Outlier detection for Non Normal and Multivariate data

AEC RW Europe, Bordeaux, October 9, 2025

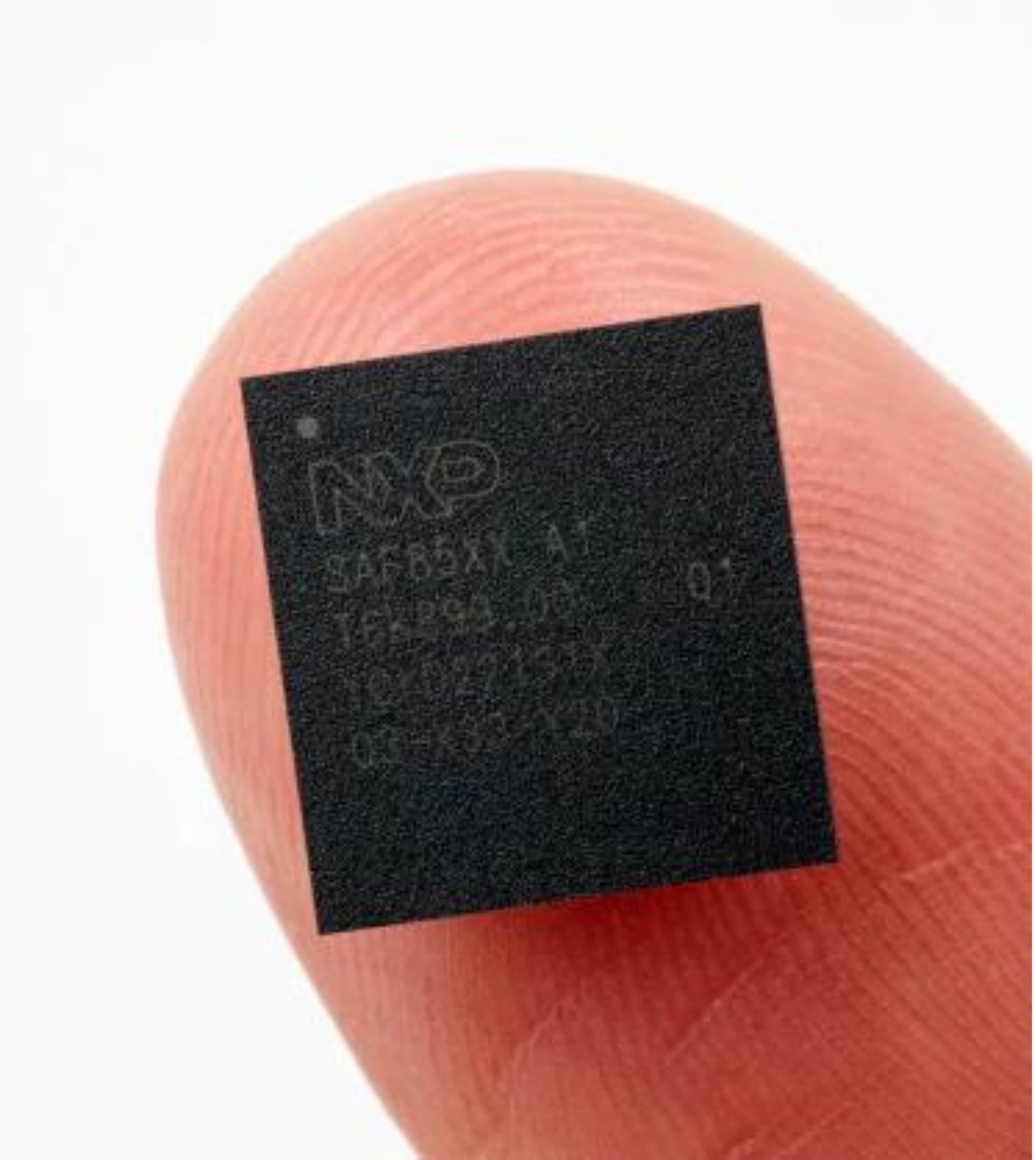
François Bourlon*, François Bergeret**

*   NXP Semiconductors, Toulouse, France
** Ippon Innovation, Toulouse, France

Presentations

- François Bergeret
  - PhD in Statistics and founder of ippon innovation
  - 50 publications and one book on industrial statistics
  - Six Sigma Black Belt

- François Bourlon
  - Senior Industrial Engineer at NXP
  - Automotive Radar Product Line
  - Six Sigma Black Belt



