Designing Unbalanced Assembly Lines: A Simulation Analysis to Evaluate Impacts on Work-In-Process Results

Rogerio Flores Da Silva*, Renelson Ribeiro Sampaio**, Francisco Uchoa Passos***
*(Ppgmcti, Faculdade Senai-Cimatec, Bra)
** (Ppgmcti, Faculdade Senai-Cimatec, Bra)
*** (Ppgmcti, Faculdade Senai-Cimatec, Bra)

ABSTRACT
This article investigates the impact of controlled imbalance levels on assembly lines, and its effects on two important performance indicators: throughput and work in process (WIP) level. Using a five workstations line simulation, with different degrees of imbalance and different configurations, we could conclude that there is a relationship between extra capacity added to non-constraints and average WIP level and line throughput. Simulation revealed that, using bowl shape configuration, the higher the imbalance, the higher the throughput, with less WIP. These results allow proposing new studies to create a framework for evaluating the feasibility of investments in extra capacity vis-a-vis those gains in resources efficiency.

Keywords: assembly lines, capacity, simulation, unbalanced.

I. INTRODUCTION
Designing production systems and allocating capacity has always been an important issue in industrial engineering [1]. And we cannot discuss the performance of the assembly line in a broad perspective if we exclude throughput and inventory results. As a matter of fact, designing an assembly line is extremely relevant either for productivity [2] as for inventory control [3]–[7].

In systems design there are different approaches on specifying a new assembly line, with distinct impacts on throughput and inventory. A possible way of differentiating these approaches considers the form that capacity is distributed and used in the assembly line [8]. So, one of the basic issues in the assembly lines design is the amount of capacity which is required in order to achieve the expected throughput.

It’s important to pinpoint that an assembly line consists of several workstations in series, with different configurations, but mostly with buffers in between the workstations. So, another issue addressed is the amount of buffer that will be necessary to guarantee the operation of the line. As the capacity usually refers to expensive and costly resources, like equipment or labor force, the process of designing an assembly line tries to find the minimal necessary capacity. But the same concern refers to the buffers, which are costly as well. This brings up the concept of assembly line balancing problem – ALBP [9].

The ALBP consists of optimally partitioning the total workload among the workstations with the objective of minimizing total idle time, therefore, maximizing resources efficiency [10]. So, the main objective of production line designers is to increase the efficiency of the line by maximizing the ratio between throughput and required costs [11].

The problem is that most of the real processes are not like the truly ideal balanced line and the probability of achieving perfect balance or even near-balance is low. Imbalance occurs because of different mean times among workstations, different variabilities, failure, etc. [12].

So, despite of considerable effort on ALBP research, [9], [11], [13]–[16], a lot of work has pointing to better performance results with unbalanced lines [6], [8], [10], [12], [17]–[25].

This paper uses a simulation to evaluate the impacts of unbalancing assembly lines on work in process - WIP inventory and line throughput. Using a simple five steps assembly line, an analysis is made on the imbalance level and consequences on work in process inventory and line throughput. The next section addresses the problem exploring some important research already done on imbalance capacity and the measurements that are going to be used to quantify imbalance level. Then, the subsequent section presents the structure of the simulation and how the questions are organized. The results of the simulation and a proposal of an algorithm to help production designers are then stated. Finally, conclusions are presented, with suggesting topics for future research.
II. UNBALANCED LINES

The idea of using unbalanced line is far from new. In less than a decade after the definition of the assembly line balancing problem [9], some scholars have already supported the idea of using non balanced configurations [26]–[28]. The unbalanced line is, of course, that one which is not perfectly balanced. It can happen by different mean cycle times - CT, variability or breakdown among workstations. Naturally, real world operations are more frequently found in unbalanced, unpaced and asynchronous lines than balanced, paced or synchronous ones.

Nevertheless, a major milestone in this field appears when an analysis of lines having up to four stations with exponential work-time distributions revealed that optimal throughput was obtained with unbalanced lines. In that case, the best configurations were when slower stations were positioned at the end of the line and the faster ones towards the center [29]. This was called the bowl phenomenon. This counter-intuitive finding [12] was followed up by several investigations analyzing various conditions of imbalance, line settings and distribution of frequency, among other variations.

Some controversy and conflicting results were pointed out during bowl phenomenon studies [30], most of them related to the production rate, idle time of the work or suitability to real scenarios. In some cases, inverted bowl shape was suggested as a better configuration [31], [32].

