The Integrated Room Layout for a Semiconductor Facility Plan

Jaewoo Chung and Jaejin Jang

Abstract—Layout study is important because layout largely determines the initial investment and production efficiency of a plant as compared with other downstream activities. Especially for semiconductor production, the continued miniaturization of chips requires increased investment and operating cost. While approximately one billion U.S. dollars was necessary for a new semiconductor fab in 1995, manufacturers today need to invest two to three times that for the same type of facility.

This research presents a new integrated room layout for a semiconductor fab. Machine tools are laid out in four large rooms. The tools in each room are connected by one overhead hoist transporter/overhead shuttle loop. This layout increases direct interbay transportation, product mix change flexibility, and routing flexibility and reduces transportation time and initial investment cost. The performance of this new layout is compared with the integrated bay layout that is favorably considered and being used in the literature and industry. This comparison uses data provided by International SEMATECH.

Index Terms—Layout, material handling, room layout, semiconductor.

I. INTRODUCTION

SEMICONDUCTOR fabrication (fab) for microelectronic devices, which is one of the most complex manufacturing processes in existence today, faces a variety of challenging and complex problems. Business strategies, market demands, and advances in process technology continue to make it difficult for a fab to integrate. Consequently, there will be continued pressure to improve the initial investment, the lot cycle time, equipment utilization, and production yield.

Layout study is important, especially for semiconductor production, because layout largely determines the initial investment and the production efficiency as compared with other downstream activities. In addition, layout is difficult to modify once it is set up. The continued miniaturization of chips requires increased investment. While approximately one billion U.S. dollars was necessary for a fab in 1995, manufacturers today need to invest more than two billion dollars [1] or more for the same type of facility.

This research presents the integrated room layout (IRL) for a semiconductor fab as an alternative to the bay-based layouts. The integrated bay layout (IBL) [2]–[6] and ballroom layout [7] are proposed by industry to increase the flexibility and productivity of the traditional bay layout. This research increases the flexibility and productivity of fabs even further.

The IRL mainly tries to reduce the complicated interbay transportation by putting more machine tools in rooms that are larger than bays and connecting the tools in the same room by using one large loop of combined overhead hoist transporter (OHT) and overhead shuttle (OHS). The indirect interbay transportation requires many steps of material handling by such devices as OHTs, OHSs, stockers, and intermediate parts. Putting bottleneck tools together can reduce material handling and increase the product routing flexibility and utilization of tools with less work in process (WIP).

The performance of the IRL is compared with that of IBL, which is favorably considered and being used in the literature and industry. This comparison uses models of processing tools, material handling systems, and product routing based on data from International SEMATECH [8].

Section II reviews the interbay transportation in a fab, Section III reviews fab layout research from academia and industry, Section IV presents the new IRL, and Section V describes an evaluation of the IRL.

II. FAB OPERATIONS: CROSSOVER AND INDIRECT TRANSPORTATION

Typical wafer fabrication has about 18 to 30 processing layers consisting of more than 300 total processing stages. Each layer needs one litho operation, which is the bottleneck in a fab [Fig. 1(a)]. Many different types of tool groups, which consist of about 500 processing tools, are repeatedly used to fabricate the layers. Material handling for front-opening unified pods (FOUPs) that carry wafers is needed between processing stages [Fig. 1(b)]. In a modern 300-mm fab, material handling is highly automated, using stockers, OHTs, OHSs, FOUPs, and mini-environments.

The bay-based layout often puts tools of the same type in multiple bays. Litho tools, each of which can fabricate only a few different types of layers among 80 to 100 layers from four or five product types simultaneously existing in a fab, are placed in multiple bays. When a litho tool does not have any lot to process, sometimes a lot is obtained from another litho bay. This crossover is one type of interbay transportation along with the interbay transportation between two different stages of processes. There are three possible path types of transportation between bays: 1) OHT > stocker > OHS > stocker > OHT; 2) OHT > stocker > OHT; and 3) OHT. Type 1), which requires two OHTs, one OHS, and two stockers, is the least efficient and...
of cell layout, the dedicated cell layout shows more benefit of reduced setup time, improved yield, and simpler material flow than the others. However, the cell layout lacks flexibility because it assigns tools into cells based on process routings of a few product types. Agrawal and Heragu [14] provide more reviews. Although simulation provides a lot of benefits, the long lead time of analysis and the corresponding costs have been recognized as main difficulties. Pillai et al. [15] and Mackulak et al. [16] present new approaches to reduce the lead time by using standard and dynamic simulation models.

