

Scheduling Dedicated Lithography Equipment

Dr. Robert C. Leachman

Dept. of Indus Eng & Oper Res
U C Berkeley

Dr. Vincent Lin

Leachman & Associates LLC

Dr. Paolo Palezzato

Cypress Semiconductor, Inc.

What's happening in semiconductors

- **Moore's Law:** the number of transistors quadruples every four years
- **Rock's Law:** the cost of a fab doubles every four years
- **Result 1:** equipment pushed to technological limits, more process instability, more scheduling complexity
- **Result 2:** fewer advanced fabs and fewer fab operators, fabless/foundry partnerships
- **Result 3:** fabless firms do naïve planning, foundry must do real planning and scheduling and do it fast

Semiconductor supply chain

(Develop process technology – 1 year)

- **Wafer fabrication – 6 weeks**

- Test chips on wafer – 1 week

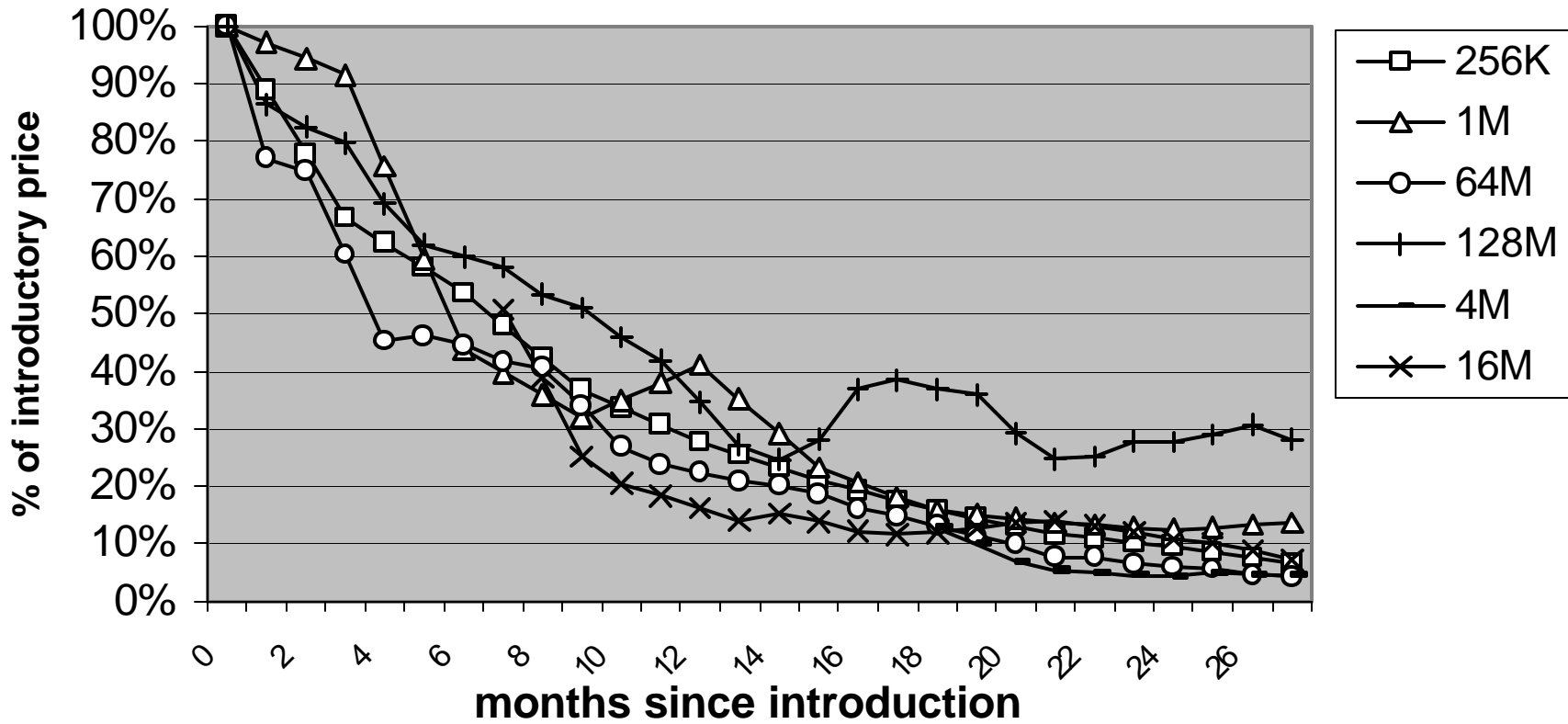
(Ship overseas – 0.5 week)

