# Use of Knowledge Discovery from Wafer Fab Data for Yield Improvement

Patrick J Carroll

RFMD Fab
7628 Thorndike Rd, Greensboro, NC 27409
pcarroll@rfmd.com, 336-678-5395

**Abstract**

The process of producing RF components from integrated circuits made from compound semiconductors involves many steps. It starts with the growth of the semiconductor boule. Other steps include substrate creation, epitaxy, wafer/circuit fabrication, die singulation and assembly with other components. Each step is quite complex involving multiple operations and a large amount of important data, including: process control information, electrical test data, time of operations, tools used, conditions, batches, shifts and operators. These steps may occur at various sites around the globe. The data may be stored in various databases, using various protocols. The information may be stored in ways that may make sense to the writer of the information, not necessarily the reader of it. In addition to the information involved with making the component, there is feedback from the customer. Can properties from the very earliest steps of the process affect the performance and quality of the end product? They can and our understanding of important factors is growing all the time. It is important therefore that knowledge that can be gained from relationships in the data can become known so that it can be used by an engineer to improve performance and yield, or to understand the root cause of a problem. The use of Data Mining and Knowledge Discovery from wafer manufacturing data can be described as a four step process: Obtain & Assemble, Fix & Filter, Analyze and Test. The results of following such a process in a fab producing RF products from III-V semiconductors will be shown.

Data retrieval for almost any project is considered to be a very lengthy step but it doesn't have to be. Commercial systems can be purchased that provide a common platform from which to access data and perform the necessary data analysis. However, they can be quite costly and may be inflexible, not allowing the user to tailor the software to his/her specific needs. At RFMD, JMP® is used as a standard platform for data analysis. We therefore decided to develop a JMP® script to retrieve this data, which we call the Universal Data Query (UDQ). The chief value of the UDQ is that the user does not need to know all the intricacies of the many data tables and databases to obtain useful, properly linked data. Queries of the individual tables have already been optimized to be efficient. Special knowledge needed to transform the data into a useful form is built in. More detail on this concept will be described in the paper.

Once the data is obtained the accuracy of the data is an issue. Manufacturing data is notorious for being full of anomalies. On common problem is non-standard data entry. Fixing this usually requires special knowledge. An advantage of the UDQ is that this special knowledge can be built in. For example, the same parameter called by different names in two different fabs can be standardized. Another problem is outliers. The UDQ can handle gross outliers, known to be caused by testing errors

but there are also data that lie clearly outside the usual distribution. The danger of filtering is that important information can be lost. There, however, is also danger in not filtering data, in that important relationships can be hidden. If one wants to learn what most of the data is saying, and the analysis is to be done in semi-automated fashion, there must be outlier filtering that does not require parameter by parameter study. There are, of course, many common methods used to filter data such as excluding beyond a specific number of standard deviations or mean absolute deviations or clipping off percentiles. We use a novel outlier screen, which will be demonstrated.

Analysis of the data once obtained should be performed such that significant relationships to a particular issue are not missed. JMP®, of course has many built in data analysis techniques that can be used to reveal relationships in the data set. We however, have developed additional techniques using JMP® scripting that find the most statistically significant relationships to any issue in numeric and categorical data. In addition, JMP® scripts have been developed that daily scour the data for trends and correlation. Some of each of these will be described.

Once relationships are revealed only the first layer of the onion has been pealed. But the process brings much-needed knowledge to any improvement project. It is invaluable toward efficiently narrowing down the most likely branches of a cause-effect diagram. Sometimes the most likely cause becomes blatantly obvious from the data mining results themselves. But often further carefully designed experiments are needed to convert correlation to eventual causation.

By quickly uncovering knowledge from the data in this way and then guiding and tracking projects using a DMAIC approach, true improvements have been obtained in the manufacturing process that lead to better sustained yields. The results of using such a process in the RFMD fab are illustrated in the chart of Known Good Die (KGD) scrap rate below. Other examples will also be shown.