In the literature, an analytical method has been proposed for a semiconductor fab facility plan to reduce the lead time. Hop et al. [17] explore a queuing network model using a two moment decomposition approach. They argue that a properly developed analytical model performs as accurately as the simulation model with much less effort. They focus on the capacity determination of toolsets. For general products (not only semiconductors), Solberg and Nof [18] analyze the performance of four typical layout configurations (product, process, carts, and conveyor layouts) using queuing network theory. Benjaafar [19] measures the performance of different layouts using WIP-based modeling. He shows that the layout obtained by the flow rate model using the quadratic assignment problem (QAP) [20] can be WIP-infeasible or poor in terms of WIP performance. He also shows that empty vehicle travel can be reduced by placing the most frequent tools together regardless of the amount of material flow between these tools. However, this research does not consider congestion of material handling vehicles. If placed together, tools with short processing times, such as the metrology tools, can cause a severe traffic jam and deadlock of automated transporters. Careful control logic is required.

B. Layout from Industry

The semiconductor industry also presents layout alternatives. While the academic research is mostly conceptual and often considers only a part of many important factors, industry presents more complete and practical layouts, though with less emphasis on methodology.

Pillai et al. [4] explain Intel’s layout and automated material handling for a 300-mm fab. They address very detailed and practical aspects of fab layout, such as building design, construction, tool layout, automation, production equipment, and fab operations. They also consider bay width, chase move-in width, and stocker location, demonstrating the advantages of integrated bays, such as reduced stacker demand, increased system reliability, reduced automated material handling system (AMHS) cost, and reduced interbay transportation. A ballroom layout is proposed by PRI Automation, Inc., which deletes walls of the traditional bay layout for operators to transfer carriers directly between tools without having to go through the center spine [7]. It is argued that travel distances can be reduced if carriers are moved diagonally instead of orthogonally.

Recently, the IBL, which connects two to three bays using one OHT loop, has gained popularity since the implementation of 300-mm wafer fabs. This layout is evaluated under different process flows and different levels of integration levels. The IBL is shown to have advantages in stocker utilization, the number of vehicles required, and the total traveling distance and lead-time.

III. LITERATURE REVIEW

A. Academic Research

Approaches for fab design can be classified into three groups: layout design methodology, simulation evaluation, and analytical evaluation on layout alternatives. A design methodology for a conventional bay layout is proposed [9]. The methodology consists of two steps of assigning tools to bays and determining tool locations within the bays by using Mixed Integer Programming and Dynamic Programming. Designing fab layouts and material handling systems are closely related, and Peters and Yang [10] integrate these two activities into a single step using the space filling curve (SFC). They also provide a new method to design the crossover turn table, which builds a short cut in the vehicle loops, using a network flow formulation. The Rectilinear Steiner Tree (RST) approach is used to develop optimal single-spine and double-spine overhead track layouts to minimize travel distance of wafers [11]. They conclude that the simplicity and shorter flow distance of the spine layout make it suitable for the 300-mm fab.

There have been many simulation approaches for the evaluation of the AMHS for different fab configurations. Hase et al. [12] and Geiger et al. [13] propose a few different types of cell layout that differ from each other in the degree of dedication of tools to the cells and compare them with a conventional bay layout. They compare alternatives under different levels of tool breakage, tool utilization, transfer time between stations, and tool setup times. In the analysis, among the different types of interbay transportation (see [6] for more details). Type 2), which does not need the center spine, is often used when the destination is an adjacent bay. Type 3), which uses just one OHT loop, is called direct transportation and is the most desirable. The inefficiency of types 1) and 2) reduces routing flexibility of a lot in a bay and sometimes causes starvation of a tool even when there are lots in a different bay that can be processed by the tool.

Most common for interbay transportation is the IBL, which connects two to three bays using one OHT loop, is called direct transportation and is the most desirable. The inefficiency of types 1) and 2) reduces routing flexibility of a lot in a bay and sometimes causes starvation of a tool even when there are lots in a different bay that can be processed by the tool.
of material handling devices. This layout is being used in industry. The conveyor layout utilizes the high transportation capacity of a conveyor [21]; however, it lacks transportation flexibility.

SEMATECH proposes a few types of fab layout in its phase I and phase II reports considering many important aspects of layout [22], [8]. The phase I report proposes layouts for 300-mm fabs, considering footprint, cycle time, inventory level, and intrabay OHT. It considers three different types of bay layout: farm layout (tools of the same groups are placed in the same or adjacent bays), hybrid layout (metrology processes are distributed over bays), and modified hybrid layout (equipment is laid out to facilitate processing four to six contiguous process stages in the same bay). The modified hybrid layout could very well be the configuration of choice. It has a 13% reduction in cycle time over the farm layout and requires the smallest footprint, fewest bays, and lowest factory area per wafer start. It is also the best in tool count, tool utilization, and inventory level. However, the hybrid layout is recommended in the report because of the inflexibility of the modified hybrid layout. The phase II report describes the follow-up activities of the phase I report. This report describes the environment used in setting up the experimental results. The models are run to compare the results of variations of the layout, nonproduct wafer loading, hot lot loading. down time of the automated material handling equipment, size of the internal buffers at batch tools, and the number of stockers.

