Design Principles for Machining System Configurations

P. Spicer, Y. Koren (1), M. Shpitalni* (1), D. Yip-Hoi
NSF Engineering Research Center for Reconfigurable Machining Systems
University of Michigan, Ann Arbor, Michigan, U.S.A.

Abstract
Until recently, mass producers have relied on long serial lines of dedicated machines in order to machine products at high volumes. Today, as the cost of CNC machining centers decreases, more and more high-volume manufacturers are examining the alternative of shorter lines with more machines configured in parallel. This paper presents the main principles for selecting the right machining system configurations. We propose a classification of systems and show that only symmetric configurations are of interest to industry. We compare four classes of systems: pure serial lines, pure parallel lines, short serial lines arranged in parallel, and short serial lines arranged in parallel with the ability to move products between the lines (i.e., with crossover). Specifically, we compare the different configurations in terms of throughput, line balancing, machine investment cost, and capacity scalability. Finally, we introduce an upgradable multi-spindle reconfigurable machine as a cost-effective alternative for system scalability.

Keywords:
Machining, system layout, design principles.

1 INTRODUCTION
Prior to the early 1990's, high-volume manufacturers (e.g., the automotive industry) enjoyed a stable growing market with long product lifetimes. Thus, there was no urgent need for these manufacturers to search for alternatives to the dedicated machining systems they were using for producing their machined components. Furthermore, computerized numerically controlled (CNC) machines—the main building blocks of flexible systems that offer the alternative solution—were excluded due to their high cost, low reliability, and low productivity. As a result of increased global competition in the 1990's, manufacturing companies began facing more frequent and unpredictable market changes. These changes include the rapid introduction of new products, abrupt changes in product demand and mix, and more frequent modifications to existing products. To stay competitive and to accommodate these changes, manufacturing companies began seeking manufacturing systems that enable a rapid response to market changes. Until recently, high-volume manufacturers have not had a real alternative to dedicated machining systems. This situation has now totally changed for two reasons: (1) Cost reductions in the price of CNC machines, and (2) the introduction of improved technologies. These technologies include controller improvements, high-speed spindles, linear motors, and multi-spindle machines. Another influencing factor is the possible use of 4-axis machines, which offer much better part accessibility than 3-axis machines, and are not significantly expensive when integrated into a high-volume system.

Multi-axes CNC machining centers equipped with automatic tool changers enable a manufacturer to assign many machining tasks to a single machining center and change them when reconfiguration is needed. This opens up the potential for a large variety of possible system configurations. Consequently, manufacturers are now looking toward reconfigurable systems whose functionality and production capacity can be changed exactly when needed [1].

For the above reasons, high-volume manufacturers of machined components have been moving away from serial-dedicated machining systems and have begun to install systems based on CNC machining centers. This move from long serial system configurations with a small number of machining tasks assigned to dedicated machines (stations) toward CNC-based systems offers many options in selecting the right configuration. Each of these options, however, should be able to produce the production rate needed in high-volume manufacturing. The following example illustrates the many options.

Imagine a component with 15 different machining tasks. It could be configured, for example, into a system of 15 machining centers (stations) arranged in serial (Figure 1(a)). In this case, each machining center would perform one machining task. The other extreme case is arranging all the machines in parallel (Figure 1(b)), assuming that each machine can perform all the machining tasks. Of course many machines must operate in parallel in order to achieve the desired rate of production. Another choice may be to shorten the system to only five machines in serial, with three lines to supply demand (Figure 1(c)). In this case, each machine would perform three machining tasks on average. Likewise, there are numerous other possible configurations.

Koren, Hu, and Weber [2] have demonstrated that the system configuration (the arrangement of the machines and the interconnection among them) has a significant impact on six key performance criteria: 1) investment cost of machines and tools, 2) quality, 3) throughput, 4) capacity scalability, 5) number of product types, and 6) system conversion time. They also mentioned that the choice of system configuration is not trivial. In fact, a recent research study suggests that the number of configurations that can be created with n machines is greater than $2^{**}(n-1)$, and that the design of the configuration of a machining system is a non-polynomial (NP) complex problem (i.e., exponential problem). Hence, there is a need for system design methodologies, guidelines, metrics, and principles to help in the selection process.
of the right configuration and in the design of manufacturing systems.

