D. Park, et al., Int. J. of Design & Nature and Ecodynamics. Vol. 11, No. 3 (2016) 258–267

REDUCING TEST TIME FOR SELECTIVE POPULATIONS IN SEMICONDUCTOR MANUFACTURING

D. PARK, M. SCHULDENFREI & G. LEVY
Optimal+ Inc. Israel

ABSTRACT

As the semiconductor industry prepares for the Internet of Things, one of the major challenges it will face is to maintain quality levels as the volume of devices continues to grow. Semiconductor devices are moving from items of convenience (PCs) to necessity (smartphones) to mission-critical (autonomous automobiles). One aspect of manufacturing operations that can, and must change, in the face of ever-tightening quality requirements is how to test the devices that are shipped into the end market more efficiently while maintaining very high levels of quality. One of the ways to achieve these diametrically opposed goals is through the use of Big Data analytics. Semiconductor manufacturing test today is a 'one size fits all' process, with every device being made to go through the same battery of tests. Devices that initially do not pass are retested to be sure they are not bad, but what about the devices that are 'exceptionally good'? Testing devices that are so 'tight' in their tolerances that statistically they will easily pass any remaining test intended to catch marginal devices is a waste of time and manufacturing resources. Using Big Data analytics within a manufacturing environment can enable companies to establish a 'Quality Index' where every individual device can be 'scored' independently. If that device achieves a high-enough quality score, it can be 'excused' from any further testing to accelerate overall manufacturing throughput with zero impact on quality. This paper will show how semiconductor companies today are putting Big Data solutions in place to improve overall product quality and simultaneously reducing their manufacturing costs by using data they already have in their possession.

Keywords: automobiles, manufacturing, quality, semiconductor, test.

1 INTRODUCTION

Semiconductors (a.k.a. integrated circuits or silicon chips) are at the core of just about every electronic device manufactured today. They provide logic, communications, power management and many other capabilities that are key to the tools we use every day. The range of products that are dependent on semiconductors is growing all the time and includes computers, tablets, cell phones, medical devices, cars (particularly in self-driving cars), aircraft and so on. To enable these applications, semiconductor companies manufacture hundreds of billions of chips a year.

Yet, although the IT world is a-buzz with 'Big Data', it has been slow to arrive into the arena of semiconductor manufacturing. On the face of it, this is quite surprising because semiconductor manufacturing has many facets that make it ideal for Big Data applications.

In this paper, we will discuss current and future applications of Big Data in semiconductor manufacturing, with particular emphasis on the topic of quality – ensuring that electronics manufacturers get chips that work as expected.

This paper is part of the Proceedings of the International Conference on Big Data (Big Data 2016)
www.witconferences.com

© 2016 WIT Press, www.witpress.com
ISSN: 1755-7437 (paper format), ISSN: 1755-7445 (online), http://www.witpress.com/journals
DOI: 10.2495/DNE-V11-N3-258-267

2 SEMICONDUCTOR MANUFACTURING 101

2.1 Semiconductor companies

The manufacture of semiconductors is a highly specialized and complex industry, involving expensive equipment and facilities and a high level of expertise. This has caused semiconductor companies to evolve into four broad categories.

Fabless: Companies who design chips but outsource manufacturing to third-party companies. (e.g. Qualcomm)

Foundries: Companies who manufacture semiconductor devices for fabless customers. (e.g. TSMC)

OutSourced Assembly and Tests (OSATs): Companies who package silicon devices made by foundries (typically into the black plastic casing we are all familiar with) and test the devices prior to shipping to customers. (e.g. Amkor)

Integrated Device Manufacturers (IDMs): Companies who design their own chips and then manufacture them at their own facilities. (e.g. Intel)

2.2 The manufacturing, assembly and test process

This section describes the semiconductor manufacturing process. For brevity, we have kept the description simple but in fact, the complete process is much more complex and varied.

2.2.1 Wafer fabrication

Semiconductors are usually manufactured in 'wafers' – disks of silicon with a typical diameter of '20 or 30' containing hundreds or thousands of chips (known as 'die') in a grid layout. Each chip may contain millions or billions of microscopic transistors that together provide the functionality of the chip.

By convention, wafer fabrication plants (fabs) build wafers in 'lots' containing 25 wafers apiece [1].

2.2.2 Wafer sort

Due to the complexity of chip design and the minute nature of the circuitry, any minor variation in the quality of the underlying silicon or the process used to create the chip can create anomalies that can cause performance variations or even failures in the chip's functionality. For this reason, once a silicon wafer has been manufactured, every chip on the wafer must be tested to ensure it functions as expected.

