Wafer Admission Control for Clustered Photolithography Tools

Kyungsu Park and James R. Morrison
KAIST
Dept. of Industrial and Systems Eng.
San Francisco, USA
July 12, 2010
Presentation Overview

- System Description: Multi Cluster Tools
- Motivation
- Methodology: Flow Line
- Wafer Admission Algorithms
- Performance Evaluation
- Concluding Remarks
System Description: Multi Cluster Tools

- Deterministic process times
- Process time can vary by product
- Scanner is the system bottleneck
- Setups between lots (reticle changes, module adjust setting)
Motivation

• For throughput in clustered photolithography tools
  – Pre-scan buffers minimize loss of scanner throughput
  – Opportunistic wafer admission

• Unnecessarily early wafer admission
  – High buffer level
  – Increased wafer residency time
  – Increased cycle time of lot
  – Loss of yield
  – Deployment capacity loss
Objective of Our Research

- Wafer admission algorithms that guarantee no loss of throughput while minimizing wafer residency time
- Find optimal admission time
Methodology: Flow Line

- Abstract to an easier problem → Flow Line
  - Common to abstract intractable problem (e.g. Production control)
  - Robots may be very fast in comparison to process times
  - Dedicated robots for each process/module
  - New equipment designs reduce the need for complex robot scheduling\(^1\)

- Flow lines can predict cycle time and throughput with high fidelity (99% throughput accuracy when not robot limited)\(^2\)

- Provide guidance for robot control algorithms

Methodology: Flow Line

- Deterministic service time for $P_i$
- One module can hold at most one wafer
- Buffers can be modeled by process module with zero process time
Wafer Admission Algorithms

**Step 1.** Calculate completion times of each wafer in opportunistic wafer entry

\[ X_1(i, 1) = \max \{a(i), X_{p+1}(i-1, W_{i-1})\} + \tau_{\text{setup}} \text{ if } c(i) \neq c(i-1), \]
\[ X_1(i, 1) = \max \{a(i), X_2(i, W_{i-1}-R_1+1)\} \text{ if } c(i) = c(i-1), \]
\[ X_m(i, w) = \max \{X_{m-1}(i, w) + \tau c_{m-1}, X_{m+1}(i, w-R_m)\} \]

for \( i \neq 1 \) and \( m \neq 1, \)

\[ X_{s+1}(i, 1) = \max \{X_s(i, 1) + \tau c_s + \tau_{\text{align}}, X_{s+2}(i, W_{i-1}-R_{s+1}+1)\}, \]
\[ X_M(i, w) = \max \{X_{M-1}(i, w) + \tau c_{M-1}, X_M(i, w-R_M) + \tau c_M\}, \]
\[ C_M(i, w) = X_M(i, w) + \tau c_M \]

**Step 2.** Compute latest release times for each wafer

\[
\text{Max } \sum_{w=1}^{W} X_1(w).
\]

Subject to

\[ X_1(i, 1) \geq \max \{a(i), X_{p+1}(i-1, W_{i-1})\} + \tau_{\text{setup}} \text{ if } c(i) \neq c(i-1), \]
\[ X_1(i, 1) \geq \max \{a(i), X_2(i, W_{i-1}-R_1+1)\} \text{ if } c(i) = c(i-1), \]
\[ X_m(i, w) = \max \{X_{m-1}(i, w) + \tau c_{m-1}, X_{m+1}(i, w-R_m)\} \]

for \( i \neq 1 \) and \( m \neq 1, \)

\[ X_{s+1}(i, 1) = \max \{X_s(i, 1) + \tau c_s + \tau_{\text{align}}, X_{s+2}(i, W_{i-1}-R_{s+1}+1)\}, \]
\[ X_M(i, w) = \max \{X_{M-1}(i, w) + \tau c_{M-1}, X_M(i, w-R_M) + \tau c_M\}, \]
\[ C_M(i, w) = C^*_M(i, w) \]
Wafer Admission Algorithms

• Notation for algorithm
  – \( X_i(w) \): Entry time of wafer \( w \) to module \( i \)
  – \( C_i(w) \): Completion time of wafer \( w \) from module \( i \)

• Key stages of algorithm
  **Step 1.** Calculate completion times of each wafer in opportunistic wafer entry \( C^*_M(w) \)
  **Step 2.** Compute admission times for each wafer
    – Subject to \( C_M(w) = C^*_M(w) \)
    – Objective function
      \[
      \text{Max } \sum_{W} X_1(w) \quad \Rightarrow \quad \text{Latest admission time}
      \]
      \[
      = \text{Minimum average residency time}
      \]
Performance Evaluation: System

- 3 lot types randomly (Average train size = 3)
- Lot size = 24
- Setup time: Uniform(120, 240)
- Reticle alignment: Uniform(240, 420)
- 10 different loading levels (JIT, 0.9, ..., 0.1)
- 1,800 lots x 10 replications

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>number of modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>92</td>
<td>70</td>
<td>52</td>
<td>2</td>
</tr>
<tr>
<td>P2</td>
<td>126</td>
<td>111</td>
<td>84</td>
<td>3</td>
</tr>
<tr>
<td>P3</td>
<td>33</td>
<td>48</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>P4</td>
<td>65</td>
<td>86</td>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td>P5</td>
<td>47</td>
<td>32</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Buffer</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>P6</td>
<td>100</td>
<td>75</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td>P7</td>
<td>29</td>
<td>16</td>
<td>41</td>
<td>1</td>
</tr>
<tr>
<td>P8</td>
<td>141</td>
<td>111</td>
<td>81</td>
<td>3</td>
</tr>
<tr>
<td>P9</td>
<td>50</td>
<td>70</td>
<td>60</td>
<td>2</td>
</tr>
<tr>
<td>P10</td>
<td>23</td>
<td>26</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>P11</td>
<td>96</td>
<td>87</td>
<td>69</td>
<td>3</td>
</tr>
</tbody>
</table>
Performance Evaluation: Cycle Time and Throughput

- Throughput = average number of completed lots per unit time
- Cycle time = duration of time that a lot is inside the tool
- Achieve maximum throughput
- Much less cycle time (e.g. 90% loading: 31.35% cycle time reduction)
Performance Evaluation: Buffer Level

- Buffer level = average number of wafers occupying the buffer
- Reduction in wafer residency time
- Low buffer size may improve space efficiency of tools
- Much less buffer level
  (e.g. 90% loading: 66.57% buffer level reduction)
Performance Evaluation: Residency Time

- Reduced wafer residency time → Yield improvement
- Much less residency time
  (e.g. 90% loading: 53.90% residency time reduction)
Performance Evaluation: Deployment Opportunity

- Opportunity time = queuing time of a lot before the lot enters a tool
- Much more deployment opportunity time
  (e.g. 90% loading: 13.62% increase in opportunity time)
Performance Evaluation: Performance Characterization

- Opportunistic wafer entry: Maximum throughput
- Admission algorithm: Minimized residency time within maximum throughput

(Loading level = 90%)
Concluding Remarks

- Wafer admission control algorithm
  - Minimize wafer residency time within maximum throughput
  - Improve lot cycle time, buffer level and deployment opportunity
  - Allow practical considerations (e.g. Pre-scan setup, reticle alignment process)
  - Provide guidance to wafer transfer robot controller

- Higher train level, more improved tool performance

- Future direction
  - Incorporate robot movements inside the tools