Most of the studies were done analyzing scenarios with lines up to eight stations [6], [33]–[46], but there were analysis evaluating longer lines as well [47], [48]. Some studies focused on buffer allocation [33], [35], [36], [38], [42], [47]–[52], others had the objective of maximizing the throughput [6], [37], [43], [48], others still analyzed cost of work in process [40], [42].

In terms of methodology, big majority of the mentioned studies used mathematical analysis and algorithmic development, while others used simulations to test their hypotheses [6], [41], [43]–[46].

It’s is important to mention that Just in Time – JIT and Kanban principles also considers that idle time is not the main problem in an assembly line. As long as the flow of material is continuous, line balancing is not essential. Extra capacity in some stages is usually considered in JIT lines. For example, [53] argue that Japanese managers allow 12% to 18% additional capacity in their production systems to guarantee Kanban. Another methodology that uses the imbalance principle in its premises is theory of constraints – TOC [25], [54]–[56]. The idea of managing some few constraints in the production line implicitly establishes the existence of unbalanced lines, with bottlenecks and non-constraints. TOC proposes the drum-buffer-rope as a methodology which requires a “drum” with a smaller capacity, even if we need to force the existence of this bottleneck [57].

Despite the unquestionable advantages of unbalanced lines in specific conditions, most of the research about assembly line design is still focusing on the traditional ALBP, mainly due to prevalence of the fordist strategies [12], [42]. As human resources are always involved in assembly lines operation and in order to support production in the twenty-first century, with its speed, flexibility and performance needs, more must be known about how to cope with line imbalance for a range of performance objectives.

2.1 Measurements: quantifying the imbalance

Evolution of ALBP studies demanded sets of measurements in order to evaluate effectiveness of each new method, so we can use some of these measurements also to analyze unbalanced lines. There are typical measurements already proposed in earlier studies [13]–[15], [58]. Two groups of measurements can be highlighted on these studies:

- Measurements of efficacy – Assessment measurements;
- Measurements to evaluate how difficult is the problem to solve – Measurements of difficulty.

These measurements are used to evaluate the efficacy of a method for ALBP, categorizing the types of problem according to the level of difficulty.

We are not going to use measurements of difficulty in this work, since our proposal is not balancing the lines. Assessment measures evaluate how balanced is the line after solved the problem, or finished the design. Line efficiency (LE) is a measure of the percentage line utilization, where CT is the cycle time, m is the number of stations and S, the total time required for executing the tasks assigned to stations:

\[
LE = \sum_{i=1}^{m} t_{av}
\]

Balance efficiency (BE) is a measure of how evenly the work is distributed to workstations. Here \( t_{av} \) represents the average workstation time across the entire line. For ALBP, the goal is to get \( BE = 100\% \), corresponding to identical workstation loads.

Smoothness index (SI) is a measure of relative smoothness of a given line, after balanced. Here, for ALBP, the goal is to get \( SI = 0 \), indicating a perfect balance. Some authors argue that SI, being dimensional, is influenced by individual problem values, making inter problem comparison meaningless [58].
As the capacity of a workstation can be obtained by the inverse of CT, these measurements also work in reverse; the higher the imbalance, the smaller BE and SI.

To evaluate impacts of imbalance, we are going to calculate BE using constraint cycle time, instead of the average cycle time of the line. As the constraint cycle time is the one which limits the throughput of the line, it’s so important to this work having this cycle time as the reference value.

III. SIMULATION

Research on balancing or unbalancing assembly lines has used simulation models very often [6], [41], [45], [46], [59]. Simulation is a technique for performance and reaction analysis of a system [30] and what made it very useful for assembly lines is exactly the difficulty of producing analytical models for real world lines. Another benefit of the simulations is to allow different reactions and systems behavior, making it possible to construct hypothesis and analyzing the impact of each variable alone. This is exactly what is expected on this work.

The idea is to model a game used on theory of constrains – TOC tool teaching: the dice game. The dice game [60]–[63] is used to present the principles of drum-buffer-rope – DBR technique on trainings and classrooms. Basically, it simulates a five step simple process, with stochastic behaviour, represented by the dice. It was chosen to this work for its simplicity, yet easy-to incremental progression to more complex models.

At the first part of the game, participants are stimulated to forecast the productivity of the line after 10 production periods in a 5 workstation line. Each station line represented by one ordinary six-sided dice and each production period represented by one roll of the dice, as shown on Fig. 1. Triangles represent the buffer and workstations, the dice, represented by squares.

