Lecture Notes 15: Wafer Fab Modeling

Prof. Marvin Nakayama
CS 661: Simulation

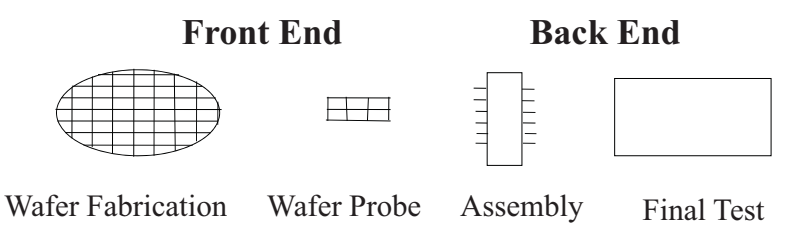
Road Map

- ***What did we learn so far?*** Process-based Simulation Modeling; Process Design and Resource Pooling
- ***What next?*** Manufacturing Systems and Scheduling Decisions
  - Description of the Wafer-Fabrication Process
  - Main Managerial Challenges
  - Performance Measures
  - Simulation Model
  - Some Key Insights
  - A Unified View of Job-Shop Scheduling
  - Conclusions
- ***Objectives:***
  - Simulation of high-tech production process
  - The role of scheduling in production processes

The Big Picture

- ***Semiconductor Manufacturing:***
  - Development of VLSI technologies is one of the most significant technological development of last twenty years.
  - Led to development of integrated circuits and computer revolution.
  - The fabrication of these electronic components involves a very complex process.
  - Effective management and control of this process can provide a competitive advantage
    - Crucial in the presence of heavy competition from abroad (Pacific Rim).
- ***Objective:*** To understand the basics of the semiconductor manufacturing process and develop simulation-based methods for improving the management of this process.

Production Process



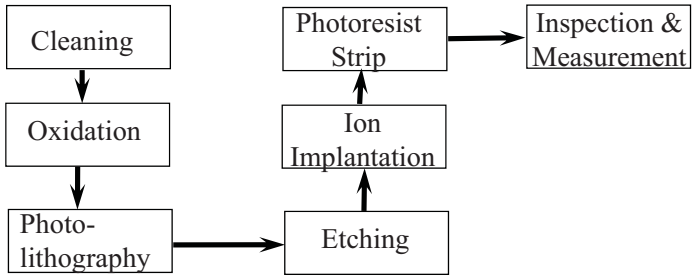
- ***Wafer Fabrication:*** Produce required circuits in layers
- ***Wafer Probe:*** Test and separate individual circuits
- ***Assembly:*** Place circuits in chips.
- ***Final Testing and Downgrading***

The Wafer Fab

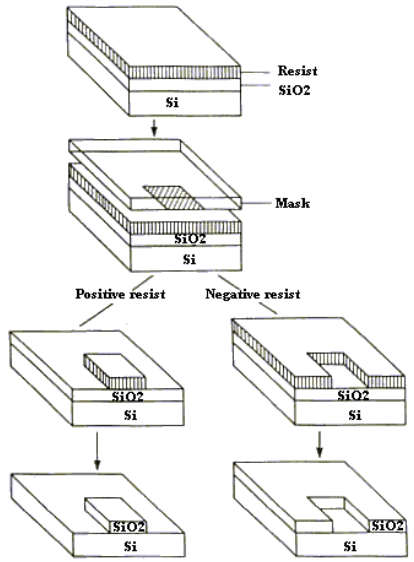
Overview:

- Circuits are created on silicon wafers through a process of photolithography and ion implantation.
- The process proceeds in layers, and when all the layers are completed, the wafer is cut into individual chips (all the chips are identical).