Agenda

Semiconductor Industry Quality challenges

Half-Sigmas, a robust metric for non-normal data

Outlier detection on **Multivariate data**

Conclusion

Introduction

Quality & Reliability: Detection of Statistical Anomalies in Electrical Data

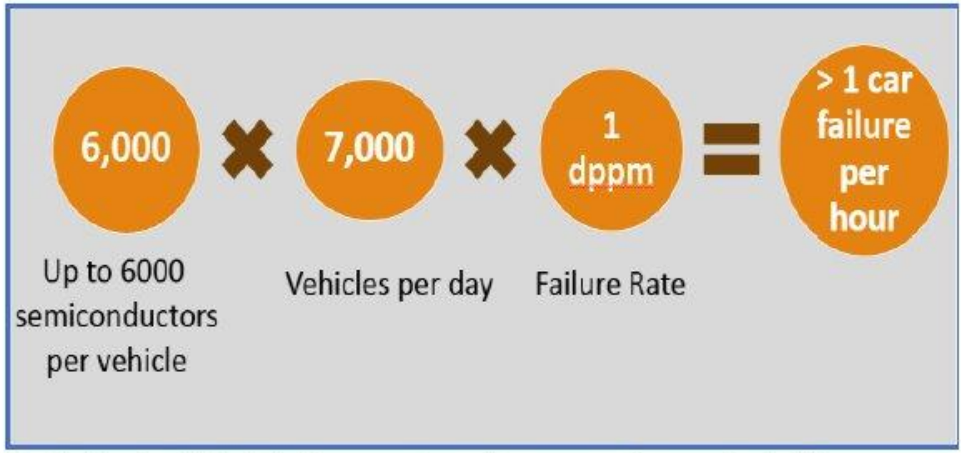
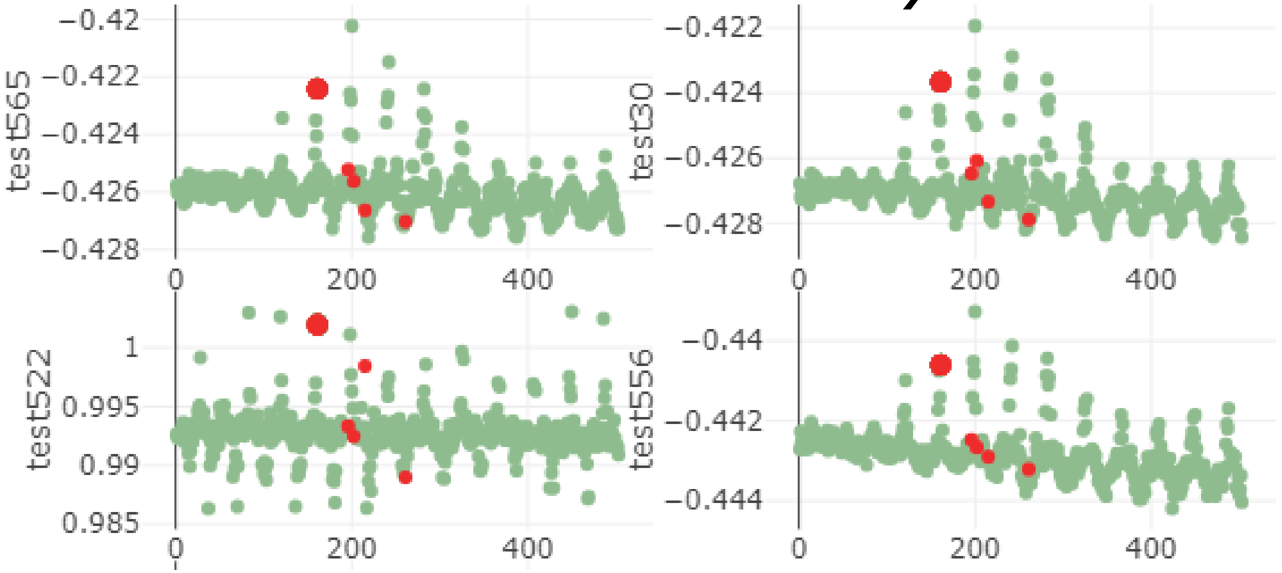


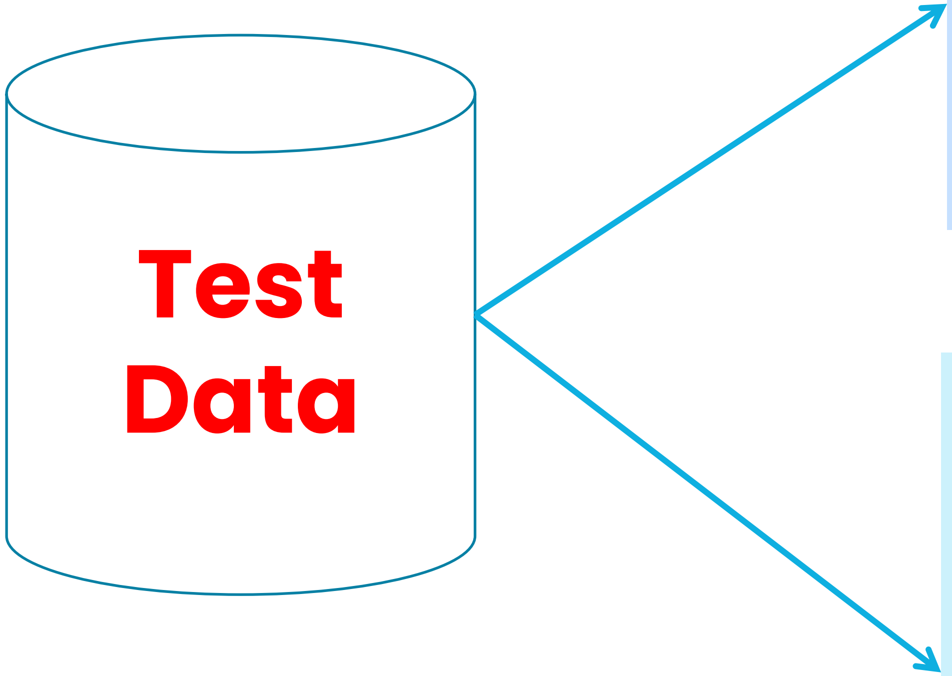
Fig 2: Single digit defective ppm is no longer good enough, 1st European AEC Reliability Workshop (Munich), BMW, October 17th, 2018

Automotive Quality target : < 1ppm

Outlier detection is key



Statistical approach



Product & Test engineers

Analyze measured values of parameters such as

-response times
-leakage currents
-power consumption
-breakdown voltages
-gains…

And compare them to targets & specs

Statisticians

Process all available data to identify anomalies

Such anomalies may be related

- to a product anomaly
-or to the test process (thus the need to clean data)

Results may provide key information

- outlier products
-product similarity (low anomaly)
-key contributors to anomalies

