

Reliability Embedded Test Engineering (RETE)

Abdelmalek Cherchari¹, Raman Gera¹, Florian Spitzbarth¹, Mathias Lipowski¹, Daniel Kriesten¹,
Luca Moriconi², Alessandro Maseri²

¹ Fraunhofer ENAS, Technologie-Campus 3, 09126 Chemnitz, Germany,

² ELES Semiconductor Equipment SpA, Zona Bodoglie, 148/1/Z, 06059 Todi (PG), Italy
raman.gera@enas.fraunhofer.de

Abstract:

The semiconductor production and testing facilities face a major hurdle in the form of intermittent and latent failures occurring in system-on-chip (SoC) devices. Modern HTOL and Burn-In test systems are utilized to perform standardized reliability screening tests (JESD47, JESD22-A108) [2][3] to achieve “Zero-Defect” semiconductor production. The test systems generate large volumes of data from multiple sources (ATE, burn-in boards, equipment, and environment logs and on-chip diagnostics), generally stored in formats such as STDF, CSV, etc. Due to such complex data ingestion and focus on the yield of pass/fail assessments, subtle multi-parametric degradation and failures remain undetected, causing burn-in and HTOL conditions to be unnecessarily stringent.

The Reliability Embedded Test Engineering (RETE) [1] is a project that focuses on developing an AI-driven reliability analytics software that is completely integrated with ELES Test-for-Reliability (TfR) framework and ART burn-in test systems. Core elements of this software consist of a unified test data format, an automated data validation and an analytics toolchain library that combines conventional statistics with specialized machine learning modules for failure prediction and lifetime estimation. The goal is to turn HTOL/ELFR process from a time and resource intensive screening step into a predictive reliability instrument.

Key words: Reliability testing, burn-in, HTOL, data analytics, machine learning.

1 Introduction

Major application for SoCs in automotive, data center and industrial fields call for extreme reliable lifetimes and safety requirements to ensure accurate operations.

The ELES ART test equipment facilitates a range of standard qualification test capabilities (HTOL, ELFR, PTC, Burn-In, Test during Burn-In). However, there is an industry-wide gap for a single, uniform data structure platform that combines reliability screening, equipment logs and STDF wafer acceptance test data. Linking of these data files with data from Design-for-Test (DfT) instruments such as ring oscillators, thermal and current sensors, BIST signatures and diagnostic registers will be vital to understand drift/degradation mechanisms and intermittent failures. RETE addresses this gap by providing unified data test format and AI supported tool chain that will aid in design and process improvement.

2 Reliability Test Data and Current Challenges

Modern reliability and test environments acquire and generate data from multiple sources:

- ATE measurements and wafer test results, usually stored in STDF or industrial proprietary formats.
- Burn-in/HTOL measurements.
- Equipment and environmental logs in a data trace file including temperature data, alarms and board level diagnostics.
- On-chip monitors such as BIST signatures and ring oscillators.

In many testing and reliability flows, these data sets are analyzed separately with inconsistent levels of data quality leading to several challenges:

- Lack of a unified data model.
- Correlation of latent failure with early parametric indicators.

- Difficulties in failure traceability across different test equipment and lots.
- Underutilization of metadata for drift analysis, outlier detection and reliable lifetime modeling.

Prolonged HTOL or testing at elevated stress voltages can overcome these limitations to provide reliable quality of devices. This industrial approach increases cost and often influences test optimization workflow. Therefore, a solution is required that improves outgoing device quality with increase in test efficiency.

3 Objective and Methodology

RETE is designed around four key objectives:

1. Provide a unified data infrastructure that strengthens reliability data.
2. Implement a structured and organized analytics workflow tool for HTOL and ELFR test strategies with advanced outlier analysis tools.
3. Integrate native AI/ML capabilities to support failure analysis, anomaly detection and lifetime estimation.
4. Demonstrate industrial validation and impact using the ART test systems.

3.1 Unified Data Infrastructure

The root of RETE is the test data (e.g., ETDF), that can ingest and systemize diverse test file formats including:

- Test data files from wafer and burn-in test.
- Equipment log files/ trace files from the HTOL systems.
- CSV files containing environmental telemetry.
- On-chip diagnostic data from DfT/DfR structures (e.g., current monitors, BIST, ring oscillators)

The data is imported via the software and converted to a centralized database, combining all details and metadata using key identification parameters such as device under test (DUT ID), lot number, wafer ID, stress conditions.