Since chip packaging is expensive, it is common to test the chips while still on the wafer in a process commonly known as 'wafer sort'. Testing is carried out by placing the wafers in a 'test cell' consisting of a 'tester' and a 'prober'. A tester tests each die by running a test program that exercises the functionality of the chip and tests various performance characteristics such as speed and power consumption. The prober is a robot that moves the wafer around below the tester, ensuring that every die is tested.

The test program tests each die by running electricity through various contact pads on the die and measuring the output on other pads. In a typical test program, many thousands of tests areconducted, each resulting in a numeric measurement and a pass/fail indicator.

At the end of the test process, the chip is assigned a 'bin', a numeric value analogous to a final score. If all of the measurements are within an acceptable range, the chip is given a passing score.

There may be more than one possible passing score, determined by the performance characteristics of the chip. For example, a perfect 'bin 1' score will typically be given to a chip running at optimum speed with low power consumption, while a slower chip running at a higher power consumption rate will be given a less-than-perfect bin.

Chips that fail one or more tests will be assigned a failing bin. The specific bin number can be used to categorize the specific issue that caused the chip to fail.

Often, wafers are tested multiple times. For example, a chip destined to be part of a high-warranty component in an automobile will need to be tested at different operating temperatures, to ensure that it will perform well under extremely fluctuating temperature conditions.

The final result of the wafer sort process is a 'wafer map', a digital file containing the final test result of each die, enabling the subsequent assembly process to separate good chips from chips that should be discarded.

2.2.3 Assembly

Once the wafers have been tested, they are sent to 'Assembly'. This is where the die is cut from the wafer, bad devices are discarded and good ones are packaged into a plastic housing. The assembly process also involves connecting the chip to the metal 'pins' (or legs) that will subsequently be used to connect the chip to a Printed Circuit Board (PCB).

2.2.4 Final test

Once packaged, the chips must be tested again to ensure that the packaging process did not break anything in the chip. The testing process is similar to wafer sort, and using the same type of tester. However, instead of a 'prober', a new type of equipment, a 'handler', is used to insert the packaged chip into the tester for testing and then to separate the chips that pass the test from those that fail.

Like in wafer sort, each chip is subjected to thousands of individual tests in order to determine its quality. These results are then distilled into a 'bin' that represents the final quality of the part.

2.3 Problems with the manufacturing process

2.3.1 Testing issues

Paradoxically, despite all of the advanced technology being used to build chips, the process has a surprising number of pitfalls that can impact the results. Often, problems such as build-up of dust in the tester, programming errors in the test program or incorrect handling of test equipment by operators can cause the tester to return incorrect results for some of the devices.

By applying analytics to the data collected during the test process, it is possible to detect many of these issues.

2.3.2 Supply chain issues

Fabless companies save themselves tremendous capital investments (billions of dollars) by outsourcing their manufacturing. However, by its very nature, outsourced manufacturing creates issues that make it difficult for the fabless company to guarantee the efficiency of the process and the quality of the chips that are produced. Fabless companies do not have the transparency in their suppliers' operations, which is needed to effectively manage them. Moreover, the OSATs and foundries are concerned with cutting costs and do not typically invest the effort the fabless company would expect in driving efficiency and quality.

One of the key challenges facing fabless companies is getting test data out of the supply chain fast enough so that actionable insights can be derived from the data. For example, if it takes 24 hours to

get the data and another 24 hours to analyze it, a problem with a tester may go undetected for 2 days, potentially causing the tester to incorrectly evaluate thousands of chips.

Interestingly, although IDM companies own their supply chain, they often suffer from the same transparency issues as the fabless companies. IDMs often own factories in different locations around the world and the local factory managers are protective of their methods and process. This means that engineers at the company's HQ do not necessarily really know what is going on at the factory.

3 APPLYING BIG DATA TO SEMICONDUCTOR MANUFACTURING

3.1 'Big Data' in semiconductor manufacturing

3.1.1 How big is big?

As might be apparent from the description of the manufacturing process, a lot of data is generated and collected along the way. For example, at each testing stage, each chip is subjected to thousands of electrical tests, each generating a measurement and a pass/fail indication. Each wafer may contain thousands of die so a single wafer is generating millions of measurements.

