Paper 230

Project RAPTOR: Implementation and Exploitation of a Manufacturing Data Warehouse

Anne Brooks, Texas Instruments Inc., Dallas Texas
Roger Thompson, JJT Inc., Dallas Texas

Abstract

As the cost and complexity of semiconductor manufacturing processes increase, companies such as Texas Instruments, Inc. face increasing pressure to improve the yield of devices produced. Demonstrating, and maintaining, high manufacturing yields is a critical and ongoing effort, which offers significant returns on investment. Small variations in yield can directly impact cost and revenue by millions of dollars. Furthermore, the ability to quickly ramp yield on new technologies can provide a tremendous competitive edge in the marketplace during lucrative early stages of a product cycle.

A crucial task in yield improvement efforts is thorough data analysis. TI's Raptor project was launched to provide a standardized architecture for data analysis that could be deployed to multiple wafer fabs, assembly sites, and product groups. The project encompassed designing and building multiple data warehouses, as well as deployment of JJT Inc.'s Maestro software, a metadata-driven analysis system developed with Base SAS®, SAS/AF®, SAS/GRAPH®, SAS/STAT®, SAS/QC®, SAS/INSIGHT®, and SAS/ACCESS® to ORACLE software in a UNIX environment. This paper outlines the project's system architecture, capability requirements and factors unique to the industry that drove the system architecture decisions. Examples of analysis made possible by the Raptor project will also be explored.

Semiconductor Manufacturing

Successful yield management includes minimizing the frequency and impact of yield excursions, as well as continuing to drive baseline yield improvement. To do this, semiconductor manufacturers need the ability to quickly identify and resolve yield-limiting problems. Semiconductor manufacturers collect an immense amount of data about the manufacturing process and the finished device (called die) performance. Combined, this data can provide valuable insight into the distribution and root cause of yield loss mechanisms.

Typically, all data collected can be classified into 3 major types:

- Lot-level Data
- Wafer-level Data
- Die-level Data

The data collection starts in the wafer fabrication (fab) plant as lots (wafer batches containing 24 individual wafers) move through a manufacturing sequence that may contain several hundred steps. At the lot-level, WIP (Work In Progress) systems collect processing information and in-line measurement data for hundreds of individual processing steps. At the wafer-level, fault detection systems collect and store machine sensor and processing information. At both the die and wafer level, defect inspection data are collected and stored. In addition to lot-, wafer- or die-level data, machine maintenance information is collected.

The finished wafers are then subjected to a variety of electrical tests. Parametric test structures (simple, non-production devices) are tested on various areas of the completed wafer to verify compliance to process specifications. The actual production die is then subjected to multiprobe testing. Based on the test results, each die on the wafer is then assigned a bin classification. More than one "good" bin is possible as well as many "fail" bins. Die-level parametric data may also be collected.

Additional process and test data is collected at the assembly sites, where finished wafers are sawn into individual die and packaged.

Problems with Data in Current State

Fortunately, the investment has already been made to collect and store the data mentioned above. However, the data in its current state presented several obstacles to data analysis.

One of the largest barriers to effective yield analysis is the fact that the data is located on multiple systems with dissimilar operating environments. Data for each wafer fab and assembly site is maintained in separate databases. Even in a single wafer fab, TI's WIP system stores data in a hierarchical IBM IMS database on the mainframe, while electrical test and bin-sort data are stored in an ORACLE database on a UNIX server.

Furthermore, both the IMS and Oracle test databases are production databases used for running the factory and test floor. The data models for both systems are optimized for transactional or operational systems. The sheer volume and organization of the data in these systems presented a daunting challenge for any type of data analysis.

In addition to the difficulties in accessing and merging the necessary data for yield analysis, the typical user lacked an easy to use analysis software package. While the SAS® System was widely available, most organizations had only one or two users with the expertise necessary to perform efficient analysis.

