribution warehouses install country-specific power supplies and power cords).

The blurring of the lines between warehousing and manufacturing raises interesting questions. How does

Figure 2: In a spine layout, products move along a main artery through the plant. Linked to the spine are mini-assembly lines owned by independent suppliers who attach additional modules as needed.
Multichannel Manufacturing

The increased emphasis on quick-response manufacturing and minimum finished-goods inventory has led many manufacturers and suppliers to invest in additional capacity, often by running parallel production lines. For example, in Newark, California, Solectron, a large contract manufacturer, has a plant with 24 production lines capable of assembling everything from pagers to laser printers (Engardio 1998). By having duplicate flexible production lines shared across products, companies hope to ensure a seamless flow of material. Depending on downstream congestion, products can move in and out of neighboring production lines, creating multiple paths, or channels, minimizing queueing and congestion. EFTC, a manufacturer of electronic goods and components, also uses multichannel manufacturing (McHale 1999). An EFTC executive describes the production process as “small production lots moving to any of the standardized production points on the parallel production lines, passing from one line to wherever it is necessary to break bottlenecks and keep products rolling.” Sun Microsystems uses a similar concept for its line of desktop workstations (Feare 1997). Sun’s facility has three identical lines or cells. Each cell has two mirror image sides, which can be turned on or off, giving Sun up to six parallel production lines. As long as flow patterns and product routings do not change significantly, parallel and linear production lines, similar to those at EFTC or Sun, would provide flexibility and reduce cycle time.

Other companies have achieved multichannel flexibility by transforming functional layouts into distributed layouts, disaggregating large functional departments into two or more subdepartments distributed throughout the plant. Duplicating departments increases the likelihood of finding an efficient path through the system for each job. Other examples of distributed layouts include the fractal layout configurations introduced by Venkatadri et al. (1997) in which a plant is partitioned into several identical cells to which workloads can be allocated dynamically. Designers of multichannel systems face such challenges as determining how many duplicate paths to have and how to organize the resource duplicates on the plant floor.

Scalable Machines

In the last few years, there has been a concerted effort in the metal cutting industry to develop machines that are highly flexible and scalable and that can perform many functions and be adjusted for various capacities. The functionality and efficiency of the machines can easily be upgraded by plugging in additional modules or acquiring additional software. The multinational Initiative on Intelligent Manufacturing Systems (http://www.ims.org) is leading such an effort, supported by a conglomerate of Japanese, US, and European machine tool makers (Ikegaya 2000). The National Science Foundation Engineering Research Center on Reconfigurable Machines at the University of Michigan (http:...
is carrying out a parallel effort, focusing on building machines that can be quickly adjusted for changes in product mix or volumes, for example, machines can be quickly upgraded by adding spindles, axes, tool magazines, or controllers (Koren et al. 1999). If successful, such efforts could lead to facilities that use one machine for most processing with little material handling and movement. Because a machine can be rapidly configured for different mixes and volumes, changes in production requirements would have little effect on layout.

A commercial product that already has some of these capabilities is the TRIFLEX machining center, marketed by Turmatic Systems. The center allows simultaneous machining using up to seven machining units with the possibility of retrofitting additional ones. It can accommodate automatic loading and unloading systems and can be integrated into similar or different machine systems. A single machining unit can be fitted to a long base slide, enabling the sides of a workpiece to be machined in one station and the front face in another. Therefore, five-sided machining is possible, even with only two machining units fitted.

Such scalable machines could transform layout design. If material movement became minimal, factory layouts would be greatly simplified and their design would be less important. Emphasis in factory design would then likely shift from the detailed design of each processing department to the higher level integration of these departments (for example, integrating machining with assembly or assembly with inspection and packaging).

