Scheduling Revisited Workstations in Integrated-Circuit Fabrication

Paul J. Kline
Semiconductor Process and Design Center
Texas Instruments Incorporated
P.O. Box 655012, M/S 3635
Dallas, TX 75265
kline@csc.ti.com

Abstract

The cost of building new semiconductor wafer fabrication factories has grown rapidly, and a state-of-the-art fab may cost $250 million or more. Obtaining an acceptable return on this investment requires high productivity from the fabrication facilities.

This paper describes the Photo Dispatcher system which has been developed to make machine-loading recommendations at a set of key fab machines. Dispatching policies that generally perform well in job shops (e.g., Shortest Remaining Processing Time) perform poorly for workstations such as photolithography which are visited multiple times by the same lot of silicon wafers.

The Photo Dispatcher evaluates the history of workloads throughout the fab and identifies bottleneck areas. The scheduler then assigns priorities to lots depending on where they are headed after photolithography. These priorities are designed to avoid starving bottleneck workstations and to give preference to lots that are headed to areas where they can be processed with minimal waiting. Other factors considered by the scheduler to establish priorities are the nearness of a lot to the end of its process flow and the time that the lot has already been waiting in queue.

Simulations that model the equipment and products in one of Texas Instruments's wafer fabs show the Photo Dispatcher can produce a 10% improvement in the time required to fabricate integrated circuits.

Introduction

Texas Instruments has a number of integrated-circuit (IC) wafer fabs which produce many different chip types. Depending on the type of chip on a wafer, fabricating that wafer will place very different demands on the processing equipment. Planning systems are used to produce weekly wafer start-plans that are within the capacity of the fab equipment. These planning systems do not develop a schedule for when each wafer will visit each machine group; instead they try to ensure that no more than a week's worth of work is started for all machine groups.

While good start plans have helped avoid some of the problems of machine overloading and late orders, these problems have not totally disappeared in the wafer fabs. Machine breakdowns, rework, etc., make it inevitable that production rarely proceeds as smoothly as desired. The manufacturing staff reacts to these disruptions by reprioritizing lots of wafers to expedite lots that are behind schedule or to cure workload imbalances on equipment. We developed the Photo Dispatcher to investigate the possibility of automating the scheduling of a set of key machines in the photolithography area.

The photolithography area was chosen because lots continually revisit this area during their processing, so improved scheduling in this area should have wide-reaching impact on the wafer fab. Figure 1 shows a typical process flow for producing a bipolar device with seven pattern steps. As shown in the figure, all of the pattern steps are performed on the same set of projection printers, Printers Grp. A scheduler for the Printers Grp would impact this device seven times as opposed to a scheduler for Depos Grp 6 which would would impact this device only once.

Previous Research

One approach to scheduling the Printers Grp would be to generate a Gantt chart each shift that shows which lots should be processed on which projection printers at which times. We decided not to take this approach because we felt that these schedules would quickly become obsolete because of projection printer breakdowns, unpredictable lot arrivals, and the unpredictable need to rework lots whose first patterning was unsatisfactory. This approach might also be difficult to scale up to schedule all the machine groups in the wafer fab because of the large number of machines (400+) and lots in process (400+).

An alternative is to wait until it is time to load a free machine and then decide which lot to load based on what is in queue and current fab conditions. Dispatch policies, which have been studied extensively in Operations Research (e.g., Panwalker & Iskander, 1977), are one way to make this decision. First-in-First-Out

Process Step	Step Type	Equipment
First Oxidation	Layer	Furnace Grp 10
DUF Pattern	Pattern	Printers Grp
DUF Diffusion	Dope	Furnace Grp 20
Epitaxial Deposition	Layer	Epi Reactors
Second Oxidation	Layer	Furnace Grp 10
Isolation Pattern	Pattern	Printers Grp
Isolation Diffusion	Dope	Furnace Grp 32
Base Pattern	Pattern	Printers Grp
Base Diffusion	Dope	Furnace Grp 53
Emitter Pattern	Pattern	Printers Grp
Emitter Diffusion	Dope	Furnace Grp 60
Contact Pattern	Pattern	Printers Grp
Aluminum Evaporation	Layer	Alum Evap Grp
Metal Pattern	Pattern	Printers Grp
Deposit Overcoat	Layer	Depos Grp 6
Bonding Pad Pattern	Pattern	Printers Grp
Electrical Test	Test	Tester Grp

Figure 1: All of the pattern steps are performed in the photolithography area of the fab on the same equipment, Printers Grp. Wafers fabricated using the process flow illustrated make seven visits to Printers Grp.