Given all of the obstacles discussed above, a steep learning curve existed for anyone needing to extract and manipulate yield data. While some local solutions existed to address these problems within Texas Instruments, the company lacked a single user-friendly system to support engineering analysis. Understanding the importance of data driven yield improvement, Texas Instruments has partnered with JJT, Inc. to develop and implement a new framework, called Raptor, designed specifically for engineering data analysis.

Raptor Requirements

The main objective for launching the Raptor project was creation of a new framework designed to increase the efficiency and sophistication of data analysis to support yield improvement. The scope of the Raptor project encompassed design and construction of the necessary data warehouses, installing Maestro software for analysis, reporting and metadata management, development of specific analysis and reporting capabilities geared toward semiconductor manufacturing, as well as user training for data warehouse orientation and Maestro software.

Requirements of the Raptor framework are discussed below:

User requirements

A primary requirement of Raptor is to provide a **single, easy to use interface** which provides many users access to all data sources while shielding them from the underlying data structure and methods for data access, extraction, and management.

The system must also be capable of meeting a **variety of analysis needs.** Raptor should provide production reporting and batch mode execution through a scheduler for **routine analysis** needs. It should also provide the capability for **on-demand, ad-hoc analysis** needed for trouble-shooting and hypothesis testing. Lastly, it should also provide **complex analysis** capability needed to reduce large volumes of data into useful information.

Administrative requirements

From an administrative standpoint, Raptor must provide the ability to support **distributed data warehouses, data marts and multi-tiered data warehouses** containing highly dynamic data. These data sources should employ a data model, which provides an **optimal balance of accessibility, efficiency, and maintainability.** In addition, Raptor must be designed to facilitate easy metabase maintenance.

Within the context of this project, distributed data warehouses are defined as a collection of independent data warehouses, which may be centrally managed, but without a central core of data. The data model for the data warehouses and use of Maestro are common to each wafer fab facility. However, they are maintained and managed independently and have no overlap of data in the data warehouses.

Multi-tiered data warehouses can be defined as combinations of enterprise data warehouses and data marts. The data marts are fed from operational sources through the enterprise data warehouse. Additional data sources providing data that does not exist in the enterprise data warehouse are fed directly to the data mart. This capability is critical to support device product groups. While the enterprise data warehouses for each wafer fab provide data for all devices manufactured at that facility, a device managed by a product group may be manufactured in multiple fabs. The data mart for the device product group would need a transparent and integrated view of multiple wafer fab data warehouses.

Organizational requirements

The Raptor project began as a joint development project involving two wafer fabs. Because the system will eventually be used by many organizations across the company, the system should provide a level of **standardization** to facilitate deployment and support by a central organization while allowing some degree of **customization** to meet individual organizations' needs. In addition, Raptor should be both **scalable** to meet the need of both small and large organizations, as well as **extendible** to new data sources.

Raptor System Description

The Raptor system architecture is shown below.

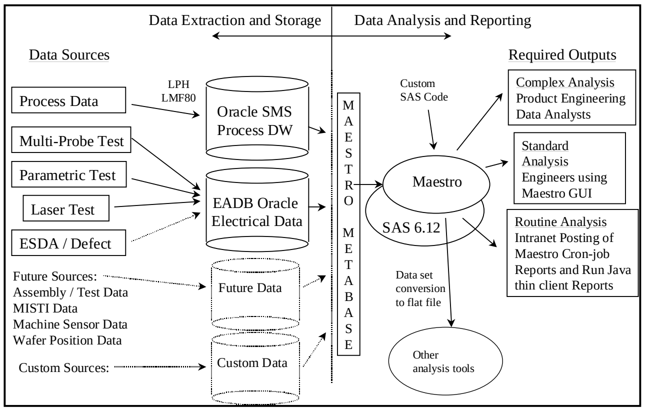


Figure 1 - Architectural overview of the Raptor system

Listed along the left side of Figure 1 are several types of data, which are important to yield improvement analysis. Nightly extractions from each of these data sources are loaded into one of two distinct data warehouses.

