Production Planning & Control,
Vol. 17, No. 7, October 2006, 664–677

Taylor & Francis
Taylor & Francis Group

Developing a 3D layout for wafer fabrication plants

B. GOLANY*, A. GUREVICH and E. PAZ PUZAILOV

Faculty of Industrial Engineering and Management, The Technion, Israel Institute of Technology, Haifa, 32000 Israel

Wafer fabrication plants (FABs), are typically arranged in two dimensional (2D) layouts on a single floor. These layouts imply various constraints on the work-in-process (WIP) and the material handling systems. In contrast, automated storage/retrieval systems (AS/RS) are arranged in 3D aisles, where each aisle is served by a robotic arm that moves back and forth between the aisle's entrance point and its storage locations. This paper offers an AS/RS-based 3D layout for FABs. First, a general 3D layout design problem is formulated and a heuristic algorithm is developed to solve it. Then, the proposed layout is evaluated with respect to its 2D counterpart, and the scenarios in which the former may outperform the latter, more conventional, layout are specified. Finally, the performance of the proposed 3D layout is compared to a corresponding 2D layout through a simulation which is based on actual data taken from the semiconductor industry.

Keywords: Layout planning; 3D layouts; Quadratic assignment problem; Simulated annealing; Simulation; Wafer fabrication plants

1. Introduction

Designing the layout of a plant is a challenging task that has a major impact on the plant's future performance. The majority of production floors are designed and built on 2D planes having only a single floor. These layouts create various constraints on the handling of work-in-process (WIP), especially in automated material handling environments.

The rationale that characterises most of the existing models for determining plant layouts is to place the production machines (or departments) so as to minimise the total material handling cost (Domschke and Krispin, 1997). Two common model formulations in this area are the quadratic assignment problem (QAP) (Koopmans and Beckmann 1957) and the ABSMODEL (Heragu 1992). The main difference between them is the question of whether the optional sites for the machines are predefined (QAP) or not (ABSMODEL). Various approaches were proposed to derive optimal solutions to these formulations. Some were based on branch and bound techniques (Gilmore 1962, Lawler 1963), others relied on various linearisation methods (Lawler 1963, Kaufmann and Broeckx 1978, Adams and Johnson 1994). But, since the QAP was proven to be a NP-hard problem, these methods can solve only small size problems (Sahni and Gonzalez 1976).

The inability to solve realistic layout problems led to the development of numerous heuristics techniques. These can be categorised into construction algorithms (e.g. Gilmore 1977, Tompkins and White 1984, Golany and Rosenblatt 1991, Sivaraman 1999), which try to build the solution from scratch, and improvement algorithms (e.g. CRAFT by Buffa *et al.* 1963) which start with an initial solution and try to improve it by switching the locations of pairs of machines. Most improvement algorithms belong to the class of greedy heuristics whose basic principle is to implement at each stage the change that yields the largest immediate improvement in the solution.

Some production floors are arranged in multi-floor structures. In such layouts the production machines are assigned to the different floors and the WIP flows among them according to the production sequence.

*Corresponding author. Email: golany@ie.technion.ac.il

Production Planning and Control
ISSN 0953–7287 print/ISSN 1366–5871 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/09537280600901103

Different transportation devices are employed to move the WIP within and between the floors. In such settings, WIP is moved within a certain floor in much the same way as it is moved in an ordinary 2D layout, while material handling between the floors is made by dedicated devices such as conveyors, lifts, elevators, etc. It is important to note that in order to transport WIP to a machine that is located in another floor, three transportation devices are typically required: first, an inter-floor device is used to move the WIP from its current location to a stocking point that connects the floors, second, an intra-floor device takes it to the destination floor and finally, another inter-floor device takes it to its target machine. The algorithms that were proposed to solve these problems are essentially extensions of the single-floor algorithms (e.g. CRAFT-3D by Cinar 1975, SPACECRAFT by Johnson 1982, MULTIPLE by Bozer *et al*. 1994, SABLE by Meller and Bozer 1996). Two efficient algorithms were presented by Patsiatzis and Papageorgiou (2003) who used integer linear programming and solved the problem by iterating between a master problem (to determine the number of floors and the machines assigned to them) and a sub-problem (to determine the location of the machines within each floor). Suzuki *et al*. (1991), introduced implementation of a multi-floor layout in a chemical plant in Japan. A special case of 3D layouts was introduced by Barbosa-Povoa *et al*. (2002), locating the production facilities in a continuous 3D space and employing a mixed integer linear programming formulation to determine their optimal locations.

The semiconductor industry produces computer components and communication equipment. In 2000 the industry's sales were over \$204 trillion. A typical production floor of a FAB consists of 30–50 machine types with up to 350 machines in total. The cost of each machine may run from \$100 000 to \$7 000 000 (Qu *et al*. 2001).

Currently, FABs are constructed in 2D layout in a 3-level 'sandwich-like' structure, which has characterised this industry from the very beginning (Chang and Chang 1998). The upper layer is the 'attic' that houses the air conditioning facilities whose purpose is to ensure a clear laminar flow of air to facilitate the clean room environment. The middle layer is the production floor itself (the clean room), where the actual fabrication takes place. This layer contains all the machines and this is where all the lots are processed. The bottom layer is the sub-FAB facilities' floor where all the utilities/pumps and the other feeding equipment are located. The maintenance of these utilities is performed at the bottom layer with little or no interruption of the actual fabrication process, thus enabling high throughput while preserving high air quality in the clean-room environment.

Various approaches were proposed to determine the layout of the middle layer (Campbell and Ammenheuser 1999, Quinn and Bass 1999, Yang *et al*. 1999, Yang *et al*. 2001, Hsieh and Hung 2004). For the most part, these approaches converged to a basic layout that is composed of several areas (known in this industry as 'bays'), where the production machines are located, and a central aisle that connects them (figure 1 presents the middle layer of a typical semiconductor plant). In practice, there are two plausible approaches for the allocation of machines to bays in 2D FABs (Geiger *et al*. 1997). The first one seeks to minimise the time between successive operations by allocating machines according to their process sequence. The second approach, which is more common in the semiconductor industry, is to place the machines according to their type so as to facilitate easier maintenance operations. There is a clear distinction between these two approaches when the implementation is restricted to a 2D layout. However, as shown later, this distinction is irrelevant in the 3D layout, which is proposed here, since the movement among machines can be done faster and more efficiently while utilising a single buffer queue, as opposed to the three queues which are required in the 2D layout, as described in detail below.