Large semiconductor manufacturers are testing millions of chips a day in multiple operations and generating tens or even hundreds of gigabytes of test data. This data has tremendous value for the semiconductor manufacturer and, therefore, the companies endeavor to collect and process it as quickly as possible.

3.1.2 Characteristics of semiconductor test data

The world of Big Data is characterized by what is referred to as the three 'V's [2] – Volume (the excessive amount of data being generated), Velocity (the speed at which it is being produced) and Variety (multiple types of data and generation points). Yet, for semiconductor manufacturing, Variety is less of an issue. The vast majority of the data collected in the test process is parametric measurement data, typically large volumes of floating point numbers. This data is highly structured and can be handled using both SQL-like techniques as well as NoSQL batch techniques.

3.1.3 Volume and velocity of data

As explained above, it is important that test data is collected and analyzed in near-real time in order to enable the manufacturer to take action on the results of the analytics. For example, it may be necessary to stop a tester and recalibrate or clean it in order to reclaim lost yield. The earlier the problem can be determined, the more effective the fix will be.

3.2 Usages for semiconductor big data

There are three primary vectors that we look at when analyzing semiconductor data:

- **Yield**: The ratio of good parts out of the total number of parts manufactured.
- **Productivity**: How effective the test process is (e.g. removing unnecessary retesting of parts or skipping tests that never fail)
- **Quality**: Ensuring that no bad parts get shipped to customers.

All of these vectors are impacted directly by test. Yield and productivity are often related. For example, dirt in the probecard used in a tester can reduce yield by causing parts to fail and at the same time it can negatively impact productivity by requiring parts to be retested multiple times.

Quality is a more complex issue because by nature, increased quality requires more testing and higher cost-of-test, as we explain in the next section.

4 REDUCING TEST TIME WITHOUT COMPROMISING QUALITY

4.1 Measuring quality

There are various ways by which quality is measured in the semiconductor industry, but perhaps the most important is 'Defective Parts per Million' (DPPM). This measure counts the number of parts that fail *after* being shipped to customers. The lower the DPPM, the fewer the customer complaints, but naturally, it comes at a price – more stringent testing that raises the cost of test.

For critical applications such as medical devices, aerospace and automotive companies strive to lower DPPM to the absolute minimum. In fact, in recent years, the target defined by automotive companies has dropped below 1 DPPM [3] and these companies now demand rates in 'Defective Parts per *Billion*' (DPPB).

Even for applications that are not mission critical, such as consumer electronics, typical DPPM targets have traditionally been in the hundreds or thousands but have now been lowered to tens or lower. Companies are more focused on quality because in this era of social media, bad publicity generated by faulty products is causing them to lose market share very quickly.

So, in the race to meet higher quality targets, companies are being forced to increase the amount of testing they are performing. However, as we will show, it is possible to *decrease* the amount of tests being conducted without sacrificing quality by leveraging Big Data.

4.2 Introducing burn-in

One of the most expensive stages in semiconductor test is known as 'burn-in' and it attempts to determine the reliability of chips over time [4]. This is where parts are placed in an oven and stressed for many hours or even days at a high temperature to simulate the wear and tear typically endured by a chip over months or years [5].

4.2.1 Burn-in reduction

Since burn-in is such a slow process, it is extremely costly. However, by correlating earlier cheaper testing processes against the results of burn-in test for large numbers of chips, it is possible to find predictors of burn-in failure and use those predictors to separate parts into two groups, those which require burn-in and those which do not, or those which need a longer burn-in testing time than others [6].

4.2.2 Parameters that impact burn-in

There are a wide number of parameters that must be taken into consideration when analyzing parts to identify predictors of burn-in failure. These include many factors such as:

- Wafer Geography – die near the edge of the wafer are generally less reliable than those in the centre of the wafer
- Die Neighborhood – die that are surrounded by large numbers of failing die on the wafer (see Fig. 1)
- Parametric Outliers – die with individual test results that are border-line and statistically significantly different than the rest of the population (see Fig. 2)

