

# Data Analysis for Yield Improvement using TIBCO's Spotfire Data Analysis Software

Andrew Choo, Thorsten Saeger  
 TriQuint Semiconductor Corporation  
 2300 NE Brookwood Parkway, Hillsboro, OR 97124  
[Andrew.Choos@tqs.com](mailto:Andrew.Choos@tqs.com) (503) 625-9381, [Thorsten.Saeger@tqs.com](mailto:Thorsten.Saeger@tqs.com) (503) 615-9365

## Abstract

**TIBCO Spotfire [1] is powerful, flexible data analysis software that has wide applications in various industries, including semiconductors. At TriQuint, the software has been deployed and in use for 5 years and has seen increasing use among various engineering groups for a wide range of investigations related to design, manufacturing test, wafer fabrication and production metrics. In the paper, we describe the general features of TIBCO Spotfire and examples of analysis using this software.**

## INTRODUCTION

Data analysis is an integral part of any yield improvement activity. At TriQuint, yield improvement activities span the entire manufacturing process, from wafer fabrication to diesort test, module assembly and final test. A key component of these activities is data analysis and the tools used to analyze data. At TriQuint, the data analysis tools available include in-house developed browser based tools for quick data review, as well as stand-alone software packages such as MS-Excel, Cornerstone, JMP and MiniTab for more detailed data analysis. While these tools offered powerful analytical capabilities, they are not without limitations.

The main limitations encountered include the the amount of data that can be analyzed at a given time and the restriction of working with singular data sets in an analysis session. A singular dataset refers to data that is obtained a particular manufacturing step, such as diesort test. Relating data between different manufacturing steps for example, when working with singular datasets then becomes an onerous task that can involve using multiple data analysis files or using more than one tool to extract, manipulate and merge data so that the analysis can be performed. Situations such as this can prove to be very time consuming and results in more time spent extracting and manipulating data, than time spent analyzing the data. Ultimately, this affects the efficiency with which problems can be analyzed

and the ability to quickly drive to root cause understanding.

TIBCO's Spotfire data analysis software offers a very rich set of features and capabilities that very elegantly overcome the limitations experienced using other tools. This tool has been in use at TriQuint for the last five years and has found increasing use as the preferred data analysis software. With Spotfire, datasets are quickly and easily extracted, transformed, manipulated, merged and analyzed usually within a single session. In terms of the amount of data that can be worked with in a session, it is not uncommon to extract and work with datasets that are greater than 2 million rows of data at a time. Compare this to the Microsoft Excel's sixty five thousand row limitation and the difference becomes apparent quickly.

The results of the analysis in a Spotfire session, whether intermediate data tables or summaries, are easily exported as data files or summarized as presentation ready slides in Microsoft PowerPoint. The efficiency realized through Spotfire has resulted in significant time savings in analyzing complex problems and understanding issues affecting yields. It has led to improved yields and the gains in lost dollars, has more than paid for the costs related to licensing and implementation of this software package.

## OVERVIEW

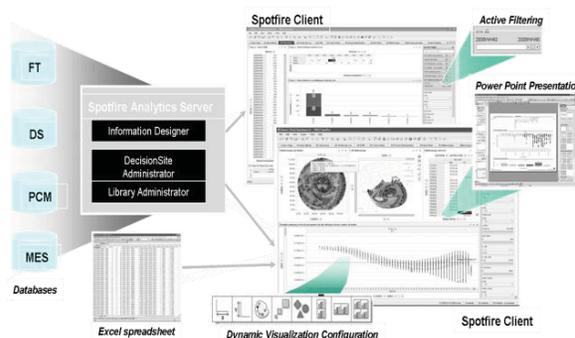


Fig 1. Overview of TIBCO Spotfire

An overview of the TIBCO Spotfire software is shown in Fig. 1. The software package is an integrated server-client based implementation that supports various hardware platforms with Solaris, Linux or Microsoft Server operating systems on the server side, and standard PC's running Microsoft Windows for the clients.

