Process Control Systems – SPC in Practice
Semiconductor Manufacturing and Chemical Engineers – BYU Oct 2011
Who is IM Flash Technologies?

- Joint venture company established January 6, 2006 between Intel Corporation and Micron Technology (IM).

- State-of-the-art 300-millimeter fabrication facility in Lehi, Utah (3-5 billion dollar investment)
  - 2.3 Million Square Feet of Building (53 acres)
  - 160,000 sq ft. of Fab Manufacturing space
  - Site sits on 2100 acres
  - 2800 parking stalls, with underground tunnels
  - 8.5 million cu. yards of dirt excavated
  - Approximately 1450 employees

Formed to **manufacture** NAND Flash memory!
IM Flash Technologies is a values-based organization and believes in operating with the highest degree of:

**Commitment**
The personal obligation of each employee to accomplish the goals and objectives of the team.

**Teamwork**
The collaborative effort that leverages the strengths of each individual while working toward common goals.

**Integrity**
Truth, respect, and confidentiality in our interactions with our fellow team members, community, suppliers, and customers.

**Execution**
The achievement of shared goals through rigorous planning, tenacious follow-through, and clear accountability.

Living these values will guide behaviors and define relationships.
Why bring two leading companies together?

• Create the best products in the world!

• Combine and leverage:
  – Micron’s leadership in NAND process and product technology
  – Intel’s Multi-Level Cell technology (MLC) and history of innovation in Flash memory!!!
  – Existing investments in R&D and customer support

• Lower process development and NAND design costs

• Allows greater productive use of prior investments in Lehi and Manassas

• Accelerate penetration into NAND, the highest growth memory product in the market
NAND Replacing Hard Disk Drives

Primary applications where NAND competes with HDD

<table>
<thead>
<tr>
<th>Current Applications</th>
<th>Emerging Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile Phones</td>
<td>Laptop Computer</td>
</tr>
<tr>
<td>MP3 Players</td>
<td>Handheld Gaming</td>
</tr>
</tbody>
</table>

NAND Flash memory is a solid-state, non-volatile (holds data in memory even when power is turned off), electrically-erasable memory that consumes less power and is used in many consumer products.
Why is NAND so important to Intel and Micron?

Total NAND Market % by Application

Apple pre-paid $250 million to Micron and Intel in return for a portion of IM Flash Technologies NAND output.
**Basic Structure of a Flash Cell**

**Cell Architecture**

I. **Control Gate** - The conductor of the wordline transistor whereby a row is selected and a voltage bias is applied.

II. **Floating Gate** - Conducting or semiconduction layer that is completely surrounded by a dielectric used to store charge and alter the threshold voltage of the device.

III. **Source/Drain** - Locations of doped silicon where voltage bias is applied to achieve conduction under the gate of the transistor.

IV. **Dielectric** - Insulator; typically ONO (Oxide Nitride Oxide).

V. **Gate Oxide** - Insulator; typically SiO₂.

VI. **Body (Substrate)** - Silicon Wafer.

**Function of a Flash Cell**

**Principles of Operation (High Level Overview)**

The following section is dedicated to delivering a high level overview of the principles of operation of Nand Flash. Do not be concerned if the material in this section is foreign to you, as it is covered in depth throughout this manual. This sectioned is designed merely to give a brief overview of the material that will be covered throughout this manual. Flash memory stores information...
How Small is Small?

Cross-section of a human hair.

100 microns

.05 microns

© 2011 IM Flash Technologies, LLC. All rights reserved.
• **90nm history**
  – 25nm problems were irrelevant OR...
  – 25nm problems were immeasurable
Process Control Systems Group (PCS)

MAINTAINING HIGH QUALITY NAND WAFERS IN A FAST-PACED MANUFACTURING ENVIRONMENT
Run to Run (R2R) Overview

- R2R is the modification of process or equipment settings between machine runs to optimize output through the use of feed forward or feedback control methods.