Typical Process Flow:



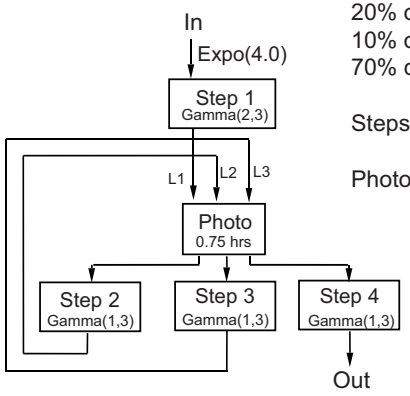
Photolithography



Major Managerial Challenges

- ***Challenge I:*** High Turnaround Times
  - 90% of the turnaround time is spent in queue.
  - The high turnaround times increase the opportunities for contamination and thus lead to low yields.
  - Important to maintain as low a turnaround time as possible.
- ***Challenge II:*** Highly Variable Turnaround Times
  - Turnaround times tend to be highly variable.
  - This makes it is difficult to quote reliable lead times to customers.
  - Important to maintain low variance in turnaround time.
- ***Challenge III:*** Due Dates
  - Meeting the due-date deadline is also extremely important.

A Stylized Wafer Fab

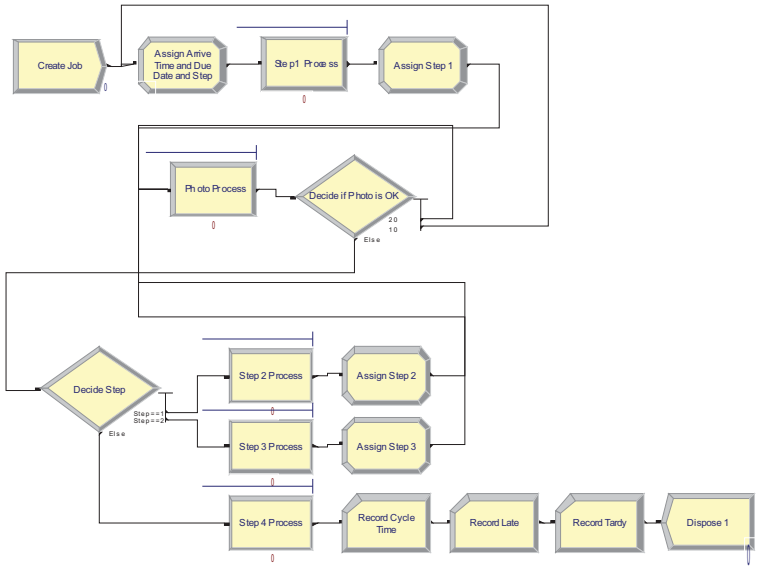


After Photo:
20% of jobs are reworked (sent to photo)
10% of jobs are scrapped (sent to Step 1)
70% continue

Steps 1-4 have 3 resource units each.

Photo has 1 resource unit.

ARENA Model



Process Management

- ***Scheduling Decisions:*** Which lot to work on next?
  - Wafer fabs operate in a rapidly changing environment
  - Scheduling decisions tend to be made in a chaotic and haphazard manner.
  - But these decisions can have a dramatic impact on the performance of the fab.

- ***Simple Scheduling Rule:***
  - First Come, First Processed
  - First Buffer First
  - Last Buffer First
  - Closest Due Date First

    What is the rationale behind these rules?

Scheduling Rules for Photo Server

- Click on `Queue` module in Basic Process Template
- In spreadsheet, look at row `Photo Process.Queue` and change `Type` to:
  - *First Buffer First*
    - Type: Lowest Attribute Value
    - Attribute: Step
  - *Last Buffer First*
    - Type: Highest Attribute Value
    - Attribute: Step
  - *Due Date*: assign higher priorities to jobs with closest deadlines
    - Type: Lowest Attribute Value
    - Attribute: Due Date

Simulation Results (with 95% C.I.)

	FIFO	***First Buffer First***	***Last Buffer First***	***Due Date***
Avg Cycle Time	45.6 (41.4, 49.8)	74.3 (61.4, 86.2)	36.3 (34.4, 38.2)	38.6 (36.3, 40.9)
Average Lateness	1.1 (0.3, 1.9)	12.8 (6.0, 19.6)	0.2 (0.1, 0.3)	0.0 (0.0, 0.1)
Percent Late	3.7 (2.0, 5.4)	17.2 (11.5, 22.9)	1.1 (0.7, 1.5)	0.0 (0.0, 0.0)

- What observations can you make here?
- Which policy is best?
- Are there any policies that appear to be uniformly bad?
- **Brain Teaser:** Does the scheduling rule affect capacity utilization?