The RETE prototype will support multi-file import, file integrity and consistency checks and automated data validation.

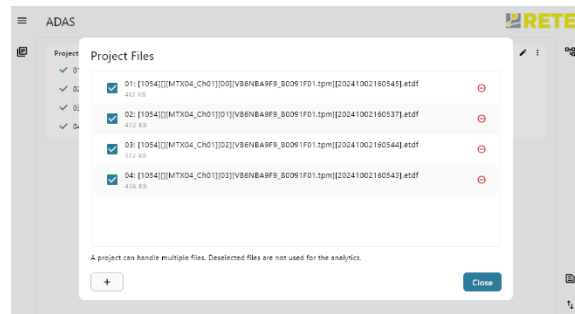


Fig. 1. Multiple file import with RETE prototype software

3.2 Analytics Workflow for HTOL and ELFR

RETE implements a structural analytics workflow adapted to JEDEC and AEC criteria for HTOL and ELFR tests. The major elements of the workflow include:

1. Data cleaning and filtering: automated removal or flagging of errors/corrupted records in the database, handling of missing values and normalization of data.
2. Part Average Testing / Outlier analysis: supports outlier identification techniques enabled through PAT, modified Z-score and multi-variable outlier analysis.
3. Statistical process control: enables the monitoring of key parameters using charts, graphs and possesses capability to identify devices across lots, wafers and chambers.
4. Drift analysis: tracks the key parametric changes over stress time enabling identification of devices with abnormal drift/failure tendencies.
5. Bin and yield analysis: assessment of the impact of outlier analysis/ reliability screening on the data to calculate yield and bin distributions.
6. Correlation and lifetime modeling: enables the correlation of device parameters and drifts with stress conditions and equipment data and enables estimation of reliability metrics such as Mean Time To Failure (MTTF) and Failure-In-Time (FIT) according to the relevant standards.

This workflow provides a consistent, flexible and configurable analysis that can be easily applied across different products, devices and lots increasing reproducibility of results.

3.3 Native AI/ML Engine

To utilize the burn-in data efficiently, RETE proposes to integrate a native AI/ML engine

combined with the analytics workflow. Major functionalities may include:

- Failure mode classification using multiple parametric drifts supporting Design-for-Reliability (DfR)
- Deep learning models for reliable lifetime estimation, anomaly detection and predictions using drift analysis
- ML-enhanced outlier analysis using historical lots and failure cases.

These capabilities are meant to assist but not replace conventional statistical techniques. Engineers are enabled to combine workflows with AI-driven insights to enhance reliability screening.

3.4 Industrial Validation of RETE

Industrial validation is carried out using ELES ART systems under realistic stress and screening conditions. The test flows comprise of various device families and methodologies including HTOL, ELFR, functional burn-in. This approach provides economic and diagnostic value to the RETE approach for data analysis.

4. System Architecture and Implementation

4.1 System Architecture

RETE software prototype is implemented as a client-server model (currently utilizing Flutter for frontend and Python for the backend) offered as a Windows desktop application structured for later transition to scalable SaaS deployment. The communication between the backend and frontend services is structured by an API. This enables the deployment of frontends and backends tailored to user- and industry-specific requirements. RETE prototype will include a file parse tool with multi-file support, filter engine and interactive plotter including statistical analysis tool widgets. RETE proposes an AI/ML engine based to operate on clean data and automate outlier analysis with known failure classifications and implements fixed analysis predefined for HTOL and ELFR tests with automated report generator creating JEDEC standard qualification reports.

4.2 Software Implementation

Frontends (clients) solely rely on the data retrieved from a backend (server) to visualize it, enabling data investigation, filtering, and analysis. With the decoupled backend responsible for the computations to achieve the aforementioned features, an appropriate scaling to the use case can be done straightforwardly. The use of the OpenAPI specification as an Interface Description Language (IDL) allows for code generation for interfaces of both the

frontend and backend to ensure a clean and structured data exchange. The generation of source code that then has to be extended with business logic and algorithms enables easier maintenance, development, and extension. RETE software prototype uses a JSON structure to build up the data analytics pipeline. The frontend enables a user to do this in a graphically appealing way by drag-and-drop input files, data analysis elements, and structured text elements for description (nodes), which make up the structure. The backend will receive the analytics pipeline through the API and executes it by producing data artifacts for each node; for example, filter the input data for the following nodes and respond with plottable data for a plotting node. This approach enables the caching of data analytics steps, ensuring faster result generation. Configuration and settings in a user and project context can be used to save individual nodes. Multiple users should be able to investigate and adapt a project and its analytics pipeline. The pipeline can be finally exported in a PDF for reporting demands.

