

REDUCING MANUFACTURING CYCLE TIME OF WAFER FAB WITH SIMULATION

Giam Kim Toh, Ui Wei Teck, Alimin Lie
*Chartered Semiconductor Manufacturing Pte Ltd, 2 Science Park Drive,
Singapore 0511*

and

Dr George Sun, Dr Wang Ming, Kelvin Kok
*CIMTEK Pte Ltd, SISIR Building, #01-65, 1 Science Park Drive,
Singapore 0511*

ABSTRACT

Wafer Fabrication process is the starting point of making any computers or integrated circuit (IC) products. It is complex. On average, a wafer needs to go through more than 300 operational steps before shipping for assembly and test. It requires high investment. This makes the control of wafer fab very challenging and the results can be very encouraging. A simulation project has been carried out in Chartered Semiconductor Manufacturing Pte Ltd (CSM) to study the current facilities and control of Work-In-Process (WIP) in order to reduce Manufacturing Cycle Time. The results have shown that the total WIP and cycle time can be reduced by 30% to 40% while throughput is maintained and stabilized. This paper describes the principle of controlling WIP in a wafer fab environment and the simulation results.

1. Introduction

Wafer Fabrication is a process to build circuit or die on a wafer one layer at a time. It involves a sequence of different kinds of operations (such as lithography and etching), with the same operation appearing repeatedly in the sequence.

After the circuit or die is being built, a protective coating of oxide material is deposited over the entire surface of the wafer. The wafer is then sent to a final lithography step which opens the windows or pads that will be used later when bonding the circuit. The final step is testing of its ability to perform the operations for which it was designed. The schematic flow of a Wafer Fabrication process is shown in Figure 1.

Chartered Semiconductor Manufacturing Pte Ltd (CSM) is the first sub-micron Integrated Circuit (IC) manufacturing company in Singapore which operates one of the most advanced IC processing facilities in South-East Asia. Certified to ISO 9002 global standards, CSM

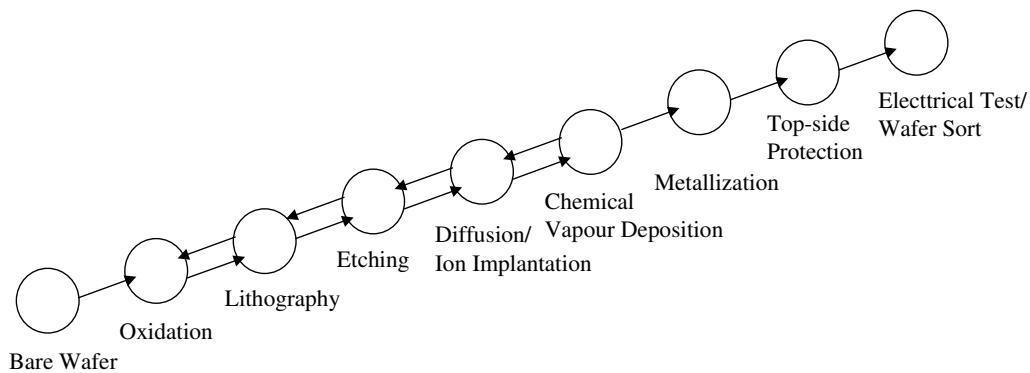


Fig. 1 Wafer Fabrication Process

offers a full line of advanced CMOS process ranging from 0.6 micron to 1.5 micron and a variety of technologies including analog, digital, mixed signal, EEPROM, high-density ROM, high speed SRAM and NVSRAM.

CSM is a member of the Singapore Technologies Venture (STV), the technology sector of the Singapore Technologies group of companies.

Since early 1994, a simulation project had been carried out in CSM. The objective of this project was to simulate the existing facilities for more accurate and effective capacity planning, and to introduce simulation technology into CSM for future process improvement. A commercial software, WITNESS, is adopted for this project. WITNESS is an interactive simulation software[1][2] developed by AT&T ISTEEL. To speed up the project and transfer simulation project knowledge, CIMTEK Pte Ltd, the sole distributor of WITNESS in South East Asia, was engaged in developing the initial model and in providing advice on the simulation analysis. Several experiments on cycle time variation with Work-In-Process (WIP) level were performed, of which results are presented here.