Portable Machines

Several equipment manufacturers are marketing portable machines that are easily and dynamically deployed in different areas of the factory as production requirements change. The TRAK QuikCell QCM-1, available from Southwestern Industries (www.southwesternindustries.com), is a compact and mobile milling machine used for small-lot, job-shop machining. It can be located close to the primary machining or turning centers producing a family of parts that requires preliminary or secondary operations on other machines. The foundation of the machine is a base casting that can be moved with a pallet jack from any side. The machine is small enough to fit through most doors, and its rigid frame does not require releveling after each move. Quick disconnects are available for electrical supply, a coolant sprayer, a power draw bar, and an air hose. Climax Portable Machine Tools (www.cpmt.com) makes machines that have the capabilities of stationary machine tools for repairing turbines, paper machinery, and heavy equipment. The portable machines go to the workplace and mount on the workpieces—instead of the other way around (that is, workpieces are stationary and movement is incurred by the machines). Hence, factories would have to be laid out to facilitate the flow of machines instead of parts.

In Northern Telecom’s facility in Calgary, Canada for manufacturing business telephone equipment, generic, modular, conveyor-mounted work cells can easily and quickly be moved from one location to another (Assembly Magazine 1996). These independent cells can be unplugged from the main assembly line and moved to accommodate different products. With frequent changes in product design, the facility uses the conveyor-mounted work cells to change tooling and layout to suit the new production and assembly requirements.

Portable machine tools require storage and retrieval. Fortunately, technology is being developed to allow easy storage and retrieval of large equipment. For example, Robotic Parking (www.roboticparking.com/tech.html) markets a modular automated parking system (MAPS) that integrates computer control with mechanical lifts, pallets, and carriers to park and retrieve large equipment in multilevel modular warehouses. Complete facilities can be constructed on lots as small as 60 by 60 feet, up to 20 stories, and above or underground. Although originally designed for parking garages, the technology is finding applications in manufacturing and warehousing. Portable machines could be maintained in a MAPS-like warehouse adjacent to the main manufacturing floor. Depending on product mix and demand, machine tools would be “picked from the shelves” and inserted in the manufacturing facility.

The shift to lighter machines is also driven by advances in materials. For example, composites are increasingly the primary choice for many components.
Aluminum composites can now replace cast iron parts and phenolics are replacing aluminum parts. These light materials can also be engineered to have excellent mechanical properties, such as hardness, heat resistance, tensile strength, and vibration absorption. Advancements in nonabrasive manufacturing processes, such as laser cutting and electron-beam hardening, are aiding the development of lightweight machining equipment. Industry is also developing permanent magnetic chucks that facilitate quick mounting and dismounting of tools, carry their own energy sources, and do not obstruct machining. With these developments in materials and processing technology, we are moving towards processing technologies that employ lightweight machine tools and can process lightweight parts. Heragu and Kochhar (1994) foresee facilities in which lightweight equipment mounted on wheels is easily moved along tracks embedded in the shop floor with universal plug points for support services, such as compressed gas, water, and coolant, dispersed throughout the plant. With such technology, it may be feasible to change layouts several times per year. With re-layout costs cut, the criterion in designing layouts then shifts from long-term material-handling efficiency to short-term responsiveness. Firms would focus on operational performance by reconfiguring layouts periodically to relieve short-term congestion and maximize throughput for current products and demand levels. The agile layout design methodology we describe later is in part motivated by this vision.

Next Generation Factory Layouts

We are carrying out research under the newly formed NSF Consortium on Next Generation Factory Layouts (NGFL). The Consortium is supported by a major grant from the National Science Foundation (NSF) and involves multiple universities and several manufacturing companies. The goal of the consortium is to explore alternative layout configurations and alternative performance metrics for designing flexible factories. Three approaches to layout design address three distinct needs of the flexible factory. The first two approaches present novel layout configurations, namely distributed and modular layouts. In the third approach, we use operational performance as a design criterion to generate what we term agile layouts.

Distributed Layouts

Distributed layouts disaggregate large functional departments into subdepartments distributed throughout the plant floor (Figure 4). Duplicate departments strategically located throughout the factory allow the facility to hedge against future fluctuations in job-flow patterns and volumes. In turn, disaggregated and distributed subdepartments reduce material-travel distances for many production flow sequences. Planners can easily find efficient flows for a wide range of product mixes and volumes. Such layouts are especially appealing when demand fluctuates too frequently to make reconfiguring the plant cost effective. In these settings, a fixed layout that performs well for many demand scenarios is desirable.