In the state of the art today, however, a systematic design methodology that determines the optimal configuration does not exist. In fact, not much has been researched about the generation of configurations, and not even about the relationships between the system configuration and its performance (because the system configuration is assumed as given). The closest related research area is facility layout planning (see [3] for a general overview). But this area focuses on the efficiency of material flow and ignores the six performance criteria in [2]. Thus, more research on the relationships between the machining system configuration and its performance is needed in order to design optimal configurations. (It should also be noted that since the problem is exponential, a better understanding of these relationships will yield better heuristics, which when used to design optimal machining systems will, in turn, enable us to deal with a reasonable number of options).

3.1 Configuration Length

In this paper, we identify several principles for the selection of machining system configuration. We introduce various typical system configuration concepts (Figure 1). We then show how set-up planning, machine design, and line balancing bound the scope of a system configuration. We then compare the different configurations in terms of line balancing, throughput performance, scalability, quality, machine investment cost and cutting tool costs. Finally, we briefly introduce a scalable reconfigurable machine tool (RMT) concept as an alternative to purchasing duplicate machines installed in parallel. Thus, this research creates new knowledge regarding governing rules and heuristics that can be used to design better configurations. Consequently, our research is not only scientifically challenging, but it is also very relevant and significant to the industry.

2 CLASSIFICATION OF CONFIGURATIONS

Machining system configurations are determined by (1) the arrangement of the machines and (2) the relations (connections) among them. Refer to Figure 1: similar machine arrangements are shown in (c) and (d); however, the connections among the machines, and thus the configurations, are different. The configuration shown in (c) represents three serial lines in parallel, while the one shown in (d) is a machining system with crossover. A configuration with crossovers allows that parts from one machine be transferred not only to a specific machine but also to one out of a set of parallel machines. Note that machines can "back-up" other failed machines in configurations with crossovers.

An important distinction is according to whether the system configuration is symmetric (also referred to as 'single-process') or asymmetric (also referred to as 'variable-process'). In Figure 1, only configuration (f) is asymmetric. In single-process (i.e., symmetric) configurations, each part goes through the same process plan and is executed on the same number of machines as all other parts, regardless of the path selected through the system. In such cases all the machines arranged in parallel perform exactly the same set of tasks. We refer to this set of tasks as an 'operation'. By contrast, in variable-process (i.e. asymmetric) configurations, a machined part may experience different process plans executed on a varying number of machines, according to the path it follows through the system. In the variable-process configuration example shown in Figure 1(f), a part may be machined by a minimum of five machines, or a maximum of nine machines. In this example many other possibilities exist.

Several factors, such as in-process inspection, quality control, process documentation, and operator training, as well as process plans and part programs, are all path-dependent (i.e., they vary according to the part machining-path). By seeking uniformity (repellion), the industry tends to avoid asymmetric, variable-process configurations. Therefore, in this paper we limit our discussion to single-process, symmetric configurations.

We further divide the symmetric configurations into five classes as shown in Figure 1, namely: (a) pure serial, (b) pure parallel, (c) parallel without crossover, (d) parallel with crossover, and (e) hybrid.

Figure 1 shows examples of these classes. In pure serial configurations, there is only one possible path a part can take through the system. In pure parallel configurations, all machines are arranged in parallel to each other. In parallel configurations without crossover, short identical serial lines are arranged in parallel, and parts are not allowed to "crossover" between the lines. In contrast, in parallel configurations with crossover, parts are allowed to "crossover" between the lines (e.g., in case of machine failures). Finally, hybrid systems are a mixture of the previous configuration types.

This level of detail is sufficient to introduce and discuss fundamental configuration principles.

3 BOUNDING THE CONFIGURATION

Although the choice of system configurations is very extensive, finite bounds can be established with respect to the minimal and maximal path lengths (number of operations or machines a part must go through) of the machining lines. In this section, we describe how process planning bounds system configurations.