- The R2R system interacts with:
  - Manufacturing Execution System
  - Recipe management
  - Automation host
  - Data collection
  - Page/email systems
R2R Control Scenarios

• Words to be aware of
  – Run
  – Lot
  – Wafer
  – Die (Dice)

• Types of Control
  – Single input/Single output (SISO)
  – Multiple input/Multiple output (MIMO)
  – Lot-Level control
  – Wafer-level control
  – Within Wafer control
## Steps to Building a Controller

<table>
<thead>
<tr>
<th>NEW CONTROLLER</th>
<th>Controller Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Controller Preparation</strong></td>
<td></td>
</tr>
<tr>
<td>Kickoff Meeting - WW Level</td>
<td>Controller Deployment Temporary ECN Submitted</td>
</tr>
<tr>
<td>FCT R2R Controller Tracking entry created</td>
<td>Test Server passive mode testing</td>
</tr>
<tr>
<td>Process Model Verification</td>
<td>Controller MES Lot Attribute requested</td>
</tr>
<tr>
<td>Software Modifications Submitted (if necessary)</td>
<td>Process Recovery and Escalation procedures defined and communicated</td>
</tr>
<tr>
<td>New Requirement Document Saved to MDM</td>
<td>Controller move to Production Server</td>
</tr>
<tr>
<td><strong>Requirements Definition</strong></td>
<td>Production Server Passive Mode Testing</td>
</tr>
<tr>
<td>PPA on R2R control for this process - Fab Level</td>
<td></td>
</tr>
<tr>
<td>Completed Requirement Document Review - Fab level</td>
<td><strong>Deployment</strong></td>
</tr>
<tr>
<td>Completed Requirement Document Review - WW Level</td>
<td>FOAK Deployment Notification</td>
</tr>
<tr>
<td><strong>Controller Development</strong></td>
<td>FOAK Testing Begin</td>
</tr>
<tr>
<td>Changes in Mfg Software as Required by E3 (Sigma Specs, GeRM Params, etc.)</td>
<td>FOAK Testing Complete</td>
</tr>
<tr>
<td>Completed E3 Controller Peer Review - Fab Level</td>
<td>Full Deployment</td>
</tr>
<tr>
<td>Completed E3 Controller Peer Review - WW Level</td>
<td>Controller and Dependencies Exported, Saved to MDM</td>
</tr>
<tr>
<td><strong>Post Deployment</strong></td>
<td></td>
</tr>
<tr>
<td>Completed E3 Controller Peer Review - Fab Level</td>
<td>Controller Description Doc</td>
</tr>
<tr>
<td>Completed E3 Controller Peer Review - WW Level</td>
<td>Post Deployment Eval</td>
</tr>
</tbody>
</table>
Performance analysis

- R2R control systems have been deployed on several process steps
- Average of 40% Cpk improvement
- Example SPC charts before and after implementing R2R controllers for two different processes
R2R Controller Performance Analysis Continued

- R2R control drives the process mean towards target
  - No R2R: $\mu = 0.18637$
  - With R2R: $\mu = -0.0075$

- R2R control also reduces process variation
  - No R2R: $\sigma = 4.16507$
  - With R2R: $\sigma = 2.50986$
Dry Etch Process Control

Process Control Schematic

- **Etch Controller with Feed-Forward and Feed Back Components**

<table>
<thead>
<tr>
<th>Feed-Forward</th>
<th>Feed-Back</th>
<th>Control Knobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>FG Thickness$_{\text{pre-etch}}$</td>
<td>FG Thickness$_{\text{post-etch}}$</td>
<td>Etch Time</td>
</tr>
<tr>
<td>CMP Dishing</td>
<td>Final Etch Depth</td>
<td>Gas flow (Etch Rate)</td>
</tr>
<tr>
<td>CG-AA Delta</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2011 IM Flash Technologies, LLC. All rights reserved.
Pre/Post Etch Comparison

- FG Thickness\textsubscript{pre-etch}
- CMP Dishing
- TunnelOx Thickness
- Etch Target
- FG Thickness\textsubscript{post-etch}
- Final Etch Depth

\( C_{\text{Target}} = \text{Vertical distance from bottom IPD to top of silicon height} \)
Pre/Post Etch Comparison

- **For pre Etch – FF calculation**
  - \[ \text{Etch Target} = \text{FG Thickness}_\text{pre-etch} - \text{CMP Dishing} - C_{\text{Target}} + \text{TunnelOx Thickness} \]
  - Etch time is determined by the Etch Target and Etch Rate for tool