The recent development of liquid crystal display (LCD) fab layout provides a good benchmarking model for the semiconductor layout research. The LCD fab also requires stacking layers of conductors, semiconductors, and insulators, and its processes are very similar to those of wafers except for what is shown in Table I. Originally, the layout of an LCD fab was the same as the layout of a semiconductor fab; the traditional bay layout with a center spine. However, its layout has been changed from the fourth generation where a generation is defined by the glass size. The difficulties in handling large glasses and space requirement forced LCD fabs to develop different types of layout.

The layout for LCD fabs has been changed from the bay layout to the stocker integrated bay layout, and recently the layout for single glass transfer. In the stocker-integrated layout, loaders of machines are embedded in stockers and wafer carriers are handled between machines by stocker cranes. Stockers are connected by conveyors. Some LCD companies have converted parts of their layouts into a transfer layout for their seventh generation LCD fabs, which fabricate glasses of about 2 x 2 m in size. These changes of LCD layouts accelerated automation of the fabs, increasing fab productivity significantly [23]. Although wafer fabs are more complex than LCD fabs, it seems that as the wafer becomes larger, the fab layout will have more to learn from the LCD fab layout.

IV. NEW IRL

A. Design Objectives

This research tries to improve the following performance indexes that were considered to be important in fab layout research [3], [4], [24]:
1) production lead time, WIP;
2) routing flexibility;
3) clean room efficiency (footprint) and initial investment cost;
4) efficiency of AMHS including direct transportation.

B. Prototype Layout and Design Principle

Fig. 2 shows a prototype IRL. This section explains its design principles.

1) Large rooms: The most distinctive characteristic of the new IRL is using large rooms, each of which is connected by one combined loop of OHT and OHS. While the bay-based layouts have more than 20 bays for 35 K wafer start per month (WSPM), the IRL has four rooms, each of which has approximately the same number of tools in five to eight bays.

2) Arrangement of litho tools: Bottleneck equipment such as litho tools is placed close together for maximum utilization. The litho tools are placed in two rooms near the center of the fab. The loops of the litho tools in different rooms share the same stockers to increase the flexibility of a lot’s tool selection and a tool’s lot selection. The prototype layout for 35 K WSPM has 42 litho tools.

3) Arrangement of nonbottleneck tools: Thin film tools are located next to litho tools because they frequently form sequential processing stages. Other tools for sequential processing stages are also placed in the same room. Metrology tools are distributed evenly to all rooms. Other nonbottleneck tools are placed to reduce carrier transportation time and secure tool maintainability.

4) Interroom transportation: Stockers are used for interroom transportation as well as temporary storage, and there is no center spine for interroom transportation. To reduce the OHT travel distance, a FOUP might be sent to the closest stocker connected to a conveyor and to the destination through a few additional stockers and conveyors. For example, a lot from a metrology tool (see (a) in Fig. 2) can be delivered to a cleaning tool (see (b) in Fig. 2) via Stocker 1, the conveyors and stockers in the center, and Stocker 2.

5) Vehicle travel path and control: All tools in a room are connected by one large OHT/OHS loop, and each loop can have as many as 20 vehicles. The vehicles are virtually zoned, meaning that not too many of them can be at the same area at the same time to prevent vehicle congestion and possible deadlock.

6) Transportation vehicle hardware: The transportation vehicles in the IRL are the combined vehicle types of interbay (OHS) and intrabay (OHT) such as what is proposed by Daifuku (Fig. 3). In IBL, OHT hoists and drops down the

---

TABLE I

DIFFERENCES BETWEEN SEMICONDUCTOR AND LCD PRODUCTION

<table>
<thead>
<tr>
<th>Item</th>
<th>Semiconductor</th>
<th>LCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of layers</td>
<td>18 - 30</td>
<td>4 - 6</td>
</tr>
<tr>
<td>Material size</td>
<td>150 mm, 200 mm, 300 mm</td>
<td>400 x 500 mm - 1 m x 2 m</td>
</tr>
<tr>
<td>Unique processes</td>
<td>Diffusion, Implantation</td>
<td>Liquid crystal injection</td>
</tr>
<tr>
<td>Lot size</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>Production capacity/month</td>
<td>10 – 40 K</td>
<td>40 – 100 K</td>
</tr>
</tbody>
</table>
carriers in a bay, and OHS transports carriers in the center spine without hoisting and dropping down. In IRL there is no center spine loop, and only one type of vehicle transports carriers between machines and stockers.

For the prototype layout, tool-size, loop width, stocker width, chase width, and maintenance space used for the prototype are referenced from the SEMATECH report [22]. The layout also references existing full size fabs with 35 K WSPM; however, no direct association between this layout and any existing fabs should be made.

Differently from the cell layout that arranges cells following the process stages of wafers, the IRL does not arrange rooms following the process sequences for higher flexibility. The IRL is also different from the ballroom layout and the IBL in that the IRL does not have a bay-type arrangement of tools.

C. Benefits of IRL

The IRL overcomes many well-known difficulties of fab layouts.