4.3 Hardware and DfT/DfR Integration

RETE implements DfT/DfR features in the DUT and on the test equipment boards to improve data analysis. These elements include scan chains, embedded thermal and current sensors, RMS detectors, current generators, BIST, ring oscillators and diagnostic registers. These elements are then accessed via the ELES architecture enabling continuous monitoring of aging indicators during the stress test for better data interpretation.

5. Advanced Outlier Analysis and Drift Monitoring

The process of detection and screening of failures/outliers cannot depend exclusively on fixed device specification limits for discrete parameters. In RETE, outlier analysis [7] follows an established analysis workflow that combines univariate and multivariate strategies making it suitable for data analysis on high volume reliability data. The methods described in the section are independent of a specific product and can be used across different test methods such as HTOL, ELFR, and functional burn-in.

In semiconductor data analysis, outlier analysis can be of two different types:

- **Outlier Scores:** Most outlier detection algorithms output a score quantifying the level of "Outlierness" of each data point. This score then can be used to rank the data points in order of their outlier tendency. This is a very general form of outlier detection.

- **Binary Labels:** The second type is a binary label denoting whether a data point is an outlier or not. Although, some algorithms return binary labels directly, outlier scores can also be converted into binary labels. This can be achieved by imposing thresholds on outlier scores to determine if the data point fits inside the limits or not.

A prerequisite for RETE analyses is that the test and data engineers apply domain knowledge and define and filter data according to same device family and product, stressed under similar conditions and test strategies within an explicit timeframe as the first step. The lot statistics, Z-scores, mahalanobis distances and drift can only be computed after the first step of data cleaning and filtering. Without the vital step of data filtering by engineers, the results of data analyses will not provide concrete inferences.

5.1 Part Average Testing (PAT)

RETE follows the part average testing (PAT) method of Automotive Electronics Guideline document AEC-Q001 [4] which defines a statistical based method as the first stage to remove outliers from AEC qualified devices. After selection of an important electrical parameter from the test data, robust mean and robust sigma values are estimated using quartiles and is computed as

$$\sigma_{\text{robust}} = \frac{Q3-Q1}{1.35} \quad (1)$$

$$\text{Robust mean} = \text{median } Q2 \quad (2)$$

Where Q1 and Q3 values are the first and third quartiles of the ranked data, respectively. Using these parameters, dynamic PAT limits are computed using the data of the current lot or wafer as,

$$\text{Dynamic PAT Limits} = Q2 \pm 6 \sigma_{\text{robust}} \quad (3)$$

Using these robust statistics, Dynamic PAT limits are established and preferred in production. Applying these PAT limits to the measured data, the user can easily identify the outliers as measured data less than lower limit or measured data more than upper limit. These devices will be classified as failures and will be removed from the good population even if their measure value lies within the formal specification limits, as recommended in AEC-Q001.

5.2 Modified Z-score Univariate Outlier Detection

For a single parameter outlier screening, RETE uses a standard statistical method known as modified Z-score, as recommended by Iglewicz and Hoaglin [5]. The modified Z-score is a method used to measure the outlier strength of a

parameter. It approximates the differences between outlier scores through median and mean absolute deviation instead of standard deviation making it less sensitive to skewed data distributions.

For a measurement x_i , the modified Z-score M_i is represented as

$$M_i = 0.6745 \frac{x_i - \tilde{x}}{\text{MAD}} \quad (4)$$

Where $\tilde{x}$ is the sample median of the parameter and

$$\text{MAD} = \text{median}(|x_i - \tilde{x}|) \quad (5)$$

MAD is the mean absolute deviation for the data that is approximately normally distributed. RETE workflow utilizes a configurable threshold to flag univariate outliers:

- If $M_i > 3.5$, it is termed as an outlier due to industrial standard cutoff based on sigma values.
- If $3 < M_i < 3.5$, the observation is termed as a potential outlier and might be qualified for further analysis.
- If $M_i < 3$, it is not an outlier.

The modified Z-score method is useful for lot-by-lot monitoring as it remains stable in existence of rare and radical values.