Classical approach for outlier detection

- PAT = Part Average Testing
  - Relies on +/- kσ, covering a given percentage of the distribution
  - Non-parametric alternative for non-normal data, using a robust sigma, proposed in AEC AEC_Q001_Rev_D
  - Based on unit test level “univariate”

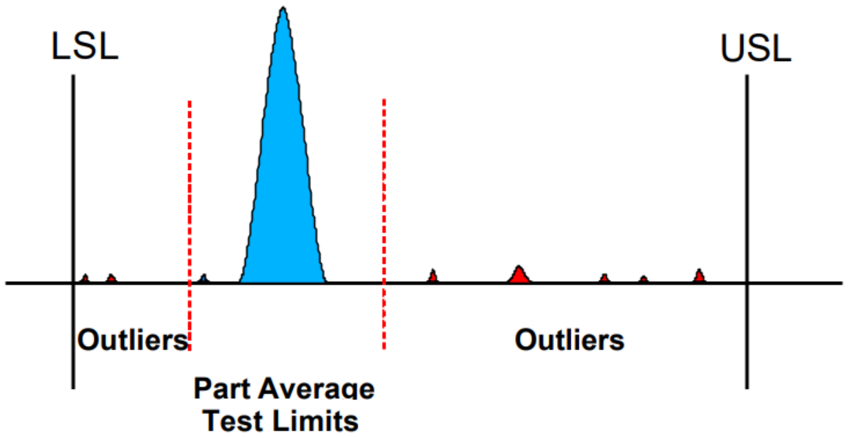


Figure 4: Graphical Representation of Part Average Test Limits and Outliers

Robust Mean = Q2 [the median]

Note 1: Q2 (Quartile 2) is the middle data point, if the sample size is an odd number. If the sample size is an even number, Q2 is the average of the two middle data points.

Robust Sigma = (Q3 - Q1) / 1.35

Note 2: The 1.35 number is inexact for sample sizes less than 20. Q1 is the point 1/4 of the way through the ranked data and Q3 is the point 3/4 the way through the ranked data.

Extracts from AEC Q001 RevD

IA and statistics for quality, yield and process control

OGAT

For Zero Defect

Multivariate outliers detection for improved reliability

YETI

Yield Optimization

Artificial Intelligence for complex processes & yield improvement

CHAM

Advanced Process Control

Identification of small signal on multivariate time serie analysis

Statistical Solutions

Software on demand

JSL, Python, R and other statistical languages coding

Data Science

ippon

Statistical expertise

On site consulting, statistical analysis, coaching & training

Improved outlier detection

Two solutions		
Distribution-agnostic methods	Can handle highly non-normal and asymmetric data	Large number of electrical tests

Two solutions

#1: **Half-Sigmas**, a robust metric for non-normal data

#2: **Multivariate method** for outlier detection

Half Standard Deviations for robust outlier screening

Outlier detection – general

- Critical in achieving zero-defect manufacturing

- Traditional methods such as Part Average Testing (PAT) often rely on the assumption of a normal distribution
  - Using thresholds based on $\pm k\sigma$ to capture a specific percentage of the data
  - However, this approach may not be suitable for non-normal distributions

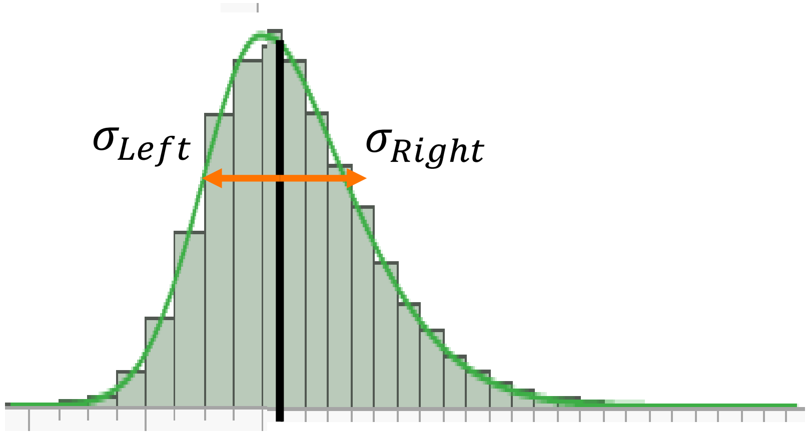
Half Moment - the idea

- Half Moment measures the dispersion on one side of a location parameter, the average for example
- By design it takes the dissymmetry into account
- There is a left half moment and a right half moment
- Underlying theory is complex, using the complex number i
- A simple alternative based on the half moment idea is the half standard deviation

Half Standard Deviation - the calculation

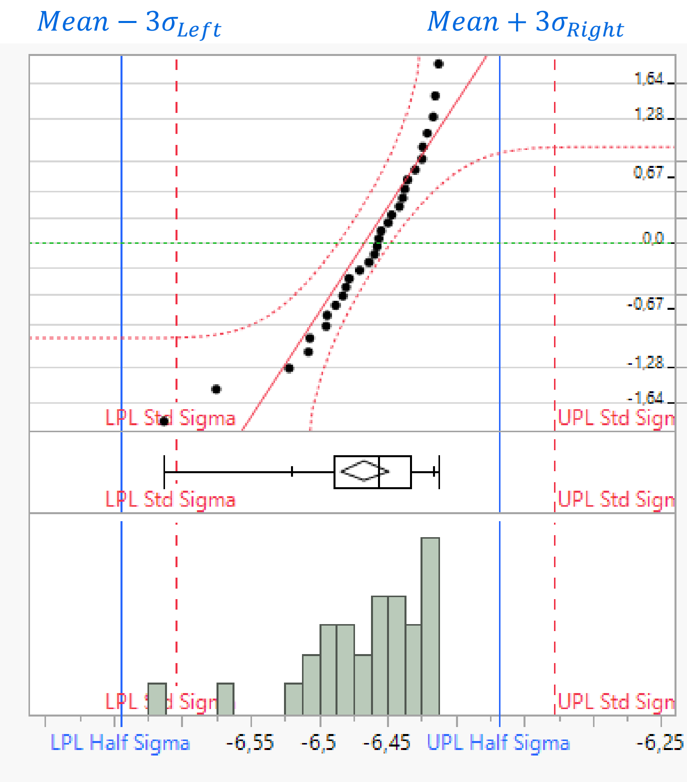
$$\sigma_{Left} = \sqrt{\frac{1}{n_L}\sum_{y_i<\bar{y}}(y_i - \bar{y})^2}$$