![Figure 1 Dice game structure](image)

As each throw has a random result varying from 1 to 6, obtaining the expected value close to 3.5 (average result), it’s expected to have a 35 pieces produced after rolling the dice 10 times at each station. But, due the variation and dependency, it’s impossible to obtain 35 pieces at the end of the line. As the game is played, WIP inventory tend to increase, bottleneck tend to change randomly and throughput remains below 3.5 pieces per roll.

At the second part of the game, phase two, the line is purposely unbalanced, adding an extra dice to four of the five stations. Therefore the capacity of the line remains the same: 3.5 pieces per roll or thirty five pieces after ten rolls. This capacity is due the limitation of bottleneck capacity. Note that this configuration is almost equivalent to an inverted bowl shape, because constraint is located in the center of the line. This configuration is used only to simplify the game, creating a single bottleneck, instead of two bottlenecks for the bowl shape.

Using two buffers to manage work in process – WIP inventory, the second part of the game allows delivering consistently thirty five pieces after ten rolls. The buffer management rule is simple and consists on establishing a maximum level to each buffer. For example, eight pieces as the maximum. If the buffer level is higher than the maximum, eight pieces, upstream workstations must be blocked. If buffer level is equal or smaller to eight pieces, the workstations must remain working. This limit is managed according to the line service level results.

Even though it is of great use for teaching purposes, the dice game does not allow further analysis because its settings are rather limited. So, in order to simulate the line, it will be used Witness Software representing dice game structure. Due to software limitations, buffer management will be done using upper limits to each individual buffer, testing each configuration with different limits.

Although the dice game is based in a random distribution, in order to be more realistic, the workstation distributions will be represented by exponential distributions considering all service times independent [42], [64]. Initially we are going to use CT = 10 minutes. Problem parameters during the simulations are:

- **AIL** = average inventory level by workstation during the simulation;
- **BE** = balance efficiency;
- **BLj** = Buffer upper limit for individual workstation \( j (j = 1, \ldots, n) \);
- **CTb** = bottleneck cycle time;
- **CTi** = non bottleneck cycle time \( i (i = 1, \ldots, m) \);
- **LE** = line efficiency;
- **LT** = Line throughput; number of pieces delivered after simulation;
- **IL** = inventory level at the end the simulation;
- **SI** = smoothness index;
- **Time** = time used to run simulation.

---

**Problem parameters during the simulations are:**

- **AIL** = average inventory level by workstation during the simulation;
- **BE** = balance efficiency;
- **BLj** = Buffer upper limit for individual workstation \( j (j = 1, \ldots, n) \);
- **CTb** = bottleneck cycle time;
- **CTi** = non bottleneck cycle time \( i (i = 1, \ldots, m) \);
- **LE** = line efficiency;
- **LT** = Line throughput; number of pieces delivered after simulation;
- **IL** = inventory level at the end the simulation;
- **SI** = smoothness index;
- **Time** = time used to run simulation.
To ensure that the effects of line startup do not influence the results, time to run simulation will consider a period to stabilize WIP and line throughput. The initial assumptions are:

- All the non-bottleneck have equal but independent CT and, the same CT distribution and same variability;
- The line produces a single type of product;
- Each task can be performed only in its specific station;
- Bowl shape will consider two identical bottlenecks, respectively at the beginning and the end of the line;
- Inverted bowl shape will consider a single bottleneck, at the centre of the line.

After this first running, another phase of the simulation will be configured using bowl shape, i.e. bottlenecks at the beginning and at the end of the line and progressive difference between cycle times. The bottlenecks will use a ten minutes cycle time and the non-constraints a configuration according to Table 1. The analysis will be done with fifty points, with different balance efficiency – BE.

Considering these previous conditions, the results of each step and each configuration of the simulation will be tabled and the results are discussed in next section.

IV. RESULTS

The model was structured with five workstations and five subsequent buffers, being the last buffer a finished product buffer. In order to represent client need in a similar configuration of dice game, it was added a last workstation, named client. The structure of the simulations is shown in Fig. 2.

Figure 2 Simulation structure

The results presented in Fig. 3 and Fig. 4 show line throughput – LT and average WIP evolution during time. They reveals that LT, in fact, is higher for unbalanced lines, according to previous studies [12], [21], [65]. Regards to WIP, unbalanced lines, with bowl shape, have a similar behaviour to balanced lines, although better results in terms of WIP level.

Figure 3 Line throughput level comparison

Figure 4 Average WIP level comparison

Inverted bowl shapes tend to have better results as well, but are more dependent of a better buffer management. Using this configuration note that, if buffer limit is higher, WIP level can exceed the level achieved with balanced lines. This is motivated by excess of capacity on non-constraints with no limit of inventory and makes the inventory management indispensable. As the first stations has excess capacity, it’s natural that buffer accumulate prior the bottleneck depending only on the buffer limit (BL).