The transportation of lots in these layouts is done by automated material handling systems (AMHS) that have a critical impact on the FABs' performance. The AMHS consist of two types of transporters: an intra-bay system, which serves a single bay, moving the lots and feeding them to the machines in that bay; and an inter-bay system, which transports lots among the different bays (Quinn and Bass 1999). Hence, the transportation of a single lot between two machines located in two different bays requires a three-way transfer between the

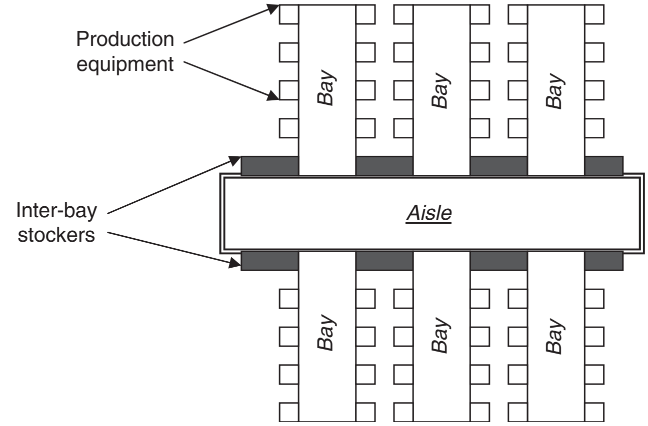

Figure 1. Schematic structure of FAB's layout.

transporters: Bay → Aisle → Bay. Since it is practically impossible to co-ordinate each succeeding transporter to wait for its predecessor to arrive, the transfers are facilitated through inter-bay stockers. This means that each lot in transfer experiences delays in three queues: first, waiting for transportation in the origin bay (first station); then in the first buffer for transportation between the bays; and finally in the buffer for transportation into the target bay where it needs to undergo the succeeding operation. This complex structure results in extended queue times and thus leads to delays in the overall throughput of the FAB. The investments in the AMHS amount to millions of dollars both during the construction phase as well as during the operational stage—not to mention the operational problems associated with scheduling the movement of the lots in the system. Therefore, the design and operations of AMHS have created a stream of ongoing research aimed at improving both the effectiveness and the efficiency of these devices (e.g. Mackulak and Savory 2001, Nazzal and Bodner 2003).

In contrast to the common practice in production floors, automated storage/retrieval systems (AS/RS) work in 3D environments, having a basic rack structure, where each aisle is served by a robotic arm that moves back and forth between the aisle's input/output point and its storage locations.

This paper offers a drastically new approach for FABs' design by proposing to construct them as AS/RS. Instead of storage cells, the approach proposed here places the FAB's machines on 'shelves' arranged in aisles served by automatic transporters. These transporters (cranes) are capable of taking WIP directly from any 'origin' machine to any 'target' machine, regardless of their location within the aisle, thus saving the need for stocking points to connect levels (or floors) and avoiding the need to switch between several transportation devices (as is required by conventional multi-floor layouts).

The rest of the paper is organised as follows. The proposed 3D layout concept is presented in section 2. Section 3 provides the formulation of the associated layout design problem and develops solution methods for its sub-problems. Then, section 4 offers the results of a simulation study which compared the performance of the proposed layout *vis-à-vis* its 2D counterparts. Finally, section 5 concludes the paper and provides directions for future research.

2. The proposed 3D layout

An important advantage of AS/RS over the AMHS discussed above is that the material movement in the former is done in a single operation using a single transporter and without requiring several intermediate buffers. Another important advantage of AS/RS lies in its relatively small area requirements. These two features are the main factors motivating our proposal of a 3D layout for FABs.

The idea is to construct an AS/RS-like structure, where the FABs' machines are placed in the 'storage cells' and a transportation device associated with each aisle moves the lots among the machines. Figure 2 demonstrates the proposed layout. Figure 2(a) shows a frontal view of the cells where the machines are located, whereas figure 2(b) provides a 3D view of an aisle, where the transporter moves along the X and Z axes.

The implementation of the proposed 3D layout would require solutions to some technical issues; these are introduced and discussed below. Here we provide only general directions for dealing with them, since the complete technical answers are beyond the scope of the current paper.

- *Vibration*. The vibration of machines located at some levels of the 3D structure may have a diverse effect on the accuracy of operations performed above or below them. This issue may have a

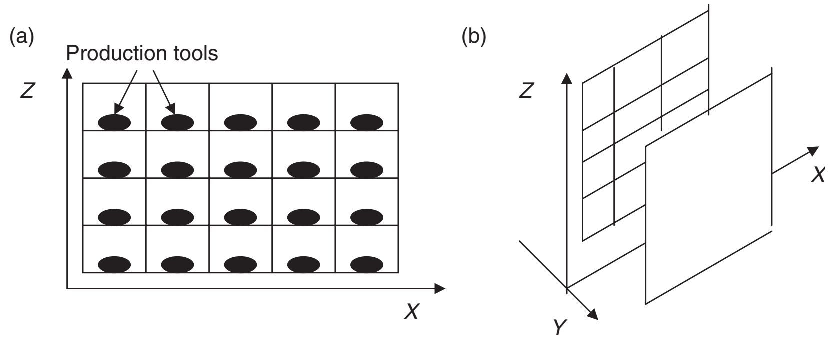

Figure 2. Schematic structure of 3D layout.

significant impact on the construction design; however, it is solvable with current construction technology (since the current 2D FABs are already constructed with multiple 'sandwich' floors). Thus, it can be expected that the relatively high vibrating machines (such as steppers and other litho equipment) will be located on the lowest level and that the beams carrying the levels above them will be based on separate trusses.

- *Sub-FAB facilities*. The sub-FAB facilities are traditionally located below the clean room floor. Locating them at the same level of the productions tools as is suggested here will require additional adjustments in the FAB design layout.
- *Attic*. In the current 'sandwich-like' production floor, the attic is responsible for the laminar flow that creates the clean room environment in the production floor itself (the middle part of the 'sandwich'). Since the proposed layout will be much higher, the attic must be re-designed to maintain a suitable clean environment and to prevent potential turbulences in the laminar flow. However, since all new generation tools (300 mm) are equipped with the mini-environment infrastructure and the lots are completely sealed during the transportation process, the required air quality which the attic has to provide is lower than in the current layouts.
- *Capacity assumptions*. The planning of the 3D FAB must take into consideration the performance parameters of the aisle's transporter to prevent it from becoming the bottleneck of the entire system. This will require installing highly reliable transporters as well as creating some redundancy. Since the transporters are much cheaper in comparison with the production machines, the investment in backup systems should be affordable.
- *Interface*. In order to ensure a fully operational transportation device the FAB must have a standard interface within all productions machines. Since this is already a prerequisite in all the new FABs' machines, this issue is already solved.

One of the main difficulties in the existing 2D FABs is the degree of fitness between the layout structure and the current machines and operational sequences. When a new FAB is planned, its layout is optimised (according to either operational or maintenance considerations, as discussed above). But, due to the ever-changing technologies, typical FABs experience several cycles of change every year. Short-term cost considerations dictate that new machines are located according to free space and therefore, the optimality of the original layout is quickly lost. 3D layouts are much less sensitive to changes and therefore will usually stay in a near-optimal status. This is due to the fact that only one movement is being done from one machine to another without the need to use buffers and additional transportation systems in the way.