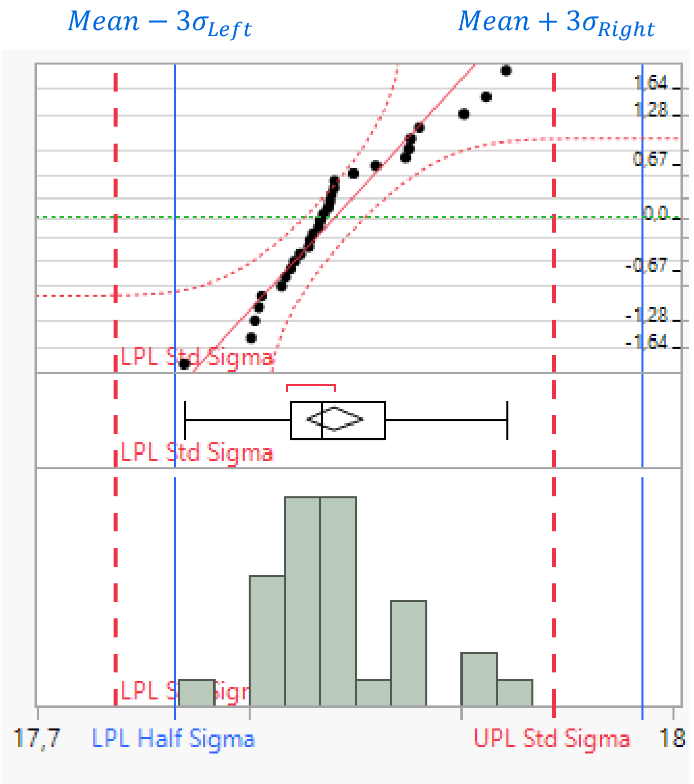
$$\sigma_{Right} = \sqrt{\frac{1}{n_R}\sum_{y_i\geq\bar{y}}(y_i - \bar{y})^2}$$

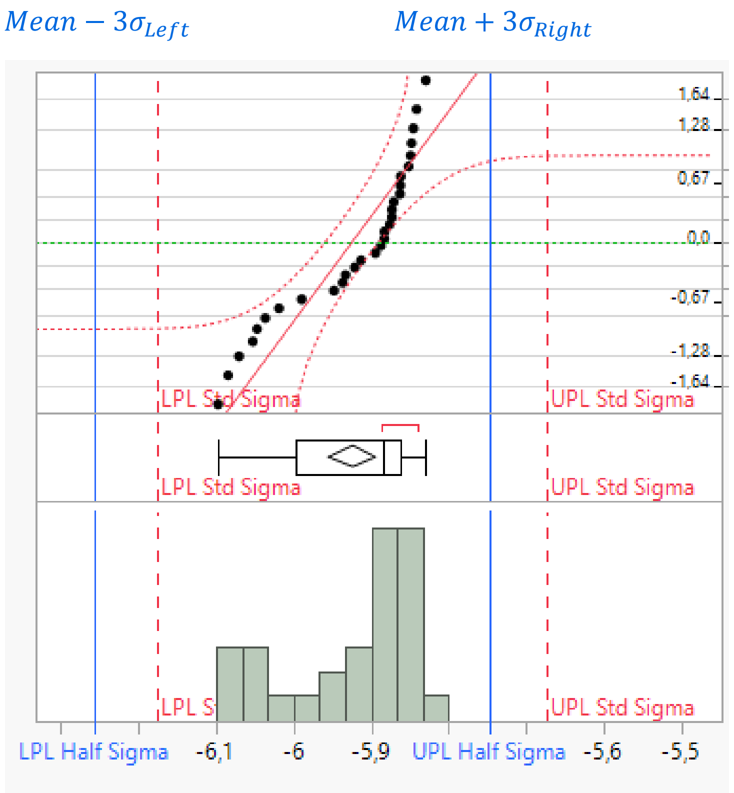


Each Half-Sigma measures the dispersion of the distribution on each side

Half Standard Deviation - examples







Half-sigma « **adheres** » to the distribution dissymetry → **enhanced robustness** for Capability assessment and outlier detection