- Packaging and final test – 1 week

(Ship overseas – 0.5 week)

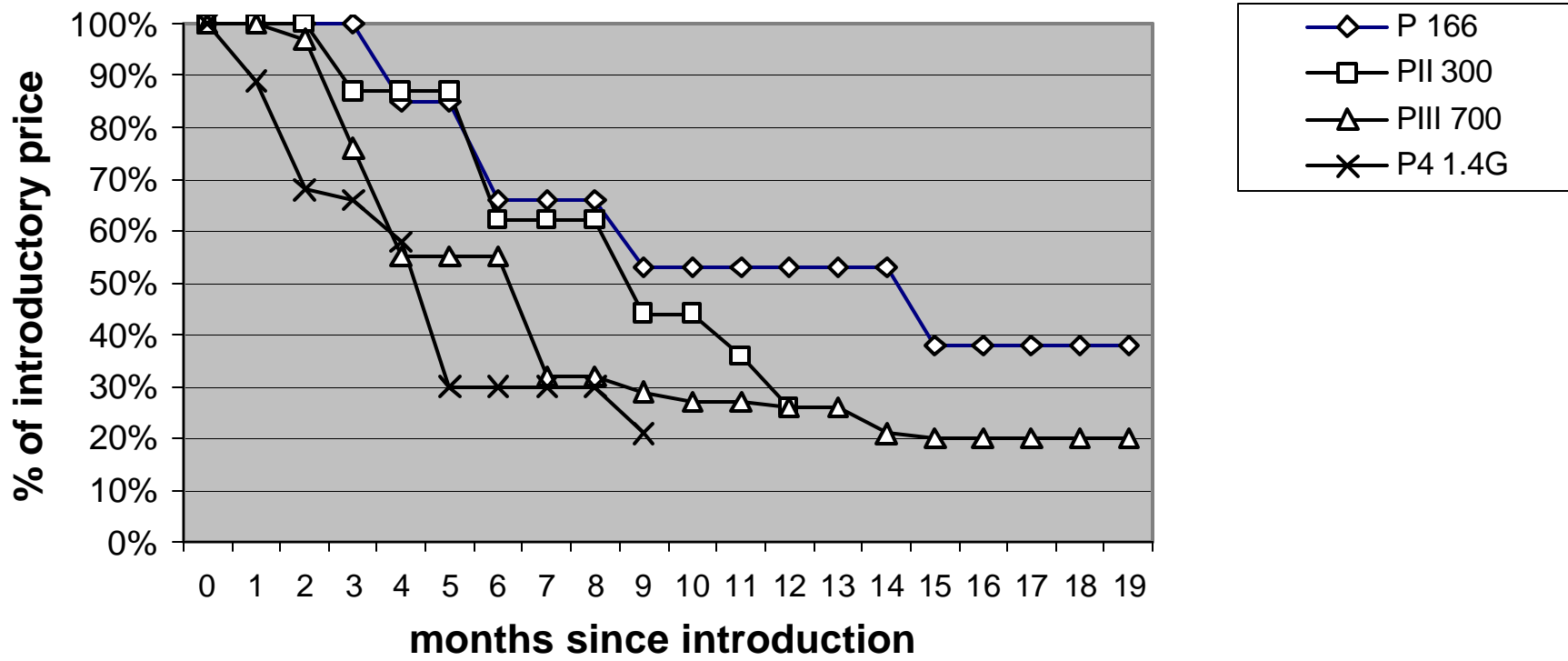
“Time is money”