2. Control WIP In A Loop Flow Environment

Wafer fab is a loop flow where wafers must be processed on the same machine again and again. Here the capacity of a machine is shared by operations of different stages. It is meaningless, therefore, to only control the total WIP in front of a machine. Instead, the waiting parts should be categorised by their operations, and the WIP of each operation must be separately controlled. To avoid the domain specific terminology, a simple example is used to explain the principle of controlling WIP in a wafer fab environment.

Figure 2 is a much simplified illustration of the loop flow production. Assume a product needs to go through nine operations on three machines. In the first three operations, the product goes through M1, M2 and M3 (first layer: L01); and is sent back to M1, goes through the three machines in its 4th to 6th operation (second layer: L02); and sent back to M1 again to go through M1 to M3 in its 7th to 9th operation (third layer: L03).

Operation	1	2	3	4	5	6	7	8	9
Machine	M1	M2	M3	M1	M2	M3	M1	M2	M3
Layer	L01	L01	L01	L02	L02	L02	L03	L03	L03

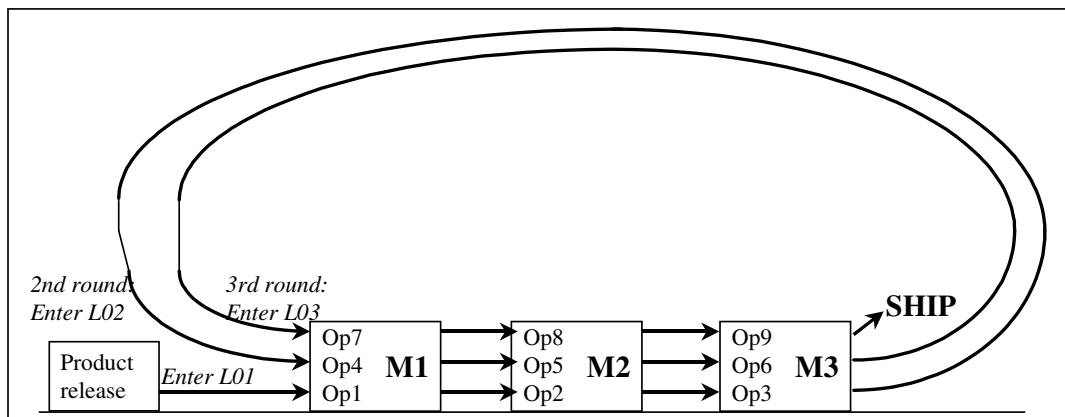


Fig. 2. A Simple Loop Flow Environment

In order to achieve a smooth output, the total WIP within each layer must be maintained at a relatively stable level. The total WIP within a layer is defined as all the lots after the first operation of this layer and before the first operation of the downstream layer, e.g. the total WIP within layer L02 refers to all the lots after Operation 4 and before Operation 7. To stabilize, the WIP for a layer must be replenished whenever a lot is out of this layer. As shown in Figure 3, for example, if a lot is pulled from "buffer in L02" to start Operation 7, another lot must be pulled in from "buffer in L01" to start Operation 4.

In real production environment, however, the number of machines each layer covers and the layers that various products go through may be different. To design a WIP control system, two tasks are essential:

- *to decide which layers to control.* For some layers, lots may not move fast enough to wait in front of the entry machine to be pulled by the next layer. It would be logical not to control these layers;
- *to determine the WIP volume for each layer.* For all the layers under control, some may go through long cycle, thus requiring more WIP staging in the layer, whereas products in other layers move fast and require less WIP staging.