5.3 Mahalanobis Distance Based Multivariate Detection and Drift Analysis

Several degradation mechanisms in burn-in test lead to correlated deviations in several parameters. Therefore, multivariate outlier analytics are more accurate and suitable as compared to separate and univariate analysis. RETE employs the mahalanobis distance outlier analysis algorithm, introduced by P.C. Mahalanobis [6] and widely used in the industry for multivariate data screening to see how far each device deviates from the main population.

For a k-dimensional vector x_i corresponding to a vector created from the selected key reliability parameters from the test data. The sample mean vector μ and covariance matrix Σ are calculated in the backend to derive the inverse covariance of the dataset Σ^{-1} . The deviation vector is computed as

$$\Delta x_i = x_i - \mu \quad (6)$$

And mahalanobis distance is evaluated as

$$D2_{m,i} = \Delta x_i^T \Sigma^{-1} \Delta x_i \quad (7)$$

Mahalanobis distance follows a χ^2 distribution with k degrees of freedom, RETE selects the value for χ^2 as 99th percentile as industrial standard. The user is also provided with

opportunity to edit this threshold for varying results. RETE classifies a device as a multivariate outlier if $D2_{m,i} > \chi^2$. The data engineer must filter the input data with proper boundary conditions to obtain a meaningful result from a homogenous population.

Since the distribution of burn-in data for RETE is already time-dependent, the data vector and calculation can be evaluated through each time interval or test index vector providing us with the mahalanobis distance for each time interval as $D2_{m,i}(t_j)$ where t_j signifies the time intervals. RETE simply derives the drift of the device by calculating the overall slope values as ΔD^2 divided by Δt . Devices with rapidly increasing mahalanobis distances are flagged as drift outliers and the values are plotted to get inferences from the data such as positive drift pattern and negative drift pattern.

6. Industrial Case Studies

To illustrate the practical applications of RETE workflow for the industry, two industrial case studies are presented. The first case study examines 30 automotive ADAS devices undergoing an accelerated voltage screening test with multiple data files. The second case study highlights an early life failure test of 288 accelerometer devices stressed at high temperatures. For both case studies, RETE imports multiple test files, data filtering is performed by the test engineer to execute outlier analysis and drift monitoring. The software is able to highlight devices that would remain undetected with conventional pass/fail analysis.

6.1 Reliability Screening: Accelerated Stress Test of ADAS SoCs

To achieve zero defect methodology of RETE, 30 Advanced Driver Assistance Systems (ADAS) SoCs were subjected to accelerated voltage stress test. The experiment ran for a short duration of time with 25°C environment temperature. The methodology used active internal test patterns to maintain a target on-die junction temperature of ~100°C. Continuous parameter telemetry was performed for cyclical

test patterns and indices tracked through built-in-self test codes, leakage current monitors, die thermal gradients (DIE0_SEN0 to SEN4), die traceability indicators (DIE0_Wfr) and core junction temperature monitor (Tjd0).

RETE performed thermal profiling via 5 localized embedded sensors (DIE0_SEN0 to SEN4) to validate the stability of the test.

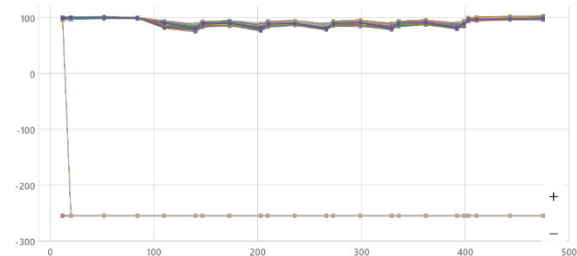


Fig. 2. Thermal sensor on die telemetry across test indices.

The software highlights the isolated infant mortality breakdown of site number 12 and site number 7 as the stress is initialized. The rest of the devices keep in the thermal range fluctuating between 74°C and 103°C creating a synchronized pattern through the test indices signifying different pattern executions. Site number 12 was categorized immediately as a hard failure. Filtering the data of this site number provides us with a better image showcasing the temperature profile for the test patterns.

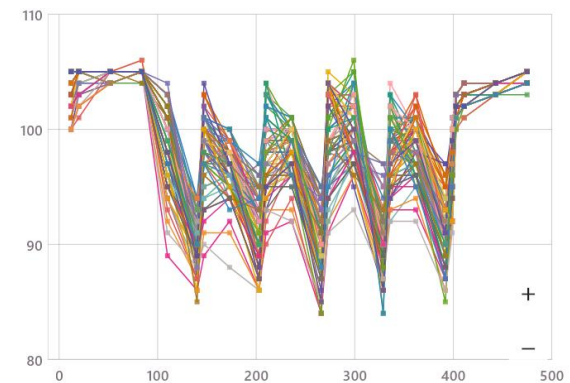


Fig. 3. Filtered thermal sensor on die telemetry across test indices demonstrating sawtooth temperature profile.