Capacity Utilizations

	FIFO	*First Buffer First*	*Last Buffer First*	*Due Date*
Step 1	0.745	0.747	0.740	0.746
Step 2	0.325	0.327	0.326	0.326
Step 3	0.285	0.286	0.286	0.285
Step 4	0.249	0.251	0.250	0.250
Photo	0.923	0.926	0.922	0.923

Observations:

- Photolithography has highest utilization (bottleneck).
- Step 1 has second highest utilization
- Step 4 lowest utilization.
- *Does this make sense?*

Unified View of Job-Shop Scheduling

Typical Scenario:

- Potential customers arrive with a request for work.
- Shop Management and customer negotiate price and delivery time.
- Production manager decides
  - Sequencing of Jobs
  - Releasing of Jobs

- ***Problem:***
  - *How should one release and sequence jobs in order to minimize throughput times and delays?*
- ***Solution:***
  - Regulating the amount of work on the shop floor for bottlenecks can substantially reduce the throughput time (and WIP inventory).

Example 1: Deterministic Release Policy

Policy: Release jobs in the system deterministically (every 4.0 min.)

	FIFO	*First Buffer First*	*Last Buffer First*	*Due Date*
Avg Cycle Time	27.1 (26.0, 28.2)	31.7 (30.9, 32.5)	25.1 (24.8, 25.4)	25.9 (25.5, 26.3)
Average Lateness	0.0 (0.0, 0.0)	0.2 (0.1, 0.3)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Percent Late	0.0 (0.0, 0.0)	0.6 (0.4, 0.8)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)

Observations:

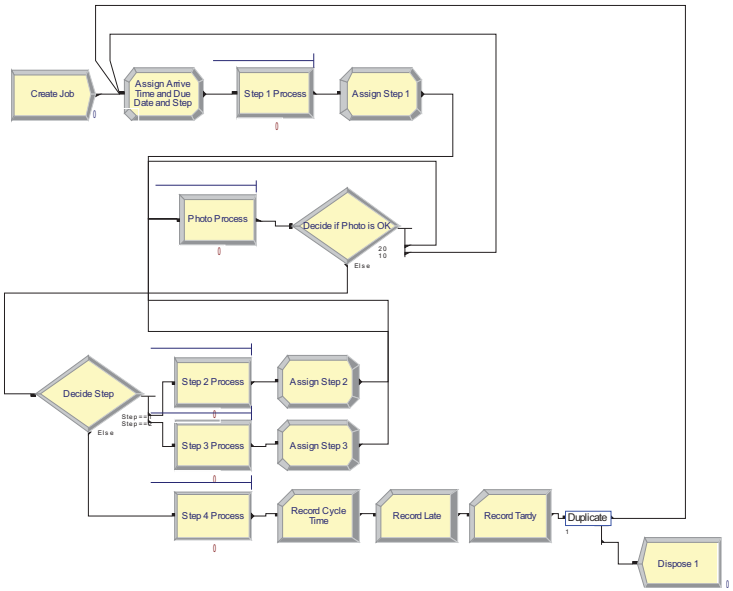
- Deterministic release policy reduces average cycle time even more.
- Slight reduction in Cycle Time compared to Random Release Policy (for Last Buffer First's Cycle Time)

Example 2: Closed-Loop Release Policy

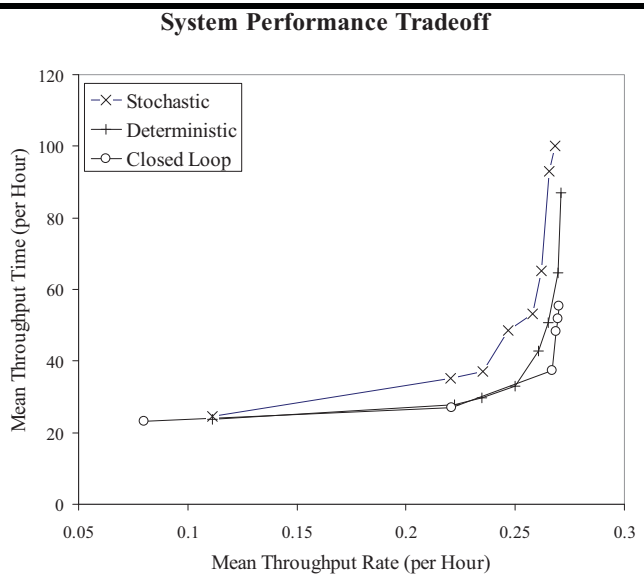
- Maintain a constant number of lots (8) in system.
- Release one lot whenever another lot exits.
- Number of lots chosen so that average throughput rate with FIFO is same as that for Expo(4.0).