Outlier detection on Multivariate data

Research work

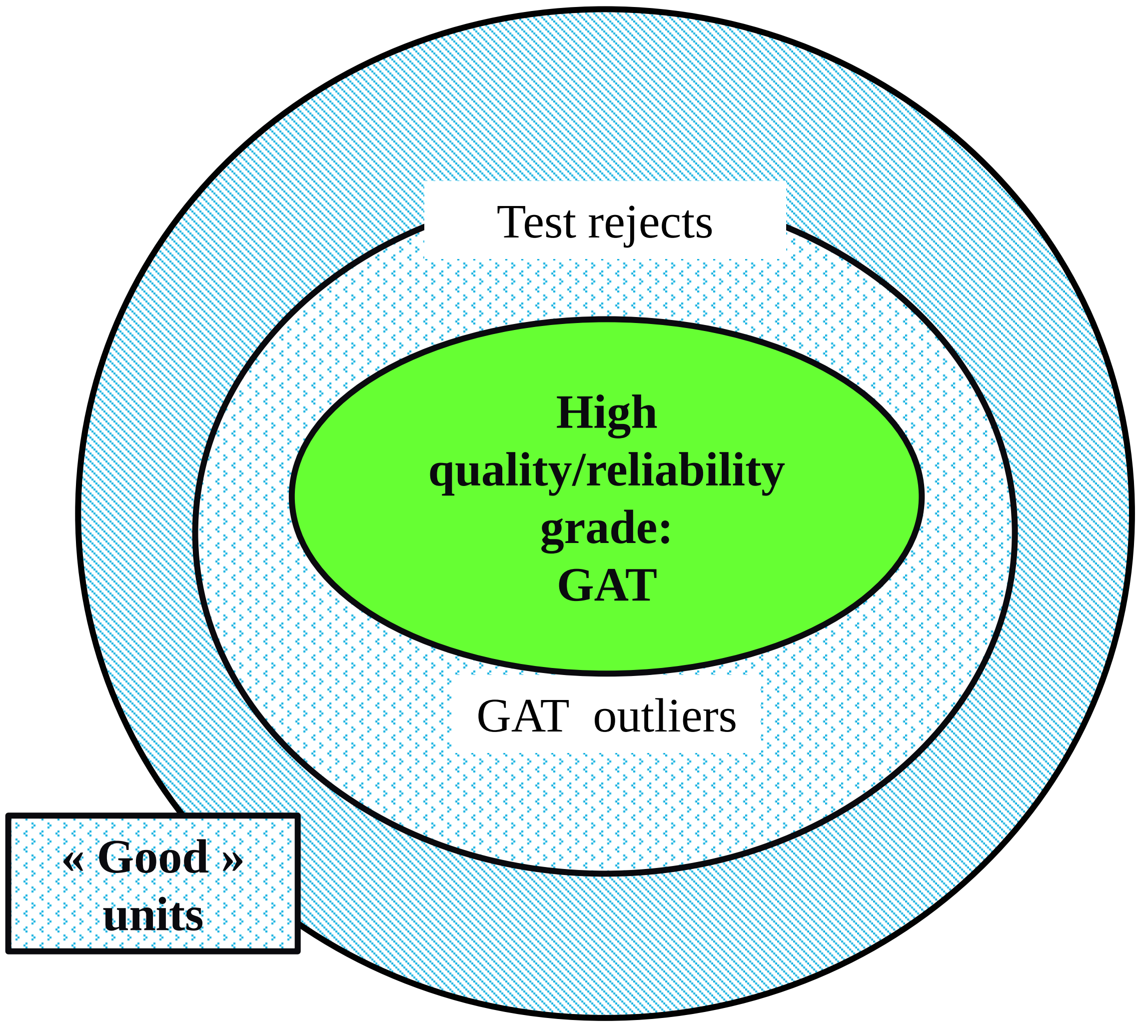
- Development
  - GAT was developed in collaboration with Toulouse University and Institute of Mathematics
  - It was the subject of a doctorate thesis

- Qualification
  - It was tested, optimized and qualified with a semiconductor industrial partner*as a part of *RESIST*
  - *RESIST* means *RESilient Integrated SysTems;* it was a 3 years European project

* Microchip: many thanks to Sophie D'Alberto, Christian Bonnin and Microchip managers for this project