3.1 Configuration Length

A machined component, such as a cylinder head, requires many separate machining tasks (e.g. surface milling, drilling, tapping, etc.). Since the tasks are performed on multiple faces at many different angles, several different set-ups are required. For example, all of the machining tasks on face "A" of the part shown in Figure 2 may be accessible in a single set-up. The part may then be repositioned and placed in another set-up to machine the features on face "B".

Figure 1: Alternative system configurations: (a) pure serial, (b) pure parallel, (c) parallel without crossover, (d) parallel with crossover, (e) hybrid, (f) variable process.

In the state of the art today, however, a systematic design methodology that determines the optimal configuration does not exist. In fact, not much has been researched about the generation of configurations, and not even about the relationships between the system configuration and its performance (because the system configuration is assumed as given). The closest related research area is facility layout planning (see [3] for a general overview). But this area focuses on the efficiency of material flow and ignores the six performance criteria in [2]. Thus, more research on the relationships between the machining system configuration and its performance is needed in order to design optimal configurations. (It should also be noted that since the problem is exponential, a better understanding of these relationships will yield better heuristics, which when used to design optimal machining systems will, in turn, enable us to deal with a reasonable number of options).
It turns out that the set-up arrangements selected in a system have a strong impact on the system configuration length. In order to clearly describe the relationship between set-up planning and system configuration, it is necessary to define three types of hierarchical set-up arrangements:

1. **Fixture Set-up:** The set of possible machining tasks that can be performed on a machine with a given part-holding fixture.
2. **Process Set-up:** The actual set of machining tasks assigned to a specific fixture set-up on a specific machine.
3. **Operation:** The set of process set-ups assigned to a single machine.

Fixture set-ups are determined by taking into account four factors:

1. Part geometry (e.g. location of machined features).
2. Fixture design (e.g. orientation and clamp locations).
3. Process conditions (e.g. feedrate and spindle speed).
4. Machine capability (e.g. a 3-axis machine can access only one face per set-up, while a 3-axis machine equipped with a 4th rotary table axis can access multiple faces and angles).

Process set-ups are determined by assigning tasks to the fixture set-ups while taking into account four factors:

1. Machining task precedence constraints.
2. Processing time requirements.
3. Line balancing.
4. The requirement for special machining, assembly or inspection processes that mandate the use of special purpose dedicated machines (e.g. line boring, bearing cap assembly, leak testing, etc.).

In typical high-volume CNC-based machining systems, manufacturers assign only one process set-up to each operation. But, when manufacturers use so-called "AB fixtures," they assign two different process-set-ups to an operation. This allows for machining on two different sides of two different parts while on the same machine.

In single-process configurations (see Section 2), the length of the configuration equals the number of operations that a part must go through. For example, the configuration length in Figure 1(a) is 15, in 1(b) is 1, and in 1(c) is 5. Furthermore, the following rules determine upper and lower bounds on the length of a single-process system configuration:

1. **The maximum configuration length** is achieved when only one machining task is assigned to each operation. For example, a complex part that requires 40 machining tasks has a maximum possible configuration length of 40 machines. This situation creates a very long system that is usually unbalanced, because no measures are taken to balance the line.

2. **The minimum configuration length** is achieved when a maximum number of tasks are assigned to each operation. For simple components, if all tasks can be done in one operation, the corresponding minimum configuration length is also one. However, for complex machined components like the cylinder head shown in Figure 2, several different operations are usually required.

Within the bounds given above, the configuration length can be changed with two basic techniques:

1. **Operation Division:** If the configuration length is shorter than the maximum, fixture set-ups can be duplicated, so that machining tasks can be assigned to more than one operation. (This can be referred to as similar operations in parallel).
2. **Set-up Combination:** If the configuration length is greater than the minimum, process set-ups can be combined together, resulting in fewer operations.

**Example:** Figure 3(a) shows an initial machining system line balance. A, B, and C denote different faces of the part, on which many machining tasks must be performed. Several machining tasks have been assigned to 3-axis CNC machines. But, the system is quite unbalanced because of fixture set-up limitations. This is because 3-axis CNCs can only access one face in each operation. Figure 3(b) shows the difference in line balance and configuration length that results from applying operation division. In Figure 3(b), operation 1 has been divided into two operations with identical fixture set-ups. This improves the line balance and increases the line length by 1. In Figure 3(c), operations 2 and 3 have been combined into one operation, because a 4-axis machine is used instead of two 3-axis machines. This improves the line balance and shortens the configuration length by one operation.