(FIFO) is an example of a simple dispatch policy.

However, for the current application, dispatch policies have the following drawbacks:

1. Many dispatch policies do not work well on revisited machine groups like Printers Grp.
2. The typical dispatch policy is myopic in the sense that it considers only the local situation and does not consider the needs of downstream machine groups.

A dispatch policy such as Shortest Remaining Process Time (SRPT) will lead to problems when applied to a revisited workstation like the Printers Grp. If the queue for Printers Grp is made up of a number of lots with the process flow shown in Figure 1, then SRPT will prefer lots that are at their last pattern step, Bonding Pad Pattern. It will only select lots at the first pattern step, DUF Pattern, if there are no other lots in the queue. This leads to long waiting times at DUF Pattern and alternating starve/glut feeding patterns for DUF Diffusion.

We have investigated other dispatch policies such as Shortest Processing Time and Slack to Due-Date, but our experience has been that they also share the defect of SRPT of being biased in favor of one or another of the pattern steps and neglecting others. The problem seems to be that these policies are "winner take all" policies as opposed to "winner take a bigger share" policies. There is nothing wrong with adding priority to lots near the end of their flows or lots in trouble with their due dates. What causes problems is that "winner take all" schemes based on these factors run the risk that low priority lots may wait forever if higher priority lots arrive fast enough. To get around this problem, the Photo Dispatcher uses a variation of a round-robin scheme which provides a "winner take a bigger share" selection.

The FIFO dispatch policy is not biased in favor of particular pattern steps, but it is myopic in the sense that it will select lots that will just have to sit at their next process step because a key machine is down. Alternatively, it may pass over lots that would help keep a downstream bottleneck machine group from starving. Goldratt's OPT system (1984, 1988) emphasised the importance of bottleneck resources in scheduling. AI scheduling systems that emphasize the importance of bottleneck resources include Smith, Fox, & Ow (1986) and Eskey & Zweben (1990). The Photo Dispatcher avoids myopic decision making by monitoring current and historical workloads throughout the wafer fab and reacting to these workloads to avoid starving bottleneck workstations and avoid sending lots to workstations where they will just sit in queue.

Approach

The Photo Dispatcher makes recommendations about which lot of wafers in the queue should be processed next by Printers Grp. The Photo Dispatcher develops these recommendations in three stages:

1. Establish priorities for processing lots at the different pattern steps.
2. Use the priorities to choose a pattern step to work on next. That is, decide whether to work on a lot that is waiting for DUF Pattern, or Isolation Pattern, etc.
3. Choose a specific lot waiting for that pattern step. There are typically several lots waiting for DUF Pattern and this third stage determines which of these lots should be recommended. While a number of different criteria have been investigated for making this lot selection, none of them have outperformed FIFO, so currently this selection is just based on time of arrival of the lots waiting for DUF Pattern.

Pattern-Step Priorities

Figure 2 gives an example of the calculation of priorities for four pattern steps. Three numbers are summed to determine priorities. The percentage of lots waiting for a particular pattern step (e.g., 20 percent for DUF Pattern) is added to a number that is based on the nearness of that pattern step to the end of the processing flow (e.g., 01, the first two digits of the step id number). Finally, a positive number (e.g., 30) is added if more work is needed at downstream work areas or a negative number is added if there is too much work. The higher the priority number for a pattern step the more lots at this step will be recommended for pro-

```
 0100 DUF PATTERN
% Lots Queued    20
Flow Position    01
Feedback         30   (Send More;
                 ----  Short Wait at
Priority         51    Furnace Grp 20)

 0600 ISOLATION PATTERN
% Lots Queued    04
Flow Position    06
Feedback        -30   (Send Less;
                 ----  Long Wait at
Priority        -20    Furnace Grp 32)

 2300 BASE PATTERN
% Lots Queued    12
Flow Position    23
Feedback         00   (No Adjustment;
                 ----  Average Wait at
Priority         35    Furnace Grp 53)


 6000 CONTACT PATTERN

% Lots Queued    05
Flow Position    60
Feedback         60   (Send Much More;
                 ----  Starving Bottleneck
Priority        125    Alum Evap Grp)
```

Figure 2: Three factors determine the priority of a particular pattern step: the fraction of lots waiting for this pattern step, the nearness of this pattern step to the end of the process flow, and feedback from downstream work areas.

cessing. The rationale for each of the three factors illustrated in Figure 2 will be discussed in turn.