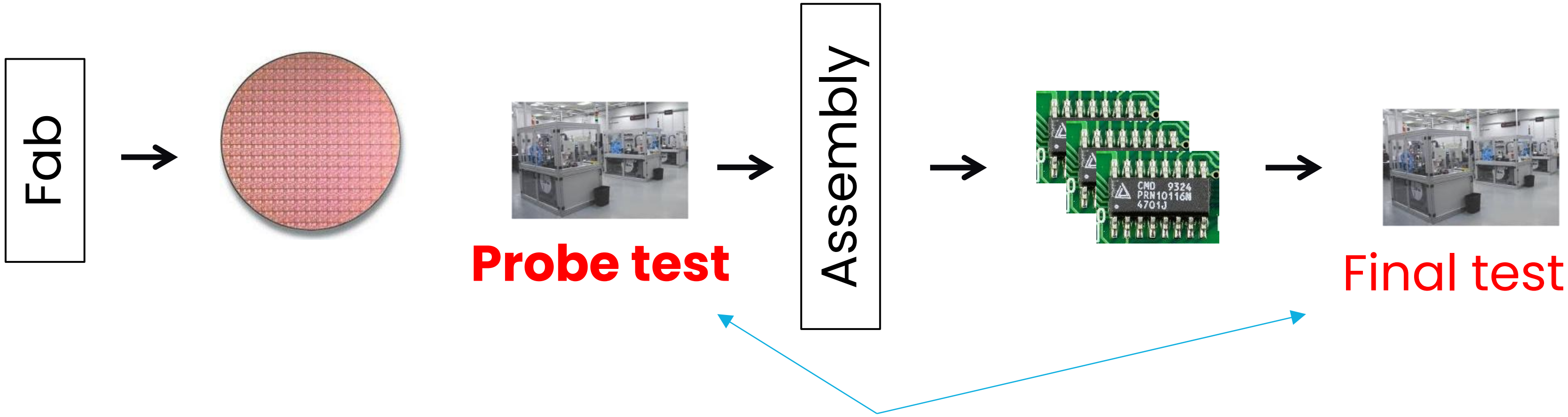
Screening in Aerospace & Automotive

We need *advanced statistical tools* for zero defect and high reliability



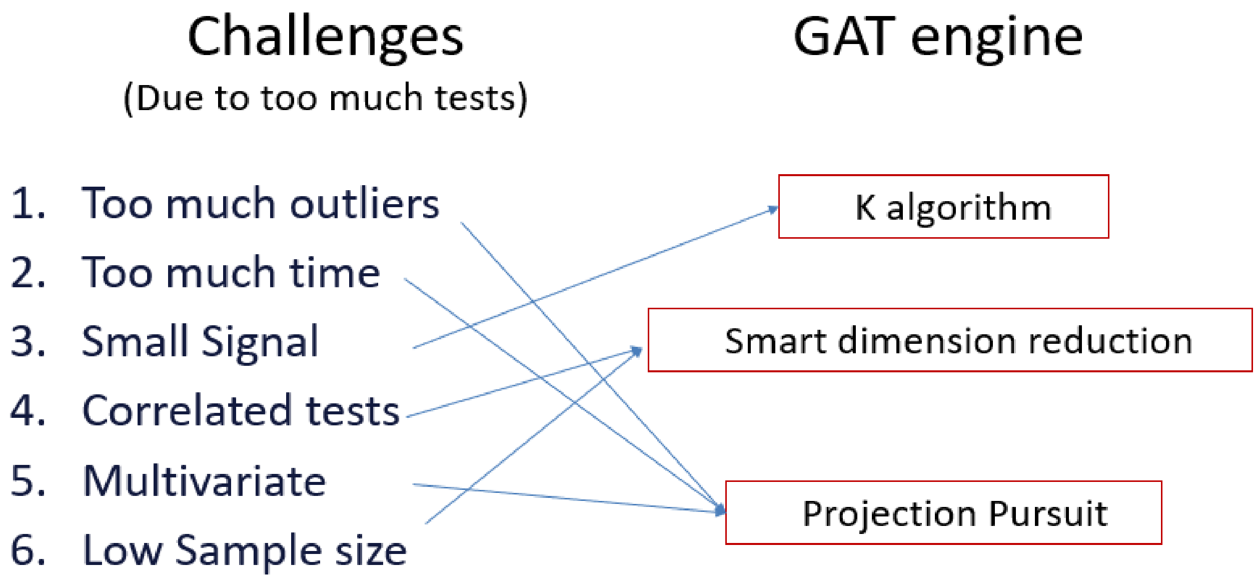
Prerequisites for standard usage

- GAT is processed after electrical testing of all units
- **There is a data cleaning step first: very important!**
- GAT is applied on units that ‘pass’ all electrical tests
- GAT requires full traceability of each unit



Can be applied at several test levels
to detect failures as early as possible

Zero Defect and microelectronics challenges



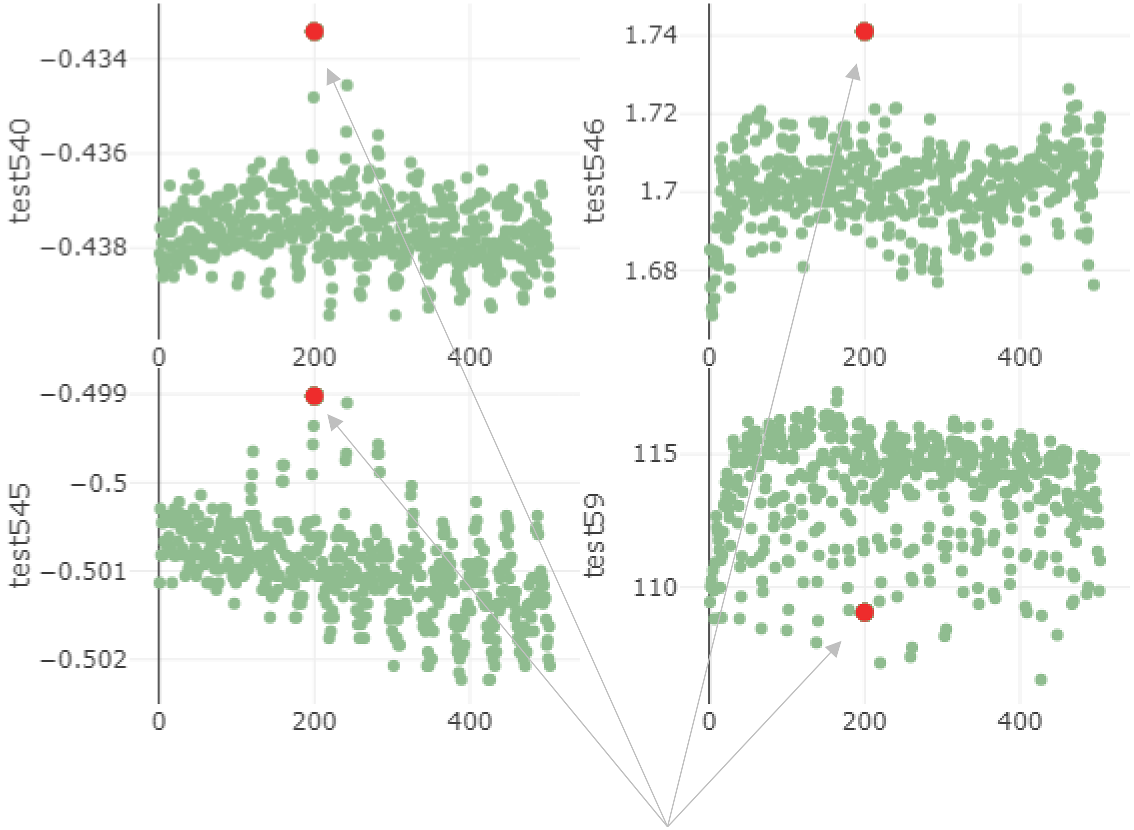
GAT algorithm has a 3-levels engine

1. **K-algorithm for a first screening**
2. **Smart dimension reduction preprocessing**
3. **Projection Pursuit to find final outliers when there are many tests**