3. Model formulation, analysis and solution

3.1 *The modelling approach*

In order to design the layout concept introduced in the previous section, one should be able to determine an assignment of the production machines to the cells in the particular aisles. This problem consists of three main sub-problems.

1. Determination of the number of aisles and assigning each machine to an aisle (a partition problem).
2. Placing machines in a single aisle (a location problem).
3. Determining the optimal regime for the transportation device (a sequencing problem).

The partition of machines to different aisles arises since, in practical terms, placing all the machines into a single aisle will overload the transporter, which will become a bottleneck for the entire system.

The answers to these three sub-problems depend on various factors such as: aisle cost, transportation device velocity, area cost, cost of each additional level, etc. All aspects mentioned above are interdependent and therefore, the best approach would be to solve them simultaneously within a global optimisation problem. However, locating machines within a single aisle is by itself an NP-complete problem and so many real-world cases, especially in FAB environments where there are hundreds of machines, cannot be solved this way. Hence, this paper follows the two-stage approach which was implemented by Kaku *et al.* 1988 and Meller and Bozer 1997, and solves sub-problems (1) and (2) separately in a serial manner. The third sub-problem (sequencing the transportation device) is not solved here mainly in order to simplify our discussion. The general idea is to show that the performance of a 3D layout is better than its 2D counterpart even without optimising the transportation policy of the former—it is clear that its optimisation will make the 3D layout even more attractive.

3.2 *Assigning machines to cells in a single aisle*

Assume that a set of n machines are to be assigned to a single aisle. For practical considerations, only machine arrangements which will result in a symmetrical aisle structure will be considered. Thus, for example, if $n=19$, it is expected that nine machines will be on one side of the aisle and 10 on the other. In an extreme case, all 10 machines will be stacked one on top of the other yielding at most 10 levels. In the other extreme case, all 10 machines will be placed next to each other in a single level. Therefore, the number of possible sites that needs to be considered in each side of the aisle is $10 \cdot 10 = 100$. More generally, the solution space consists of

$$N \equiv 2\left\lceil\frac{n}{2}\right\rceil\left\lceil\frac{n}{2}\right\rceil \text{ possible sites arranged in } \left\lceil\frac{n}{2}\right\rceil \text{possible levels.}$$

The corresponding 3D-QAP formulation (see the Appendix) assumes that there are no major size differences among machines and hence it considers a standard structure where each machine occupies a single standard cell. However, machine sizes in the semiconductors industry may exhibit significant variance. There are several ways to express the different sizes within the 3D-QAP formulation. A straightforward way would be to designate a separate aisle for each size and assign all the machines in that size category to this aisle. Then, the arrangement of the machines within each aisle is done through the 3D-QAP. However, this approach will be viable only if the total volume (i.e. number of machines $\times$ their space requirement) needed for each category, is about the same. Another way is to associate each cell and each machine with their corresponding size category and make the appropriate changes in the constraints of the 3D-QAP formulation (see the Appendix for details). A third way is to define the standard cell size according to the smallest machine category, and treat larger machines as though they are composed of an integer multiple of the standard size. By inserting very large values in the relevant entries of the 'from–to' matrix that describes the frequency of batch transfers among machines, one can ensure that the pieces of each machine will be located adjacent to one another. Since all of the above adjustments lead to variants of the basic 3D-QAP model, this will be our reference model throughout this paper.

The lemma below focuses only on the transportation cost component in the objective function of the 3D-QAP model (the proof is given in the Appendix).

Lemma 1: *A solution to the 3D-QAP may yield lower transportation costs in comparison with its 2D counterpart.*

The proof of the lemma is partially based on the trade-off between the costs of the horizontal and vertical movements. Another important trade-off which should be taken into consideration is the one that exists between the area cost and the vertical construction cost. This trade-off can be analysed similarly to the approach taken in the proof above. Clearly, as the area cost becomes more dominant, the 3D layout becomes more attractive. Conversely, when the vertical construction cost increases the 3D layout becomes less attractive.

Solving the 3D-QAP. As mentioned before, QAP is a NP-hard problem. It is clear that 3D-QAP is NP-hard as well. Therefore, we developed a heuristic approach based on simulated annealing (SA) to solve it. SA is a general-purpose heuristic that has already been implemented to solve a large variety of problems. In particular, it proved to be useful for solving 2D layout problems (Wilhelm and Ward 1987, Connoly 1990, Meller and Bozer 1996). Heragu and Alfa (1992), in a survey of different solution algorithms developed for the QAP, have found that those that were based on SA were quite promising.

Our heuristic is composed of two routines: Initiate and Solve, which operate iteratively one after another (the two routines as well as our specific choices of parameters for the SA procedure are detailed in the appendix). Initiate(m, n) gives an initial layout for an aisle with length m and height n, while Solve(m, n) improves a given layout using the SA approach. These two routines are implemented through the following algorithm which checks all possible arrangements for K production machines.

1. Let $m=1$, $n=\mathrm{K}/2$; Assign an initial object function value (OFV): Min OFV = M.
2. Call Initiate(m, n) to get an initial 3D layout.
3. Call Solve(m, n) to obtain a SA solution, starting from the initial layout obtained in step 2. Calculate the OFV of the solution found.
4. If OFV < Min OFV, set Min OFV = OFV.
5. Let $m=m+1$; $n=n-1$.
6. If $n=0$, stop. Otherwise, go to Step 2.

This algorithm was tested as follows. First, random frequency matrices were generated and for each of them a 3D layout problem was solved using our SA algorithm. Then, a large number of random solutions were generated for each of these problem instances, where the CPU time required for the random generation was set according to the time used by the SA algorithm for the same instance (on average, 21 518 random solutions were generated for each instance). The best of these random solutions was used as a benchmark to compare with the SA solution. It was found that the objective function values in the solutions obtained by SA algorithm were lower by 10–25% than the randomly generated results. [The code of the algorithms is available from the corresponding author upon request.]

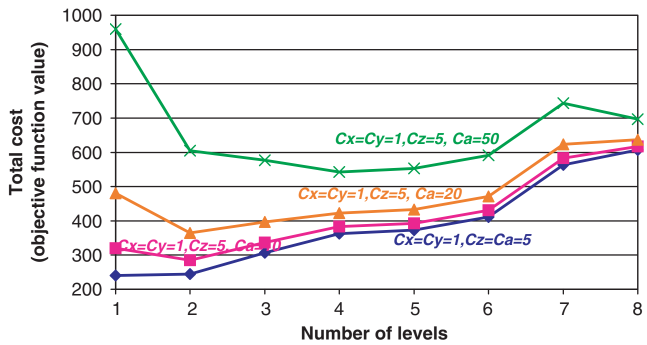

Figure 3. Objective function values of the SA algorithm for various cost data.

A frequency matrix used by Heragu (1992) (see the Appendix) was used to demonstrate the performance of the algorithm. Figure 3 shows the objective function values obtained through the proposed algorithm for this data with various combinations of cost data.