Lot-based process data is stored in the **Oracle SMS Process Data Warehouse**. Data is organized by process logpoint, operation, and parameter number. (A logpoint represents a group of sequential process steps, known as operations, used in the manufacturing process. At each logpoint-operation combination, parameter numbers corresponding to the data to be collected are specified.) The data type can be either character (i.e. process recipe used, process machine) or numeric (i.e. measurement data). Because changes in data collection are frequent, a vertical schema was selected.

A vertical schema is defined as one in which multiple rows of a table are used to store data for a single event or granular transaction. A given column is used to determine which event took place or measurement was taken and a second and possibly additional columns are used to store information about the event or value of the measurement taken. This model provides easy loading and extremely dynamic data accommodation, which is crucial in situations when a very large number of different measurements are taken and many new measurements are added on a regular basis.

Electrical test data is stored in the **Engineering Analysis Database (EADB)**. Data is organized into separate tables for die based and site based data. Because this data, like the process data, is highly dynamic, a vertical data schema was chosen as well.

The dynamic nature and volume of new data items entering the data warehouse requires strict inventory control to maintain order. This is accomplished by collecting and storing metadata in the Maestro Metabase as new types of data items enter one of the data warehouses. The Maestro Metabase provides an integrated view of all data warehouses for which it stores metadata. Seamless data access capabilities for any data item known to the metabase is provided by the Maestro software through a point and click graphical user interface for analysis and reporting.

Maestro Metabase

Metadata stored in the Maestro Metabase plays a critical role in each of the data warehouses described. The Maestro Metabase stores metadata for each data source and data item within that source in the data warehouse. However, the use of metadata within Maestro goes well beyond a data source for browsing descriptive information on the contents of the data warehouse.

The Maestro Metabase exploits metadata on behalf of the user to transparently generate SQL queries and submit them to the appropriate data sources. Data Items stored in the metabase are surfaced to the user through a GUI (Figure 2) when creating a report. Upon report execution, the Maestro Metabase determines the most efficient way to extract the required data items from the data warehouse sources, generates the SQL, and submits the query to the appropriate data sources.

Due to the dynamic nature of data entering the data warehouse, the Maestro Metabase is synchronized on a nightly basis after each of the data warehouse updates has completed. The nightly synchronization process detects new data items that entered the data warehouse and updates the Maestro Metabase making it available for the user to select for analysis and reporting.

Maestro User Interface

The statistical and graphical capabilities of the SAS System are provided in a point and click environment through Maestro's graphical user interface to support the analysis needs listed to the right side of Figure 1.

The list of unique parameters known to the Maestro metabase is presented to Raptor users in the Data Item Selection Tree (Figure 2). The Data Item Selection Tree is organized into branches in a fashion that makes it easier for users to identify the data they want.

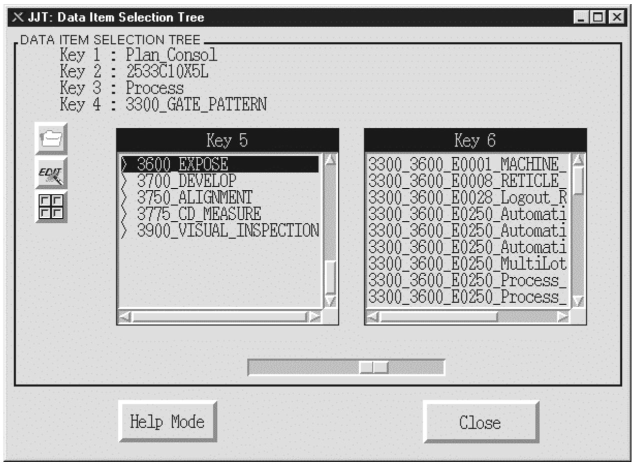


Figure 2 – Maestro Data Item Selection Tree

Reports are created in Maestro by associating specific data items to report templates (Figure 3). In the report template shown below, the data item TP_IDSN_NOM was selected by finding the data item in the Data Item Selection Tree and dragging it to the Analysis box.