% Lots Queued If the process flow in Figure 1 was used by all devices, there would be roughly equal numbers of lots waiting for the individual pattern steps. However, since there are roughly 300 different process flows, the pattern steps do not occur with equal frequency. Including a factor for the percentage of lots waiting for a particular pattern step ensures that the round-robin scheme does not penalise lots that are waiting for frequently used pattern steps.

Flow Position The second factor, nearness of a pattern step to the end of the process flow, has the effect of reducing work-in-process (WIP). Lou and Kager (1989) recommend that when scheduling revisited workstations in IC fabrication, higher priority should be given to process steps that are later in the process flow. Our experiments confirm the benefits of this practice.

Feedback From Downstream Workstations The feedback to Contact Pattern in Figure 2 shows a bottleneck, Alum Evap Grp, requesting additional work because it is starving. Figure 3 illustrates the computations performed to determine this machine group is a starving bottleneck. A software object called a *Work Monitor* is associated with machine groups in the fab. WORK-FOR-ALUM-EVAP knows how to use the current position of a lot provided by the WIP tracking system to determine whether that lot is at a process step that uses the Aluminum Evaporators. In the case illustrated in Figure 3, there is one lot consisting of 48 wafers arrived at a processing step using the Aluminum Evaporators. WORK-FOR-ALUM-EVAP also can tell if a lot has left photolithography and is on the way to a process step using the Aluminum Evaporators but has not arrived yet. Identifying lots that have been sent to the Alum Evap Grp provides an estimate of upcoming workloads.

WORK-FOR-ALUM-EVAP knows how to translate the number of wafers arrived into the time it should take the Alum Evap Grp to complete processing those wafers. In Figure 3 the 48 wafers arrived are estimated to keep the Aluminum Evaporators busy for the next .58 hrs. The following factors are considered to determine how many hours it will take to complete processing a particular set of wafers:

1. number of machines in Alum Evap Grp and their capacities
2. process time for each wafer; different devices may have different processing times
3. setup times

Work Monitors categorise workloads along four different dimensions (TYPICAL-LOAD, COMPARE-NOW-TO-HISTORY, etc.) and these dimensions are referenced by rules that determine the appropriate

```
WORK-FOR-ALUM-EVAP
Lots sent to Alum Evap Grp: 0 lots, 0 wafers

Lots arrived at Alum Evap Grp:
1 lot, 48 wafers
Est work in queue for Alum Evap Grp:
14 wafers, 0.26 hrs
Est work in process for Alum Evap Grp:
34 wafers, 0.32 hrs
Estimated work available = .58 hrs

The avg amount of work, 10.43 hrs, that
 arrives in this area is HIGH relative to the
 long-run avg (4.44 hrs, sigma=3.19)
 for all areas.
Work currently arrived in this area, .58 hrs,
 is LOW relative to the long-run avg
 (10.43 hrs, sigma=6.28) for this area.
Work currently arrived in this area, .58 hrs,
 is LOW relative to the avg currently arrived
 (3.51 hrs, sigma=3.66) for all areas.
Work currently sent to this area, 0.00 hrs,
 is LOW relative to the long-run avg
 (2.70 hrs, sigma=0.64) for this area.

The classification is:
TYPICAL-LOAD              HIGH
COMPARE-NOW-TO-HISTORY    LOW
COMPARE-TO-OTHER-MACHS    LOW
COMPARE-SENT-TO-HISTORY   LOW,
                so MUCH MORE work is needed.
```

Figure 3: An example of a starving bottleneck. The Aluminum Evaporators typically have 10.43 hrs of work; however, at the time this workload evaluation was performed there was only .58 hrs of work and none on the way. A request for much more work is fed back to Contact Pattern.

feedback (e.g., send MUCH MORE). By changing the feedback rules, a wide variety of workload regulation schemes can be implemented. Simulation experiments were run to evaluate three different workload regulation schemes: Bottleneck Starvation Avoidance, Smallest Next Queue, and Workload Smoothing. Each of these three workload regulation schemes performed well at some WIP levels, but none of the three was superior at all WIP levels. A hybrid of these approaches was devised which performed well at all WIP levels investigated.

Some approaches to scheduling using bottleneck starvation avoidance require that the there is only one bottleneck and its identity is known in advance and provided as input to the scheduler (e.g., Glassey & Resende 1988). This assumption makes it difficult to handle shifting bottlenecks which can arise in wafer fabs because of a change in product mix or the breakdown of key equipment. The work-monitoring mechanism of the Photo Dispatcher determines which workstations are bottlenecks by gathering workload histories. As the bottlenecks change over time the shift in workload histories is tracked by the Work Monitors and the priority of lots is changed accordingly.