The 3 components of the algorithm address the 6 challenges of zero-defect screening

Illustration of the "K-algorithm"

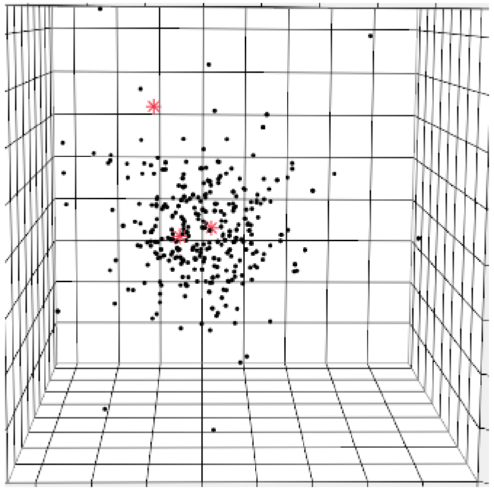
A part statistically abnormal: a potential quality or reliability issue



Usually, a "K-outlier" is marginal on *several* tests: anomalies are cumulated to provide a first anomaly score

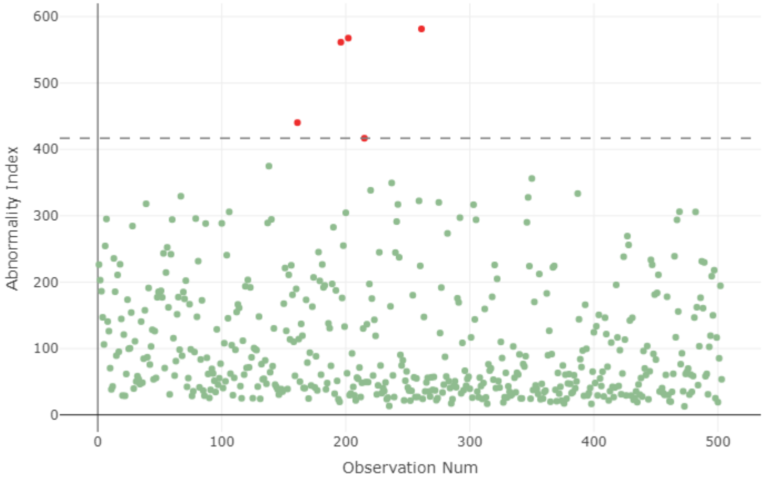
Illustration of the projection pursuit

The « best » projection



High p-dimensional data
(represented here in dimension 3)

Projection in 1-dimension



Each observation is projected from a p-dimensional space to a 1-dimensional space

$$P(x_i) = \alpha_1 x_i^1 + \alpha_2 x_i^2 + \cdots + \alpha_p x_i^p$$

Projection Pursuit vs other multivariate methods

Proposition 1. *Assume that q remains fixed and p tends to infinity, then under model (1):*

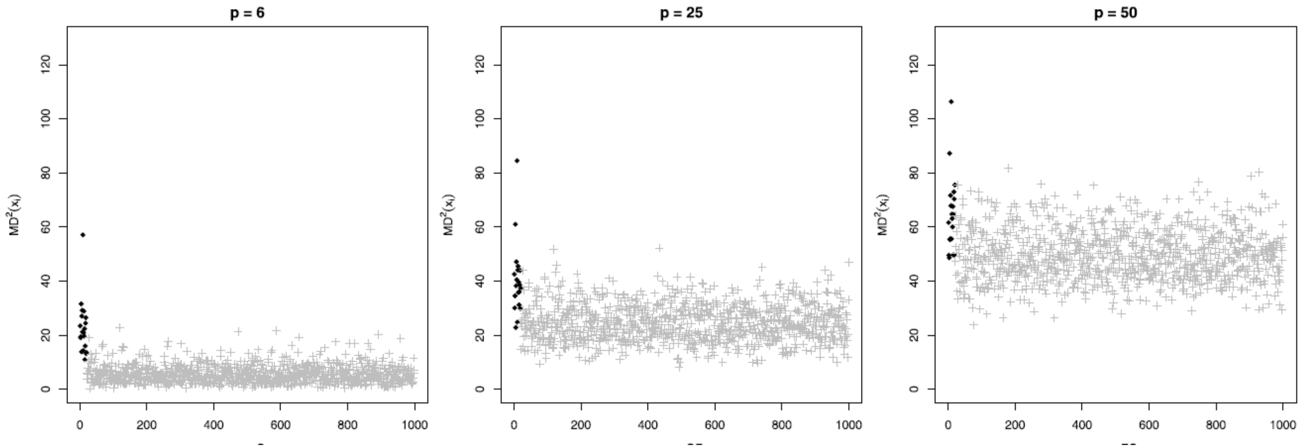
$$\frac{1}{2\sqrt{p}}\left(d^2(\mathbf{X}_{o,h}) - d^2(\mathbf{X}_{no}) - \mathbb{E}\left(d^2(\mathbf{X}_{o,h}) - d^2(\mathbf{X}_{no})\right)\right)$$

and

$$\frac{1}{2\sqrt{p}}\left(d_R^2(\mathbf{X}_{o,h}) - d_R^2(\mathbf{X}_{no}) - \mathbb{E}\left(d_R^2(\mathbf{X}_{o,h}) - d_R^2(\mathbf{X}_{no})\right)\right)$$

converge in distribution to a standard Gaussian distribution and the expectations $\mathbb{E}\left(d^2(\mathbf{X}_{o,h}) - d^2(\mathbf{X}_{no})\right)$ *and* $\mathbb{E}(d_R^2(\mathbf{X}_{o,h}) - d_R^2(\mathbf{X}_{no}))$ *do not depend on p.*

