Data Analytics for Semiconductor Manufacturing
What do we mean by “Data Analytics”? 

Analytics uses **data** to **drive decision making**, rather than gut feel or intuition. Analytics uses data to derive **insights** which drives business **actions** and therefore produces business **outcomes**.

### Analytics Maturity

<table>
<thead>
<tr>
<th>Level of Sophistication</th>
<th>Prescriptive</th>
<th>Predictive</th>
<th>Diagnostic</th>
<th>Descriptive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Competitive Advantage</strong></td>
<td><em>What should happen?</em></td>
<td><em>What will happen?</em></td>
<td><em>Why did it happen?</em></td>
<td><em>What happened?</em></td>
</tr>
</tbody>
</table>

"It would appear, Hopkins, that your gut feel was only indigestion"
Data Analytics 3.0

- Optimized workflow and decision making with prescriptive analytics
- In-Memory processing with streaming data
- Smart factory
- Enterprise system Integration

Big Data (Analytics 2.0)

- Big data (Structured + Unstructured)
- IoT (Internet of Things)
- 3Vs (Volume, Variety, Velocity) + Value
- Machine learning and Deep learning

Structured Data set based Descriptive Analysis
- Decision making from Historical Data

Data–Enriched Offerings (Analytics 3.0)
Data Analytics for Semiconductor Manufacturing Process

- Data is being logged everywhere
- Data comes from many disparate sources
- Get insight from aggregated data and make actionable decisions to
  - Improvement Yield
  - Detecting error
  - Reduction Manufacturing cost
  - Process optimize
  - Prognostics
  - equipment health/Maintenance
Workflow of Analytics for Integrated Wafer Inspection system

1. Big Data Access
   Big/Diverse Data Sources

2. Analytics
   Machine Learning

3. Integrate to Enterprise System
   Automation, Business

Data Aggregator / Reducer and Automated / Enterprise System

Process, feature data
Business and Transactional

Data Exploration  Analytics Modeling & Evaluation  Integration
Any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications.”

(Wikipedia)

“Any collection of data sets so large that it becomes difficult to process using traditional MATLAB functions, which assume all of the data is in memory.”

(MATLAB)

Big Data

- Various Data Sources from complex process
  - image, sensor, numeric, text, etc.

- Rapid Data exploration from enormous number of manufacturing equipments
  - Scalable Data processing with scalable HW

- Identify the value from uncertainties
  - Statistics, Clustering, Classification, Regression

- Ease of deployment
  - Support common development environment
    (C/C++, .Net, Java, Excel, Hadoop)
“Reading Big Data in MATLAB”

“Techniques for Big Data in MATLAB”

Load, Analyze, Discard
parfor, datastore,

MapReduce

Distributed Memory
SPMD and distributed arrays

1. Big Data Access
Big/Diverse Data Sources

Load, Analyze, Discard

MapReduce

Distributed Memory
SPMD and distributed arrays

 Researchers

Complexity

Embarrassingly Parallel

Non-Partitionable

ImageAdapter
SystemObjects (streaming data)
freadd
memmapfile
imread
database
textscan
readtable
xliff
load
matfile
API
datastore

Researchers

BIG

SMALL

Load, Analyze, Discard
Parfor, datastore,

MapReduce

Distributed Memory
SPMD and distributed arrays

Researchers

Complexity

Embarrassingly Parallel

Non-Partitionable
Why use MATLAB for Machine Learning?

- Complete environment for Machine Learning (prototyping to production)
- Verified and trusted machine learning algorithms
- Classification Learner app that lets you perform common machine learning tasks interactively
- Tightly integrated with Signal Processing, Image Processing, Computer Vision and other workflows.
- Open and flexible architecture enables a customized workflow
ASML Develops Virtual Metrology Technology for Semiconductor Manufacturing with Machine Learning

**Challenge**
Apply machine learning techniques to improve overlay metrology in semiconductor manufacturing

**Solution**
Use MATLAB to create and train a neural network that predicts overlay metrology from alignment metrology

**Results**
- Industry leadership established
- Potential manufacturing improvements identified
- Maintenance overhead minimized

"As a process engineer I had no experience with neural networks or machine learning. I worked through the MATLAB examples to find the best machine learning functions for generating virtual metrology. I couldn’t have done this in C or Python—it would’ve taken too long to find, validate, and integrate the right packages."

