SCHEDULING & PLANNING in SEMICONDUCTOR MANUFACTURING

- Problem Classification
  - Performance Evaluation, simulation
    ⇒ understand system
  - Production Planning
    ⇒ long-term, aggregate planning — WEEKS/MONTHS
  - Shop-Floor Control
    ⇒ movement or dispatch of material from station-stations — DAYS/HOURS/MINUTES

- What makes the problem hard?
  1. Complex product flows
     - many steps
     - shared equipment or "reentrant product flows"
  2. Random Yields
     - lost material
     - binning
     - engineering time/hold time
  3. Diverse Equipment
     - batch vs. single wafer
     - setup times
     - time windows
  4. Equipment Downtime — UNCERTAINTY
  5. Production/Development Comixed
  6. Data e.g. 240,000 transactions/day

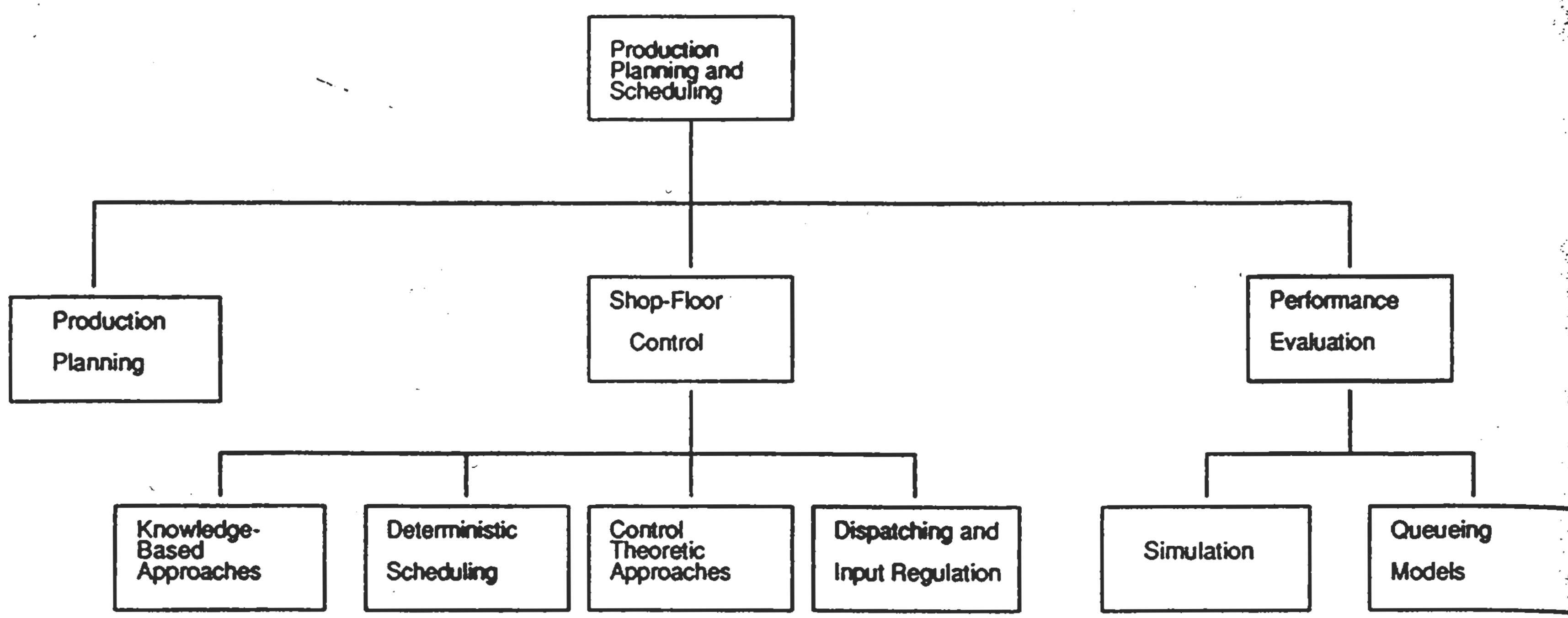


Figure 1. Classification of production planning and scheduling research

TYPICAL OBJECTIVES

- Minimize production costs (inc. inventory)
- Increase productivity
- Improve quality
- Improve delivery-time performance