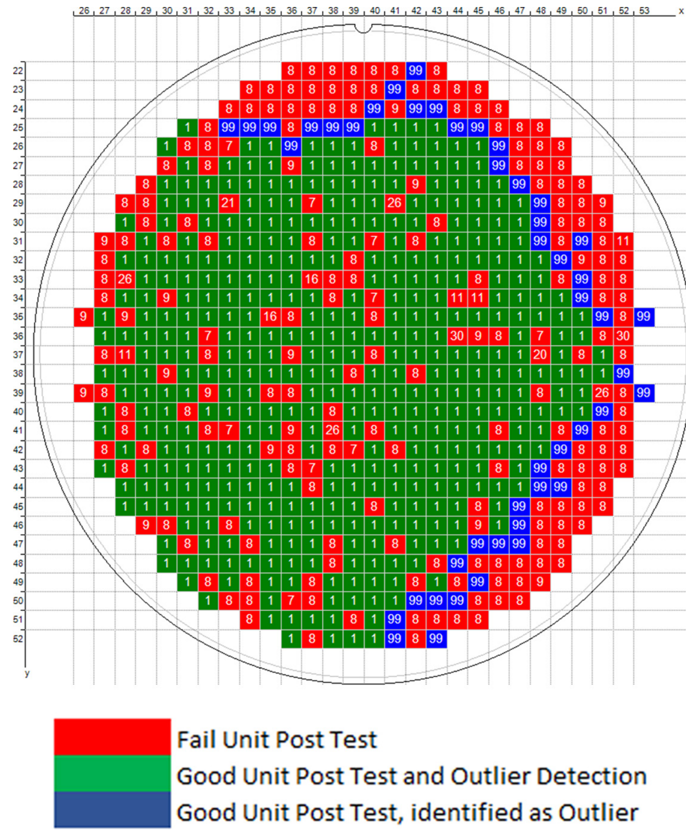

Figure 1: Geographic outlier.

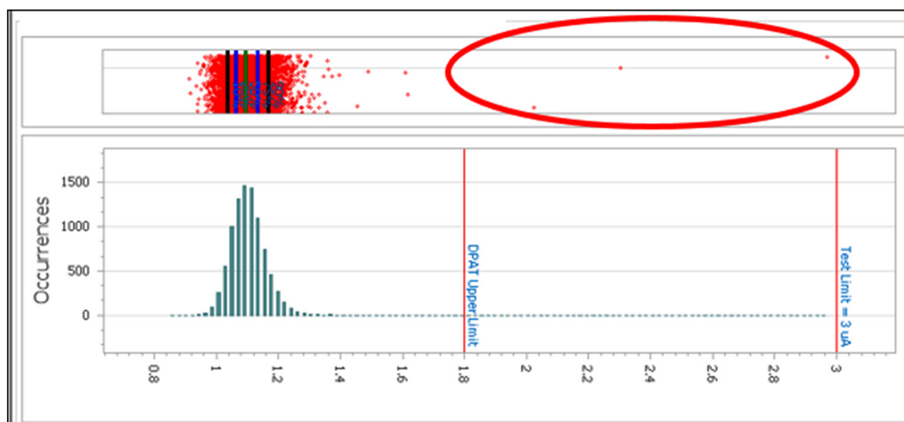

Figure 2: Parametric outliers.

- Multivariate Outliers – die where combinations of test results are statistically different than others [7]
- Low Yielding Wafers – die on wafers with unusually poor yield

4.2.3 Combinations of parameters as indicators of quality

Outliers are typically handled through implementation of industry standard outlier detection methods such as Parts Average Testing (PAT) [8]; however, this methodology looks at each type of variable individually and not at combinations of variables. If any issue is found, the part is disqualified.

For example, an engineer may define that if more than 50% of the die surrounding an individual 'good' die on a wafer are bad, the system should re-bin the die as 'bad'. Similarly, if a parametric result is more than 10 sigma away from the rest of the population, the die should be marked as bad [9].However, what about a die which has 45% bad neighbors and its parametric measurements are 9 sigma away from the population? Although it 'passed' both of the outlier detection tests described above, is it really good enough to put in a mission-critical application like a car's anti-lock braking system?

4.2.4 Quality index

The patented solution we present here is called Quality Index [10]. A Quality Index is the result of a function that applies weights to each of the various parameters described above and calculates a single value that represents the overall quality of a part. In its simplest form, the Quality Index is a simple linear expression such as:

$$QI = a_1p_1 + a_2p_2 + a_3p_3 + \dots + a_np_n$$

Where p_1, p_2, p_3, etc. are parameters like those described in the previous sections and a_1, a_2, a_3, etc. are the weights being applied to each parameter.

Using the Quality Index, we can categorize the 'goodness' of parts based on combinations of issues and catch problematic parts that do not fail individual tests.