**3.2 Configuration Width**

The configuration width is defined as the number of machines in parallel for a given configuration. For example, in Figure 1(c) and (d) the configuration width is 3. When the number of machines in parallel varies as shown in Figure 1(e) and (f) we can talk about the maximal width, which is 3 in these cases. The configuration width is a function of the required production capacity and the configuration length. Given a certain capacity requirement, the bounds on configuration width are determined as follows:

1. **The maximum configuration width** is achieved when a system configuration is at its minimum length. Each operation has a long cycle time, and many machines are required in parallel to meet capacity requirements.
2. The minimum configuration width is achieved when a system configuration is at its maximum length or close to it (such that no operation takes more time than the longest task). In this case, each operation has a short cycle time, and few machines are required in parallel to meet capacity requirements.

For well-balanced system configurations with a fixed number of machines, the total configuration solution space can be viewed as a hyperbola with length and width defined as in Figure 4.

![Figure 4. Configuration solution space for a given capacity.](image)

4 EFFECT OF CONFIGURATION ON THROUGHPUT

In [2], Koren, Hu, and Weber demonstrated that system configuration affects throughput. Specifically, they showed that for a given machine reliability the expected throughput is a function of system configuration depending on the degree to which machines are arranged in serial or parallel. In this section, we expand on their ideas. When the machine reliability (or availability) is assumed perfect (i.e., 1), the throughput is called the gross throughput. The gross throughput depends on two factors: the line balancing that was already discussed, and the impact of the part transfer time, which depends on the configuration. First, we discuss the effect of system configuration on gross throughput resulting from part material handling time. Second, we describe the effects of system configuration on expected throughput due to machine availability. Finally, we explain the effects of system configuration on the throughput probability distribution.

4.1 Configuration Effect on Gross Throughput

Gross throughput is defined as the maximum throughput of the system when all machines in the system are running. It is directly related to the machine part-to-part cycle time. Machine cycle time can be divided into two components: (1) machining time, and (2) material handling time. Material handling time takes a smaller percentage of machine cycle time in short parallel lines compared to long serial lines. Therefore, shorter configurations will have a higher percentage of time spent machining, and consequently higher throughput than long configurations.

For example, consider a part that requires $T=100$ seconds of machining tasks; and assume that it takes $T_c=15$ seconds to load/unload a part from a CNC machine. Let us assume that a manufacturer has the choice of assigning the tasks to four ($n=4$) machines arranged in parallel or serial. In the serial configuration, each machine will have a cycle time of $100/4 + 15 = 40$ seconds, or, in general terms, the cycle time is $T/n+T_c$. Thus, the gross throughput is $1.5$ parts per minute, or $90$ parts per hour.

Note that 38% of the cycle time is spent on material handling. However, in the parallel configuration, each machine will have a cycle time of $100 + 15 = 115$ seconds. In this case, only 13% of the cycle time is for material handling. Since there are four machines, the system cycle time is $115/4$, or, in general terms, the cycle time is $(T+T_c)/n$. Therefore, with four machines in parallel, the throughput is: $4(3600/115)=125$ parts per hour, which is significantly greater than the serial configuration. The ratio of the two productivities, $90/125$, is given in the general case by the equation $(T/T_c + 1)/(T/T_c + n)$, where $n$ is the configuration length. For $T>T_c$, the benefit of designing more parallelism in a system is not so significant as it appears in the above example.

4.2 Configuration Effect on Expected Throughput

Without considering the use of buffers, machine downtime has a different effect on expected throughput depending on system configuration. The most significant difference is between serial and parallel line configurations. In a pure serial line configuration, if a single machine fails, the entire system must stop because the other machines are blocked or starved. However, in a parallel configuration, if a machine fails, the system only loses a portion of its productivity because alternative paths exist for parts to be processed. For this reason, parallel systems have higher expected throughput than serial systems. Table 1 gives some normalized expected throughput values for selected system configurations and two values of machine availability.