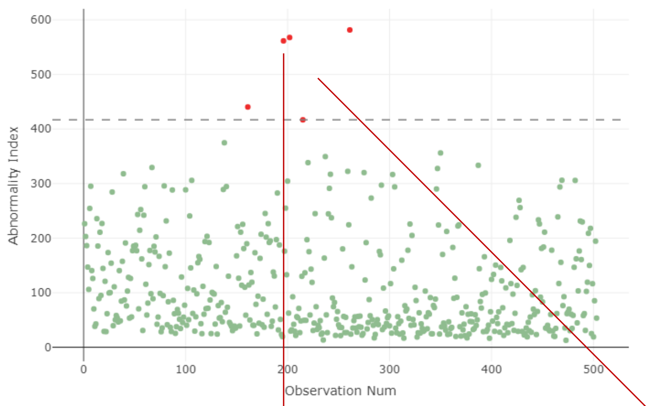
Proof that the Mahalanobis distance & Hotelling T^2 perform poorly when the number of tests (p) increases and when p is large compared to the dimension of the susbpace where the observations are outlying



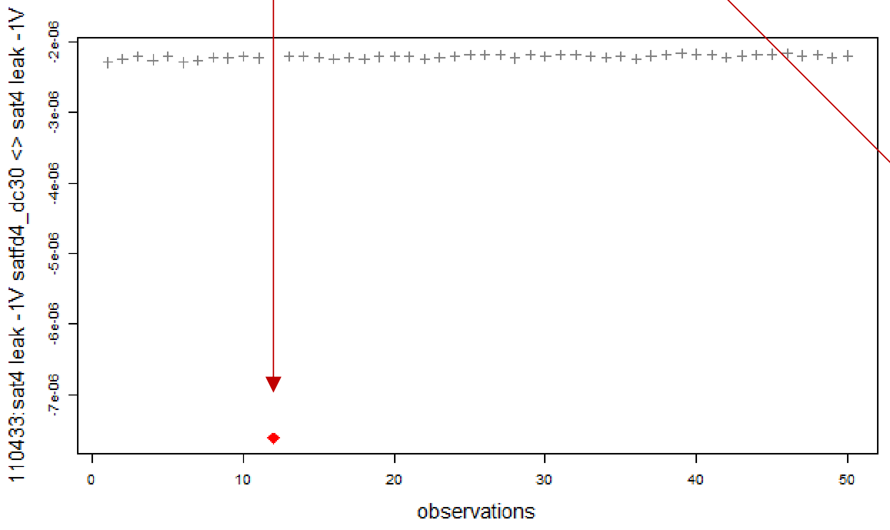
***T^2* outliers (bold) are no more detected if the number of tests *p* increases**

Standard multivariate methods like PCA or Mahalanobis distance or Hotelling *T^2* do not work when p increases!

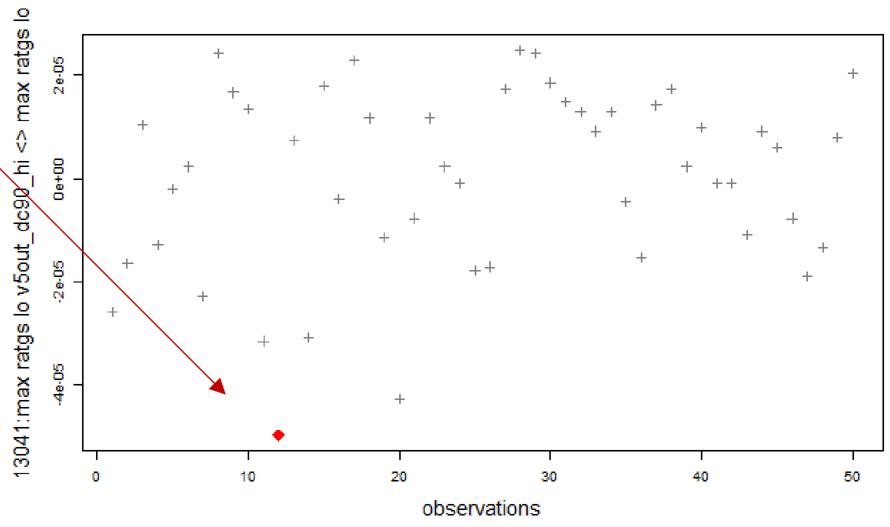
GAT at work: contributors



GAT identifies main parameters involved in the outlier detected:
We call them **contributors**
(Vdvramp, idvramp, iglhv, vg, threshold voltage & ISGS, leakage currents…)



Some are pretty 'obvious' with a clear univariate outlier



Others are harder to detect but important for analysis

Conclusion

- Two new approaches for enhanced outlier detection: **Half-Sigmas** & **Multivariate method « GAT »**
  - Finer detection compared to standard PAT (Part Average Testing)
  - Distribution-agnostic
  - Can handle highly non-normal and asymetric data

- Methods proven on real data

- GAT tested, optimized and qualified with a semiconductor industrial partner

References and ackowledgments

- Tchebychev, *Des valeurs moyennes,* Journal de mathématiques pures et appliquées, 2e série, vo. 12, 1867, p. 177-184.
- Ivan Svetunkov & Nikolaos Kourentzes, *Asymmetric prediction intervals using half moment of distribution*, ISIR conference, Budapest, 2016
- Nikolaos Kourentzesa, Ivan Svetunkovb, and Juan R. Traperoc, *Connecting forecasting and inventory performance: a complex task*, preprint
- Harry Markowitz, “Portfolio Selection,” Journal of Finance 7, No. 1 (March 1952), p. 77-91.

- Thanks to Nikolaos Kourentzesa for the idea of half moments and discussions