- **For post Etch – FB calculation**
  - \[ \text{C}_{\text{actual}} = \text{FG Thickness}_\text{post-etch} - \text{Final Etch Depth} + \text{TunnelOx Thickness} \]
  - Feed back delta between \( C_{\text{target}} \) and \( C_{\text{actual}} \) for adjustment
  - \( \text{Etch Amount}_{\text{post-etch}} = \text{Etch Rate} \times \text{Etch Time} + \text{Bias (constant)} \)
  - \( C_{\text{delta}} \) and Etch Rate determine gas flow and etch time adjustments
Post CGAA Delta Good vs. Bad
Cross Wafer Uniformity Control

- A multiple inputs and multiple outputs (MIMO FF+FB) solution has been implemented to improve Center to Edge profile and uniformity

\[
\begin{bmatrix}
CGAA\_Center \\
CGAA\_Edge
\end{bmatrix} =
\begin{bmatrix}
c_{11} & c_{12} \\
c_{21} & c_{22}
\end{bmatrix}
\begin{bmatrix}
EtchTime \\
GasFlow
\end{bmatrix}
+ 
\begin{bmatrix}
g_1 \\
g_1
\end{bmatrix}
\begin{bmatrix}
FF\_Disturb
\end{bmatrix}
+ 
\begin{bmatrix}
b_1 \\
b_2
\end{bmatrix}
\]

- Model Predictive Control is used in MIMO control system
  - MPC is one of the most advanced control technologies
    - Optimizes tuning knobs to get multiple outputs all closest to target
    - Takes into account all tuning knobs' constraints
    - Allows operation closer to constraints
    - Fully optimize use of tuning knobs within process windows
    - Minimize under and overshoot of target and hits targets in fewest runs

![Graph showing model predictive control](image_url)
Data Comparison

- **CG-AA Delta before and after controller implementation**
  - 38% Cpk improvement
The Furnace R2R Input - Output Model

The Input - output model for the furnace can be described as below,

\[ y_k = Cu_{k-1} + b_{k-1} \]

or,

\[
\begin{bmatrix}
    Top _Thick. \\
    TopCenter _Thick. \\
    Center _Thick. \\
    BottomCenter _Thick. \\
    Bottom _Thick.
\end{bmatrix}
\begin{bmatrix}
    c_{11} & c_{12} & c_{13} & c_{14} & c_{15} \\
    c_{21} & c_{22} & c_{23} & c_{24} & c_{25} \\
    c_{31} & c_{32} & c_{33} & c_{34} & c_{35} \\
    c_{41} & c_{42} & c_{43} & c_{44} & c_{45} \\
    c_{51} & c_{52} & c_{53} & c_{54} & c_{55}
\end{bmatrix}
\begin{bmatrix}
    Temp1 \\
    Temp2 \\
    DepTime \\
    Temp4 \\
    Temp5
\end{bmatrix}
+ \begin{bmatrix}
    b_{1} \\
    b_{2} \\
    b_{3} \\
    b_{4} \\
    b_{5}
\end{bmatrix}
\]

Note: the slope matrix (c11, c12, c13,..., c55) was obtained via experiment. Significant non-diagonal terms were observed (i.e. \(c_{54} \neq 0\)), believed to be due to test wafer positions relative to heater elements, and due to reactor/recipe design.
The State Equation and Output Equation,

- State Equation: \( x_{k+1} = Ax_k + Bu_k + Fw_k \)

- Output Equation: \( y_{m,k} = Cx_k + Gf_k + v_k \)

where \( x_k \) is state vector; \( w_k \) is state noise vector; \( v_k \) is measurement noise vector and \( f_k \) is feed forward components vector. \( A, B, C, F, G \) are state model parameters.

State Space mapping allows modeled disturbances to linearized and added to the control model as a separate term.
The E3 R2R framework used for R2R is the Model Predictive Control (MPC)

- Batches K-3 through K have processed in the furnace reactor and have post-metrology measurements. The Filter Horizon (or State Estimation Horizon) consists of a fixed number historical runs where the model state is known.

- Batch K+1 has not yet been processed, and constitutes the Prediction Horizon (or Controller Horizon).