⟹ HIGH THROUGHPUT & Equip. Utilization — USUAL FOCUS
LOW CYCLE TIME (Mean & Variance)

⇢ MEET DELIVERY SCHEDULES — esp. ASIC

— indirect attention: INVENTORY

PROBLEMS WITH INVENTORY

- costs \$ to create, generates no \$
- larger inventory ⟹ longer production time, more customer waiting
- more time ⟹ more vulnerable to damage & yield loss
- more time ⟹ longer interval between PROBLEM & DETECTION
- space, material-handling
- OBSOLESCENCE!

LITTLE'S LAW: $L = \lambda W$

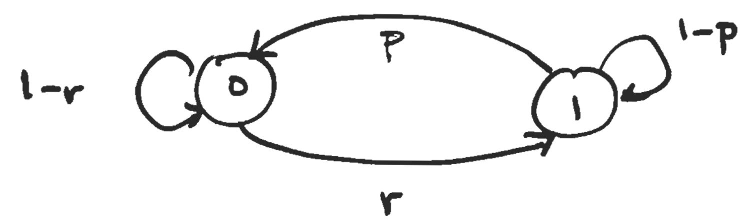
L: WIP

λ: arrival rate

W: cycle time (average time part is in process)

BASIC QUEUEING THEORY - Unreliable Machines

- Machine States: UP (1) or DOWN (0)



$p\,\delta t$: prob. of failure in δt

$r\,\delta t$: prob. of repair in δt

$\mu\,\delta t$: prob. oper COMPLETES while machine up

- Long-run production rate of machine? (Markov process)

$$p_r(0, t+\delta t) = p_r(0,t)(1 - r\,\delta t) + p_r(1,t)\,p\,\delta t + o(\delta t) \qquad (1)$$

Pr(still down) = Pr(was down, no repair) + Pr (was up, went down)

or

$$\frac{dp_r(0,t)}{dt} = -p_r(0,t)\,r + p_r(1,t)\,p \qquad (2)$$

and similarly

$$\frac{dp_r(1,t)}{dt} = p_r(0,t)\,r - p_r(1,t)\,p \qquad (3)$$

Solution: $$p_r(0,t) = \frac{p}{r+p} + \left[p_r(0,0) - \frac{p}{r+p}\right] e^{-(r+p)t}$$

$$p_r(1,t) = 1 - p_r(0,t)$$

as $t \to \infty$

$$Pr(0) = \frac{p}{p+r} \qquad Pr(1) = \frac{r}{p+r}$$

- Average Production Rate: $$P_r(1)\,\mu = \frac{r\,\mu}{p+r}$$

M/M/1 QUEUE

- Infinite storage
- Parts ARRIVE according to POISSON process

  $\Rightarrow$ interarrival times are exponentially distributed with arrival rate λ : $e^{-\lambda t}\lambda\delta t = p_r$(part arrives in $\delta t + t$)
- Service times are exponentially distributed

  service rate μ : $e^{-\mu t}\mu\delta t = Pr$(completes between t & $t+\delta t$)
- Probability of Parts in system

  n = # parts

  part arrives, part left

  $$Pr(n, t+\delta t) = p_r(n-1, t)\lambda\delta t + p_r(n+1, t)\mu\delta t + p_r(n,t)(1-(\lambda\delta t + \mu\delta t)) \quad \text{for } n>0$$

