1. Introduction

Manufacturers today face more challenges than ever before due to the highly volatile market, which creates large fluctuations in product demand. To remain competitive, companies must design manufacturing systems that not only produce high-quality products at low cost, but also respond to market changes in an economical way [1].

Reconfigurable manufacturing systems (RMS) have been suggested by Koren et al [2] as a solution to address the needs for meeting the changing product demands. This has been recognized and supported later by other researchers [3-7]. From the viewpoint of RMS, a manufacturing system should be designed in such a way that it can be rapidly and cost-effectively reconfigured to the exact capacity needed to match the market demand. The capability of manufacturing systems to adapt their throughputs to changing demands is called scalability.

Scalability is an important system design characteristic in markets with volatile demand, and its cost-effective solution requires knowledge from engineering and business [8]. Researchers at the ERC/RMS [9], have addressed system scalability since the late 1990's [10], and issued a patent that deals with strategies to change production capacity in reconfigurable manufacturing systems (RMS) [11]. They developed one of the first algorithms that address capacity scalability [12], but this early algorithm was limited to upgrading the capacity of serial lines only. A more comprehensive approach was presented in [13] where scalability was analyzed as one of the critical issues in designing large, complex machining systems. Capacity scalability may be also achieved by scaling the capacity of individual pieces of equipment [5,6,14,15,16], but the most practical approach to system scalability adding machines to existing manufacturing systems, and in this cases the original system layout design is critical for achieving cost-effective scalability [17].

A dynamic model for capacity scalability analysis in reconfigurable manufacturing systems is introduced in [4]. This dynamic model is associated with minimizing the delay in scaling the system's capacity and thereby improving the RMS performance in response to sudden demand changes. However, in this current paper we deal with optimizing the original system layout [3,5] such that adding machines when needed by the market demand will be done quickly and cost effectively. Simultaneously with adding machines, also the material handling system must be adapted to serve the new added machines. There are cases in which several AGVs form the material transport system [18], but although AGVs facilitate the part transfer to and from the new machines, AGVs are expensive and slow, and therefore are not regarded as a cost-effective solution. There are cases in which RMS are designed to produce several products simultaneously [19, 20]. In such systems the capacity design issue is more complex and it is beyond the scope of this paper.

With the advancement of machine technologies over the past decade, the production of medium-to-high volume, large size mechanical parts, such as automotive powertrain components, has undergone a transformation. Dedicated transfer lines with dedicated machine stations are being replaced with systems composed of flexible CNC machine tools. As shown in Fig. 1, this system architecture is composed of multiple parallel CNC machines at each stage, with all machines performing exactly the same machining tasks, [21]. Such configurations of parallel identical machines in each stage, with material transfer between the stages (also known as crossover) improve throughput and reduce work-in-process inventories [22].