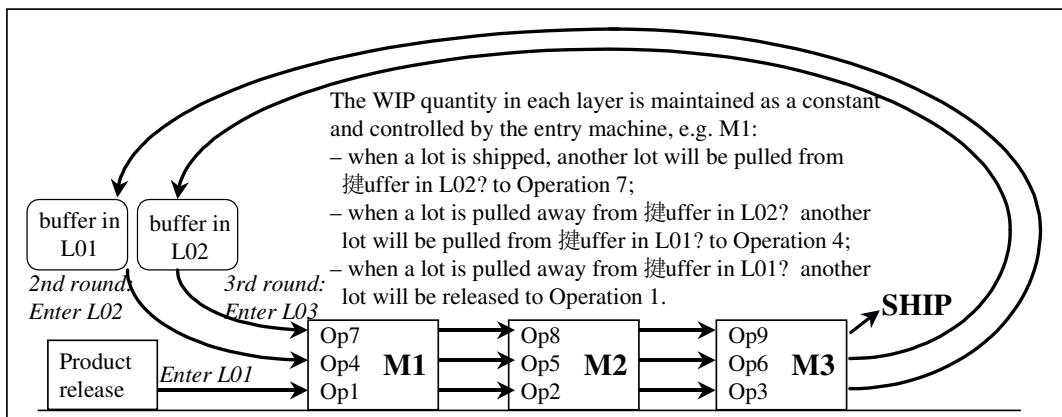


Fig. 3. Control WIP by Layer: Replenish the Layer whenever a Lot is Out

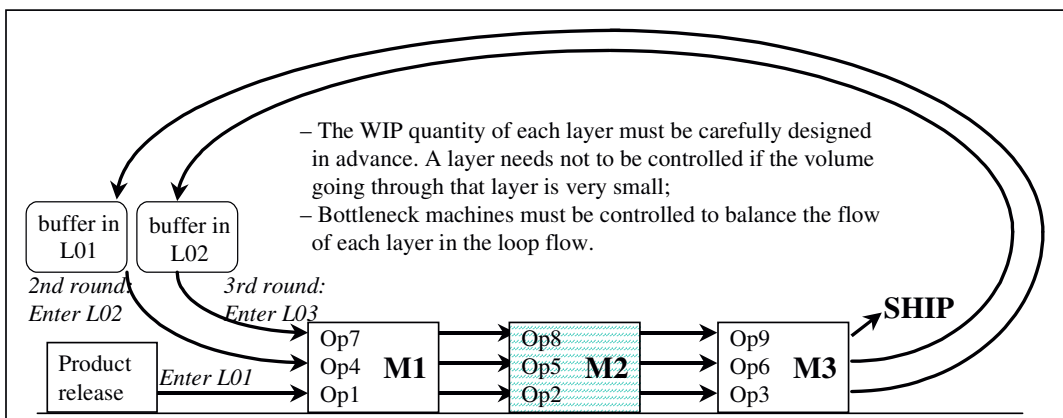


Fig. 4. Control WIP by Layer: Design WIP and Control Bottleneck

To make sure products move smoothly within a layer, bottleneck machines must be controlled to balance flow between layers. In Figure 5, it is assumed that M2 is a bottleneck. The time of using M2, therefore, must be proportionally distributed to Operation 2, 5, and 8. Dispatching rules may be devised for loading lots to the bottleneck machine, such as the layer where the WIP quantity completed by this machine is closest to its controlled volume is processed first.

There is neither generic rule in deciding which layer to control nor general relationship between the WIP per layer and the total output. The design of the WIP control system is a trial and error process. As it is too time-consuming and risky to experiment in real production, computer simulation provides an excellent platform for designing this WIP control mechanism.

3. Data Collection And Analysis

In order to build a simulation model of the wafer fab process, four types of data are required: process routings, processing times, equipment data, and wafer release. In CSM, the routing data and wafer release information are captured in a specialized data collection system. The processing times and equipment data are maintained by Industrial Engineering for capacity analysis purpose.

There are about 50 process routings in the production database. By combining the similar flows, the data of 26 active process routings were analysed and modelled in this project. For each of these processes, about 200 major operational steps (excluding measurement and inspection) were modelled in detail. The measurement and inspection operations were modelled using distribution, rather than as discrete events. The demand of a particular process may change from time to time, but it is stable for months, such as Routing 5 in Table 1. Therefore, a predetermined control mechanism would be applicable for several months, and the parameters need to be reviewed when demand pattern changes.