5 TURNING THEORY INTO PRACTICE

In order to implement an end-to-end solution to reduce burn-in in production, the main challenges are firstly, how to distill all of these parameters into a method that can be used to separate highly reliable chips from suspect chips [10] and secondly, how to implement that in a high-volume, highly distributed production environment.

A solution needs to consist of several stages, including:

- Analysis – determining the correct combinations of parameters to use to predict parts that do not require burn-in
- Simulation – validating that the selected combination correctly predicts burn-in failure by testing it against large volumes of historical data
- Deployment – enabling execution of the algorithm in a reliable manner on material as it is being manufactured and tested

5.1 Analysis

Determining the relevant parameters and weights requires execution of analytical algorithms on historical data. Once the equation has been calculated, it is possible to simulate its use on a separate

set of historical data to validate its efficiency. Ideally, none of the parts with a QI indicating high quality would fail at burn-in.

Both in the analysis [11] and simulation phases, significant data quantities are required. Due to the large number of combinations of parameters, the compute resources are significant and the process as a whole lends itself very well to Big Data technologies.

5.1.1 Example: bi-variate outliers

Bi-variate outlier analysis involves looking at all combinations of tests to find pairs with high correlation, and then investigating the highly correlated pairs for outliers.

To find bi-variate outliers, our system uses entrance criteria to filter out tests that are relevant for analysis. For example, the system may filter out tests that are not continuous or have abnormal distributions. The remaining tests are then correlated against each other.

In the example in Fig. 3, the heat map shows the value of the Spearman [12] correlation between hundreds of pairs of tests. The X and Y axes contain the list of tests and the pixel colour at each intersection depicts the strength of the correlation (green is high, red/brown is low). The engineer evaluating these tests will focus on the tests with high correlation.

In Fig. 4, a very good correlation between 2 tests is shown. The highlighted unit is a bi-variate outlier. i.e. it is not an outlier for any of the tests when each test is looked at individually, but definitely is an outlier when tests are looked one vs. the other.

It is also possible to use more advanced analytics or machine learning to categorize quality. Examples of algorithms used for this include Support Vector Machines and Principal Component Analysis [13,14].The results of the analysis will be:

- A Quality Index function
- The list of parameters the function requires
- The minimum quality threshold a part must achieve to enable it to skip burn-in

5.2 Simulation

Once the Quality Index function has been defined, it must be validated against historical data using simulation. The simulator runs the algorithm against historical wafer-sort and/or final test data and compares the results to known failures from historical burn-in data. An algorithm can be considered valid if none of the parts determined by the algorithm to be good enough to skip burn-in actually failed burn-in in reality.

To satisfy DPPM requirements, many millions of parts typically need to be included in the analysis, once again making Big Data infrastructure a requirement.

5.3 Deployment

In order to use this methodology in High Volume Manufacturing (HVM), it is necessary to distribute the QI equation into the supply chain and integrate it into the HVM process in order to get relevant input parameters from all test operations.

By integrating with the Manufacturing Execution Systems (MES) of the suppliers as well as the test cells themselves, we have been able to fully automate the process, so that parts requiring burn-in are physically separated from those that do not.

A discussion of this methodology is out of the scope of this paper.

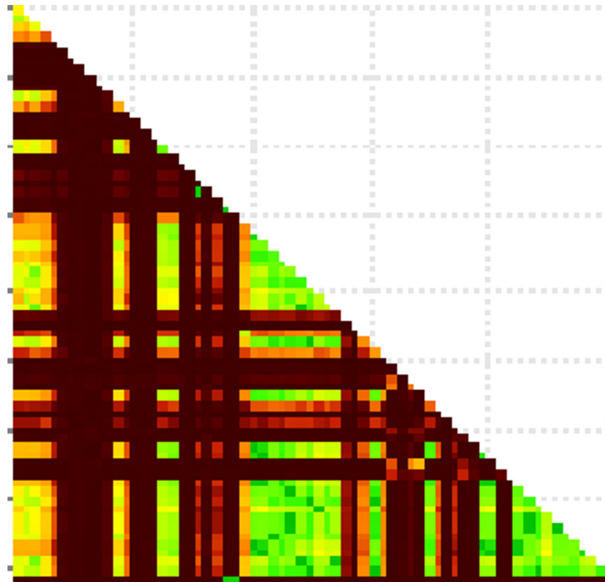

Figure 3: Heat map of correlations.