In designing a distributed layout, a firm faces several challenges. How should it create subdepartments, and how many should it have of each type? How much capacity should it assign to each subdepartment? Where should it place the subdepartments? How should it allocate workload among similar subdepartments? How will department disaggregation and distribution affect operational performance (for example, material-handling times, work in process, and queueing times)? How should the firm manage material flow, now that there is greater routing flexibility? How should it coordinate the competing needs for material handling of similar subdepartments? What performance measure should the firm use when designing distributed layouts? Should it measure expected material-handling cost over possible demand scenarios, or should it seek a measure of robustness that guarantees a minimum level of performance for all scenarios? More important, how sensitive are the final layouts to the adopted performance measure? Although duplicating departments might increase flexibility, it could also increase and diminish economies of scale (for example, operators and auxiliary resources must be duplicated). The firm must trade off the material-handling benefits of disaggregation and duplication against cost increases in other areas.

Benjaafar and Sheikhzadeh (2000) and Lahmar and Benjaafar (2002) explored some of these questions. Benjaafar and Sheikhzadeh considered situations in
which demand for products is characterized by finite discrete distributions, represented by a finite number of demand-realization scenarios and probabilities of occurrence for each scenario. Demand for products may be independent or correlated. Both cases result in scenarios consisting of different product-demand combinations, each with its own probability of occurrence. The distributions may be based on historical data or on forecasts. When the demand distributions are difficult to characterize, one can assign equal likelihood to all possible demand scenarios. Alternatively, one can aggregate the scenarios into a smaller subset that represents the range of possible demand scenarios.

From the distribution of demand scenarios, the product routings, and the product unit transfer loads, we determine for each possible demand scenario the amount of material for each product that will flow between each pair of departments. This results in a multiproduct from-to flow matrix for each demand scenario. The objective is to select a layout that minimizes expected material-handling cost over the entire set of scenarios. For each scenario, we need to determine the optimal allocation of flow among subdepartments of the same type. Thus, we have a combined layout and flow-allocation problem. Benjaafar and Sheikhzadeh (2000) describe a model for this layout-flow allocation problem, as well as an effective decomposition solution procedure. Lahmar and Benjaafar (2002) extend the model and the solution procedure to settings with multiple periods, where the layout can be reconfigured at a cost at the beginning of each period.

Benjaafar and Sheikhzadeh’s (2000) and Lahmar and Benjaafar’s (2002) experiments with distributed layouts, using both randomly generated examples and data collected from industry, showed firms could benefit from disaggregating and distributing functional departments (over 40 percent improvement in most cases). Distributed layouts provide the greatest advantage when demand is variable, particularly for layouts with large departments or many department types. If the distribution of flow patterns can be categorized a priori, including flow information at the design stage can improve layouts. However, material-handling costs can be reduced even without flow information.

Figure 4: In a distributed layout, not all equipment of the same type (represented by a particular shape in the figure) is placed in adjoining locations. Instead, equipment of the same type is either grouped in multiple clusters (partial distribution) or placed individually throughout the plant (maximal distribution).
(for example, by distributing subdepartments randomly). Furthermore, the quality of distributed layouts is insensitive to inaccuracies in the demand distribution. More important, firms can obtain most of the benefits from duplicating departments with few replicates, rarely having to fully disaggregate functional departments.

A layout that distributes department replicates throughout the plant floor can also help a firm to handle products with short runs or products with short life cycles. It can do so, for example, by quickly forming temporary cells, consisting of adjoining subdepartments, dedicated to a particular product line or job order (Figure 5). This cell is disbanded once the product is phased out or once the customer order is completed. The individual replicates are then free to participate in new cells. Drolet (1989) discussed such virtual cells. Lahmar and Benjaafar (2002) found that distributed configurations can be useful in handling production growth and contraction gracefully. For example, when products mature over several periods, the firm can avoid redesigning its facility repeatedly to accommodate product growth by using a distributed layout and adding machines to the periphery of the layout as needed. The facility can then grow almost in a concentric fashion, keeping layout space compact and maintaining efficient material handling. With this approach, the firm can modify capacity in small increments since introducing or removing capacity takes place at the periphery with the factory core remaining intact.