Users with administration privileges access the server through a custom interface on a web browser, called Information Designer, to configure connections to data sources and the information model library, with which customized scripts called information links are created to extract data. The information model library provides a simplified virtual view to users of the often complex database schema of configured data sources. The information links behave as scripts that are executed during an analysis session and are submitted to the data sources as SQL queries, and are stored on the server so that they are accessible to all users. Users do have the option of also saving customized information links locally on their PC's. One very important benefit of the Spotfire information model approach is that users do not need to understand database schemas or know how to write functional error free SQL queries in order to get the data they need.

All information links and any data analysis files that are stored on the server is managed using the Library Administrator tool, which like Information Designer, is accessed using a web browser by a user with administration rights. The library manager also allows for the creation of libraries of data analysis files that can be grouped by application in a hierarchical tree of folders and sub-folders. With analysis files stored on the server, it becomes possible to promote consistency in data analysis methodology as well as sharing between teams of users working on a problem.

Apart from preconfigured data sources defined with the Information Designer, Spotfire also allows for users to import data into a client session from Microsoft Excel spreadsheets and tab delimited text files using nothing more than drag-and-drop actions within Windows. Multiple data files can be imported into a session and linked as needed using common data fields to enable analysis across disparate datasets.

Users access the server through client programs, using client that are installed on the user's PC's. The client interface present users with a flexible, feature rich environment to create charts or visualizations of the data such as box plots, bar charts, scatter plots, pie charts and line charts. Though the

available types of charts seem limited especially when compared to other tools, the functions built into these charts make it possible to associate different data fields with chart attributes that expands the amount of information that can be incorporated into a chart. These attributes include coloring, sizing and shaping data points or markers by different data fields and hierarchical grouping of data on either chart axis. Another very powerful feature built into all visualizations is the use of custom expression to perform computations on data within a chart, without having to create new data columns in a dataset.

One key feature of the client interface is the active filtering panel. This provides a user with a view of all the data columns and the ability to dynamically filter what data is shown on all visualizations. Users have a choice to selecting different methods of filtering a data field, such as specifying ranges, using check boxes or selecting unique data values one at a time using a slider bar.

At the conclusion of an analysis, or even to capture intermediate results, the client software provides functions to export visualizations to Microsoft PowerPoint for presentations ready slides, or data tables and summaries as text files which can then be used for future analysis or incorporation into other analysis files.

## DEPLOYMENT

At TriQuint, Spotfire deployment started in our Richardson, Texas facility in 2004 with an initial license purchase for 5 users after a rigorous evaluation of the software's capability was conducted. The evaluation included benchmarking data extraction and analysis capabilities of Spotfire against existing tools. The evaluation also included a reviewing deployment costs and return on investment analysis.

In 2005, the license package for Spotfire was upgraded to site licenses at TriQuint's Texas and Oregon sites to support the increasing use of Spotfire in the manufacturing operations groups. In 2008, Spotfire released a new client called DXP that incorporated added features and capabilities. With this release, new data analysis techniques became possible and use of Spotfire further increased to other groups. In 2009, the license package was renewed and expanded to an enterprise license to cover users in all TriQuint sites in the United States, Europe and Costa Rica.

Spotfire is now a standard data analysis tool that is used across sites and enables rapid data sharing and joint analysis by teams working out of different sites around the world.

Deployment of TIBCO Spotfire involves setting up a dedicated server at each site, integration and installation of the server Spotfire server software, configuration of data sources and creation of the information model library for accessing data, setting up user accounts. Once the server setup is complete, end users are required to install client software on their personal computers and access the server using their Spotfire accounts. With the server and client installations completed, training for the users commences using pre-purchased training sessions conducted by consultants from TIBCO Spotfire.

## **DATA ANALYSIS TECHNIQUES**

One of the most powerful techniques that is easily implemented in Spotfire is data drill down. This is a data analysis technique that employs a process of analyzing data sequentially in steps, with the data viewed in each step used as criteria to view more detailed data in the next step. This technique provides for an efficient, disciplined approach for analyzing data using large multivariate datasets.