<table>
<thead>
<tr>
<th>Configuration Length</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine Availability</td>
<td>0.95</td>
<td>0.950</td>
<td>0.903</td>
<td>0.905</td>
</tr>
<tr>
<td>Availability</td>
<td>0.90</td>
<td>0.900</td>
<td>0.810</td>
<td>0.818</td>
</tr>
</tbody>
</table>

Table 1. Expected throughput of selected configurations.

The existence of crossover is another feature that has an impact on expected system throughput. Parallel systems with crossover have higher throughput than parallel systems without crossover due to the machine backup effect of the crossover. Thus, a parallel line with crossover has more possible paths for a part to take through the system than one without crossover because some machines may "back-up" failed machines. Therefore, failed machines have less opportunity to hinder the production of other machines. Values in Table 1 suggest that this effect, though significant, is relatively small compared to the comparison of parallel and serial systems.

4.3 Throughput Probability Distribution Function

Although expected throughput is a helpful metric, it does not give complete information about how a system may perform daily. For example the serial transfer system shown in Figure 5(a) may operate with a long run expected throughput of 65.6% of gross, with a gross production rate of 100 parts per hour. This means the expected throughput is about 66 parts per hour. Unfortunately, at any given time, the system runs at 100 parts per hour or zero parts per hour. Even if the customer demand is less than 66 parts per hour, the system may not be able to supply the customer anything for long periods of time (e.g., hours, or days). Therefore, it is
important to look at a system's throughput probability distribution to determine the probability that the system will meet demand requirements on a daily basis.

Just as with expected throughput, a parallel line has a more advantageous throughput probability distribution function than a serial line. Similarly, parallel lines with crossover have the same distribution but with slightly higher probability than those without crossover. Figure 5 illustrates this point with a comparison of four systems: (a), (b), (c), and (d). All are assumed to have the same gross throughput of 100 parts per hour, and all have the same individual machine availability of 90%. Note that the difference between the use or nonuse of crossover is again small when comparing systems (b) and (c).

Overall, our research suggests that for the same number of machines, short parallel systems with crossover will have the highest gross throughput, the highest expected throughput, and the best throughput probability distribution. Short parallel systems without crossover will perform slightly worse in terms of expected throughput and throughput probability distribution. Finally, long serial systems will have the lowest gross throughput, the lowest expected throughput, and the worst throughput probability distribution functions.

5 SCALABILITY
In [1] and [4], the topic of capacity scalability is introduced. We define scalability as the ability to adjust the production capacity of a system through system reconfiguration with minimal cost in minimal time over a large capacity range at given capacity increments. Designing manufacturing systems with the characteristic of capacity scalability enables management to increase or decrease production capacity quickly and cost-effectively in response to market demand. Scalability requires the investment of some extra capacity more rapidly. Examples of reconfigurable machine tools can be found in [1, 4, 5, and 6].

5.1 Effect of Configuration on Scalability
It was shown in [1] that system configuration has an impact on capacity scalability. Specifically, shorter parallel configurations have the ability to scale up in smaller increments, since each operation includes more machines in parallel to match the required capacity. That is, short configurations have fewer operations with more tasks assigned to each machine, and each machine has a high cycle time. Since a high cycle time translates to a low capacity, each machine added in parallel adds a small increment in capacity. This yields smaller increments of better resolution for capacity scaling. However, one must also take into consideration that these machines are more expensive. For the same reasons, long serial configurations have the worst scalability increments. Table 2 compares the smallest capacity increment for the systems in Figure 1(a), (b), (c) and (d) as a percentage of the total system throughput. Note that crossover has no effect on scalability.

<table>
<thead>
<tr>
<th>Minimum capacity increment</th>
<th>100%</th>
<th>7%</th>
<th>33%</th>
<th>33%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machines per increment</td>
<td>15</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 2. Capacity increment comparison.

The following two principles summarize design for scalability:
1. As the configuration width increases, it becomes easier to add small capacity increments.
2. Adding system crossovers does not affect scalability features.

