Flow Time Estimation with Consideration of Production Target-induced Operation Variability for Semiconductor Fabrication

Yu-Ting Kao\textsuperscript{1}, and Shi-Chung Chang\textsuperscript{1,2}

\textsuperscript{1}Grad. Institute of Industrial Engineering,  
\textsuperscript{2}Dept. of Electrical Engineering,  
National Taiwan University

August 20, 2013
Outline

- Daily target setting (DTS) and wafer flow
- Flow estimation with target-induced variability considered
- Approximation of the penetration time between two stages (APT-2)
  - Probabilistic switching model
  - Inter-departure time analysis for \( R_i \)
  - 2-stage flow time estimation
  - Recursion in SOPEA
- Simulation studies
- Conclusions
Daily Target Setting (DTS)

- Semiconductor fabrication: reentrant process flows
- Daily target: # of wafer to finish processing (moves) at each stage for each product in a day

Chicken-egg relationship in re-entrant
- Machine allocation and WIP availability determine target setting
- Target setting guides machine allocation and then affects WIP flows
DTS Impacts on Fab Performance

- On foundry and memory fab operations
  - 20% increase in daily moves and
  - 8% decrease of wafers-in-process (WIPs) [Chang99]
  - Inter-planner decision variability reduction [Kao10]
Need for Better Wafer Flow Estimation

- Wafer flow estimation under given target setting and machine allocation as a basis of adjusting targets and machine allocation

- Current practices
  - Moving-average of cycle times + human judgments
    - Often deviates largely from the actual of a day
  - Mean value-based flow estimation [Chang99]
    - Over-optimistic, sometime up to 40% higher

→ Need to consider variability
Flow Estimation with Target-induced Variability

- Target setting and detailed machine allocation over time introduce variability

**Challenges**

Given targets, initial WIP distribution profile, and raw processing time of individual stages

- C1. How to model the variability caused by detailed machine allocation to track targets?
- C2. How to incorporate the operation variability into the wafer flow estimation?
Probabilistic Switching Model of Machine Allocation over Time

- Supervisor’s behavior of machine allocation
  - Allocate an available machine to stage $j$ with probability $p_j$
  - $p_j$ proportional to target $T_j$
  - Stage target workload divided by the total capacity

\[ p_j = \frac{T_j \cdot \tau_j}{C_m} \]
Inter-departure Time Analysis

Coupling of the inter-departure time relation

- $R_i$: inter-departure times between $i$-th lot and $(i-1)$-th stage-$j$ lots
- $T_i$: processing time of $i$-th lot of stage $j$
- $B_i$: number of allocations of non-stage $j$ before the machine is again allocated to stage-$j$
- $\Gamma_{Bi}$: total time needed before $i$-th lot of stage $j$
Properties of $R_i$

- **Probability density function**
  
  $f_{R_i}(r_i) = \begin{cases} 
  A \lambda_j e^{-\lambda_j r_i} - B \lambda(-j) p_j e^{-\lambda(-j) p_j r_i} & \text{if } j \text{ is stage-
  j} \\
  \Gamma_i & \text{if } j \text{ is non-stage-
  j} 
  \end{cases}$

  where $A = \frac{p_j (\tau_j - \tau_{(-j)})}{p_j \tau_j - \tau_{(-j)}}$ and $B = \frac{(1 - p_j) \cdot \tau_{(-j)}}{p_j \tau_j - \tau_{(-j)}}$

- When $p_j = 0 \implies R_i = \Gamma_{Bi}$
- When $p_j = 1 \implies R_i = T_i$

- **Expected value**
  
  $E[R_i] = \tau_j + \tau_{(-j)} \left( \frac{1 - p_j}{p_j} \right)$

  - Mean processing time of one stage-$j$ lot
  - Sum of processing times for non-stage-$j$ lots
    - number of machine allocations * mean processing time of one non-stage-$j$ lot
2-Stage Flow Time Estimation

\[ PT_{j(j+1)} = PT_{jj} + \sum_{c=1}^{W_j+W_{j+1}} PT_{(j+1)(j+1)|L_{PT_{jj}}=c} \cdot P(L_{PT_{jj}} = c \mid W_j, W_{j+1}) \]

- 2 stage cascaded queue
  - \( PT_{jj'} \): time needed for the last lot in WIP of stage \( j \) to finish processing at stage \( j' \)
  - \( L_{PT_{jj}} \): number of lots present in the stage \( j+1 \) at the time \( PT_{jj} \), including the \( W_j \)-th lot arriving at time \( PT_{jj} \)
Recursion in SOPEA

- Exploitation of the recursion to approximate multiple-stage flow times from the two-stage ones

For each stage $j$

$k$ from 1 to $K$

if $k = 1$ \( PT_{j(j+k)} = APT-2(j) \)

else \( PT_{j(j+k)} = SOPEA(PT_{j(j+k-1)}, PT_{(j+1)(j+k)}) \)

For each stage $j$
Service Variability Impact

- APT-2 falls between simulated 95% C.I.s

<table>
<thead>
<tr>
<th></th>
<th>Allocating probability</th>
<th>Initial WIP</th>
<th>Processing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage j</td>
<td>1</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Stage j+1</td>
<td>[0.1,1]</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>
Arrival Variability Impact

- APT-2 falls between simulated 95% C.I.s

<table>
<thead>
<tr>
<th></th>
<th>Allocating probability</th>
<th>Initial WIP</th>
<th>Processing time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage j</td>
<td>[0.1, 1]</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Stage j+1</td>
<td>1</td>
<td>1</td>
<td>50</td>
</tr>
</tbody>
</table>
Conclusions

- Probabilistic proportional-to-target machine allocation model to capture target-induced variability
- APT-2 captured variability to estimate the time needed for daily initial wafers at one stage to finish processing at the downstream stage
- Integration of APT-2 into SOPEA, where first order effect of variability captured in two-stage penetration times propagates through the recursive calculation of SOPEA to approximate fab-wide penetration times and flows
- Approximations falling into the 95% C.I. of simulated flow time under service and arrival variability
- 14% approximation error reduction over SOPEA
BACK-UP