Drill down within a data set is possible if the data extract is large enough to make the drill down possible and meaningful. The main disadvantage of this approach is that in order to make the data drill down efficient, the initial data extract needs to be large to contain as much related information as possible, and can take a significant amount of time to extract from a database. The amount of time needed will increase when more than one dataset is being extracted to support the data analysis. This approach does not support fast and efficient data analysis, especially in working critical problems due to the heavy time requirements. This does not mean that this method does not have its applications. It is useful for data analysis of data that occurs over a longer time period, such as weekly reports, and the need for the data to be contained within a file and does not need constant refreshing to obtain the most up-to-date data.

The second approach to data drill down is to start with extracted dataset that fulfills a high level summary and analysis. Based on a selected subset of this initial data extract, more data is extracted as needed. This is then repeated as many times as is needed for the analysis. This technique relies on

multiple data extracts occurring sequentially in real time as the analysis proceeds, with each subsequent data extract being more detailed than the previous steps and is based on criteria imposed by the previous step. The benefit of this approach is that data extracts are small for the initial steps in the analysis and fast. This makes it possible to review and analyze data in real time, such as in working meetings.

With the introduction of TIBCO Spotfire's DXP client, the concept of data drill down is very easily implemented in a data analysis file and led to the creation of specialized data analysis files stored on the Spotfire server, called Dashboards at TriQuint. Dashboards are built to provide a general purpose approach in data analysis methodology, as well as support different data drill down paths depending on the needs of the user. These also serve as a basis for customized data drill downs at any point in the analysis process to extract, link and quickly perform correlation analysis across different datasets. The flexibility of the Spotfire environment also allows users to easily create new visualizations based on the extracted data or link to new data using Information Links as needed.

## **APPLICATIONS**

Use of TIBCO Spotfire at TriQuint has generally followed one of two paths. The first is the creation and deploying on the server of dashboards that are accessible by users from a library, as in the case of dashboards. The second approach results in customized analysis files that are created by users. These customized files can be stored on the server, shared network folder or on the user's computer. Both methods allow multiple users to share data and collaborate on the data analysis.

The main difference between the two types of analysis files is the scope of users. As mentioned earlier, dashboards are more general purpose in nature and therefore used by a wider group of users for broad based analysis with drill downs to detail data as needed. The customized analysis file approach is generally more focused on specific products or issues and is used by small teams working together or individually. For general purpose analyses, dashboards were built to enable analysis on different types of production test and metrics data. Production test includes Process Control Monitor (PCM) on wafers while in process in the Fab, Diesort (DS) of finished wafers and Final Test (FT) of modules. These types of analysis help drive yield improvement activities and

investigations into problems in product performance or wafer fabrication. Metrics includes wafer scraps, lot holds and known good die usage in modules.

### Example 1: Final Test Yield Analysis

An example of a dashboard is one used to review final test yields and drill downs to understand parametric yield inhibitors, during weekly yield meetings. This dashboard is run in live mode during the meeting and comprises of 4 pages of charts with each page setup as a drill down from the previous page.

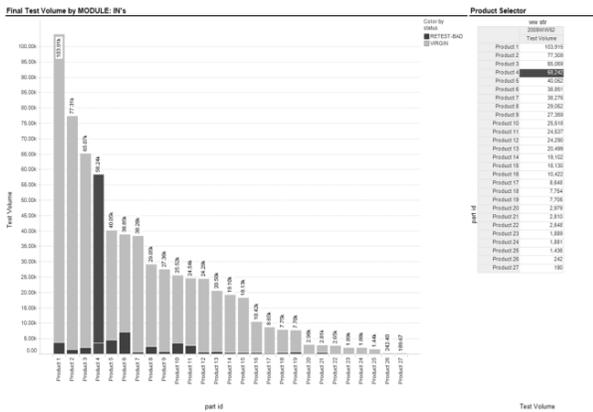


Fig 2. Yield Dashboard Page 1 showing ranked order of product test volumes.