As demonstrated in figure 3, the main trade-off exists between the area cost and the movement cost along the vertical axis. For example, for the cost data $c_x = c_y = 1$, $c_z = c_a = 5$ (where c_x, c_y and c_z denote cost per unit distance in the x, y and z dimensions, respectively and c_a denotes per unit area cost) there is no reason to build more then a single level in the aisle, since vertical movement is expensive relative to the area cost. As the area cost grows, the solution contains more levels. Thus, with $c_x = c_y = 1$, $c_z = 5$, $c_a = 10$, the solution is to build two levels and with $c_x = c_y = 1$, $c_z = 5$, $c_a = 50$, the solution is to build four levels.

Another interesting phenomenon which can be seen in figure 3 is that the cost functions are rather flat near the optimal points. This issue, however, wasn't studied any further, as it goes beyond the scope of the current research.

3.3 *Assigning machines to aisles*

In practical cases, it will probably be impossible to assign all the machines to a single aisle since their number (typically, several hundreds) will overload the aisle's transporter. A similar problem was addressed in the first stage of the two-stage algorithm developed by Meller and Bozer (1997). This first stage, denoted by Meller and Bozer as floor assignment formulation (FAF), assigns machines to floors using an integer programming model.

Adopting the FAF concept and modifying it with some minor changes, led to a formulation which is denoted here as AAF (aisle assignment formulation)—see the Appendix.

The main difference between FAF and AAF is that the latter treats the number of aisles as a decision variable which affects the objective function (with a penalty charge for every additional aisle that is activated). Solving the AAF can be done through the linearisation method suggested in Meller and Bozer (1997) or through any of the numerous partition heuristics or general heuristic techniques available for that problem. Here, we don't develop this issue any further.

4. Comparative study of layouts for a FAB

A simulation experiment of the operations in a typical FAB was developed in order to compare the 2D and 3D layouts. The experiment focused on comparing the two operational parameters which play a major role in Little's Law: WIP and the throughput time (TPT)—the average time which the lot spends in the production floor (the latter is often referred to in the semiconductors industry as average cycle time). The operational characteristics of the system were held equal for both layouts to ensure a fair comparison. The simulation was first verified by comparing the results of the 2D layout with those reported in the literature.

The main features of the simulation experiment are listed below:

- *Description of the experiment*. The basic parameter values for the experiment were taken from the SEMATECH production layout data in Campbell and Ammenheuser 1999.
- The stages of the experiment's methodology.
- *Building the 2D layout simulation model*. This stage replicates the research reported in Campbell and Ammenheuser, 1999. It serves two purposes: first, to verify that the results correspond to results reported in the SEMATECH study and, second, to establish a benchmark for evaluating the 3D layout.
- *Constructing the 3D aisle-based simulation model*. The only difference between this model and the previous one was in the structure of the production layout.

After running the two models described above, the following assumptions were analysed:

$$\begin{cases} H_0\text{: } TPT_1 = TPT_2 \\ H_1\text{: } TPT_1 > TPT \end{cases} \quad \begin{cases} H_0\text{: } WIP_1 = WIP_2 \\ H_1\text{: } WIP_1 > WIP \end{cases}$$

where $WIP_{1/2}$ is the WIP in the 2D and 3D layout processes correspondingly and $TPT_{1/2}$ the product's

TPT on the production floor in the 2D and 3D layout correspondingly.

A thorough description of the simulation model developed for this research is given in Gurevich (2004). The main assumptions are given below:

1. The production machines are divided into two aisles according the model given in (12)–(16) and located within the aisle using the model given in (1)–(4).
2. Every aisle is served by a single transportation device which has a simultaneous movement in the horizontal and vertical directions. V_x, the horizontal velocity of the transportation device, was set to 1.52 m/sec and various other values were considered for the vertical velocity $V_z = \text{Alpha} * V_x$, $\text{Alpha} \leq 1$.
3. The material handling between the aisles is done by another transportation device that transports the material between the I/O points of the relevant aisles.
4. In the 2D layout, transportation between two machines is done by two types of transporters—within each bay there is an intra-bay transporter and between two bays there is an inter-bay transporter. The connection between these transporters is done via buffers. The 2D layout assumptions replicate the research reported in Campbell and Ammenheuser (1999).
5. In the proposed 3D layout, transportation of lots is made directly between pairs of machines unless the target machine is found to be busy. To handle such cases, a buffer space (where waiting lots are stored) was defined for each aisle. When the target machine becomes available, the lot is transported from the aisle's buffer to the relevant machine.
6. In the 300 mm layouts, buffers are practically eliminated since the WIP travels on a mono-rail device that passes through the entire layout until the machine where its next production stage takes place is ready to download it. In order to represent this layout and compare it with the proposed 3D layout all the constraints associated with the transportation resources and buffer capabilities in the initial 2D layout were removed.
7. Various levels of reliability parameters of the transportations devices (using different mean time between failure (MTBF) levels) were tested.
8. The processing times on the machines were assumed the same as in the Campbell and Ammenheuser (1999).

Figure 4 presents the main results of this study—the WIP over the simulation time for the 2D and 3D layout types. The upper line shows the WIP accumulation in the 2D layout. The three remaining lines show the WIP accumulation for the 3D layout under the three values assumed for transporter's vertical speed (V_z), expressed in relation to its horizontal speed (V_x). $V_z = \text{Alpha} * V_x$.

It can be observed that for the Alpha values presented here, there is a substantial gap in WIP levels across time (a 30% decrease) in favour the 3D layout. There are Alpha values where the system does not stabilise (e.g. Alpha < 0.5). Figure 5 below shows an analysis of the sensitivity of the results to the transporter's reliability.

Here, too, one can see clearly that the 3D layout has a significant advantage over the 2D model in all the MTBF levels examined. According to Little's Law, WIP reduction means a reduction of the TPT at a similar rate. The question is what elements of the TPT are shortened and by how much. Table 1 below presents the simulation results regarding the TPT components in the two layouts (the numbers in parentheses are associated with the 300 mm 2D layout).

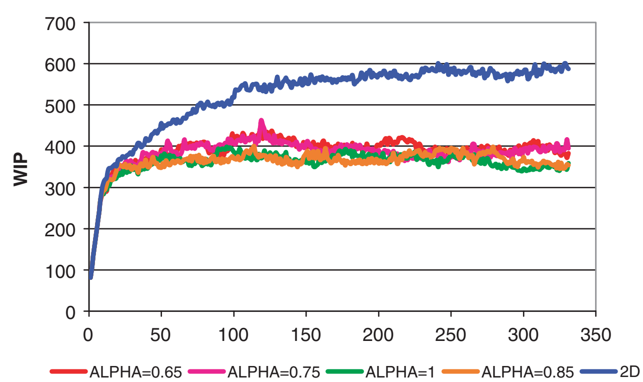

Figure 4. WIP accumulation across time.