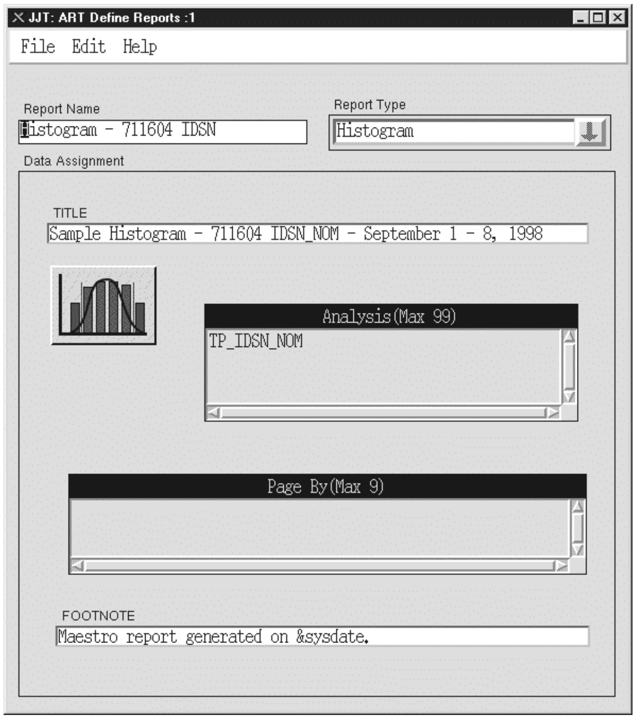


Figure 3 – Maestro Report Definition Template

Constraining or filtering data extracted for a particular analysis is accomplished by specifying simple or compound phrases in the constraint editor. In Figure 4, two simple phrases have been specified for the report definition in Figure 3. Data will be extracted for analysis where TP_IDSN_NOM measurements for device M2+/F711604 were taken between 25AUG1998 and 09SEP1998.

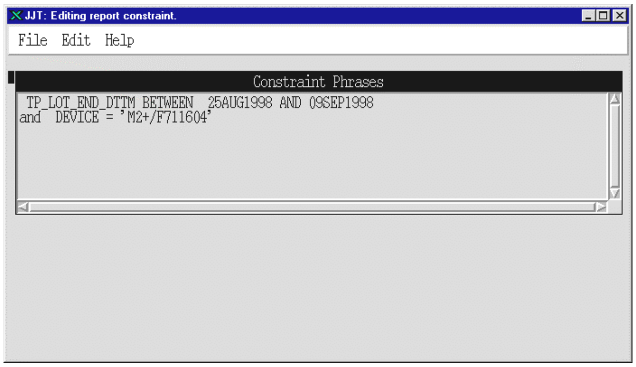


Figure 4 – Maestro Constraint Editor

Exploiting the Data Warehouse

The Raptor system has been implemented in two Texas Instruments wafer fabs to support engineering analysis needs. Routine analysis needs are supported through daily batch scheduled reports that provide detail and summary information for all material completing parametric and functional test during the previous day.

One fundamental report is the cumulative wafer map shown in Figure 5. The wafer map provides a plot of yield as a function of die position. The wafer map signature often provides vital clues as to the actual fail mechanism. This analysis can be performed on any functional test bin for a single lot, group of lots, or all lots probed during a specified time frame.

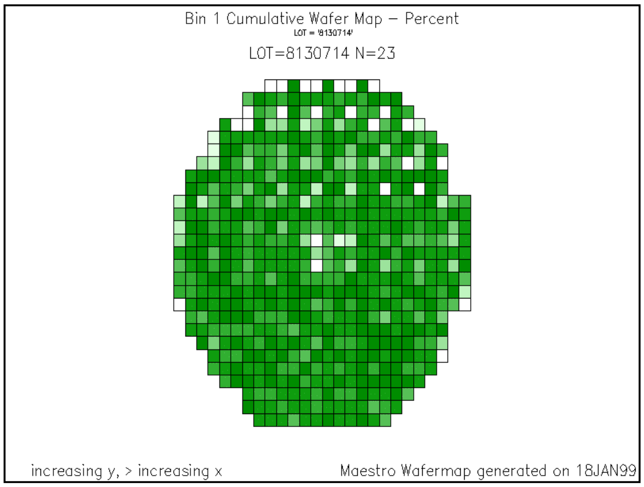


Figure 5 – Sample Wafer Map Report Output

Another type of routine analysis is functional bin Pareto analysis to identify the most frequently occurring fail bins for a specified lot, group of lots, etc. An example is shown in Figure 6.