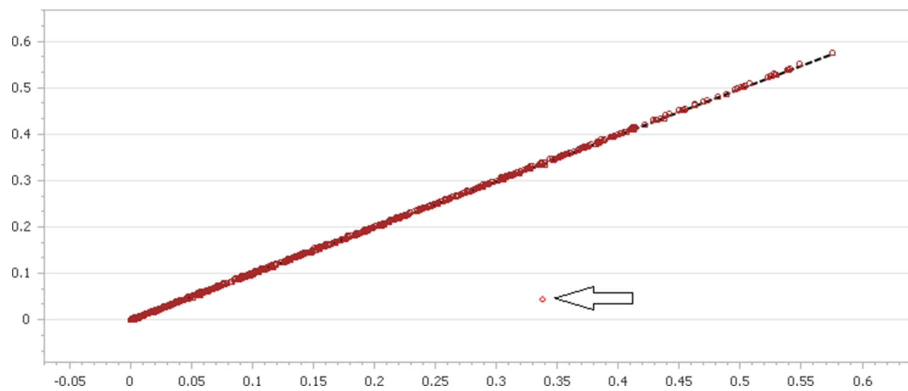

Figure 4: Bivariate outlier.

6 CONCLUSIONS

In this paper, we have shown how semiconductor companies are able to leverage Big Data analytics to dramatically reduce the cost of test, without compromising quality. The key requirement is the ability to process very large volumes of structured parametric measurement data in order to correctly predict which parts will not fail at the next test step and can therefore skip the step completely.

This novel methodology is now in production at several large semiconductor companies and they are reporting up to 80% reduction in burn-in costs.

6.1 Future research

In the examples shown in this paper, the concept of a quality index is used to grade chips based on test results and other related data. Further research in several areas will unlock more potential for increased quality and lower cost. These include:

- Defining algorithms for multivariate categorization of chip quality
- Including additional manufacturing and in-use data in the methodology
- Grading chips by quality index to target different use-cases (e.g. the highest quality parts for automotive, and lower quality for consumer electronics)
- Implementing a similar process across industries (e.g. taking data from chip to board to consumer product).

As manufacturers tap into the processing power of Big Data solutions, many more opportunities will arise to fine-tune the process and reduce cost.

REFERENCES

[1] Nishi, Y. & Doering, R., Handbook of Semiconductor Manufacturing Technology, CRC Press: Boca Raton, London and New York, pp. 33-3
[2] Laney, D., *3D Data Management: Controlling Data Volume, Variety and Velocity*, MetaGroup Research, 2001.
[3] Hyman, P.B., *IC Makers Face Hurdles in Auto Infotainment*, available at http://electronics360.globalspec.com/article/4248/ic-makers-face-hurdles-in-auto-infotainment.
[4] Tobias, P. & Trindade, D., *Applied Reliability*, Section 2 – Bathtub Curve for Failure Rates, Chapman & Hall, pp. 36–37, 1995.
[5] JESD22-A108C, *Temperature, Bias, and Operating Life*, JEDEC Standards, available at https://www.jedec.org/
[6] Nahar, A. & Daasch, R., Burn-in reduction using principal component analysis. *Intl. Test Conf*, 2005, paper 7.2.
[7] Rencher, A., *Methods of Multivariate Analysis*, 2nd edn., WILEY-INTERSCIENCE, Section 4.5.2 – Outliers in Multivariate Samples, New York, pp.101–105, 2002. http://dx.doi.org/10.1002/0471271357
[8] Automotive Electronics Council – Q001 Rev. D, *Guidelines for PAT Average Testing*.
[9] Madge, R., Rehani, M., Cota, K. & Daasch, W.R., *Statistical Post-Processing at Wafersort – An Alternative to Burn-in and a Manufacturable Solution to Test Limit Setting for Sub-Micron Technologies*, VLSI Test Symposium, pp. 69–74, 2002. http://dx.doi.org/10.1109/vts.2002.1011113
[10] Patent US7340359, *Augmenting semiconductor's devices quality and reliability*.
[11] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R., *Cross Industry Standard Process for Data Mining Methodology*, CRISP-DM 1.0 Step-by-step data mining guides.
[12] Daniel, W.W., Spearman rank correlation coefficient. *Applied Nonparametric Statistics*, 2nd edn., PWS-Kent: Boston, pp. 358–365.
[13] Timm, N., *Applied Multivariate Analysis*, Section 8 – Principal Component, Canonical Correlation, and Exploratory Factor Analysis, Springer, pp. 445–476.
[14] O'Neill, P., Production multivariate outlier detection using principal components. *Intl. Test Conf*, 2008.