**Modular Layouts**

Modular layouts are hybrid layouts for systems with complex material flows that cannot be described as functional, flow line, or cellular. Several of the emerging trends in industry are leading to such configurations. For example, the automobile industry builds modular factories around flow-line-like cores with connected supplier production lines in various forms. Firms that delay product differentiation also use layouts that combine product, process, and cellular features. Irani and Huang (2000) were first to introduce the concept of layout as a network of basic modules. They assumed, at least in the short term, a known product mix and fairly stable demand. As the mix and demand change, some modules are eliminated and others added. With such modular layouts, manufacturers can scale their activities up or down quickly. In their research on modular layouts, Irani and Huang (2000) sought to answer the following fundamental questions: Could a layout other than the three traditional layouts better fit the material flows of multiproduct manufacturers? Perhaps a combination of the three traditional layouts? Could a network of layout modules provide a metastructure for designing multiproduct manufacturing facilities in general? Would grouping and arranging resources into modules corresponding to specific traditional layouts minimize total flow distances or costs?

Irani and Huang (2000) designed a modular layout for a Motorola facility (Figure 6). The company wanted to assess the feasibility of changing the layout in one of their semiconductor fabs from functional to cellular. The functional layout comprises seven bays (or process departments): diffusion, etching, film deposition, implant, photolithography, metrology, and backend. Motorola provided four product routings representative of the fab's product flows. The authors found that a cellular layout would not be viable because it would require duplicating equipment and processes. However, a visual string-matching analysis of the routings

---

**Figure 5**: A distributed layout can be used to quickly form temporary (virtual) cells, consisting of adjoining subdepartments, dedicated to a particular product line or job order. The cell is disbanded once the product is phased out or once the customer order is completed.
revealed that different pairs of routings had substrings of operations that were identical or had many operations in common. Based on this observation, they designed a new layout (Figure 6) that combines the three traditional layouts. In this layout, all pairs of consecutive operations in all the product routings are performed in the same layout module or in adjacent modules, where a layout module is a group of machines whose flow pattern is characteristic of a traditional layout. The authors have since studied samples of product routings obtained from published data from industry and found that product routings often have common substrings of operations that could be aggregated into modules.

**Design Procedure for Modular Layouts**

Irani and Huang (2000) showed that material flow in any multiproduct facility can be decomposed into a network of layout modules, each module representing part of the facility. A module is a group of machines connected by a material-flow network with a well-understood flow pattern and method for designing its layout (Figure 7). For example, the flow-line and cell modules have a part family focus. The flow-line module aggregates routings that are identical, whereas the cell module aggregates routings that have similar machine sequences. In contrast, the functional layout
module is a group of machines that do not process products with similar routings. However, the material-flow pattern in its from-to chart could correspond to an acyclic digraph, as in an assembly or disassembly line or, in the worst case, a completely connected digraph.

In the ideal solution, each product would be completely processed on a dedicated flow line, but that would entail significant investment in equipment. A practical approach would be to maximize the number of consecutive operations in a family of routings that are performed in the same module. To find such a structure, Irani and Huang (2000) employed the method of string matching and clustering used in genetics, molecular chemistry, and biology. At the core of this approach are the concepts of common substrings and residual substrings in a product routing. A common substring consists of consecutive operations that two or more operation sequences have in common. Residual substrings are the substrings of operations that remain after all the common substrings are extracted. For example, in operation sequences $S_a(1\rightarrow2\rightarrow3\rightarrow4\rightarrow7\rightarrow8)$ and $S_b(1\rightarrow2\rightarrow5\rightarrow6\rightarrow7\rightarrow8)$, the common substrings are $1\rightarrow2$ and $7\rightarrow8$. The residual substrings are $3\rightarrow4$ and $5\rightarrow6$ in sequences $S_a$ and $S_b$, respectively. Given the sample of routings for products manufactured in the facility, Irani and Huang (2000) first extract the common substrings between all pairs of routings. Next, they compute the frequency with which each common substring occurs in the routings. They then aggregate similar substrings and create a layout module for each cluster of substrings. Finally,
they eliminate modules that do not meet criteria for machine utilization or constraints on machine allocation and duplication among multiple modules. The typical result from using this approach is a facility layout that is a network of dissimilar modules. In the example (Irani and Huang 2000), the layout consists of a cell module (M2), two patterned flow modules (M1, M4), a flowline module (M3), and a functional module (M2) (Figure 8).