The first page of the dashboard shows a summary of product test volume for the week with products in ranked order from highest to lowest volume, as in a Pareto chart. A product selector panel is used to select the product of interest. In fig. 2, product 4 is selected. This sets the data to be viewed in the next step of the data drill down in the next page, shown in fig. 3.

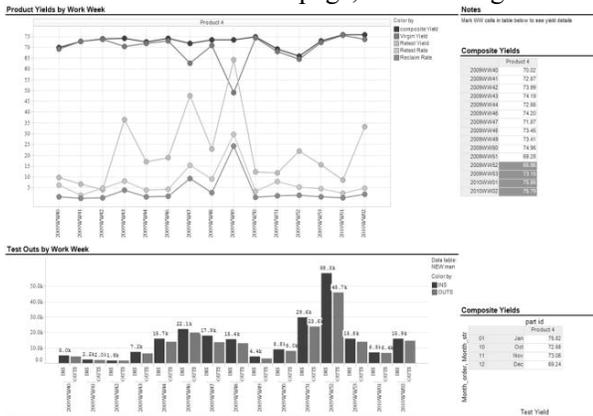


Fig 3. Yield Dashboard Page 2 showing yields for the selected product.

The second page of the yield dashboard shows yield and volume summaries for product 4. In the Composite Yields tables, yields are calculated on the basis of work weeks and by month. The expanded yields are shown in the Product Yields by Work Week line chart. Five yield lines are calculated as follows:

- Overall composite yields for the week, which is a roll up of the total yields for all tested and passing parts through the initial (Virgin) and re-test (RETEST-BAD) test steps.
- Initial (Virgin) test yields for each week.
- Re-test yields of failing parts at Virgin test for each week.
- Retest rate of tested parts for each week. This is the ratio of the number of failing parts to the number of parts at initial test.
- Reclaim rate of failing parts for each week. This is the ratio of number of passing parts at re-test to the number of parts at initial (Virgin) test.

The fourth chart is a bar chart summary of test volume trends for product 4. To drill down to the next page, the time range of four most recent work weeks is selected in the Composite Yields summary table. The detailed results are shown in page 3 of this dashboard.

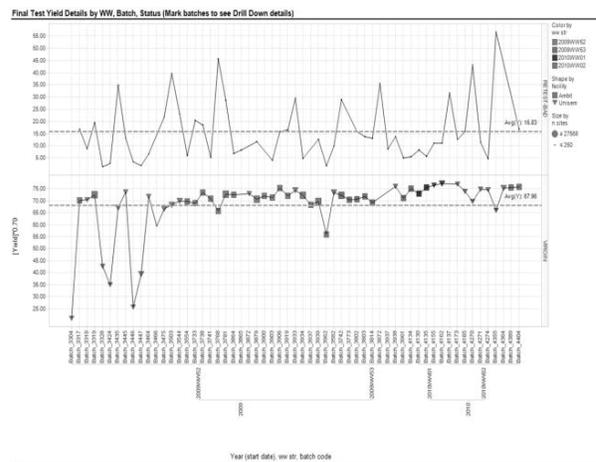


Fig 4. Yield Dashboard Page 3 of showing product yields by lot at Virgin and RETEST-BAD test steps.

The third page of the yield dashboard shows a detailed summary of individual lot yields for product 4 at 2 test steps for the time frame selected in page 2.

Virgin test is the initial test of the lot and RETEST-BAD is the retest of failing parts from the initial test. The time frame selected in page 2 of the dashboard is also used to set the data drill down summary shown in pages 4 and 5.