1) Higher product mix flexibility: Without considering product mix change (and tool failures), the product layout or the transfer layout could be the choice for the best productivity. However, product mix changes are frequent and many companies are trying to accommodate diverse product types. Bay-based layouts are popular because of their functional arrangement of tools for efficient product mix change despite their generally recognized lower productivity than the product layout.

The IRL takes more advantage of the functional layout than bay-based layouts, especially for the bottleneck litho tools; more litho tools are put together, being effectively connected by intraroom loops and stockers.

2) Higher routing flexibility for tool selection: In the bay-based layouts, the machines for the same process stages can have slightly different process capabilities, as in the
3) Reduced WIP and cycle time: In the IRL, bottleneck tools often placed in multiple bays, resulting in increased intrabay transportation. The IRL reduces this intrabay transportation of wafers and increases routing flexibility.

4) Increased direct transportation: Direct transportation is more important when we consider the dedication of litho tools to layers. The litho process is so sensitive that one specific layer can be processed only by a few dedicated litho tools. Engineers approve tools for a specific layer after a series of fastidious tests. As the width of a circuit gets smaller, the tool control will be tighter, and the tool dedication is expected to be more restrictive in the future. This dedication limits the wafer’s routing flexibility for litho tools. Due to the large number of layer types and the division of litho tools to multiple bays and frequent change of product mix, the litho tools that can process the same layers are often placed in multiple bays, resulting in increased interbay transportation. The IRL reduces this interbay transportation of wafers and increases routing flexibility.

5) Efficient space usage (footprint): The IRL does not use the center spine, improving the footprint of a fab.

6) Reduced initial cost for material handling systems: The room layout requires fewer stockers and vehicles than the bay-based layouts. In the IBL, stockers are placed near the entrance of bays often with low utilization. There can be as many as 33 stockers in a full scale 35-K WSPM fab. The room layout requires fewer stockers with higher utilization. Fig. 2 shows 25 stockers. Fewer vehicles are also needed due to the increased direct transportation. Also, more vehicles in a vehicle loop increase the chance of having idle vehicles at a closer location of carrier loading.

7) More efficient nonproduct wafer (NPW) handling: Although bay-based layouts are functional, their bays are arranged following the general flow patterns of wafers. However, the flow patterns of NPW handling are different from product wafers and complicate the material flows. The increased direct transportation in the IRL facilitates the material flow of the NPW and reduces its WIP. Especially, the recent devices with higher integration require more precise process control, therefore requiring more NPWs. The effective flow control of NPW is becoming more important.

D. Vehicle Path Control

One possible concern about the IRL is the control of many transportation vehicles in a loop. However, while bidirectional AGVs were popular for intrabay transportation, unidirectional OHTs and OHSs are more popular recently for inter- and intrabay transportation due to their increased safety and larger capacity for transportation, thereby reducing the complexity of large vehicle loop control. Theoretically, bidirectional vehicles can process more carriers; however, careful use of by-passes on the paths is supplementing the disadvantages of unidirectional paths. Also, procedures for bidirectional vehicle control without deadlock are available in the literature [25], [26].

V. Quantitative Analysis

A. Input Data

In this section, the new IRL is compared with the IBL. The fab model references the SEMATECH reports and existing full size fabs with 35-K WSPM with some simplification assumptions and modifications for proprietary information. The model considers major processing stages such as diffusion, oxidation, CVD, sputter, photolithography (litho), implant, wet bench, plasma etch/strip, and metrology. The dedication levels of litho tools are also considered. Such important characteristics of tools such as tool size, throughput rate, and tool count given in the SEMATECH report are used in this analysis. Each wafer has 21 layers. Table II describes layer types of a sample product based on the process sequences given in the SEMATECH report. This model fab has four product types, whose layer
types are given in Table III. While the SEMATECH report considers a 25-K WSPM fab, we consider 35-K WSPM to represent recent full size fabs. For simplicity, this analysis does not consider nonproduct wafer (NPW), which can be reflected in the analysis in similar ways described in this section.

### B. Material Flow Time

This section calculates the material flow time of wafers to compare the material handling in the IBL and the IRL. The material flow time consists of a vehicle’s loaded travel time, a vehicle’s empty travel time, and a lot’s handling time spent in stockers. It excludes the time on processing machines and waiting time for processing machine assignment in stockers. For the product mix given in Table III, the proper number of vehicles in a vehicle loop is first determined, and the average utilization level, average loaded travel time, and average empty travel time of vehicles are calculated. The travel distances between loading and unloading locations are measured along the vehicle travel loop given in the layouts. The vehicle travel time includes an additional 20% of a vehicle’s uninterrupted travel time to reflect traffic congestion on the travel paths and five seconds of acceleration and deceleration time for precise positioning of a vehicle. The loaded travel time includes the loading and unloading time of carriers. Table IV shows the characteristics of the material handling systems from the SEMATECH report.