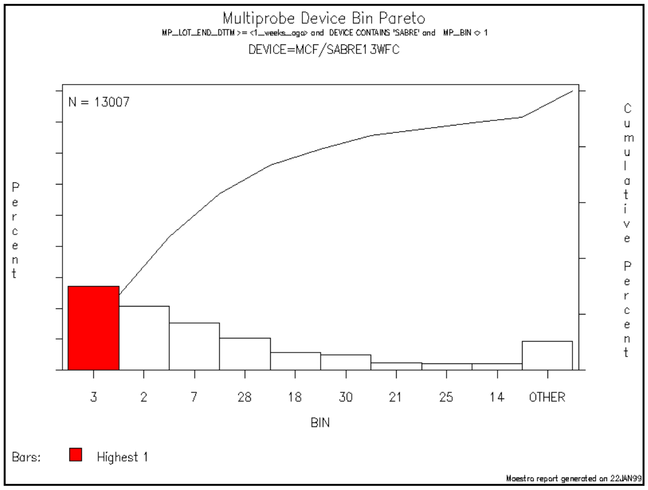


Figure 6 – Pareto Report Output

Parametric data indicates how specific device components are performing on a given wafer, and therefore provide important information with respect to functional yield results. Parametric distribution plots, as well as yield and parametric trend charts, are included as routine Raptor analysis. All of these daily reports are posted to a web server for easy access by product engineering to monitor parametric and yield performance.

Once a new yield loss signature is identified, fab engineers can use Raptor analysis to help isolate the root cause. Especially valuable during the troubleshooting process is the ability to perform analysis of electrical fail data with respect to data collected during the manufacturing process. Parametric and yield data can be easily plotted as a function of process date and time and equipment at any manufacturing process step, allowing the engineer to test for relationships between known events at that process step and the change in electrical test results. In Figure 7, parametric data (collected at the end of the manufacturing process) is plotted by lot number as a function of time through a specific manufacturing process.

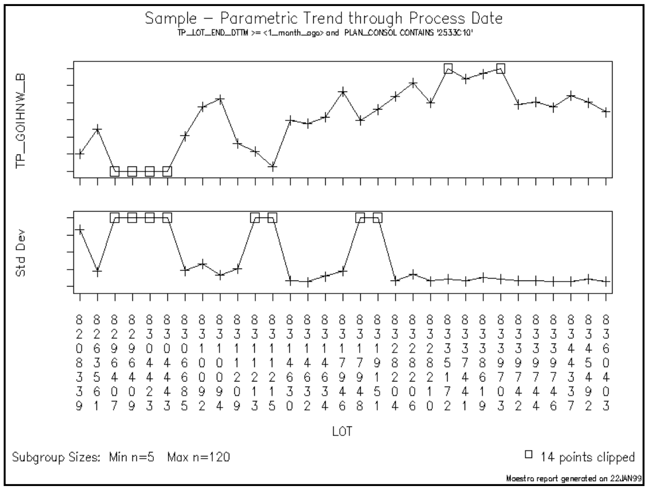


Figure 7 – Sample Control Chart Report Output

The ability to use categorical process data such as equipment, operator, etc, as grouping variables during electrical test data analysis supports hypothesis testing. For example, parametric data is plotted against furnace position during a specific process step in Figure 8. The results indicate that material processed through furnace position 5 has somewhat lower values of the parameter plotted than the other furnace positions.