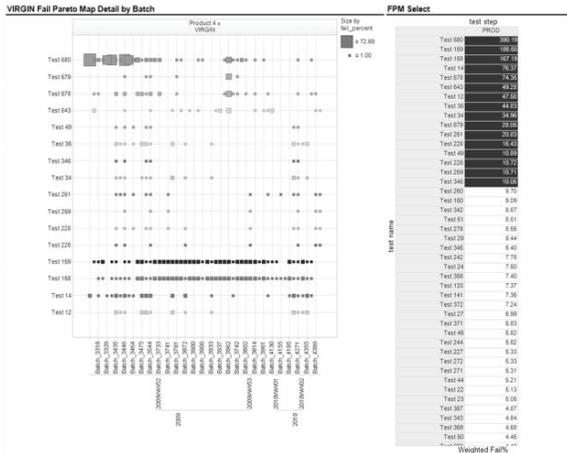


Fig 5. Yield Dashboard Page 4 showing the Virgin Fail Pareto Map for product 4 over 4 weeks.

The fourth page of the yield dashboard is shown in Fig 5. The fail pareto map is a TriQuint invention that adapts the standard fail pareto chart into a 2-D scatter plot format into Spotfire and uses its features to incorporate failure rate by each test into the marker size for each test on each batch of parts. The displayed information can span many batches of modules, representing several tens of thousands of parts. With the ability to actively filter displayed failure rates, the fail Pareto map allows for a very quick visual view of yield inhibiting test parameters and failure trends at production test of parts representing anything from days to weeks. In the figure shown above, the tests are ranked by failure percentage in the FPM Select summary table. The top failing sixteen parameters are selected and the details are shown on the scatter plot on the left. Only the failure rates greater than 1% is shown. The horizontal trend lines for Test 159 and Test 158 show that there is a consistent low level of yield loss to these two tests. For Test 680, the large markers for several lots in the early part of work week 52 indicate an excursion in yield loss to this test. This information is used to drive a yield improvement investigation by the responsible engineering teams to understand why this occurred and what corrections are needed to prevent future occurrences.

### Example 2: Fab Tasks

Due to its capability to connect to multiple data sources Spotfire has proven to be very valuable for process engineers. It is used for many applications like process control, throughput optimizations or yield and failure analysis. The examples shown below are just a small set of all the different ways Spotfire is used in Fab at TriQuint Oregon.

The chart in Fig. 6 represents a snapshot of lots from one of TriQuint’s HBT technologies how they move through the wafer fab. The abscissa shows the move in time stamp of a lot at an operation which is plotted on the ordinate. By connecting the data points each lot shows up as a line.

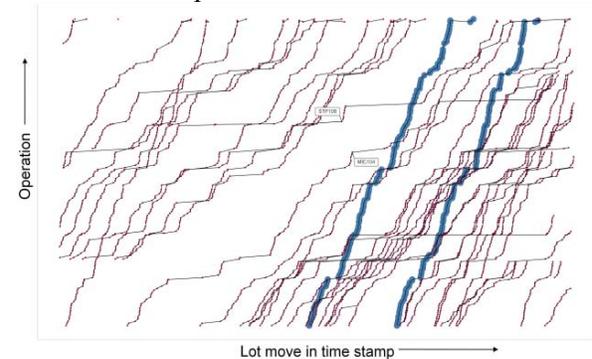


Fig 6. Lot Movement Waterfall plot.

The slope of each line represents the average speed a lot is moving through the wafer fab. As one will notice, horizontal lines mean a delay at a tool; gaps between lines represent idle times. Industrial engineers use this visualization to optimize throughput. The data set includes meta information, e.g. tools, operators, etc which can be visualized through labels, data point shape, size or color. As an example two labels are added with tool identifiers. Additional data can be added and visualized the same way. Examples are PCM test results for specific test, die sort yield information or special attributes like engineering splits and alike. The two emphasized lots have been such splits. As one can see, their slope is more steep compared to the other lots indicating as quicker turn in fab.

Attributive information can also be used to separate data into trellises. As an example, this function is used for trend charts for a daily test of a standard resistor on four inline parametric testers on two different locations.