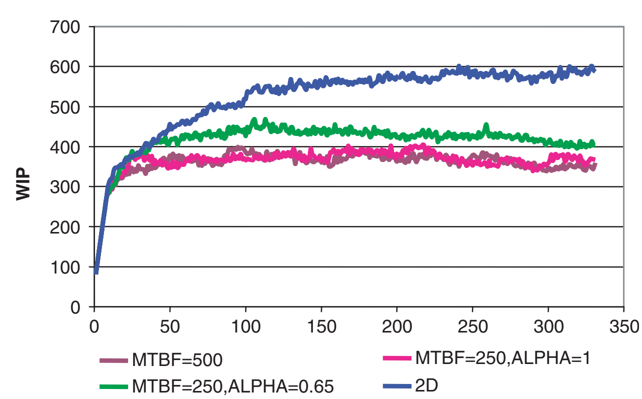

Figure 5. WIP across time with different reliability parameters of the transporter.

Table 1. Throughput time components.

	The 2D layout	The 3D layout	Reduction time (in hours)	Reduction time (%)
Processing (in hours)	190 (190)	190	0	0%
Batching time (in hours)	24.3 (27)	23.6	0.7	0.03%
Transportation (in hours)	8.8 (8)	1.8	7	80%
Waiting for transportation	257 (210)	123	134	52%
Waiting for machine repair (in hours)	30 (25)	23.6	6.4	21%
Total TPT (in hours)	510.1 (460)	362.0		

The most significant reduction in the total TPT time occurs in the waiting time for transportation followed by the transportation time itself. The explanation for this phenomenon lies in the fact that in the 3D layout under investigation, the buffers, which by nature obstruct WIP flow, were in fact eliminated. In the 2D layout, every move between two tools was done through *at least* one buffer, in those cases where the two machines were located in the same bay, and sometimes through two buffers, when the two machines were located in two separate bays. In the 3D layout, the transportation between the two machines was carried out in fact through *at most* one buffer, where the WIP must wait only if the target machine is occupied. This led to a significant reduction in the waiting time for transportation, which occurs mostly in those buffers we removed in the 3D layout. We can also see the reduction in transportation time, which is obtained mainly because of the reduction in the physical dimensions of the layout. However, this reduction has a minor effect on TPT reduction.

The results of 2D layout which represents the more advanced 300 mm FABs indicate a significant reduction in the waiting time for transportation compared with the conventional 2D FABs. This reduction was obtained by using the more advanced material handling devices that eliminated the need for lots to spend any time in buffer. However, compared with the proposed 3D layout (where transportation is usually done directly from point to point), the lots in the 300 mm FAB do stay relatively longer periods of time in transit as a result of the circular nature of the handling system.

Suppose there are 10 machines which are visited in a circular (cyclic) pattern by the handling system. A lot that needs to be transported from machine 10 to 9 will have to travel first through 9 machines $(10, 1, 2, \ldots, 9)$ before it reaches its destination. Another delaying effect may be caused when machine 10 was busy when the lot was near it but was released immediately after the lot has passed by it. Then, the machine will stay idle until the lot completes a full circle and reaches it again.

5. Discussion

This paper presents a novel idea of constructing an AS/RS structure for production layout. According to both the analytical and the experimental results presented here, the proposed layout seems to be a promising alternative to the existing 2D layouts especially in places where area costs are high. Regardless of the area costs, it is demonstrated that there are plausible scenarios in which the new layout will save on transportation costs and will yield better overall throughput. The simulation experiment indicates the potential to reduce throughput time and work in process by approximately 30%. Even if only a fraction of that number will eventually be achieved in real-world applications, the savings will amount to billions of dollars.

Future research is needed on the technical aspects of the 3D layout (some of which were presented in section 2). Once these aspects are treated in depth, a real world case can be modelled and analysed in detail. Also, it is clear that in practical environments there are certain acceptable layout dimensions, especially concerning the number of floors, which will require some adjustments in the heuristic procedure that was presented in section 3. Other directions for future research include testing various heuristics to solve the AAF model, attempting to replace the SA core of our heuristic with other known procedures, and more.

In summary, the main contributions in the paper are:

1. Proposing a brand new concept for the design of multi-floor production plants.
2. Demonstrating the potential effectiveness of the 3D layout.
3. Developing supporting tools for the planning of 3D layouts, based on customisation of existing algorithms.
4. Offering a comparative study of 2D and 3D layouts with real data taken from the semiconductors industry.

This study indicates that a 3D layout is a potentially valuable approach for reducing material handling costs, work in process levels and throughput times.

Appendix

3D-QAP formulation

$$\text{Min} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} f_{ij} d_{kl} x_{ik} x_{jl} + h\left(\max_f\left(\sum_{i=1}^{N}\sum_{k\in\{M_f\}} S_i x_{ik}\right)\right) \quad (1)$$

s.t

$$\sum_{i=1}^{N} x_{ik} = 1, \quad k = 1, 2, \ldots, N. \quad (2)$$

$$\sum_{k=1}^{N} x_{ik} = 1, \quad i = 1, 2, \ldots, N. \quad (3)$$

$$x_{ik} \in \{0, 1\}, \quad i, k = 1, 2, \ldots, N. \quad (4)$$

where

- x_{ik} Decision variable, equals to 1 if machine i is assigned to the site k and 0—otherwise.
- d_{kl} Distance between two sites—k and l.
- f_{ij} Frequency in which lots are transported between machines i and j.
- S_I Area required for machine i.
- $\{M_f\}$ Set of sites associated with floor f ($f = 1 \ldots \lceil n/2 \rceil$), defined a-priori.
- $\max_f(\sum_{i=1}^{N}\sum_{k\in\{M_f\}} S_i x_{ik})$. The area required for the 3D layout, assuming that it is determined by the biggest floor area.
- $h(S)$ Area cost function.

3D-QAP with multiple size machines

Suppose there are $l = 1, \ldots, L$ possible machines' sizes and let $\{A_l\}$ and $\{B_l\}$ represent, respectively, the machines and sites sets associated with size l. In this case

$$N \equiv 2\sum_{i=1}^{L}\left\lceil \frac{|A_l|}{2} \right\rceil$$

($|S|$ denotes the cardinality of the set S). Using this notation, the 3D-QAP formulation is adjusted by replacing constrains (2) and (3) with the constraints given in (5) and (6):

$$\sum_{i\in A_l}^{N} x_{ik} = 1, \quad k \in \{B_l\}, \ l = 1, \ldots, L \quad (5)$$

$$\sum_{k\in B_l}^{N} x_{ik} = 1, \quad i \in \{A_l\}, \ l = 1, \ldots, L \quad (6)$$

Proof of Lemma 1

Let (7) and (8) be the objective functions for QAP and 3D-QAP, respectively.

$$Z_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} f_{ij}(c_x d_x(k, l) + c_y d_y(k, l)) x_{ik} x_{jl} \quad (7)$$

$$Z_2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} f_{ij}(c_x d'_x(k, l) + c_y d'_y(k, l) + c_z d'_z(k, l)) x_{ik} x_{jl} \quad (8)$$

where c_x, c_y, c_z are the movement cost per unit distance along axis x, y, z; d is the distance matrix between the sites in the 2D plant layout problem; d' is the distance matrix between the sites in the 3D plant layout problem.