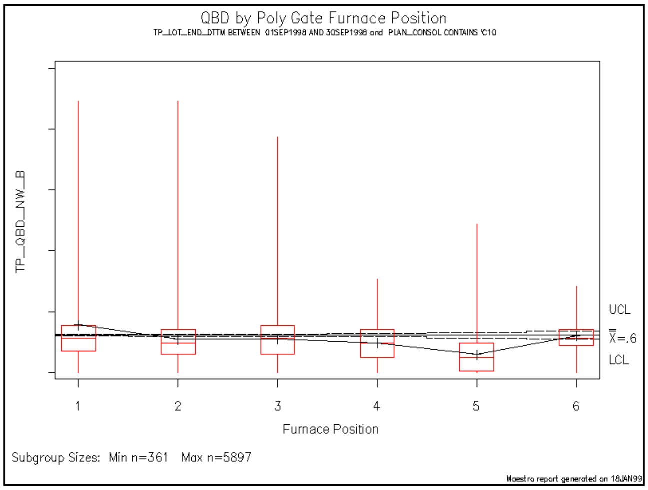


Figure 8 – Sample Box and Whisker Plot Report Output

Using the Maestro interface, users can create their own data items, defined in terms of global data items, for use in their analysis. This feature is especially useful during the evaluation of a manufacturing process change to improve yield or reduce costs. The evaluation process begins with experimental split lots, in which different groups of wafers receive different process treatments. User defined data items that assign wafer numbers to various process treatments can be created and used in experimental split lot analysis. In Figure 9, cumulative distribution plots grouped by the split variable help the engineer analyze the impact of the various experimental process conditions on the parametric results.

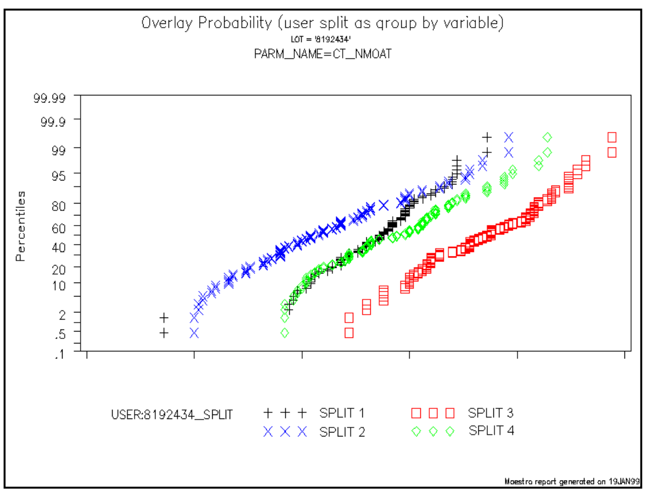
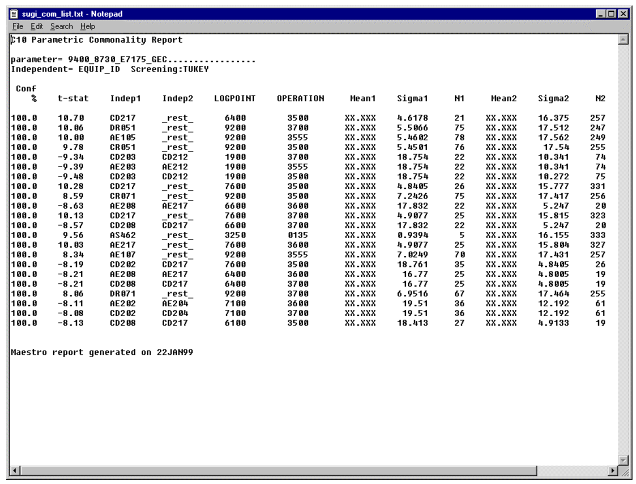


Figure 9 – Sample Cumulative Distribution Plot Report Output

When the analysis of a single or group of experimental lots indicate a change is favorable, engineering usually runs a larger trial group of lots under the proposed process conditions before releasing the change to production. User defined data items are used during this analysis to divide material into current (or baseline) and experimental groups. The definitions of these groups do not have to be based on lists of lots or wafer numbers, but could be defined in terms of any appropriate global data items (such as equipment id numbers and specified time periods).