Using Priorities to Select a Pattern Step

Priorities are recomputed periodically based on the current workloads at downstream machine groups and the number of lots queued for each pattern step. The priorities influence subsequent selection of pattern steps by controlling a round-robin scheme.

Each time a machine in Printers Grp is ready to be loaded, the round-robin advances to the next pattern step in the cycle. The priority assigned to this pattern step determines whether a lot is chosen from this pattern step or whether the round-robin immediately advances to the next pattern step. Priority numbers are rescaled to range from 5 to 100 and pattern steps with priority 5 have a lot selected one out of every 20 times the round-robin stops there (5%), pattern steps with priority 50 have a lot selected every other time the round-robin stops there, etc.

When the priorities are as shown in Figure 2, a sequence of selections produced by this round-robin might be Contact Pattern, DUF Pattern, Base Pattern, Contact Pattern, Contact Pattern, DUF Pattern, Contact Pattern, ... By interleaving pattern steps in this fashion there is limit on how many times a low priority pattern step can be passed over before it is selected.

Results

The Photo Dispatcher was tested using a simulation of one of Texas Instruments' wafer fabs. The fab in question manufactures a wide variety of Bipolar and BiMOS devices. The simulation modeled all of the 103 machine groups (410 machines) in the fab and all of the process steps needed to fabricate any of the roughly

INITIAL WIP	FIFO		Photo Dispatcher	
	CT	Output	CT	Output
290	339	140	328	140
400	482	132	427	136
675	680	138	604	141
1000	1068	132	975	136

Table 1: In simulation experiments, the Photo Dispatcher produces an improvement over a FIFO policy in the average number of hours needed to complete processing a lot of wafers (i.e., cycle time or CT). The same lot starts were used in all simulations, so WIP levels were manipulated by varying the number of initial WIP lots in the fab at the start of simulation.

300 different ICs produced. There are roughly 50 process steps in the recipe for a typical IC in this fab and these process steps are broken down in the simulation to 165 separate operations each of which requires the use of another machine. The simulation included random machine breakdowns, but this was the only random component as processing times were assumed deterministic, lot transport time was not modelled, and there were no operator limitations.

Table 1 shows the Photo Dispatcher has better average cycle-time performance in simulations than a FIFO policy. FIFO does not separate the decision of which lot to process next on Printers Grp into a pattern step selection followed by a lot selection. Instead it merely finds the lot that has been waiting longest for the printers and starts that lot. This is the default behavior of the simulator and it is also used when exercising the Photo Dispatcher for machine groups other than the Printers Grp.

The size of the cycle-time improvement depends on the amount of WIP in the simulated fab. At WIP levels above 400 lots, the Photo Dispatcher reduced cycle-times by roughly 10% with greater output in terms of finished lots per week. Since the fab being simulated currently operates at WIP levels in excess of 400 lots, a 10% cycle-time improvement was projected from the use of this scheduler. Results to date suggest that flow position is the most important factor in producing cycle-time improvements of the three factors combined in Figure 2.

Summary

The Photo Dispatcher provides an effective scheduling approach for revisited workstations such as photolithography in IC fabrication. It takes advantage of the unique opportunity that these revisited workstations provide to shift workloads from one downstream area to another and to reduce WIP by speeding up processing on lots near the end of their process flows.

While the Photo Disptacher was developed with the intention of installing it in Texas Instruments' wafer fabs, to date it has not been used for real-time scheduling of fab operations. However, its Work Monitor capability has been used to a limited degree to analyse production problems in fabs.

References

Eskey, M. and Zweben, M. Learning search control for constraint-based scheduling. *Proceedings Eighth National Conference on Artificial Intelligence (AAAI-90)*, 1990, pp. 908–915.

Glassey, C.R. and Resende, M.G. Closed-loop job release control for VLSI circuit manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, Vol. 1, 1988,

Goldratt, E. and Cox, J. *The Goal.* North River Press, 1984.

Goldratt, E.M. Computerised shop floor scheduling *International Journal of Production Research.* Vol. 26, 1988, pp. 443–455.

Lou, S.X.C. and Kager, P.W. A robust production control policy for VLSI wafer fabrication. *IEEE Transactions on Semiconductor Manufacturing*, Vol. 2, No. 4, 1989.

Panwalker, S.S. and Iskander, W. A survey of scheduling rules. *Operations Research*, Vol. 25, 1977, pp. 45–61.

Smith, S.F., Fox, M.S., and Ow, P.S. Constructing and maintaining detailed production plans: Investigations into the development of knowledge-based factory scheduling systems. *AI Magazine*, Fall 1986, pp. 45–61.