  part arrives & part leaves

  and I.C. / boundary

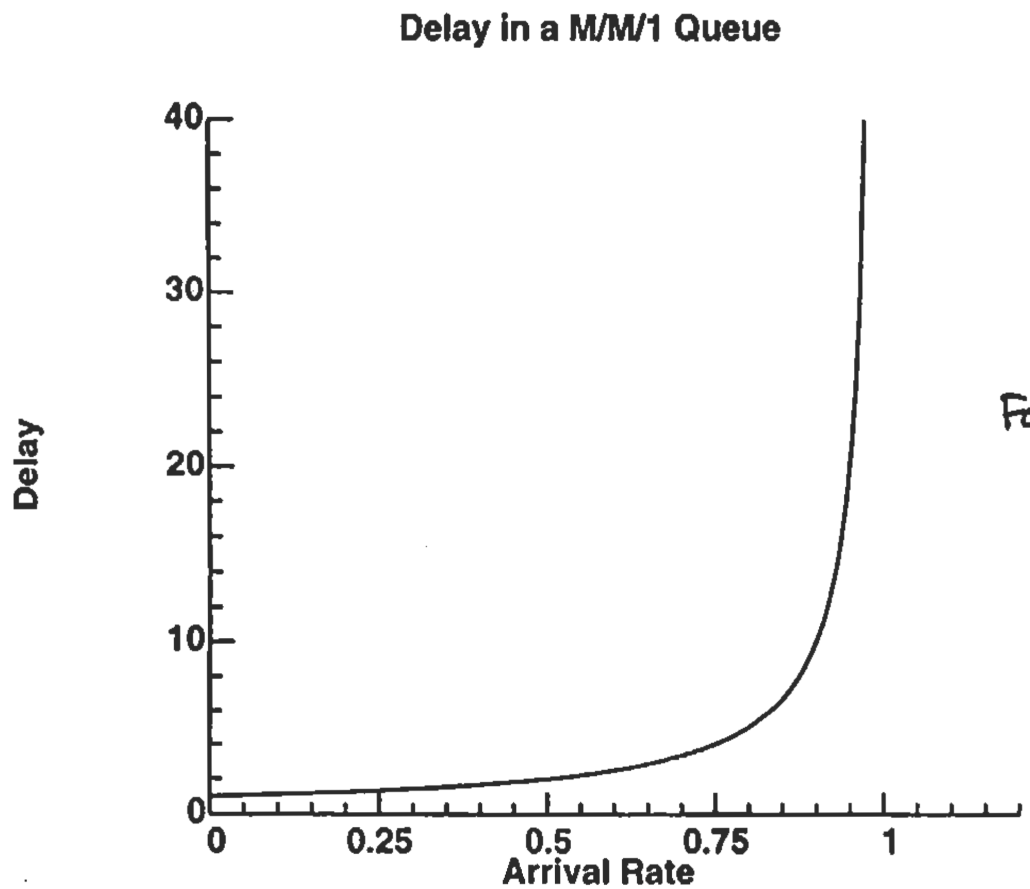
  $$Pr(0, t+\delta t) = p_r(1,t)\mu\delta t + p_r(0,t)(1-\lambda\delta t)$$
- Solution & Steady state <u>distribution</u>

  $$Pr(n) = (1-\rho)\rho^n, \quad n \geq 0 \qquad \text{if } \rho < 1$$

  $$\rho = \frac{\lambda}{\mu}$$

  $$\bar{n} = \text{avg. \# parts in system} = \sum_n Pr(n)\cdot n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$$

  Average Delay $\quad w = \dfrac{1}{\mu-\lambda}$



For $\mu = 1$

Delay *versus* Arrival Rate

- $\rho < 1 \Rightarrow \lambda < \mu$; Arrival rate > process rate
  $\Rightarrow$ UNSTABLE

  # parts $\sim (\lambda - \mu) t$

- CAPACITY of System is $\underline{\underline{\mu}}$
  $\Rightarrow$ greatest rate at which parts can enter & leave

- WIP (Work in Process) or inventory in <u>increases</u> <u>dramatically</u> as $\lambda \to \mu$