If X^* is the optimal solution which minimises (7), it is a feasible solution for the 3D layout problem, with objective function (8). The following simple operation proves an existence of cost values for which $Z_1(X^*) \geq Z_2(X^*)$.

Assuming that the movement is done along the main axis only, it is possible to represent the layout as a graph, with machines located on its vertices. Figure 6 shows the graph representations of 2D and 3D layouts. To obtain a 3D counterpart for the given 2D layout one needs to 'fold' the 2D layout across the imaginary line drown in the 'middle' of the 2D layout. The transformation of a 2D layout to its 3D counterpart amounts to drawing additional edges, connecting pairs of vertices which are equally distant from the imaginary line drown in the middle of the layout (this line divides the vertices to two sets—V and U). The distances from this line to any column b are denoted as $s(b)$. The relevant values of $s(b)$ are demonstrated in figure 6 above the diagrams.

Denoting the site where machine i is located by $w(i)$, we can write the objective functions mentioned above as follows:

$$Z_1 = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(c_x d_x(w(i), w(j)) + c_y d_y(w(i), w(j))) \quad (9)$$

$$Z_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(c_x d'_x(w(i), w(j)) + c_y d'_y(w(i), w(j)) + c_z d'_z(w(i), w(j))) \quad (10)$$

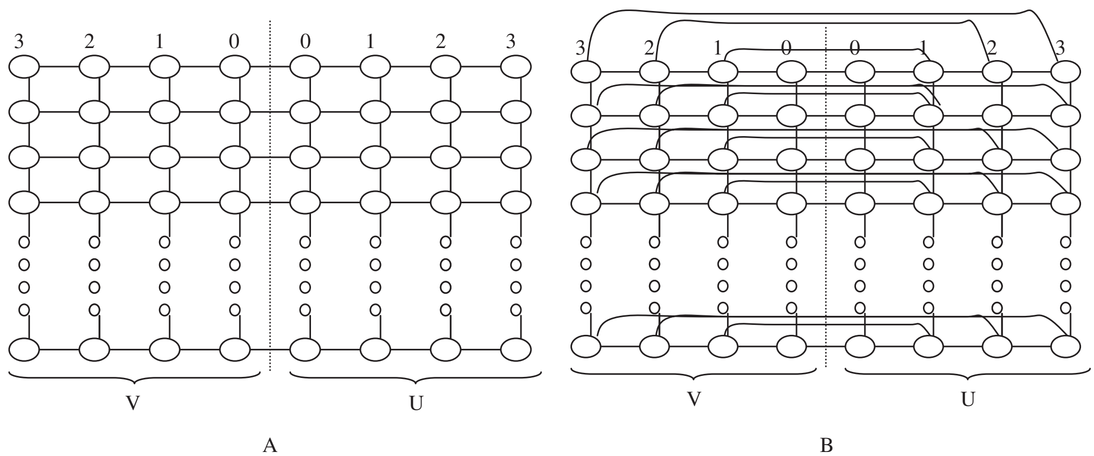

Figure 6. Graph representations of the 2D and 3D layouts.

Developing (10) we get,

$$Z_2 = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(c_x d'_x(w(i), w(j)) + c_y d'_y(w(i), w(j))$$

$$+ c_z d'_z(w(i), w(j))) = \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij} c_x d_x(w(i), w(j))$$

$$+ \sum_{i\in V}\sum_{j\in U} f_{ij} c_y + \sum_{j\in U}\sum_{i\in V} f_{ij} c_y + \sum_{i\in V}\sum_{j\in U} f_{ij} c_z |s(j) - s(i)|$$

$$+ \sum_{j\in U}\sum_{i\in V} f_{ij} c_z |s(j) - s(i)| + \sum_{i\in U}\sum_{j\in U} f_{ij} c_z d_y(w(i), w(j))$$

$$+ \sum_{j\in V}\sum_{i\in V} f_{ij} c_z d_y(w(i), w(j))$$

In the case $c_y = c_z$, we obtain:

$$Z_2 \le \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij} c_x d_x(w(i), w(j)) + \sum_{i\in V}\sum_{j\in U} f_{ij} c_z d_y(w(i), w(j))$$

$$+ \sum_{j\in U}\sum_{i\in V} f_{ij} c_z d_y(w(i), w(j))$$

$$+ \sum_{i\in U}\sum_{j\in U} f_{ij} c_z d_y(w(i), w(j)) + \sum_{j\in V}\sum_{i\in V} f_{ij} c_z d_y(w(i), w(j))$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}(c_x d_x(w(i), w(j)) + c_y d_y(w(i), w(j))) = Z_1$$

In general, assuming $c_x = c_y$, we obtain the existence of cost values (c_z and c_x) for which the 3D layout is preferred to the 2D one with respect to transportation cost. For example, the values for which the following expression holds:

$$\frac{c_z}{c_y} \le \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij} d_y(w(i), w(j)) - \sum_{i\in V}\sum_{j\in U} f_{ij} - \sum_{j\in U}\sum_{i\in V} f_{ij}}{\sum_{i\in V}\sum_{j\in U} f_{ij}|s(j) - s(i)| + \sum_{j\in V}\sum_{i\in V} f_{ij}|s(j) - s(i)| + \sum_{i\in U}\sum_{j\in U} f_{ij} d_y(w(i), w(j)) + \sum_{j\in V}\sum_{i\in V} f_{ij} d_y(w(i), w(j))} \quad (11)$$

Thus, if (11) holds the 3D layout is guaranteed to have lower transportation costs than its 2D counterpart. However, nothing can be said if (11) does not hold. □

Algorithms for 3D layout generation

Initiate (m, n)

This algorithm generates an initial layout by employing a so-called 3D filling curve to determine the order of in which sites are selected to be occupied with machines. The idea of filling curves was proposed by Bozer *et al.* (1994) and here we adapt it for our 3D environment. Figure 7 illustrates the implementation of the 3D filling curve.

The algorithm starts in the first floor with one of the cells closest to the I/O point. This cell will be the first site to fill. The second site is the cell in front of it. Afterwards the curve climbs to the next floor and continues in the same manner until the first column of cells in all m floors is occupied. Then, the algorithm starts again at the first floor in the cell adjacent to the cell in which the previous iteration started.

The machines that are assigned to the cells are chosen randomly from the set of unassigned machines—the reason for this is that an initial solution has a minor affect on the results of SA-based algorithms (Meller and Bozer 1996).