	FIFO	*First Buffer First*	*Last Buffer First*	*Due Date*
Avg Cycle Time	22.5 (22.4, 22.6)	23.0 (22.9, 23.1)	22.1 (22.0, 22.2)	22.4 (22.3, 22.5)
Average Lateness	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Percent Late	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)

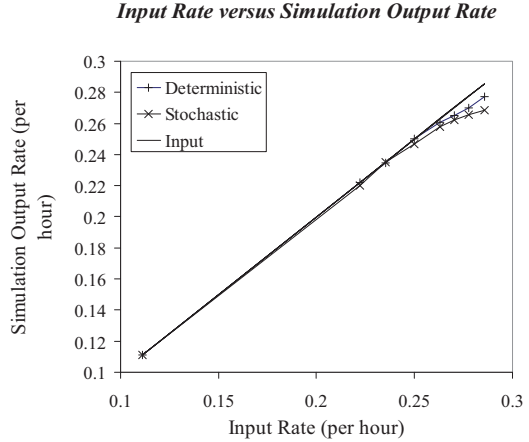
Closed-Loop Release Policy



System Trade-Off (Last Buffer First)



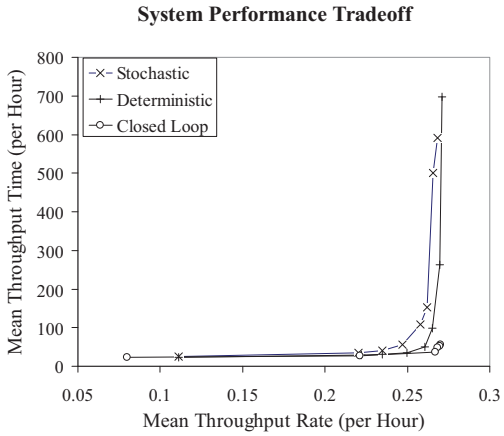
An Important Caveat



Paradox: When the input rate exceeds 0.24 items per hour, the output rate is less than the input rate. Is this a simulation-related fluke?

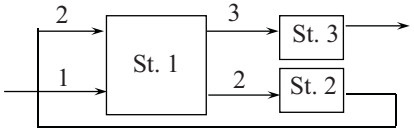
An Important Caveat and Warning

- ***Warning:*** When the processing network is heavily loaded, one needs very long simulation runs to get reliable simulation estimates.



Example 3: Workload-Regulating Policy

- State-of-the-art wafer-release policy.
- Best-practice in semiconductor manufacturing industry.
- Briefly, this policy works as follows:
  - Focus on bottlenecks (a wafer fab can have more than one bottleneck).
  - When sum of expected amount of work in bottlenecks falls below a threshold, release an additional wafer.
  - To compute expected amount of work in each station, we must take into account future visits to the station.
- Example: *Suppose each station has 1 server and all stations are currently busy. What is the expected amount of work for station 1?*



Conclusion

- Wafer Fabs are a paradise for simulation modeling!
- They represent complex systems where wrong decisions can be extremely expensive (in terms of lead times and productivity).
- General Guidelines:
  - Schedule wafers based on a Shortest-Remaining-Processing-Time discipline – e.g., Last Buffer First.
  - Try to push work out of the system as much as possible.
  - Try to release wafers into the system according to a deterministic policy.
  - Workload-regulating policies represent "best-practice" in semiconductor industry. They can reduce turnaround times by 10% compared to the simple deterministic release policy.