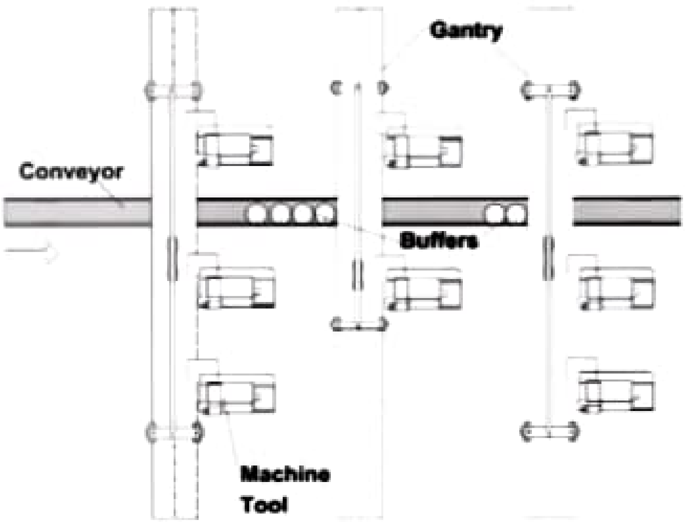


Figure 1: Schematic Layout of Production Lines with Conveyors and Gantries

Each manufacturing system is designed with a specific capacity in mind to fulfill a planned forecasted demand. However if the forecast for an annual product sale is between 250,000 and 300,000 units, marketing dictates building a capacity for 300,000 units. Therefore, even if a system is optimally designed, capacity may be still wasted when the real demand is significantly lower than the full planned capacity. When considering the entire life cycle of a manufacturing system the periods in which the system is operated at the full capacity are usually short [23]. If, however, the investment in the excess capacity (for 50,000 units in this example) could be delayed until it is actually needed, the system lifetime cost can be significantly reduced. A system design-for-scalability means that a manufacturing system is designed in a way that enables a rapid capacity upgrade to meet a larger demand, exactly when needed.

This paper introduces a practical design-for-scalability method for reconfigurable manufacturing systems comprised of reconfigurable and/or CNC machines tools. A scalability planning methodology is presented to determine the most economical way to add machines to an existing system to match a new market demand. It does so through concurrently changing system configuration and rebalancing the system. The reminder of this paper is structured as follows: Section 2 defines the system scalability and describes the concept of incrementally scaling system capacity. Section 3 introduces a mathematical formulation to minimize the total number of machines to be added by concurrently reconfiguring and rebalancing the system. Section 4 proposes heuristic algorithm based on a genetic algorithm. Section 5 presents case study to validate the proposed approach. Conclusions are presented in Section 6.

2. Defining System Scalability

To adapt the throughput of manufacturing systems to the fluctuations in product demand, the system capacities must be adjusted quickly and cost-effectively. Capacity scalability of manufacturing systems is a necessary characteristic needed for rapidly adjusting the production capacity in discrete steps, allowing thereby a given system's throughput to adjust from one yield to another to meet changing market demands. We define system scalability, in percentage, as:

System Scalability = 100 – smallest incremental capacity in percentage

If the minimal capacity increment by which the system output can be adjusted to meet new market demand is small, then the system is highly scalable. For example, if a serial line (Fig. 3a) needs to increase its production capacity to satisfy a larger market demand, an entire new line must be added. The step-size of this addition doubles the production capacity of the system. Mathematically, the minimum increment of adding production capacity in a serial line is 100% of the system, i.e., adding a whole new line, making the scalability of a serial line 0%. Doubling the line capacity will be expensive because there is no guarantee that the extra capacity will ever be fully utilized, risking a substantial financial loss. Thus, zero scalability means that in order to increase the system capacity, the entire production line must be duplicated.

Dedicated lines do not have scalable capacity and cannot cope with large fluctuations in product demand. This challenge can only be met by flexible or reconfigurable manufacturing systems which are composed of singular CNC machines, as these systems are scalable in small increments accomplished by adding individual machines can be added as a need arises.

Similar scalability calculations for the other systems in Fig. 3 show: Configuration **b** has a scalability of 50% and Configuration **c** has 67%. Configurations **d** and **e** have a scalability of 84%; the highest possible for 6-machine configurations. A minimum increment of only one sixth of the system (16%) —in these cases, one machine— can be added to increase system capacity; for example, a machine can be added to stage 2 of Configurations **d** as shown in Fig. 3d.