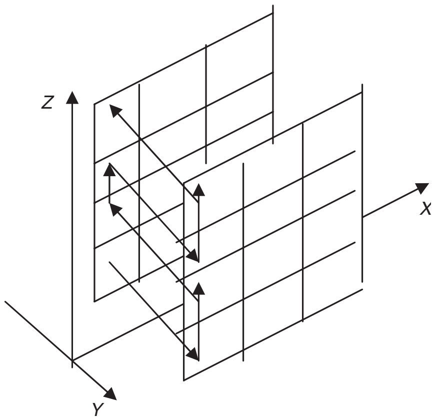

Figure 7. Generating an initial layout using the 3D filling curve method.

Solve (m, n)

1. Set $i := 1$; $j = 1$; $T_{\text{Iteration}} = T_s$; Iteration: = 1;
2. Set $j := j + 1$;
3. If $\Delta\text{OFV}(i, j) > 0$
   (a) Set $U := \text{UNIF}(0, 1)$;
   (b) Set $R = \text{Exp}(-\Delta\text{OFV}(i, j)/T_{\text{Iteration}})$
   (c) If R > U – go to step 4a, Else – go to step 5
4. Else
   (a) Set $k := i$; $i := j$; $j := \text{k}$;
5. If $j < N$ – go to step 2,
6. Else,
   (a) Set $i := i + 1$.
   (b) If $i <$ N – go to step 2,
   (c) Else, Set Iteration: = Iteration + 1
7. If Iteration < Limit
   (a) Set $i := 1$, go to step 2;
8. Else,
   (a) If $T_{\text{Iteration}} < T_f$ – stop,
   (b) Else, set Iteration: = 1, go to step 7a.

Where $\Delta\text{OFV}(i, j)$ is the improvement in objective function as the result of sweeping machine i with machine j; Iteration is the iteration counter in the same temperature; T_s is the starting temperature; T_f is the final temperature; Limit is the maximal number of iterations using the same temperature value.

Choosing parameters for SA algorithm

The specification of a SA algorithm includes the initial temperature (T_s), the final temperature (T_f), the cooling scheme and the number of iterations we stay at certain temperature.

There are various techniques to derive those parameters; many of them are given in Anderson and Vidal (1993). Since this is not the core of our research, we used the following simple approach:

1. The initial and the final temperatures were selected among several tested values.
2. The maximal number of iterations within the same temperature (Limit) was set to 50 (after testing several values and finding that changes within the range of 10–50 didn't cause any significant influence in the final solution.)
3. The cooling scheme that was used multiplies the current temperature with the parameter α. After testing various parameter values, the value $\alpha = 0.85$ was chosen.

The following table presents the objective function values for the case of $c_x = c_y = 1$, $c_z = 5$, $c_a = 10$ solving the problem with frequency matrix given in the text. We have chosen to use the parameters highlighted in the shadowed line in the following table.

			Floors number							
T_s	T_f	Limit	1	2	3	4	5	6	7	8
150	0.01	10	320	285	337	383	393	431	583	617
300	0.01	10	317	285	336	383	393	419	629	623
600	0.01	10	321	285	333	394	397	419	584	619
1000	0.01	10	320	285	335	383	391	419	584	626
150	0.02	10	320	285	337	383	419	431	583	617
300	0.02	10	317	285	341	403	393	419	584	619
600	0.02	10	321	297	333	394	401	419	586	619
1000	0.02	10	320	285	335	383	419	427	584	626
150	0.01	50	316	285	335	383	391	419	583	623
300	0.01	50	316	285	333	383	401	419	583	617
600	0.01	50	314	285	333	383	391	419	583	619
1000	0.01	50	319	285	333	383	391	419	583	617

Frequency matrix (Heragu 1992)

$$F=\begin{bmatrix}
0&1&0&8&0&2&3&0&0&0&0&0&5&2&0&0\\
&0&1&1&1&0&0&0&0&0&0&0&0&0&2&0\\
&&0&2&0&0&0&0&0&0&0&0&0&0&0&1\\
&&&0&4&14&11&0&0&0&0&0&0&0&0&0\\
&&&&0&1&0&0&0&0&0&0&0&0&0&0\\
&&&&&0&3&0&0&3&0&0&0&0&0&0\\
&&&&&&0&5&5&9&8&2&0&0&0&0\\
&&&&&&&0&8&0&0&0&0&0&0&0\\
&&&&&&&&0&0&0&6&0&0&0&0\\
&&&&&&&&&0&5&5&0&5&1&0\\
&&&&&&&&&&0&4&0&0&0&0\\
&&&&&&&&&&&0&2&10&5&0\\
&&&&&&&&&&&&0&0&0&0\\
&&&&&&&&&&&&&0&0&0\\
&&&&&&&&&&&&&&0&0\\
&&&&&&&&&&&&&&&0
\end{bmatrix}$$

Aisle assignment formulation

$$\text{Min} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{R}\sum_{l=1}^{R} c_m f_{ij} d_{re} y_{ir} y_{je} + c_r \sum_{r=1}^{R} z_r \tag{12}$$

s.t.

$$\sum_{r=1}^{R} y_{ir} = 1, \quad i = 1, 2, \ldots, n \tag{13}$$

$$\sum_{i=1}^{n} h_i y_{ir} \le \overline{T}_r, \quad r = 1, 2, \ldots, R \tag{14}$$

$$m \sum_{i=1}^{n} y_{ir} \le z_r, \quad r = 1, 2, \ldots, R \tag{15}$$

$$y_{ir} \in \{0,1\}, z_r \in \{0,1\} \quad i = 1, 2, \ldots, n; \quad r = 1 \ldots R \tag{16}$$

- c_m Movement cost between the aisles (per unit distance).
- c_r Cost of activating aisle r.
- f_{ik} Frequency of transporting batches between machines j and k.
- d_{re} Distance between aisle r and aisle e.
- $\overline{T}_r$ Time window available when a single transportation device works in aisle r.
- h_i Transportation device time needed by machine i (may be evaluated assuming average movement time and taking into consideration frequency matrix).
- y_{ir} A binary variable whose value of 1 if machine i is assigned to aisle r and 0 otherwise.
- z_r A binary decision variable whose value of 1 if we aisle r is activated.
- m A small positive number.

References

Adams, W.P. and Johnson, T.A., Improved linear programming-base lower bounds for the quadratic assignment problem. In *Quadratic Assignment and Related Problems*, DIMACS Series on Discrete Mathematics and Theoretical Computer Science 16, edited by P.M. Pardalos and H. Wolkowicz, pp. 43–75, 1994 (AMS: Providence, RI).

Anderson, K. and Vidal, R.V., Solving the quadratic assignment problem. In *Applied Simulated Annealing*, Lecture Notes in Economics and Mathematical Systems, edited by R.V. Vidal, 1993 (Springer-Verlag: New York, NY).

Barbosa-Povoa, A.P., Mateus, R. and Nocais, A.Q., Optimal 3D layout of industrial facilities. *Int. J. Prod. Res.*, 2002, **40**(7), 1669–1698.

Bozer, Y.A., Meller, R.D. and Erlebacher, S.J., An Improvement-type layout algorithm for single and multiple-floor facilities. *Manage. Sci.*, 1994, **40**(7), 918–932.