DRAM price trends

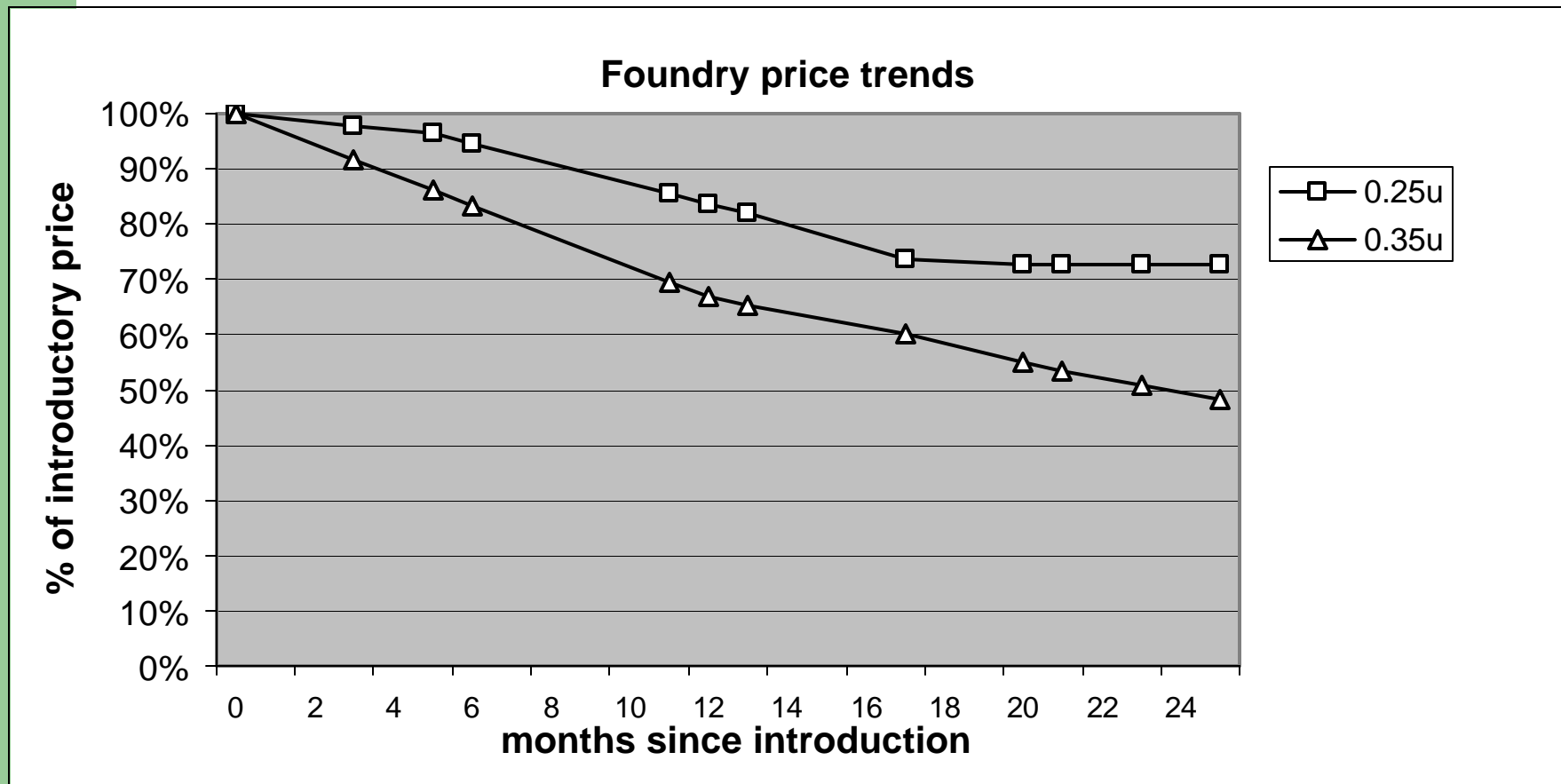


“Time is money”

Intel's price trends



“Time is money”



The economic value of *cycle time* in the supply chain

- Let $P_i(t)$ denote the sales price of device i at time t . $P_i(t) = P_{0i} \exp(-\alpha t)$, where P_{0i} is the current price.
- Let $W_i(t)$ denote the wafer starts, $Y_i(t)$ the yield, $CT_i(t)$ the manufacturing cycle time, and H_i the remaining product life.
- Then the remaining lifetime revenue is

$$\int_0^{H_i} P_{0i} e^{-\alpha[t+CT_i(t)]} W_i(t) Y_i(t) dt$$

Lithography

- **Lithography represents the greatest capital expense in an advanced fab**
- **Strategies to stretch machine capabilities:**
 - each manufacturing lot must be processed using the same lens (machine) at 3-10 “critical layers”
 - at other layers, the lot may only be processed by one or several machines that have been “matched” to the critical machine
- **Lithography accounts for ~25-40% of the wait-time portion of fab cycle time**
 - e.g., 20 mask layers, 6-8 hours avg. wait time per layer