But the most important result of this phase is when changing level of imbalance affects WIP average. Fig. 5 shows inventory behaviour with different CT on non-constraints. As bowl shape had better performance in terms of more regular WIP profile, we tested only bowl shape lines with different levels of imbalance. Results demonstrated that as we increase the level of imbalance, less WIP is found in line. We can assume that it’s due to the increase of inventory turns.

Figure 5 WIP level with different MT and BL

WIP level, on the other hand, is not strongly influenced by buffer limits, but demonstrates a slightly increase when buffer limit is increased.

These behaviours pointed out to some perceptions:

- Buffer limit is very important with adopting inverted bowl shape, in order to avoid excess of WIP inventory. But bowl shape already limits this possibility using a bottleneck at the beginning of the line;
- Inventory level on perfect bowl shape tend to increase only before last station, but the average WIP tend to decrease as imbalance gets higher;

The last phase of the simulation used a long period of time, 2400 minutes, enough to stabilize the results of WIP. The bowl shape was tested in different levels of balance efficiency – BE, according to TABLE 1. Fig. 6 represents the evolution of WIP level and throughput when imbalance level increases. Note that average WIP reduced from 2.65 pieces by buffer (left axis), when the line was perfectly balanced (BE = 100%), to 1.25 by buffer with maximum imbalance (BE= 60.8%). The effect was positive in throughput as
well, increasing from 173 to 184 pieces (right axis) produced in 2400 minutes.

<table>
<thead>
<tr>
<th>Step</th>
<th>Bnk 1</th>
<th>Bnk 2</th>
<th>Bnk 3</th>
<th>Bnk 4</th>
<th>Bnk 5</th>
<th>Client</th>
<th>BE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>9.9</td>
<td>9.9</td>
<td>9.9</td>
<td>10</td>
<td>99.2%</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>9.8</td>
<td>9.6</td>
<td>9.8</td>
<td>10</td>
<td>98.4%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>9.7</td>
<td>9.4</td>
<td>9.7</td>
<td>10</td>
<td>97.6%</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>5.1</td>
<td>0.2</td>
<td>5.1</td>
<td>10</td>
<td>60.8%</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 6 WIP and throughput evolution**

If imbalance level, here represented by different MT in non-constraints, clearly impacts inventory level and throughput, managers can use extra-capacity to reduce WIP on their industries. Depending on the inventory cost and extra-capacity cost, why not adding extra capacity either by equipment investments or increasing labour force to reduce WIP?

This hypothesis was already addressed [66], with the proposal of an analysis on the return of investment – ROI, comparing the investment on extra-capacity on non-constraints to the reduction in queue time and inventory cost and the increase in gain.

\[ (4) \]

So, if extra capacity on non-constrains reduces WIP level, it’s possible to measure impact on cost reduction and compare with investment done. The question to be answered is: How much resources efficiency is obtained by imbalance level?

**V. CONCLUSION**

The use of simulation can bring useful insights for assembly line designers, especially in terms of workstations capacity. Simulating different models can, not only reveals trends, but presents important clues to establish relationships between line configurations and line performance.

Most of the research done in this field uses simpler models, what makes both balancing and unbalancing processes distant from practical works. A study pointed out that, regard to the astounding academic effort, with 312 different papers about balancing problem, only in 15 were identified work with real world assembly lines [67]. This reveals a huge gap between research and real world configuration problems. Thereunto, simulation models have a great utility, because it is easier to modelling more practical scenarios.

Using simulation in a five stations single line, we could note a clear influence of imbalance level, when using bowl shape, on WIP level and throughput. These findings reveal a possibility of specifying an algorithm to determine the reduction on inventory cost caused by investment in extra capacity.

This algorithm might consider the impact on return of investment (ROI) before and after the imbalance, evaluating impact on gain, due to the throughput increase and the impact on inventory cost, due to the WIP reduction. The impact of this algorithm might be very important not only on designing new assembly lines, but it can be considered on improving performance of existent ones as well.

Then, this work suggests a framework to help on investment decision for production assets based on the trade-off: cost reduction with less WIP and better throughput versus cost of investment in new assets. Future studies can bring a wide range of new possibilities on this extremely relevant field.

**REFERENCES**


[37] G. A. Vouros and H. T. Papadopoulos, Buffer allocation in unreliable production lines using a knowledge based system,


[64] F. S. Hillier and R. W. Boling, Finite queues in series with exponential or Erlang service times—a numerical