Table 1. Demand by Routing in Percentage

	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
Routing 1	12.7	10.7	19.2	10.5	14	7.5	7.9	8.8	10.5	7.5	12.7	12.7
Routing 2	5.7	6.4	9.7	12.1	9.4	7.7	6.1	13.9	13.1	14.4	4.7	22.3
Routing 3	0.5	0.2	2.9	1.8	5.6	4.8	8.6	16	7.3	12	12.1	14.7
Routing 4	0	0.1	0	0.1	1.1	7	7.9	0	0	0	0	0
Routing 5	0	0	6.4	9.8	0	9.2	27.2	24.1	24.4	23.9	27.3	15.7
Routing 6	1.4	8.3	0	0.8	7.7	5.9	6.6	7.7	6.3	4.8	7.4	8
...												
Routing 26	0.3	0.1	0.1	0.1	0.1	0.4	1.6	1.8	2	2.1	1	0.3

The operational specification is termed as a Recipe in wafer fab. One machine is able to process multiple Recipe types, and the processing times are machine and Recipe dependent. Because operations of many process routings may require the same recipe, the processing times and their related machine and Recipe are stored in separate files in order to reduce the total number of records and simplify the data maintenance.

Equipment data included the machine quantity, the number of wafers that it can process simultaneously, down times, setup times and preventive maintenance. Information of minor processes such as Bake, Sink, Inspection, Measurement, etc. were collected and

modelled. Wafer release information included the process routings that the wafer follows, the lot size (number of wafers per cassette), and the date when it is released.

Since simulation is a trial and error process, initial analysis was carried out to determine the layers to be controlled and the WIP volume by layer.

3.1 Select Layers to Control

Based on a set of criteria, 13 out of a total of 33 layers were selected for control. The criteria to establish layers to be controlled are:

i). *Volume* that go through each Layer. It is more effective to control a layer with large volume than one with small volume. Table 2 shows the historical data of volume percentage that go through the selected 13 layers. Because the types of wafer to release is well controlled by Planning Department, it can be clearly seen that the volume to process through each layer is very stable.

Table 2. Volume being Processed by the 13 Selected Layers in Percentage

	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May
L02	7.2	7.3	7.3	7.4	7.3	7.3	7.3	7.2	7.3	7.2	7.3	7.2
L03	7.2	7.3	7.3	7.4	7.3	7.3	7.3	7.2	7.3	7.2	7.3	7.2
L04	7.2	7.3	7.3	7.4	7.3	7.3	7.3	7.2	7.3	7.2	7.3	7.2
L09	2.2	2.1	2.1	2	1.9	1.8	1.4	1.8	1.1	1.4	1.7	1.4
L22	7.1	6.7	7.3	7.3	6.7	6.9	6.8	6.7	6.8	6.8	6.7	6.6
L25	14.8	15.6	15.2	15	14.8	15.1	14.6	14.3	14.4	14.3	14.1	14.3
L27	9	9.4	9.7	9.7	9.7	10.1	9.9	10.1	9.7	9.8	9.1	10.2
L28	7.2	7.3	7.3	7.4	7.3	7.3	7.3	7.2	7.3	7.2	7.3	7.2
L29	6.4	6.8	6	5.9	5.9	5.8	5.5	4.9	6	5.6	5.3	5.5
L30	6.4	6.8	6	5.9	5.9	5.8	5.5	4.9	6	5.6	5.3	5.5
L31	6.4	6.8	6	5.9	5.9	5.8	5.5	4.9	6	5.6	5.3	5.5
L32	0.9	0.6	1.3	1.5	1.3	1.5	1.7	2.3	1.2	1.6	1.9	1.7
L33	7.2	7.3	7.3	7.4	7.3	7.3	7.3	7.2	7.3	7.2	7.3	7.2

ii). *Frequency* that a Layer is involved in major routings. The value of a layer in Table 2 indicates the progress of production. It does not, however, represent the actual circuit layer being deposited on the wafer. In other words, products may go through different layers and different sequence of layers. It is more effective to control the layer that is gone through by most process routings than those involved in a limited number of minor routings.

iii). *Current WIP* by Layer. If there is a high WIP in current production for a layer, it is necessary to control. This can be applied to the simulated WIP as well.

3.2 Decide Volume of WIP by Layer

The determination of the WIP volume by layer involves two steps: i). Decide the target total WIP in process; ii). Distribute the WIP over the selected layers by the formula: WIP volume of this layer = target total WIP x WIP volume ratio of this layer.

Three methods were used to determine the WIP volume ratio for the selected layers: i) Proportional to the volume that go through each layer; ii) Proportional to the number of functional areas each layer visits; iii). Proportional to the number of operational steps each

layer visits. The proposed ratio of WIP volume by Layer based on the above methods are shown in Table 3. If the total WIP is designed to be 15K and Method 1 is used, for example, the WIP volume to be controlled in layer L02 will be: $15K \times 7.2\% = 1.08K$.