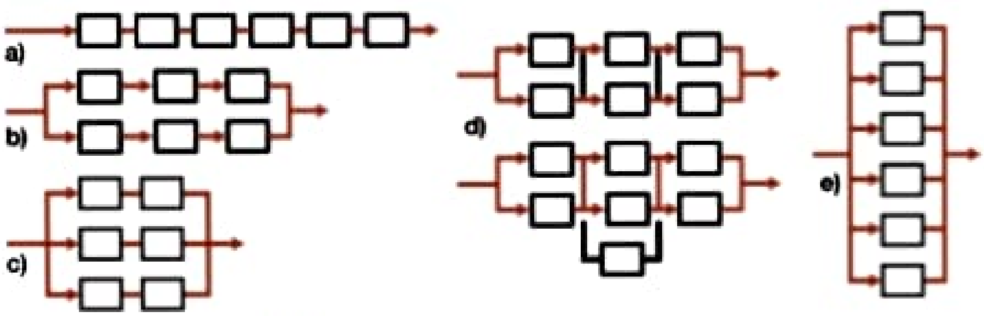


Fig 3. Five scalable configurations

In this example, the configuration depicted in Fig. 3c of two stages with three machines per stage, might be a compromise between reasonable scalability and investment cost. In this case, if a product requires machining on both the upper and side surfaces, the three machines in the first stage might be 3-axis vertical milling machines, and the three machines in the second stage might be 3–axis horizontal milling machines. Conversely, in a parallel system, all six machines in Fig. 3e must be 5-axis milling machines – making the system much more expensive. In the system in Fig. 3c, capacity scalability must be performed in steps of 33.3% by adding one vertical machine and one horizontal machine, rather than in steps of 16.6% as with the parallel configuration. Adding a step of 16.6% in Fig. 3e in practice means adding one 5-axis machine with a large tool magazine that does the whole part processing.

To conclude, in general, the smallest scalability adjustment steps can be accomplished when the original system is purely parallel (e.g., Fig. 3e). However, the initial cost of a parallel system is the highest of all system configurations. In parallel configurations, each machine must perform all the manufacturing tasks needed to complete the part. Therefore, each machine must have the entire set of tools needed to produce the whole part and should also be able to perform more functions, for which more axes of motion are needed. As a result, the capital cost per additional volume increment added to a parallel configuration is the highest of all configurations.

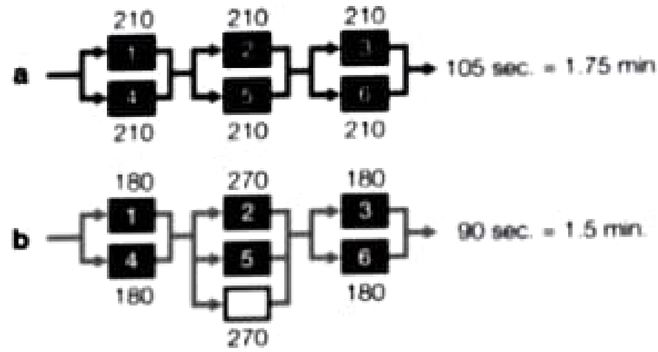
The following example clarifies the option of adding a small incremental capacity.

Example: On a system composed of six machines, as shown in Fig. 4, we have to process a part that requires 21 machining tasks of 30 second each, totaling 630 seconds, or 10.5 minutes, needed to machine each part. The required demand is 274 parts per 8-hour shift, namely 480 minutes. Therefore, the required cycle time is 480/274= 1.75 minutes/part.

a. Design a scalable system configuration.
b. After one year, the demand has grown, and 320 parts per shift are needed, reducing the cycle time per part to 1.5 minutes/part. How many machines should be added, and what is the new configuration?

The cost-effective scalable system configuration is depicted in Fig. 3d, which is shown in detail in Fig. 4. Here, each machine does seven tasks, totaling 210 seconds per machine. When the demand grows to 320 parts/day, seven machines are needed. Only Configuration d yields the cost-effective solution by adding the new machine to Stage 2. One task is shifted from Stage 1 to Stage 2 so each machine in Stage 1 operates for 180 seconds on the part, and another task is shifted from Stage 3 to Stage 2; each machine in Stage 2 will then operate for 270 seconds on each part, as shown in Fig. 4b.

