Use of Knowledge Discovery from Manufacturing Data for Yield Improvement

Patrick J Carroll
Yield & Integration Engineering Manager
RFMD, Greensboro NC

CS ManTech, Boston MA
April 24, 2012
The Scientific Method

• Taught early in science education
• Method used to advance science, prove or disprove theories
• In CS manufacturing industry, as scientists & engineers with device knowledge
  – May be used to tackle a manufacturing yield challenge
  – Observe the problem, Hypothesize the cause, Start experiments
  – May or may not result in root cause discovery
  – Cost of experiment & time
The Six Sigma Method

- Define
- Measure
- Analyze
- Improve
- Control

• For good manufacturing problem solving a more prescribed method is used
• DOEs are encouraged as opposed to One Factor at a Time experiments
• Experiments in a fab can be costly & time consuming
• Therefore, it is important to have the most complete knowledge possible before beginning that process
• This can be obtained using Knowledge Discovery from Data
KDD – Data Mining

- KDD
  - Knowledge Discovery in Databases
  - Knowledge Discovery and Data Mining
- From Wikipedia
  - The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining).
- It is not Hypothesis Testing
- No Presumptions Made
- Learning what knowledge can be extracted from the data

KDD can be described as these steps

1. Assemble
2. Fix
3. Explore
4. Validate
### Assembling Manufacturing Data for RF Products Made from Compound Semiconductor

<table>
<thead>
<tr>
<th>Level</th>
<th>Data Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boule</td>
<td>Supplier, Run #, Location</td>
</tr>
<tr>
<td>Substrate</td>
<td>Slice #, Mobility, EPD</td>
</tr>
<tr>
<td>Epitaxy (MBE,MOCVD)</td>
<td>Tool, Run #, Layer Thickness, Doping, Date</td>
</tr>
<tr>
<td>Wafer Fab (Patterning, Metals, ...)</td>
<td>Tools, Shifts, Operators, Process Control Data</td>
</tr>
<tr>
<td>Electrical Tests (PCM)</td>
<td>Rsh, Rcon, Capacitance, ...</td>
</tr>
<tr>
<td>Die Tests (KGD)</td>
<td>Circuit Measurements</td>
</tr>
<tr>
<td>Assembly (Die, Wiring, Molding , ...)</td>
<td>Bonding data, Components, Inspection, Location</td>
</tr>
<tr>
<td>Package Test</td>
<td>Final electrical test data</td>
</tr>
<tr>
<td>Customer Quality</td>
<td>Customer feedback, DPM, Field returns</td>
</tr>
</tbody>
</table>

- Data in multiple data bases across the globe.
- Data in various formats.
- Can the properties of the boule affect the performance and quality of your phone?
- How can we process all this data to solve a problem?
- Should we buy a system?
Assemble using The Universal Data Query (UDQ)

- Before the UDQ – All manual
  - Individual queries – Quality varied
  - Issues joining data
  - Issues with ‘bad’ data

- What is the UDQ? A single software tool to collect and join information from all relevant data storage sites
  - Written in JMP Scripting Language (JSL)
  - Obtains a list of lots of interest (Wafer or Package)
  - Queries tables from any of dozens of database tables located on various servers
  - Data is joined in JMP

- What are the advantages?
  - User doesn’t need special knowledge of each data source
  - Each table can be queried most optimally
  - Special knowledge to transform data built in
  - Known issue in data can be fixed
  - All data obtained quickly
Filter

- Automated data exploration requires automated filtering to prevent overlooking relationships

- A Common sense algorithm to filter outliers is Normalized Ordered Differences (NOD)

- Some standard ways to deal with outliers
  - Eliminate outliers for just cause term-by-term
    - Most Conservative
    - Most impractical for automation and a large set of terms
  - Apply a statistical test (e.g. +/- 3 sigma)
    - Tails are “cut-off” of legitimate non-normal distributions
    - Outliers may have a great effect on sigma
  - Use inner percentile of data (e.g. 5th to 95th percentile)
    - Data is eliminated from every term whether it is outlying or not
NOD Explained & Demonstrated

1,2 and 3 Mean Absolute Deviations from the Median Y

Before

After 2.5 NOD 3x
Explore – Finding Relationships

- JMP provides a number of techniques to explore data (e.g. CART, Best Models)
- We have developed a number of additional scripts to analyze and visualize the most statistically relevant relationships
  - Best Least Squares Fits to all numeric values arranged in order
    - User determines Degree of Fit
    - $R^2$ threshold
    - Display characteristics
  - Best 2 Parameter Fits

LSQ_Demo.avi
Explore Kruskal-Wallis

- Script available to generate Best Fit to Categorical Data using:
  - ANOVA [ Prob (F Ratio) ]
  - Kruskal-Wallis [ Prob (ChiSq) ]

- What is Kruskal-Wallis (KW)?
  - Kruskal-Wallis is a rank sum test
    - Data is evaluated by rank (smallest to largest)
    - The ranks are summed for each category
    - Expected ~ equal mean score if randomly distributed
    - A ChiSq test is done to determine if the difference in the sums is significant

- Why I like Kruskal-Wallis
  - Avoids some ANOVA assumptions about the data (e.g. normality)
  - Outliers have very little effect

Example of a Oneway Analysis using Kruskal-Wallis Rank Sum Test
With all the knowledge obtained from the data a better hypothesis can be made for the root cause of a manufacturing issue.

A experiment or DOE may still be needed to confirm in order to move from correlation to causation.

But the knowledge obtained may be sufficient to make a needed correction and confirm with manufacturing data.

Much more likely occurrence now.
Example – Improved Product Yield related to Epi Improvements

- Important New Product with wide Large Yield Variation due partially due to linearity fails
- UDQ used to join data from package product test to substrate
- Best Relationships to Linearity found among all data
- Several parameters found including a key parameter (Xquotient) related to MBE
- Yield improvement resulted from better screen of wafer starts
- Studies of MBE process resulted in improved wafer Quality:
  - Much improved process capability to linearity
  - Improved capability to other critical parameters
Conclusion

• In order to quickly identify likely causes to manufacturing issues, a KDD approach should be taken to Assemble, Fix and Explore existing data.
• The output of the process can be fully automated for common issues.
• By using the KDD process at RFMD, many opportunities for yield improvement have been found.
• A result of that is illustrated in the improvement shown in the composite Known Good Die (KGD) scrap rate of all technologies.
Acknowledgements

• The author would like to thank the wafer fab staff at RFMD for their support and contributions to yield improvement.

• Special thanks to the Yield Improvement Team

• Special thanks also to Yik Yang, Dain Miller and Nathan Conrad for contributions to the Universal Data Query.