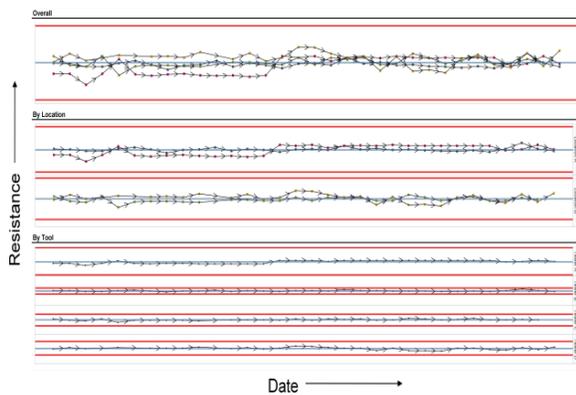


Fig 7. Parametric test trend chart.

The top chart shown in Fig 7. combines the result of the daily test in one trend chart. With the help of this visualization the test engineer can immediately assess the matching of all four testers. The control limits shown are calculated dynamically using the “customs expressions” functionality of Spotfire. The algorithm used is described in [2] and allows the detection of an out-of-control condition of a tester quickly.

The same data set can be separated into location (middle chart) and the individual tester (lower chart) using the trellis function. The control limits are recalculated based on this sub group. The middle chart is used to detect if environmental changes like temperature and humidity causes an effect on test results. The gap between the control limits of the lower charts shows the test engineer right away which tester has the best or worst repeatability. While the gap is approximately the same for Tester C and D, Tester B shows the smallest and therefore best repeatability.

### Example 3: Diesort Wafermap Analysis

We have also used the power of Spotfire’s visualization capabilities to analyze diesort data, especially wafermaps and the spatial correlation of test parameters to actual die positions on wafers. An example is shown in Fig 8. A scatter plot is used to plot the measured data of a wafer at diesort test. Where markers in the scatter plot are set to be colored by bins and the result shows a large area of high currents starting at the center of the wafer and extending to the left. The histogram of the measured parameter is shown with lines for mean and +/- 3 standard deviations for the data. By selecting the upper

distribution of the histogram, shown as the very dark colored markers, it is easily seen where the higher current die is located on the wafermap.

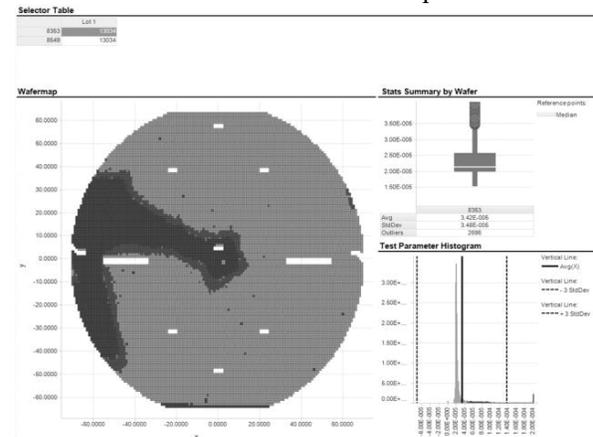


Fig 8. Diesort Test Wafermap and Histogram

### Conclusion

The TIBCO Spotfire software package is powerful and flexible from a user’s perspective and enables quick and effective analysis of large amounts of data. The package is also flexible enough through its data import and export functions to allow it to be used in conjunction with existing data analysis tools. The seamless integration of TIBCO Spotfire and other tools gives TriQuint engineers access to a rich set of powerful data analysis tools to analyze data that is not just confined to yield improvement activities, but to other areas such as new product design characterization, process experiments and manufacturing metrics. In terms of costs, use of the software has results in significant time savings in analyzing problems and enable yield improvement efforts at TriQuint with subsequent realization in cost savings.

To quote a senior manager at TriQuint: “Spotfire allows us to do things that other tools wouldn’t allow us to do”.

### References

- [1] TIBCO Spotfire: <http://spotfire.tibco.com/>
- [2] Guidelines for Part Average Testing, AEC - Q001 - Rev-C, Automotive Electronics Council.