  $\Rightarrow$ true of <u>ALL</u> systems with waiting

YIELD & THROUGHPUT

- Assume: single server queueing system

  λ - arrival rate
  μ - service rate

  $W \triangleq$ cycle time ~ Exponentially distributed w/. mean $(\mu-\lambda)^{-1}$

- For yield, assume indep. poisson r.v. w/. mean d : e^{-d}

  Let $d = aW$ ~ defect rate a (in time)

- Combining: D = # die / wafer
  Y_W = wafer yield (# good die on wafer)

$$\bar{Y} = E(Y) = \int_{0}^{\infty} (\mu-\lambda) e^{-(\mu-\lambda)W} D e^{-aW} \, dW$$

$$= \frac{D(\mu-\lambda)}{\mu-\lambda+a}$$

- Thruput Rate T = $\dfrac{\text{mean good die} \div \text{time}}{\text{\# die / wafer}}$

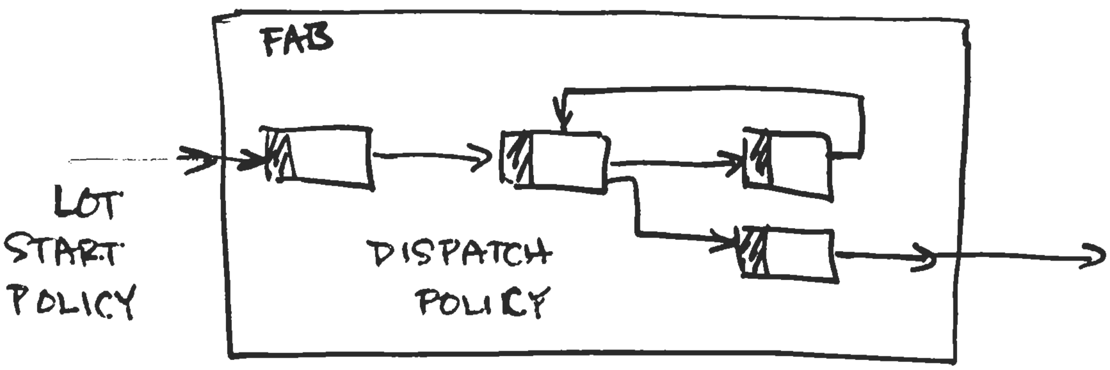
  So
$$T = \lambda \bar{Y} = \frac{(\mu - \bar{W}^{-1})\bar{W}^{-1}}{\bar{W}^{-1}+a}$$

- Solve for mean cycle time, and maximizing

$$\bar{W}^* = \frac{1}{\mu} + \sqrt{\frac{a+\mu}{a\mu^2}}$$

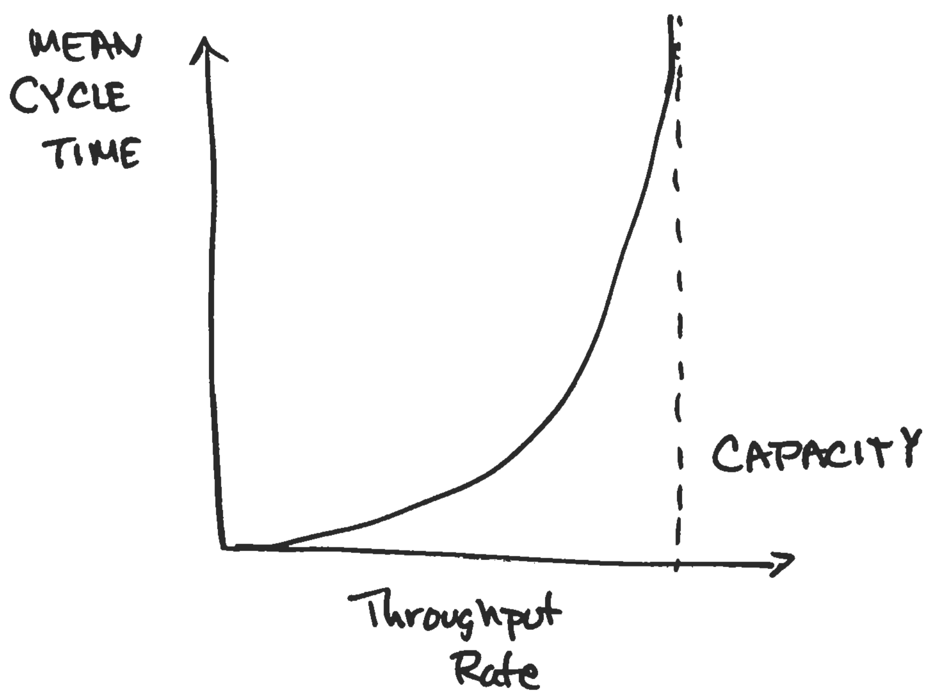
$\bar{T}^*$ = Max Capacity w/. start rate $\lambda^* = \dfrac{\mu + T^*}{2}$