Lithography

- Example small litho dept: 3 scanners (S1, S2, S3), 3 DUV steppers (D1, D2, D3), 5 I Line steppers (I1, I2, I3, I4, I5)
- Example alternative photo routes for a high-volume device:

Layers	Route 1	Route 2	Route 3
1 and 3	I1	I3	I4
4, 5 and 7	I1 or I2	I3 or I4	I4 or I5
2, 6, 8-12 and 13	S1	S2	S3
14, 15 and 16	D1	D2	S2
17	I1 – I5	I1 – I5	I1 – I5

Lithography

- Despite the matching restrictions, the photo process occasionally fails and requires engineering effort to restore stability
- As a risk-mitigation strategy, it is desirable to spread the production of a given device across several photo routes

Lithography capacity

- Capacity for new starts depends on the device mix
- Capacity for new starts each week depends on the WIP status
- Photo wait times are sensitive to the allocation of the planned weekly starts to photo routes and to days of week

Rest of fab

- In contrast, capacity of the other process areas in the the fab is expressable by technology (i.e., all lots of all devices in same technology have same set of qualified machines, same process times, etc.)
- To control wait times in the rest of the fab, linear production rates by technology are desired

Planning and scheduling architecture – fabless firm

- Use commercial LP-based tools such as i2, Adexa, MIMI, Avere, etc.
- Plan *starts requests* - Obtain estimates of cycle times and yields from foundry periodically, obtain WIP snapshot on-line
- Incorporate *start commits* from foundry as if it is additional WIP
- Quote delivery dates based on WIP-out projections

Sophisticated planning and scheduling architecture - foundry

- LP-based planning system to analyze weekly wafer start requests and/or die out requests
 - Dynamic capacity analysis of workloads from WIP and new production
 - Rigorously analyze lithography and other capacities
- IP-based fab starts scheduling system to allocate the accepted starts to machines and to days of week

Fab starts scheduling problem

- **Given:** time-varying machine capacities, qualified photo routes, process times, total manufacturing lots to be started this week, WIP snapshot, estimated cycle times and yields
- **Variables:** what day of week and what photo route on which to start each manufacturing lot
- **Multiple objectives:** balance workloads of photo machines, linearize starts rates by technology, spread out device starts across photo routes

Integer programming model

Constraints:

- Workload on machine i on day $t =$ workload from WIP arriving on day t and workload from allocated starts arriving on day t
- Overload of $(i,t-1) +$ workload of $(i,t) \leq$ capacity $(i,t) +$ overload (i,t)
- Other accounting constraints and variables (e.g., define total overload by machine type)

Goal programming approach

First, we minimize 10 different objectives independently:

- Max overload of any photo machine on any day
- Total of daily overloads on scanners
- Total of daily overloads on DUV steppers
- Total of daily overloads on I Line steppers
- Max total workload on any scanner
- Max total workload on any DUV stepper
- Max total workload on any I Line stepper
- Max starts on any day of technology 1
- Max starts on any day of technology 2
- Sum over devices of max starts of device i on a single route

Goal programming approach

- Afterwards, we solve for an objective to minimize the weighted deviation from optimality of each of the 10 goals
 - Try to simultaneously minimize overloads, equalize workloads, linearize starts by technology, and minimize risk of too many lots of same device assigned to same route

Implementation and results

- Fab switched from manual scheduling of fab starts to the IP-based fab starts scheduler
- Application is fully automated, takes about 40 mins to run using CPLEX (5K constraints, 1.5K integer vars, 11 IPs)
 - IP's are solved to first integer solution within 1-2% of optimality
- Six weeks later, **the average photo wait time per lot at the fab had declined by 1.5 days** for both major technologies in production, while fab start volume was slightly increased

What is 1.5 days reduction worth?

- For the case-study fab's current technology and product portfolio,

$$\sum_i \left(\int_0^{H_i} P_{0i} e^{-a[t+CT_i-1.5]} W_i Y_i dt - \int_0^{H_i} P_{0i} e^{-a[t+CT_i]} W_i Y_i dt \right) = \$515,000$$

- The revenue gains will be even larger for future technologies

Summary

- Industry trends have increased the value of cycle time compression in the supply chain
- Optimization in the semiconductor industry is now being employed at the supply chain level, at the foundry planning level, and, starting with this application, at the scheduling level