In addition to answering specific questions regarding yield data, the Raptor system can be used to support complex analysis. Using reports developed specifically for the semiconductor industry and which utilize the power of SAS, engineers can identify the questions to be asked. One example is a report that will systematically perform statistical comparisons of the distributions of a specified data item with respect to the manufacturing equipment used at each processing step. Report output includes a result list sorted by statistical significance (Figure 10) along with comparative distribution plots of the data (Figure 11).


sugi_com_list.txt - Notepad

File Edit Search Help

C10 Parametric Commonality Report

parameter= 9400_8730_E7175_GEC.................
Independent= EQUIP_ID Screening:TUKEY

Conf %	t-stat	Indep1	Indep2	LOGPOINT	OPERATION	Mean1	Sigma1	N1	Mean2	Sigma2	N2
100.0	10.70	CD217	_rest_	6400	3500	XX.XXX	4.6178	21	XX.XXX	16.375	257
100.0	10.06	DR051	_rest_	9200	3700	XX.XXX	5.5066	75	XX.XXX	17.512	247
100.0	10.00	AE105	_rest_	9200	3555	XX.XXX	5.4602	78	XX.XXX	17.562	249
100.0	9.78	CR051	_rest_	9200	3500	XX.XXX	5.4501	76	XX.XXX	17.54	255
100.0	-9.34	CD203	CD212	1900	3700	XX.XXX	18.754	22	XX.XXX	10.341	74
100.0	-9.39	AE203	AE212	1900	3555	XX.XXX	18.754	22	XX.XXX	10.341	74
100.0	-9.48	CD203	CD212	1900	3500	XX.XXX	18.754	22	XX.XXX	10.272	75
100.0	10.28	CD217	_rest_	7600	3500	XX.XXX	4.8405	26	XX.XXX	15.777	331
100.0	8.59	CR071	_rest_	9200	3500	XX.XXX	7.2426	75	XX.XXX	17.417	256
100.0	-8.63	AE208	AE217	6600	3600	XX.XXX	17.832	22	XX.XXX	5.247	20
100.0	10.13	CD217	_rest_	7600	3700	XX.XXX	4.9077	25	XX.XXX	15.815	323
100.0	-8.57	CD208	CD217	6600	3700	XX.XXX	17.832	22	XX.XXX	5.247	20
100.0	9.56	AS462	_rest_	3250	0135	XX.XXX	0.9394	5	XX.XXX	16.155	333
100.0	10.03	AE217	_rest_	7600	3600	XX.XXX	4.9077	25	XX.XXX	15.804	327
100.0	8.34	AE107	_rest_	9200	3555	XX.XXX	7.0249	70	XX.XXX	17.431	257
100.0	-8.19	CD202	CD217	7600	3500	XX.XXX	18.761	35	XX.XXX	4.8405	26
100.0	-8.21	AE208	AE217	6400	3600	XX.XXX	16.77	25	XX.XXX	4.8005	19
100.0	-8.21	CD208	CD217	6400	3700	XX.XXX	16.77	25	XX.XXX	4.8005	19
100.0	8.06	DR071	_rest_	9200	3700	XX.XXX	6.9516	67	XX.XXX	17.464	255
100.0	-8.11	AE202	AE204	7100	3600	XX.XXX	19.51	36	XX.XXX	12.192	61
100.0	-8.08	CD202	CD204	7100	3700	XX.XXX	19.51	36	XX.XXX	12.192	61
100.0	-8.13	CD208	CD217	6100	3500	XX.XXX	18.413	27	XX.XXX	4.9133	19