Different from a vehicle’s loaded travel time, a vehicle’s idle travel time to pick up carriers depends on the empty vehicle selection rule. When there are multiple idle vehicles for a new transportation request, the nearest idle vehicle (NIV) is used in this analysis. If there are multiple parts waiting for a vehicle when a vehicle becomes idle, the first come/first served (FCFS) rule is used. Egbelu and Tanchoco [27] report that the selection of a part to transport by an idle vehicle impacts the performance of a material handing system less than the selection of an idle vehicle by a part. An estimation of the means and variances of a vehicle’s empty travel time and a vehicle’s loaded travel time, and a part’s mean waiting time for vehicles under the NIV and FCFS rules can be obtained from [28] and [29]. The procedure uses the distances and travel frequencies between ports. In the current research we do not consider the probability of balking of lots to stockers due to the lack of buffer space at the destination tool. Suh et al. [30] present a procedure to consider this balking.

1) Integrated Bay Layout (IBL) Case: Fig. 5 shows the test IBL model that has 23 bays, nine OHT loops, and a center spine. The first six rows of Table V are the indexes of loops and input values of the calculation. Using the product mix of equal ratio for the products given in Table III, the hourly traffic rates of carriers between ports given in the fourth row are calculated. The hourly traffic rate is the number of loaded travels of vehicles within a loop per hour, which is used as a weight factor to aggregate the results in the table into one index. Also, the hourly transportation rate to stockers is calculated. Table VI shows a part of the traffic frequency table in a litho loop of the IBL. In the table, s1, s2, s3, and s4 represent stockers. When there are multiple possible destination machines for a process stage, they are selected randomly with each destination having equal probability of being chosen.

The seventh row of Table V shows the number of vehicles in each loop. There are two different possible indexes for the determination of the number of vehicles in a loop: the average vehicle utilization and the average waiting time of parts for vehicles after transportation request. While the average waiting time shows the long-term performance of the material handling system, the average vehicle utilization is more closely related to the peak time performance of the material handling system. In this table, the number of vehicles in a loop is determined so that the average vehicle utilization is no more than 70%. The vehicle utilization includes the loaded travel time, empty travel time, carrier loading/unloading time, and traffic congestion time. In this test, the waiting time of carriers for the assignment of vehicles is very small and is included in the vehicle’s empty travel time. A total of 88 vehicles are used in the fab, including the

### TABLE II

**Types of Process Sequence of Layers (Example of Product A; Also See Table III)**

<table>
<thead>
<tr>
<th>Layer No.</th>
<th>Process sequence</th>
<th>Layer type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OCLMDW/DMDW</td>
<td>L1</td>
</tr>
<tr>
<td>2</td>
<td>OLMIDW</td>
<td>L2</td>
</tr>
<tr>
<td>3</td>
<td>LMDW</td>
<td>L3</td>
</tr>
<tr>
<td>4</td>
<td>OCLMDWD</td>
<td>L4</td>
</tr>
<tr>
<td>5</td>
<td>OLMIDW</td>
<td>L2</td>
</tr>
<tr>
<td>6</td>
<td>LMDWCDW</td>
<td>L5</td>
</tr>
<tr>
<td>7</td>
<td>LMDW</td>
<td>L3</td>
</tr>
<tr>
<td>8</td>
<td>LMIDW</td>
<td>L6</td>
</tr>
<tr>
<td>9</td>
<td>CWCMLMDW</td>
<td>L7</td>
</tr>
<tr>
<td>10</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>11</td>
<td>COLMDWM</td>
<td>L9</td>
</tr>
<tr>
<td>12</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>13</td>
<td>COLMDWM</td>
<td>L9</td>
</tr>
<tr>
<td>14</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>15</td>
<td>COLMDWM</td>
<td>L9</td>
</tr>
<tr>
<td>16</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>17</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>18</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>19</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>20</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
<tr>
<td>21</td>
<td>CLMDWM</td>
<td>L8</td>
</tr>
</tbody>
</table>

### TABLE III

**Frequencies of Layer Types by Product Types**

<table>
<thead>
<tr>
<th>Prod. Layer</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>L2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>L3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>L4</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>L5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>L6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>L7</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>L8</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>L9</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Authorized licensed use limited to: Kyungpook National University. Downloaded on July 22, 2009 at 01:41 from IEEE Xplore. Restrictions apply.
center spine. Using the numbers of vehicles in the loops, the average utilization, loaded travel time, and empty travel time of vehicles are calculated following the procedure given by [28] and [29]. The total stocker time is the time of carriers spent in stockers for one hour. It is calculated by the hourly traffic rate into stockers given in Table V multiplied by \(20 + 7 + 7 = 34\) s per stocker use obtained from Table IV. The sum of flow times of parts in a loop for one hour is \((\text{hourly transportation rate}) \times \text{total stocker time}\), which is given at the last row of Table V. The total flow time of carriers in the fab for one hour is the sum of flow times of each loop including the center spine and is \(238\text{,}778\) s = 66.77 h.