- The Filter Horizon is used to estimate the model states, while the Prediction Horizon is the range over which the controller set points are tuned via the controller block.
States Estimation – Filter Horizon

\[
\min_{\{w_k, v_k\}} \left( J = \sum_{k=-1}^{N-1} w_k ' Q w_k + \sum_{k=0}^{N} v_k ' R v_k \right)
\]

Subject to:

\[
\begin{align*}
    x_0 &= x_0 + w_{-1} \\
    x_{k+1} &= A x_k + B u_k + F w_k \\
    y_k &= C x_k + G f_k + v_k \\
    x_{\text{min}} &\leq x_k \leq x_{\text{max}} \\
    w_{\text{min}} &\leq w_k \leq w_{\text{max}} \\
    v_{\text{min}} &\leq v_k \leq v_{\text{max}}
\end{align*}
\]

- Q and R are configurable weighting matrices used as part of the optimization cost function (analogous to lambda in EWMA control).
- The optimization function allows states to be calculated such that state noise and measurement noise are minimized while remaining within established model constraints.
- The state estimation block updates model states based on historical run data using the control model and the state estimation optimization equation.
Controller - Prediction Horizon

\[
\begin{align*}
\min_{\{u_k, \ldots, u_{k+p-1}\}} & \left( \sum_{i=k+1}^{k+p} (y_i - y^t)' Q (y_i - y^t) + u_i' R u_i \right) \\
\text{Subject to:} & \\
x_{i+1} &= A x_i + B u_i \\
y_i &= C x_i + G f_i \\
u_{\text{min}} &\leq u_i \leq u_{\text{max}} \\
u_{\text{-Delta}} &\leq (u_i - u_{\text{default}}) \leq u_{\text{+Delta}} \\
y_{\text{min}} &\leq y_i \leq y_{\text{max}}
\end{align*}
\]

- Q and R are additional configurable weighting matrices allowing additional dampening by minimizing the predicted output errors and input moves.

- There are cases where it is necessary to use optimization based on higher cost of input moves when an explicit control model solution is outside of operating window constraints.

- In short, the controller block calculates the new recipe settings by driving the predicted metrology towards target for the Prediction Horizon, based on the model states.
• Wafers are not always measured at post-metrology in the same sequence in which they were processed in a given furnace.

• State estimation uses the process run sequence to estimate the furnace states instead of the post-metrology run sequence. Such a method is critical for processes with large or rapidly changing state disturbances.
• **Non threaded Run to Run control**
  - Sharing information among different threads (e.g. minority part states could be continuously updated through majority part.)

• **Performance monitoring and self tuning**
  - Good controller performance monitoring metric can alert potential need for R2R tuning, which is the key to maintaining near optimal performance
  - Controller’s weighting factors can also be changed dynamically based on performance metric

• **Run to run control through virtual metrology**
  - Virtual Metrology (VM): FDC data and upstream metrology information can be utilized to predict metrology data
What We Look for in a Process Control Engineer

- **Group Demographics**
  - 54% ChE
  - 23% EE
  - 15% ME
  - 8% Other
  - 15% PhD
  - 23% MS
  - 62% BS

- Demonstrated dedication to IMFT's core values of commitment, integrity, teamwork, and execution
- Enthusiasm for SPC, process control, and data
- Proven ability to be effective in verbal and written communication
- Strong knowledge of DOE and data analysis
- Detailed understanding of control loops and control systems
- Thorough knowledge of fab processes and tools
- Proven ability with fab software applications
- Strong analytical, problem-solving skills
- Self-Starter/Strong Work Ethic
- Ability to develop, teach, present, and understand technical subjects
- Understanding of fab data flow and tool/host communication
- Programming ability
Where do I go from here?

- Look into potential of working with IMFT engineer as a mentor for a project, or better yet work toward an internship at IMFT
- Shore up your statistical knowledge. Especially helpful are applied/DOE type statistical classes
- Look into advanced control classes. Industry trend is leaning heavily toward state space solutions, which are not covered to great extent in introductory control classes
1. Which two companies combined to form IMFT?
2. What type of memory does IMFT produce?
3. What does ‘PCS’ stand for?
4. What’s feature size of current technology in mass production at IMFT?
5. What do SISO and MIMO stand for?
6. What’s feedback control?
7. What’s feed forward control?
8. What’s the indicator for performance analysis?

Bonus: Name one of the core values at IMFT.