Fig. 4. When demand grows, the initial system, ***a***, is cost-effectively scaled-up to Configuration ***b*** to meet the new demand



The initial capital investment in the system configuration in Fig. 4 is a bit higher than one in a serial line, because the material handling system is more complex. However, the extra capital investment is similar to buying an insurance premium for a future event which is likely to occur. If the demand does rise, the system can easily be scaled up and the new demand can be supplied in a short time, at a minimum additional investment. If the demand is unchanged during the lifetime of the system, a small capital investment on the more sophisticated material handling system was lost.

In this example, twenty-one equal tasks were needed to complete the part and the system with three stages and two machines per stage was perfectly scalable. In general, we will obtain similar scalability results in a symmetric configuration which has m stages and n machines per stage if the number of equal tasks needed to complete the part is:

$$(n.m + 1)m.$$

System design Fig. 4 the system designers must leave an empty space reserved for possible addition of a seventh machine and an extended material transport system to the spot.

3. Formulation of Scalability Planning Problem

We propose below a method to determine the most cost effective system reconfiguration to meet a new market demand. To perform system scalability planning, many factors need to be taken into consideration. These include a detailed process plan, setup plan, the machine capability, and the number of spots reserved for adding machines at each stage of the original system configuration. When reconfiguring an existing manufacturing system, simultaneous reconfiguration planning and system rebalancing attempts are needed to maximize the capacity of systems. In this section, an optimization model is proposed for the scalability planning. The solution to the model will be discussed in section 4.

3.1 Assumptions

The following assumptions are made based on the current industry practice in he powertrain industry.

- A multi-stage system with configuration as shown in Fig. 1 is considered. Parts are moved from one stage to another through conveyors and delivered to different machines within a stage using gantries.

- The number of stages will remain unchanged during the reconfiguration process. This way, the system setups will not be changed to avoid adjustment of the process plan, thereby minimizing impacts on the product quality.

- All the machines within the same stage perform exactly the same tasks.

- There are reserved spaces to add new machine in each stage and the gantries can be extended to deliver parts to the newly added machine(s).

3.2 Inputs

Scalability planning requires the following four types of inputs:

- <u>Configuration information</u>
  Number of stages L;
  Number of machines in each stage N_i, where $i = 1,2,...,L$;
  Maximum number of machines allowed in each stage M_i, $i = 1,2,..., L$, which is restrained by the capability of material handling system.

- <u>Stage Characteristics</u>
  Each manufacturing stage usually has limited capabilities which are defined by a group of key characteristics of the stage. These include machine tool capability such as functionality, power, accuracy and machining ranges, and fixture capability such as face accessibility, which defines the faces that are accessible by the cutting tool. When a set of tasks are assigned to a stage, the necessary capabilities must fall into the key characteristics of the stage, otherwise the task allocation is invalid.

  Assuming the number of key characteristics of each stage is K, a capability matrix S stores all possible key characteristics of each stage.

  $$S[i,j] = d : d \text{ is } j_{th} \text{ key characteristic of stage } i \text{, where } 1 \le i \le L, 1 \le j \le K$$

- <u>Manufacturing tasks</u>
  - *Task Precedence Tree*: manufacturing tasks must be performed at a certain order. This tree defines sequential constraints between tasks. Each task in the precedence tree can only be performed after all its parent tasks have been completed. A two-dimension binary matrix $Pre[1 \cdots N, 1 \cdots N]$, where N is the number of tasks to be processed, is used to represent the precedence tree.

    $$Pre[i,j] = \begin{cases} 1, \text{if task i must be performed before task j} \\ 0, \text{otherwise} \end{cases}$$
  - *Task Key Characteristics:* These include task type, access direction, dimension, accuracy and power needed to perform the task. For a task to be assigned to a stage, its key characteristics must fall in the key characteristics of the stage. Assuming the number of key characteristic of each task is R, a task key characteristic matrix K is used to store the key characteristic of each task.

    $K[i,j] = f : f$ is the j_{th} key characteristic of task i. Where $1 \le i \le N$ and $1 \le j \le R$.