2) IRL Case: With the same test data, the performance of the IRL is estimated. It is expected that the flow times of lots in the four rooms will be similar because they have a similar number of tools and traffic intensities; the material flow time in the upper-left rooms is analyzed.

In Table VII, the second column shows the performance of a room when we have the same number of vehicles as in the case of IBL. The wafer’s flow time for the four rooms is \(25631\times4/3600 = 28.4\) h, a 58% reduction from the IBL case.

In the third column, the vehicle utilization level is kept at 70% or lower. The required number of vehicles for the four rooms is 40, a 55% reduction from the IBL case, and the flow time for the four rooms is \(29\text{,}003\times4/3600 = 32.2\) h, a 52% reduction from the IBL case.

This analysis does not consider the detailed path planning of vehicles. However, the room layout has more vehicles in a loop than the bay layout, and it is expected that more vehicles in larger loops in the IRL will allow larger performance improvement from a more advanced vehicle control logic such as the Look-Ahead control logic [31] and bottleneck machine first (BMF) rule [32] than the smaller loops. This analysis does not consider the use of conveyors. It is expected that the conveyor use will reduce the requirements for vehicle transportation and
TABLE V
MATERIAL FLOW DATA AND PERFORMANCE IN BAYS OF IBL (CS: CENTER SPINE, *: BAY USES OHTS AND CENTER SPINE USES OHS) (TIME UNIT IS SECONDS)

<table>
<thead>
<tr>
<th>Loop no.</th>
<th>PVD/ RTP</th>
<th>Litho</th>
<th>Litho</th>
<th>Fum</th>
<th>P-Strip</th>
<th>CMP/ Implant</th>
<th>CVD</th>
<th>CS</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Tools</td>
<td>31</td>
<td>26</td>
<td>28</td>
<td>58</td>
<td>57</td>
<td>49</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Hourly traffic rate</td>
<td>110</td>
<td>64</td>
<td>73</td>
<td>371</td>
<td>187</td>
<td>109</td>
<td>109</td>
<td>217</td>
</tr>
<tr>
<td>No. of Stockers</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>Hourly traffic rate to stocker</td>
<td>54</td>
<td>30</td>
<td>36</td>
<td>167</td>
<td>72</td>
<td>41</td>
<td>41</td>
<td>104</td>
</tr>
<tr>
<td>No. of vehicles*</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>20</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Avg. vehicle utilization (%)</td>
<td>60</td>
<td>60</td>
<td>70</td>
<td>66</td>
<td>66</td>
<td>67</td>
<td>67</td>
<td>66</td>
</tr>
<tr>
<td>Avg. loaded travel time</td>
<td>78</td>
<td>77</td>
<td>78</td>
<td>89</td>
<td>79</td>
<td>82</td>
<td>82</td>
<td>87</td>
</tr>
<tr>
<td>Avg. empty travel time</td>
<td>19</td>
<td>24</td>
<td>25</td>
<td>17</td>
<td>16</td>
<td>27</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>Total stocker time</td>
<td>1.820</td>
<td>1.020</td>
<td>1.224</td>
<td>5.685</td>
<td>2.448</td>
<td>1.380</td>
<td>1.380</td>
<td>3.541</td>
</tr>
<tr>
<td>Hourly flow time</td>
<td>12,504</td>
<td>7,524</td>
<td>8,832</td>
<td>45,019</td>
<td>20,238</td>
<td>13,403</td>
<td>13,403</td>
<td>27,026</td>
</tr>
</tbody>
</table>

increase the availability of empty vehicles at nearer locations for transportation requests, further reducing the empty travel time of vehicles.

C. Utilization of Litho Tools

This section compares the utilization of litho tools in the IBL and IRL by queueing modeling to find the effect of collocating them in the IRL. In the IBL, the litho tools are placed in two loops, each of which has dedicated stockers to it. This dedication requires a separate queue for each loop and crossover of lots. The litho tools in the IRL are arranged in two loops that share the common stockers, resulting in one central queue for all litho tools. A bay and a room are complicated queueing systems, and we model these as approximated $G/D/1/K$ systems using simplification assumptions, as will be explained in the following.

In the model, the processing time of a lot (25 wafers) is one hour and is almost the same for all layer types (or layers). There are 42 litho tools, four product types, and 21*$4=84$ layers. We assume a litho tool can process eight layers (tool versatility), and each layer has four litho tools that can process it (machine dedication = $42*8/84 = 4$). The hourly arrival rate of lots to the litho stage is $35 K * 21 layers/720 h = 40.833$. The hourly arrival rate per machine is $40.833/42 = 0.972$, and the hourly arrival rate per layer is $40.833/84 = 0.486$. The average utilization of all litho tools is (arrival rate)/(service rate) = $0.972/1 = 97.2\%$, and about 2.78% of tool time is lost on average due to starvation; however, if we can increase the utilization of litho tools by putting them together, the production rate of the fab will be increased.