5.2 Multi-spindle Scalable Machine Tool
As an alternative to adding machines in parallel for scalability, we introduce a new multi-spindle reconfigurable machine tool concept designed specifically for scalability. Figure 6 is an illustration of the proposed machine tool. Note that it is designed with the option to have between one and four spindle units. While today it is possible to buy two-spindle or even three-spindle machines, these machines are not reconfigurable and the buyer must purchase all three spindles even if they are not needed at the time of purchase. In contrast, with the proposed multi-spindle machine, a single-spindle machine can be purchased at the outset and later converted to a two-spindle machine when the market requires higher product demand. This has the same impact on capacity as if a whole single-spindle machine had been added to the system. But the change occurs at a lower cost than that of purchasing a whole new machine.

In addition to improving scalability, multi-spindle reconfigurable machine tools have the following impact on machining systems:
1. Multi-spindle reconfigurable machines increase individual machine throughput and can reduce system investment cost for a given demand requirement.
2. Multi-spindle reconfigurable machines reduce reconfiguration time, because spindles are easier to install/remove than bases of machines.
3. Multi-spindle machines reduce system floor space, because fewer machines are required to meet capacity.
4. On the other hand, multi-spindle machines will be less reliable than single spindle machines because of increased complexity.

5. In conclusion, the optimal design of scalable systems that include multi-spindle reconfigurable machines is dependent on the system design in previous and successive time periods.

![Multi-spindle reconfigurable machine tools](image)

Figure 6: Multi-spindle reconfigurable machine tool.

6 SUMMARY

In this paper, we introduced principles related to the impact of machining system configurations on various aspects of manufacturing and production. The importance of this research lies in its dealing with varying configurations of the machining systems. This is as opposed to the traditional studies that assume a certain given configuration [7]. The results obtained are important to industry as they can serve as a type of guidelines for designing machining systems configurations.

Overall, for the same number of machines, pure parallel configurations have the highest gross throughput, the highest expected throughput, and the best throughput probability distribution. In addition, they have the best characteristics for scalability. Yet, parallel systems have more quality streams, which lead to a deterioration in the uniformity of the products. Pure parallel systems fit relatively low-volume scenarios. However, even in these cases the feasibility of pure parallel systems depends on part complexity, task precedence requirements and machine accessibility fit to the part structure. Next to pure parallel systems, parallel systems with crossover have the best performance in each area discussed, except for the number of large quality streams. Parallel systems with crossover are followed closely by parallel systems without crossover. Finally, pure serial systems have the worst performance with respect to throughput and quality.

Fewer machines need to be purchased with short parallel configurations because of their improved throughput, thus lowering investment cost. But the higher cost of these machines with more axes of motion may offset gains in system investment cost. Also, shorter configurations require more cutting tool investment. Therefore, cutting tool costs will also offset gains in system investment cost. Consequently, it is important to consider the cost of machines and cutting tools in the system configuration decision.

The multi-spindle and other reconfigurable machine tool types may offer significant advantages in scalability, investment cost, and floor space. However, the reliability of multi-spindle machines may offset some of their benefits. Furthermore, scalable system designs that include multi-spindle reconfigurable machines will require new design methodologies. These design methodologies should include numerical methods for finding optimal configurations that minimize investment and reconfiguration costs.

Finally, when considering the issues of reconfiguration and scalability, the structure of the control system must be considered. We have encountered cases in which the changes needed in the control system were the main impedence to reconfiguration. Open-architecture systems [8,9] have the potential of reducing reconfiguration time and cost, and should be considered as an enabler to system reconfiguration for scalability.

The findings of this research indicate that more flexible, reliable and productive systems should be based upon shorter lines and more flexible, reconfigurable machines. Subject to technological constraints (e.g., precedence graphs, physical space available, etc.), optimal systems can be obtained by trading of investments in machines, cutting tools and quality streams against shorter lines. Furthermore, the layout of the system and its final performance would be strongly affected by the material handling systems. Thus, flexible and configurable handling systems (without buffers) are becoming a more and more important factor in configuration design.

7 ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Engineering Research Center for Reconfigurable Machining Systems (NSF Grant EEC95-92125) at the University of Michigan and the valuable input from the Center's industrial sponsors.

8 REFERENCES