Maestro report generated on 22JAN99

Figure 10 – Sample Parametric Commonality Report Output

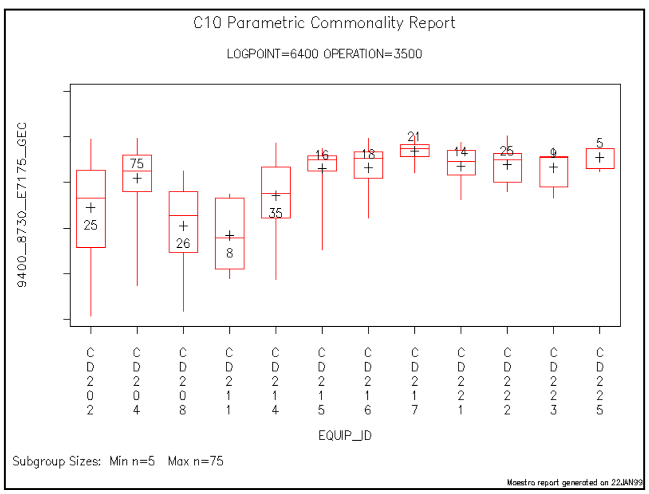


Figure 11 – Sample Parametric Commonality Report Output

The same report can also be used to seek out the most significant parametric differences between treatments on experimental lots. Another report currently under development allows the user to perform regression analysis between large groups of parameters and identify the pairs with the highest degree of correlation. All of these examples help the engineer efficiently sort through large volumes of data to identify relationships worthy of further investigation.

Future Plans

Many opportunities exist to enhance the current Raptor system. Database optimization and tuning will focus on decreasing response time for the most common types of analysis. The addition of summary data is also being considered to increase analysis efficiency and flexibility.

Integration of additional data sources is planned to expand analysis capability. Measurement data taken during functional test, such as leakage current levels is of considerable value in yield analysis, but present special challenges due to its volume and changing nature. Final test data collected at overseas assembly facilities is also necessary to drive final packaged device yield improvement.

Maestro software development activities are focused on both administrative and user-targeted improvements. Efforts to streamline and enhance metabase maintainability are underway. Improvements visible to users will include enhancement of report definition interfaces, increased options for report output management, and development of additional report types.

Deployment plans call for the installation of Raptor analysis systems in additional wafer fabs. Systems to integrate data sources from multiple wafer fabs and assembly facilities are also planned for specific product groups.

Conclusion

Texas Instruments, Inc. and JJT, Inc. have partnered together to develop Raptor, an engineering analysis system which integrates multiple ORACLE and SAS data warehouses with JJT's SAS System based Maestro software for data warehouse exploitation. Raptor provides an end to end data warehouse solution which enables engineers to access, manipulate and analyze a variety of data to support yield improvement activities. TI has implemented Raptor in two wafer manufacturing fabs and plans to deploy the system to additional sites.

Acknowledgements

The authors would like to acknowledge the technical contributions of the following Raptor team members:

TI Raptor team - Mike McFadden, Jim Hallas, Barb Albrecht, Clayton Parker, Minick Rushton, Charlie Martin, Sebastian Thompson, Bernie Rivera, Steve James, Mitch McCabe, Troy Hedges, Choudor Laksminarayan, Marie Abenes, Glenn Schuette, Ed Russell, Aftikar Aslam and Stacy Ajouri.

JJT Inc development team - John T. Stokes, Tim Feeley, Linda Hiserote, Jonathan Stokes, Carlton Andrews, Kim Hiserote, Vivian Badillo, Donna Johnston and Erich Stokes.

They would also like to thank the following members of TI management for their support and encouragement: Jose Ramon, Bob Orr, Clint Montgomery, Don Russell, Venu Menon, Nai-Yung Chen, Doug Verret, and Doug Baker.

Authors

The authors are available to answer any questions at the addresses below.

Anne Brooks
Texas Instruments
13536 North Central Expressway, MS 347
Dallas, TX 75243
972.927.7795
Email: a-brooks1@ti.com
www.ti.com

Roger Thompson
JJT Inc.
912 Clermont Avenue
Dallas, TX 75223
Phone: 214.320.2200
Email: roger@jjt.com
www.jjt.com

Trademark Information