Buffa, E.S., Armour, G.C. and Vollmann, T.E., Allocating facilities with CRAFT. *Harv. Business Rev.*, 1964, **42**(2), 136–158.

Campbell, E. and Ammenheuser, J., 300 mm factory layout and material handling modeling, Phase 2 Report, *International SEMATECH*, 1999.

Chang, C.K. and Chang, S.K., A layer based layout approach for semiconductors fabrication facilities. *IEEE/SEMI Semiconductor Manufacturing Conference*, 1998, pp. 385–390.

Cinar, U., Facilities planning: A systems analysis and space allocation approach. In *Spatial Synthesis in Computer Aided Building Design*, edited by C.M. Easton, pp. 19–40, 1975 (Wiley: New York, NY).

Connoly, D.T., An improved annealing scheme for QAP. *Euro. J. Op. Res.*, 1990, **46**(1), 93–100.

Domschke, W. and Krispin, G., Location and layout planning—A survey. *OR Spektrum*, 1997, **19**(3), 181–194.

Gilmore, P.C., Optimal and suboptimal algorithms for the quadratic assignment problem. *SIAM J. Appl. Math.*, 1962, **10**(2), 305–313.

Geiger, C.D., Hase, R., Takoudis, C.G. and Uzsoy, R., Alternative facility layout for semiconductor wafer fabrication facilities. *IEEE Trans. Compon., Packag. Manuf. Techn., Part C*, 1997, **20**(2), 152–163.

Golany, B. and Rosenblatt, R.J., A distance assignment approach to the facility layout problem. *Euro. J. Op. Res.*, 1997, **57**(2), 253–270.

Gurevich, A., 2004. Developing 3-D layout for fabrication plants. M.Sc. thesis, The Technion – Israel Institute of Technology (in Hebrew).
Heragu, S.S., Recent models and techniques for solving the layout problem. *Euro. J. Op. Res*, 1992, **57**, 136–144.
Hsieh, S. and Hung, C.R., Feasibility study of modular plant for 300 mm-IC fabrications. *J. Intel. Manuf.*, 2004, **15**(2), 233–244.
Johnson, R.V., SPACECRAFT for multi-floor layout planning. *Manage. Sci.*, 1982, **28**(4), 407–417.
Kaku, B.K., Thompson, G.L. and Baybars, I., A heuristic method for multi story layout problem'. *Euro. J. Op. Res.*, 1988, **37**(2), 384–397.
Kaufmann, L. and Broeckx, F., An algorithm for the quadratic assignment problem using benders' decomposition. *Euro. J. Op. Res.*, 1978, **2**, 204–211.
Koopmans, T.C. and Beckmann, M.J., Assignment problems and location of economic activities. *Econometrica*, 1957, **25**, 53–76.
Lawler, E.L., The quadratic assignment problem. *Manage. Sci.*, 1963, **9**(4), 586–599.
Mackulak, G.T. and Savory, P., A simulation based experiment for comparing AMHS performance in a semiconductor fabrication facility. *IEEE Trans. Semicon. Manuf.*, 2001, **14**(3), 273–280.
Meller, R.D. and Bozer, Y.A., A new simulated annealing algorithm for facility layout problem. *Int. J. Prod. Res.*, 1996, **34**(6), 1675–1692.
Meller, R.D. and Bozer, Y.A., Alternative approaches to solve multi-floor facility layout problem. *J. Manuf. Syst.*, 1997, **16**(3), 192–203.
Nazzal, D. and Bodner, D.A., A simulation-based design framework for automated material handling systems in 300 mm fabrication facilities, in *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P.J. Sánchez, D. Ferrin and D.J. Morrice, 2003, pp. 1351–1359.
Patsiatzis, D.I. and Papageorgiou, L.G., Efficient solution approaches for the multi-floor process plant layout problem. *Indust. Eng. Chem. Res.*, 2003, **42**(4), 811–824.
Qu, P., Skinner, G.E. and Mason, S.J., Sizing a pilot production line using simulation, in *Proceedings of the 2001 Winter Simulation Conference*, edited by B.A. Peters, J.S. Smith, D.J. Medeiros and M.W. Rohrer, 2001, pp. 1188–1193.
Quinn, T. and Bass, E., 300 mm Factory Layout and Material Handling Modeling, Phase 1 Report, *International SEMATECH*, 1999.
Sahni, S. and Gonzalez, T., P-Complete approximation problems. *J. Assoc. Comp. Mach.*, 1976, **23**, 555–565.
Suzuki, A., Fuchino, T. and Muraki, M., Equipment arrangement for batch plants in multi-floor buildings with integer programming. *J. Chem. Eng. Japan*, 1991, **24**(6), 737–742.
Tompkins, J.A. and White, J.A., *Facilities Planning*, 1984 (Wiley: New York, NY).
Wilhelm, M.R. and Ward, T.L., Solving quadratic assignment problems by simulated annealing. *IIE Trans.*, 1987, **19**(1), 107–119.
Yang, T.H., Rajasekharan, M. and Peters, B.A., Semiconductor fabrication facility design using a hybrid search methodology. *Comp. & Indust. Eng.*, 1999, **36**(3), 565–583.
Yang, T.H., Su, C.T. and Hsu, Y.R., Systematic layout planning: a study on semiconductor wafer fabrication facilities. *Int. J. Op. & Prod. Manage.*, 2001, **20**(11–12), 1360–1372.

Boaz Golany is the Dean of the Faculty of Industrial Engineering and Management (IE&M) at the Technion - Israel Institute of Technology where he holds the Samuel Gorney Chair in Engineering. He is also a Senior Research Fellow at the Austin-based technology think tank IC^2. Dr. Golany received his BSc in IE&M from the Technion in 1982 and a PhD from the University of Texas at Austin in 1985. Dr. Golany has published over 70 papers in peer-reviewed journals and 15 book chapters in the areas of efficiency and effectiveness analysis, supply chain management, project management, and more. He has held editorial positions for the *Journal of Productivity Analysis, IIE Transactions*, and *Omega*. He also served as vice president of the Israeli Operations Research Society during 1996–2000.

Alexey Gurevich has BSc degree in IE&M and MSc in Operations Research and Systems Analysis, both from Technion. He currently serves as senior researcher in the Center for Systems Analysis in the Israel Defense Forces' Planning Branch. In his position he mainly concentrates on employing operations research tools to solve practical problems. In addition, he also teaches in Israeli colleges and universities.

Emanuel (Emil) Paz Puzailov is a veteran Industrial Engineer at Intel Corporation, currently acting as the Operations Finance Manager of Intel's FAB18 which is located in Israel. Mr. Puzailov holds a BSc from the Technion (1998) and an MBA from the Ben-Gurion University (2001). He has accumulated significant experience in various aspects of the Semiconductors Industry including layout planning, capacity planning, cycle time analysis & reduction activities, and optimization of scheduling activities.