JOB RELEASE & SCHEDULING



- QUESTION: HOW DO JOB RELEASE & DISPATCH POLICIES AFFECT
  - MEAN CYCLE TIME

  AS A FUNCTION OF THROUGHPUT RATE ?

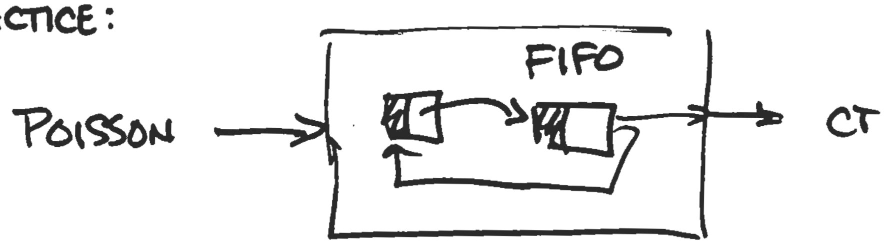


Wein, "Sched. Semi. Wafer Fab", TSM 1(3) 1988.

- Study CYCLE TIME for
  - INPUT CONTROL
  - SEQUENCING RULES
- Approach: SIMULATION (SIMAN)
- Proposal: New <u>WORKLOAD REGULATION</u> Policy

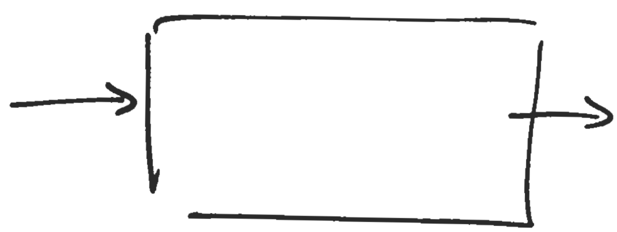
vs.

(1) COMMON PRACTICE:



(2) ALTERNATIVE (local knowledge only)

CONSTANT



+ reduced variability

(3) CLOSED-LOOP (uses loading info. from fab)

+ reduced mean & variability

RESULTS

① ACTUAL vs. THEORETICAL CYCLE TIME

- Poisson/FIFO:

Fab 1 :	1.9X
Fab 2 :	2.6X
Fab 3 :	3.8X

~ smaller than usual in industry.

② FAB 1 (1 true bottleneck – Station 14)

a. SPRT w/. Poisson ⇒ good improvement 13 – 16%
significant?

b. Best improvements by DIFFERENT INPUT POLICIES

⇒ 41.8%

c. WR input & any dispatch
much better.

NOTE: w/. WR, dispatch doesn't matter as much

③ FAB 2 (2 bottleneck case)

a. Under Poisson input ⇒ dispatch can matter a lot

b. Again, DET, CL, or WR ⇒ bigger impact

④ FAB 3 (Multiple bottlenecks)

a. Poisson ⇒ dispatch only marginal help
b. IRegulation does help.