1) Modeling IBL Case: Let 21 out of the 84 layers (25%) be processed only in the first loop, loop1, another 21 (25%) only in the second loop, loop2, and the rest of the 42 types (50%) in both loops. Then, the loop1 processes $21 + 42 = 63$ layers. Let PH1 be a litho tool in loop1 that can process layers 1 to 8: layers 1–3 being processed only in loop1, and layers 4–8 in both loop1 and loop2 approximately according to the average arrival rates of layers.

Here, we first notice that it is very unlikely that any two litho tools in a loop can process many common layers; the chance of two machines being able to process three or more same layers is $\sum_{i=3}^{8} (s C_2 (s-i))/s^2C_8 = 0.057$ under a simplification assumption of equal chances for layers to be selected. In this case, even when PH1 is starving, most other tools will have lots to process. We model this system as a queueing system with a single server.

One possible way of WIP control for PH1 is that if there are not enough lots of layers 1–8 in a loop, we allow no litho tools other than PH1 to process these lots, which is being used for load-balancing in industry. This operation is depicted in Fig. 6, where $K$ is the threshold and the boxes represent litho tools, ignoring the possible processing of multiple layers of 1–8 by a tool. Because half of the layers, 4–8, are processed by loop2, the arrival rate of layers 1–8 to loop1 is $0.486*3 + 0.486/2*5 = 2.673$, exceeding the process rate of PH1. The arrivals that correspond to the exceeding process capacity of PH1 will be processed by tools other than PH1, as is shown in Fig. 6 when the number of waiting lots is larger than $K$. This consumption will be quick because many other tools can process layers 1–8. Here, we model this operation as balancing—the lots balk to tools other than PH1 if the queue length of PH1, including the lot under processing, is larger than $K$. We approximate this queueing system as a $G/D/1/K$ system. The service time is deterministic and one hour.

This $G/D/1/K$ system is a special case of the $G^A/C^B/c/K$ system, the queueing formula of which is given
by Lee [33], [34] based on diffusion approximation. It is known that the diffusion approximation is good under heavy traffic, which is the case for this model. A part of the formula is used for the queuing formula to calculate the probability of part starvation of PH1 except for the variance of interarrival time, $V_B$.

2) Modeling IRL Case: In the IRL, all 42 litho tools are located along OHT loops connected to the common stockers. The stockers provide space for the central queue, and the system is again modeled by a $G/D/1/K$ queuing system with the modified hourly arrival rate of layers 1–8 being 0.48688 = 3.888. We again have all the necessary information except for the variance of interarrival time of lots $V_R$.

3) Calculation of Means and Variances of a Lot’s Interarrival Time to Litho Tool: Here, we find the relationship between the means and variance of interarrival times to PH1 in the IRL and IBL. Let $Y$ be the interarrival time of lots 1–8 to PH1 in loop1 of the IRL and $X_i$ (i = 1, 2, 3, . . . ) be the interarrival time of lots 1–8 to PH1 in the IRL. Then, $Y = \sum_{i=1}^{N} X_i$, where $N$ is a random variable. If we assume independence between $X_i$, it is known that $E[Y] = E[X]E[N]$ and $V[Y] = E[N]V[X] + V[N]E^2[X]$ [35, p. 66] where $E$ and $V$ represent the expected value and the variance of a random variable, respectively. Here, $E[X]$ = (mean interarrival time of lots 1–8 in the IRL) = 1/(3.888) = 0.257 and $V[X] = V_R$.

In the IBL, when layers 1–3 arrive at the litho stage, the lots are sent to loop1 in the IBL. Let us consider two possible flow control policies for layers 4–8: 1) the lots are sent to one of the two litho loops randomly with probability $p = 0.5$ and 2) the lots are sent to the two litho loops alternately.

In Case 1), if a lot of layer 1, 2, or 3 arrives at the litho stage (with probability of 3/8), it is sent to the first loop; and if a lot of layer 4, 5, 6, 7, or 8 arrives (with probability of 5/8), half of it is sent to the first loop. Then, $N$ follows a geometric distribution with parameter $3/8 + 5/8/2 = 0.688$. $E[N] = 1/0.688 = 1.4545$, and $V[N] = 0.312/0.688^2 = 0.659$. Now, $E[Y] = E[N]E[X] = 1.4545 \times 0.257 = 0.373$, and $V_Y = V[N]E^2[X] = 1.4545V_R + 0.659 \times 0.257^2 = 1.4545V_R + 0.0164$.

In Case 2), the distribution of $N$ depends on the type of the last lots arriving at PH1. Table VIII shows the distribution of $N$. In the table, lot type $A$ is layers 1–3 and lot type $B$ is layers 4–8. $Bo$ is a $B$ lot that comes to loop1, and $Bx$ is a $B$ lot that goes to loop2. The first column is the possible prior arrival patterns of lots to PH1, and the second column shows its probability. In the table, “$A\ldots A$” means one or more consecutive arrivals of $A$ lots. According to the table, the probability for $N$ to be unity is $Pr(N = 1) = 0.5455 Pr(N = 2) = 0.4545, E[N] = 1.4545$, and $V[N] = 0.2479 V_B \equiv V[Y] = E[N]V[X] + V[N]E^2[X] = 1.4545V_R + 0.2479 \times 0.257^2 = 1.4545V_R + 0.0164$.