**Agile Layouts**

In facilities that permit frequent reconfiguration, layouts could be designed to maximize operational performance rather than to minimize material-handling cost. As production-planning periods shrink, factories shift their focus from long-run cost efficiency to short-term responsiveness and agility. Such performance measures as cycle time, work-in-process (WIP) accumulation, and throughput become especially important. Unfortunately, capturing the relationship between layout configuration and operational performance is difficult. Meller and Gau (1996a) reviewed over 150 papers on factory layout and found only one paper on the subject. Recently Benjaafar (2002) introduced an analytical model capable of capturing the relationship between layout configuration and operational performance. He embedded the model in a layout-design procedure in which the design criterion can be one of several measures of operational performance. Heragu et al. (2000) expanded Benjaafar’s (2002) model to include set-up time, transfer, and process batch size and developed a method that can estimate operational performance measures of functional and cellular manufacturing systems.

**Design Procedure for Agile Layouts**

To capture the effect of layout on operational performance metrics, such as cycle time, WIP, and throughput rate, Benjaafar (2002) modeled the manufacturing facility as a central-server queueing network and each processing department as a multiserver queue with general distribution of product-processing and interarrival times. The material-handling system operates as a central server in moving material among departments. Benjaafar (2002) assumes that the material-handling system consists of discrete devices (for example, forklift trucks, human operators, and automated guided vehicles). The distances material transporters travel are determined by the layout configuration, product routings, and product demands. In determining the transporter travel-time distribution, he accounts for both empty and full trips made by the material transport devices.

Using the model, he showed that layout configuration does indeed have a direct impact on operational performance, often in unpredictable ways. For example, minimizing full travel can cause empty travel to increase, which, in turn, can increase congestion and delays. Thus, placing departments in neighboring locations, even though no material flows directly between them, may reduce empty travel enough to reduce overall use of the material-handling system. For example, empty travel to and from departments is highest for those visited most frequently. Placing these departments close together, although there may be no direct flows between them, could significantly reduce empty travel. Likewise, placing departments with high intermaterial flows far apart may be beneficial (Figure 9).

Benjaafar (2002) showed that, in general, a design criterion based on average travel distances is a poor indicator of operational performance. In fact, a layout that is optimal with respect to full travel could be operationally infeasible (that is, it could produce infinite

![Figure 8: The original layout of the facility has been decomposed into a network of different layout modules with minimum intermodule material flows.](image-url)
WIP accumulation). Similarly, two layouts that are optimal with respect to full travel could have vastly different WIP values. Because conventional approaches tend to optimize the average distance traveled by the material-handling system, they do not account for the variance in these distances. Distance variance, however, partly determines how much congestion a layout exhibits. More important, it was shown that congestion is not necessarily monotonic in the average distance traveled by the material-handling system. A layout that reduces average distances but increases variance could increase overall congestion. Similarly, a layout that reduces variance, even if it increases average travel distances, could reduce congestion. In practice, travel-time variance often depends on the material-handling system when material-handling is automated. Therefore, designers need to pay special attention to material handling configurations that minimize not only mean but also variance of travel distances (Figure 10).

Realizing the importance of these indirect effects, many companies are designing layouts that minimize dimensional asymmetries and reduce empty travel. For example, Volvo designed its Kalmar plant in Sweden as a collection of hexagon-shaped modules where material flows in concentric lines within each module (Tompkins et al. 1996). Lucent is experimenting with layouts in which shared processors are centrally located in functional departments and are equidistant from multiple dedicated cells within the plant. Variations of the spine layout, with departments along a common corridor, have been implemented in industries ranging from electronic manufacturing to automotive assembly (Tompkins et al. 1996, Smith et al. 2000). Layout configurations that minimize dimensional asymmetries and reduce empty travel are also found in nonmanufacturing applications. For example, both the spine and star layouts are common configurations for airports. Spine and T-shaped layouts are also popular designs for freight and cross-docking terminals (Gue 1999).