Table 3. Ratio of WIP Volume to Control by Layer

Method	L02	L03	L04	L09	L22	L25	L27	L28	L29	L30	L31	L32	L33
1	7.2	7.2	7.2	2.7	7.2	14.3	9.8	7.2	5.5	5.5	5.5	1.8	7.2
2	10.4	4.4	13.6	3.3	4.9	14.2	12.9	8.8	3.3	3.2	3.3	1.7	5.0
3	9.7	4.7	9.4	3.6	5.9	11.5	9.8	11.0	6.4	5.5	4.8	2.2	6.3

4. Simulation Results

The simulation model is completely data driven. The process routings, processing times are read in from data files. As the simulation clock advances, the wafer release information is read in and wafers are released into the model on a daily basis.

To start with, the simulation model of exiting facilities was developed. It took about one month for data collection and analysis, and two weeks for initial model building. Results of the production performance generated by the model were presented to Senior Management and Engineers and Managers from Industrial Engineering, Planning and Information Technology. The data and assumptions used in the simulation were better understood by various parties involved. This also served as a model validation process as the simulation results of output, cycle time, machine utilization, WIP were compared with the actual production.

Following that, the principle for controlling WIP by layer in a loop flow environment were presented. The layers selected to control and WIP volume of controlled layers were discussed during the meetings. The data as shown in Table 3 were modelled, and dozens of results were generated from simulation experiments.

Interesting results were drawn from the experiments. Out of the 3 methods to determine the WIP volume by layer, the first method yielded the best result. For all the simulation runs, the wafer movement in certain layers were so slow that no wafer could be pulled in to the next control layer when replenish was required, that is, the number of layers to control could be reduced.

It was noticed from the simulation results that not only would the cycle time be reduced, other benefits could be achieved by controlling WIP by layer. The simulation results consistently showed that the cycle time was more stabilized for all product types. The first chart in Figure 5 displayed the distribution of overall cycle time when WIP was not controlled, where the standard deviation of cycle time could be as long as 7.1 days, which reflected the existing production. In a WIP control system as shown in the second chart of Figure 5, the standard deviation of cycle time could be reduced to 5 days even if the WIP was the same. Along with a more stable cycle time, the output over several months was also more predictable.

Figure 6 shows that in the designed WIP control system, output could be maintained when total WIP was reduced by 30%. Output starts to drop if the total WIP is too low. The overall manufacturing cycle time is continuously shortened if total WIP is reduced. When the output is maintained at the optimum level, the cycle time could be reduced by 40%.