Emil Schmitt-Weaver
ASML

Link to user story
Machine Learning Workflow

Step 1: Training

Step 2: Evaluation

Iterate
Overview of Machine Learning in MATLAB

Unsupervised Learning
- K-means
- Fuzzy K-means
- Hierarchical
- Neural Network
- Gaussian Mixture
- Hidden Markov Model

Supervised Learning
- Decision Trees
- Ensemble Methods
- Neural Network
- Non-linear Reg. (GLM, Logistic)
- Linear Regression
- K-means
- Fuzzy K-means
- Hierarchical
- Naive Bayes, Nearest Neighbors
- Gaussian Mixture
- Hidden Markov Model

2. Analytics
Machine Learning
Integrate Analytics Application to Enterprise System

Typical Workflow

- Separated Storage; Local Drive, Network Drive, Database

Business
- Head of BU
- Domain experts
- (Data)

Business Units
- Analytics Group
- Head of Analytics
- Domain experts
- Data scientists

Process Control System
- IT
  - IT Mgr.
  - Enterprise Architect
  - Programmers

Re-coding

Increase

- Performance
- Yield rate
- Efficiency...

Decrease

- Time consumption
- Operation overhead
- Error
- Pitfall…

3. Integrate to Enterprise System
- Automation, Business
Deployment for Analytics Integration

- Integrated into production IT environments without recoding or creating custom infrastructure
- Packaged as compatible components with a wide range of systems (Java, .NET, Excel, Python & C/C++)
- Shared as standalone applications or run from: web, database servers, desktop, enterprise applications, etc.
Solutions for Web/Enterprise Application Servers
MATLAB Production Server

- Most efficient path for creating enterprise applications
- Deploy MATLAB programs into production
  - Manage multiple MATLAB programs and versions
  - Update programs without server restarts
  - Reliably service large numbers of concurrent requests
- Integrate with web, database, and application servers
Integrate Analytics Application to Enterprise System

Suggested Workflow

Out of Memory Data Access for Massive Data (Hadoop, DB)

Data from manufacturing process
- Wafer
- Oxidation
- Photo Lithography
- Etching
- Overlay Metrology
- Scanner / SEM image
- Fab/Probe/Assembly/Test Yields
- Economic market data
- etc

Analytics Integration Without Re-coding

Results/ Reporting system (C# and Java)
- Process Optimization
- Yield Improvement
- Rapid processing
- Increase efficiency
- IP management

Test deployment in R&D part

Request Broker

MATLAB Production Server

Enterprise System/ Process Automation

Prototyping Analytics solution

Excel Add-in

Hadoop

Standalone Application

C/C++

Java

.NET
Summary: MATLAB Data analytics in Semiconductor Industry

1. Big Data Access
   Big/Diverse Data Sources
   - Data Exploration/Visualization
   - Big data Access
     - Datastore, MapReduce
     - HDFS support
     - DB, No SQL support

2. Analytics
   Machine Learning
   - Scale Up Computation
     - Multicore Desktop
     - Cluster / Cloud Computing
     - GPU computing
     - Task Parallel / Data Parallel
   - Analytics / A.I
     - Statistics and Machine learning
     - (Global)Optimization
     - Neural Network toolbox

3. Integrate to Enterprise System
   Automation, Business
   - System Integration
     - Stand-alone Application
     - Multi-language target library
     - Hadoop integration
     - Web service
     - Excel Add-in
     - Hadoop
     - Standalone Application
     - C/C++, Java, .NET, Spark
     - MATLAB Compiler
     - MATLAB compiler SDK
     - MATLAB Production Server