**Research Challenges**

Several research challenges remain. In designing distributed layouts, designers of the current models assume that the number of department duplicates and the capacity of each duplicate are known. In practice, facility designers must make these decisions before developing a layout. Current models do not account for the cost of disaggregating and distributing departments nor do they capture the economies of scale associated with operating consolidated departments. The infrastructure typical of a single consolidated department in a job shop (for example, operators, computer control systems, loading and unloading areas, and waste-disposal facilities) must be duplicated in a
same order were not synchronized. To address this problem, one would need to capture setup minimization in the objective function or place additional constraints on flow allocation to prevent order splitting.

For modular layouts, several important issues need to be addressed: (1) After identifying all common substrings, one would need to aggregate several of the substrings into a single module to minimize machine-duplication costs based on a measure of substring dissimilarity and a threshold value for aggregating similar substrings. This is related to the problem of determining the optimal number of modules in the final layout. One idea is to develop measures of connectivity and transitivity of the directed graph we obtain from aggregating a set of common substrings. (2) We need to establish feasibility criteria for allocating machines to several modules subject to machine availability and criteria for minimum machine utilization. An iterative loop should be incorporated in the design to absorb any module rejected because of these criteria. (3) The current approach treats each residual substring as a sequence of operations performed on machines located in process departments. It seems logical to cluster these substrings and aggregate their machines into cell modules based on user-defined thresholds for string clustering. (4) We must compare the performance of this new layout with those of flowline, cellular, and functional layouts for the same facility.

For agile layouts, we need models that account for different routing and dispatching policies of the material-handling system. These models could then be used to study the effects of different policies on layout performance. Furthermore, we could use the queueing model to evaluate and compare the performance of classical layout configurations under varying conditions. We might identify new configurations that are more effective in achieving small WIP levels. In particular, identifying configurations that reduce distance variance without affecting average distance can be valuable. Such configurations might include the star layout, where departments are equidistant from each other, or the hub-and-spoke layout, in which each hub consists of several equidistant departments and is served by a dedicated transporter. In many applications, differentiating between WIP at different departments or different stages of the production process is useful. WIP tends

Figure 10: The star-layout configuration has a smaller variance than the loop layout, which itself has a smaller variance than the linear layout.
to appreciate in value as it progresses through the production process. We should favor layouts that reduce the most expensive WIP first, for example, those in which departments that carry out the last production steps are centrally located. Another important avenue of research is to integrate layout design with the design of the material-handling system. For example, we could simultaneously decide on material-handling capacity (number of transporters or transporter carrying capacity) and department placement, with the objective of minimizing both WIP-holding cost and capital-investment costs. We could then examine the trade-offs between capacity and WIP.

From industry, three main trends are worth highlighting again: (1) the move toward lighter and more portable equipment, (2) the increased modularization of products and the increased postponement in product differentiation, and (3) the shift to more flexible and scalable machines. The first trend could change the nature of layout design from strategic to operational (re-layout could become more frequent and more focused on supporting operational performance), requiring new layout-design tools and new design metrics. In contrast, the second and third trends could lead to simpler layouts and a reduced need for re-layout (for example, with delayed differentiation, products follow mostly a flow-line-like path). However, the above trends would most likely not apply to all industries. For many forms, neither full process reconfigurability nor product standardization will be possible (for example, industries with innovative or custom-made products). For these firms, designing layouts that are robust and able to sustain a wide range of products would remain critical.

Acknowledgments
This material is based in part upon work supported by the National Science Foundation under Grant no. DMI 9908437. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References


Drolet, J. R. 1989. Scheduling virtual cellular manufacturing systems. Ph.D. dissertation, School of Industrial Engineering, Purdue University, West Lafayette, IN.


Lahmar, M., S. Benjaafar. 2002. Design of dynamic distributed layouts. Working paper, Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN.


BENJAAFAR, HERAGU, AND IRANI

Factory Layouts