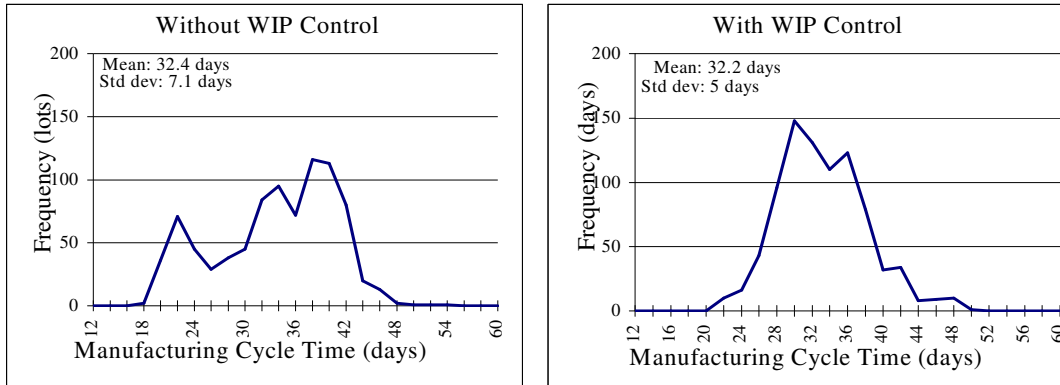


Fig. 5. WIP Control Stabilizes Cycle Time

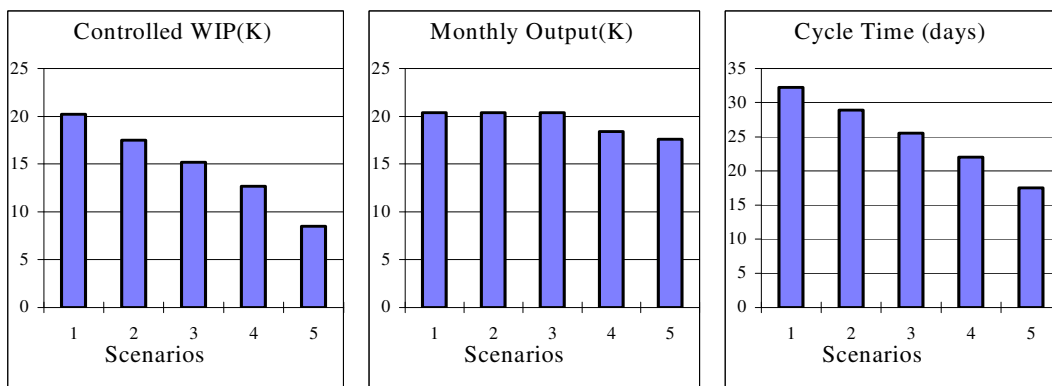


Fig. 6. WIP Control Reduces Cycle Time

5. Implementation Issues

The overall manufacturing cycle time of a wafer ranges from half a month to one and a half month. In production floor, therefore, the impact of reducing WIP on the output, cycle time, and utilization is not immediately seen. When a WIP control system is introduced, caution must be taken to introduce the system step by step to gain experience, prevent unexpected loss, and build up confidence on implementation.

Three steps were designed to introduce the system of controlling WIP by layer, which are:

1. Stabilize WIP at each layer. The first step is to control the material movement by layer, in order for the WIP of each layer to be stabilized. This would improve the manufacturing visibility because the management would be able to predict the output, time to deliver a particular product, and potential problem areas.
2. Adjust WIP quantity by Layer according to the simulated ratio. Once the WIP is stabilized, the quantity at each layer can be gradually adjusted to approach the simulated ratio. This would further stabilize the monthly output and cycle time of each layer.
3. Gradually reduce WIP by layer proportionally. This is the ultimate step in introducing a WIP control system. It requires the collaboration from Industrial Engineering, Planning,

Production, Material supply, and Information systems. In a well implemented WIP control system, the cycle time and urgent orders would be reduced, and in the meantime the output must be maintained.

6. Conclusions

The process flow in wafer fab plant is loop flow in the sense that wafers are processed on the same machine again and again. Since the capacity of a machine is shared by many stages of the same flow, it is important to control the WIP of each stage, rather than only the WIP in front of a machine.

The production stages or layers to control and the volume of WIP by layer are environment dependent. There is no general solution to this problem. Because of this complexity, simulation lends itself excellently to the evaluation of production performance, designing of the WIP control system, and improvement of operational steps. For such a large scale project, time dedication from Engineers, collaboration of various department, disciplinary of data entry, strong technical support from vendors are crucial to its success.

Controlling WIP by layer would first of all stabilize the WIP and cycle time, and improve the visibility in a loop flow environment. That is, the delivery is more predictable than in the environment when WIP is not controlled. Furthermore, if WIP is not controlled by stages, more inventory is required just-in-case of unexpected machine down times or shortage of materials. Reducing WIP in a controlled system will stabilize and maintain output, and at the same time, reduce manufacturing cycle time.

The impact of reducing WIP on the output, cycle time, and utilization may not be immediately seen on production shop floor. To gain experience and build up confidence, it is recommended to implement the WIP control system step by step.

Acknowledgement

The authors would like to thank Mr Choong Chan Yong, Vice President for Materials, Mr Steve Yeo, Planning Manager, Mr Peh Chin Sin, Information Technology Manager, for their encouragement, support and suggestions throughout the project. Thanks also go to Dr Zhuang Li of National University of Singapore for his advice on the control of WIP in a loop flow environment.

References

- [1] Pidd, M. 1988 "Computer Simulation in Managamnet Science - 2nd ed". John Wiley & Sons
- [2] Wang, M.; G. Sun; D. Wang. 1993. "Manufacturing Simulation: An Effective Tool for Productivity Improvement". In *Proceedings of 3rd International Microelectronics & Systems '93 Conference*, August. Malaysia