4) Probability of Tool Starvation: Following [33, p. 720], the probability of PH1’s starvation, $M_1$, is

$$M_1 = \frac{C_4}{C_1C_4 - C_2C_3}$$

where

$$C_1 = 1 + \left(\frac{\lambda\alpha}{2\beta^2}\right) (wK - 1) - \frac{\lambda\alpha}{\beta} (K - 1)$$

$$C_2 = 1 + \left(\frac{\mu}{\beta}\right) (wK - 1)$$

$$C_3 = (\lambda\beta)(wK - 1) + \left(\frac{\lambda}{\beta}\right) (1 - wK - 1)$$

$$C_4 = \left(\frac{\mu}{\beta}\right) (1 - w)$$

$w = \epsilon^{3/\alpha}, \alpha = \epsilon^3V, \beta = \lambda - \mu, \lambda$ is the service rate, $\mu$ is the arrival rate, $K = 5$, and $V$ is the variance of interarrival time.

These probabilities of tool starvation for the IRL and IBL are compared in Table IX for a few levels of $V_R$ and $V_B$. In this comparison, the value of $K$ is set to five. Realistically, assuming that the average inventory of wafers at the litho stage is about eight hours, we chose this $K$ value to indicate half of the average
inventory. This $K$ value is higher than the values being used in industry; however, it can supplement the baffling assumption in the model that says that the lots that see more than $K$ already waiting lots are never served by PH1. The value of $V_{P_R}$ is chosen so that the probability of starvation of PH1 is between 0 and 10%. The table shows that the collocation of litho tools can improve the production rate by 0.0 to 1.3% at the litho stage depending on the variance of the interarrival time of lots to the litho stage and the types of feeding multiple loops in the IBL. In the table, $V_{P_{H1}}$ is the variance of interarrival time calculated for Case 1) and $V_{P_{H2}}$ for Case 2). For smaller $K$ values, larger production improvement is observed in the performance test.

D. Footprint

Figs. 2 and 5 have the same number of tools and distances between tools given in the SEMATECH report. In the figures, the area of the fab is reduced by about 20% from $120^70^70 = 8400$ m$^2$ to $80^55 = 6000$ m$^2$. The reduction of the footprint reduces the fab construction cost, the clean room operating cost, and the travel distance of lots in the fab. Conservatively estimating the construction cost of an IBL fab, except for the equipment and its setup, the cost is $500$ million; the reduction is $100$ million.

E. Material Handling System

A material handling system mainly consists of stockers, OHTs, OHAs and conveyors along with control software. Tables V and VII show that, when the utilization of the vehicles in a loop is kept lower than 70%, the number of vehicles in the fab is reduced from 88 to 40 due to the reduced empty travel time of the vehicles. The tables show that the flow time of wafers in the IRL is also reduced from 238 778/3000 = 66.77 hours to 29 003/4/3000 = 32.22 hours (52% of reduction), even with this smaller number of vehicles due to the increased direct transportation and higher vehicle availability from larger fleet size. If we use conveyors, the vehicles will provide higher availability at closer locations of lot loading; however, additional time may be needed for lot transfer between vehicles, stockers, and conveyors. Careful design of conveyors will further reduce the flow time of wafers. The prototype IRL needs 24 stockers while the IBL needs 33 stockers. In the IRL, the number of stockers is reduced because of their reduced requirements for interloop lot transfer and less WIP of wafers by the central buffer effect.

VI. CONCLUSION

This paper presented the IRL for wafer production. This is a functional layout that provides even higher routing and product mix change flexibility than bay-based layouts. The most distinctive characteristic of the new IRL involves using larger rooms than bays and connecting the machines in a room by one combined OHT/OHS loop. This layout increases direct transportation of carriers between machines, significantly reducing the material handling requirements. A practical prototype of the IBL is presented based on the data provided by International SEMATECH. The performance of IRL is compared with that of the IBL, which is favorably considered in the literature and industry.

The trend of increased wafer size to 450 mm will make efficient material handling more important than before. The tight process control due to the reduced circuit thickness will make the dedication of litho machines tighter, increasing crossover of lots between bays. Tight process control will also increase the handling of NPW. This IRL will meet the need of the current and future fabs.

The four rooms in the IRL can be arranged in two floors (Fig. 8) using vertical stockers for interfloor transportation. This colonial arrangement can reduce material travel distance. Discussion of the issues for interfloor transportation and other managerial aspects of it is left for future research.

REFERENCES


[22] T. Quinn and E. Bass, 300 mm factory layout and material handling modeling: Phase I report 1999, Technology Transfer #9902368B-ENG, International SEMATECH.