Figure 4 highlights the core junction temperature of the devices i.e., Tjd0 profile across all test indices. The analysis highlights more devices undergoing electrical degradation mechanisms that were not filtered out through thermal sensor telemetry. The 25 passing devices maintained a synchronized pattern centered around 100°C stress target with fluctuations between 83°C and 103°C highlight the internal heat patterns of devices under test. A few outliers were recognized through this figure such as site number 0 that maintains a linear positive slope till 112°C. the root cause for this can be a wear out mechanism such as thermal runaway. The

test also observes a few destructive outliers related to open circuit/shorts. Multiple outliers were observed such as site number 1 and 21. These outliers lose the current leading to no self-heating and thermal profile climbing down to temperatures of 30°C to 50°C. There is sharp peak of up to 132°C at test index 403 for these outliers highlight the issues with internal power dissipation.

RETE was able to identify 3 additional outliers through the modified Z-score analysis to screen out latent defects that cannot be detected through conventional failure analysis

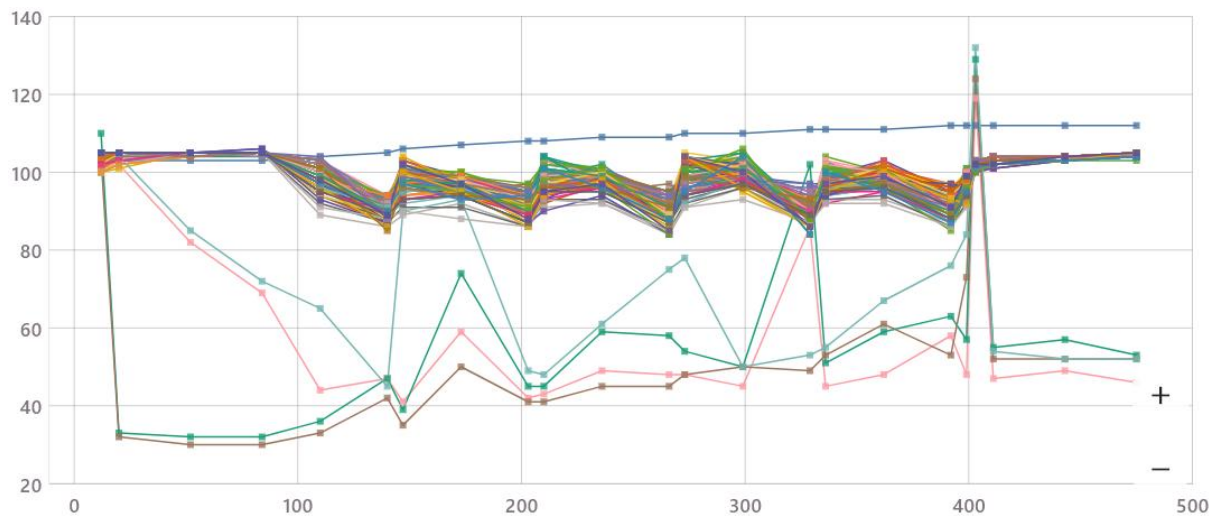


Fig.4. Transient Junction Temperature (Tjd0) Signature Analysis.

Tab. 1: Outlier Table on Thermal Telemetry

T / °C	Test Index	Z-Score	Outlier	Site Number
119	403	11.47	Definite	01
129	403	18.21	Definite	07
132	403	20.23	Definite	21
112	403	6.75	Definite	00
124	403	14.84	Definite	12

6.2 Early Life Failure Screening of Accelerometer Devices

RETE was applied to an early life failure test of 288 MEMS accelerometer devices stressed at 125°C for 24 hours. The devices belong to the same product family and were tested under identical conditions using ELES ART system. During the stress test, multiple electrical and functional parameters such as acceleration and gyroscope values were recorded in the test file.

The raw test data file was imported into RETE primarily linked via traceability parameters such as site number, DUT ID, wafer coordinates. Basic data validation checks were performed and the conventional pass/fail screening filtered out two abnormal devices at site number 80 and 96.

The traditional analysis flagged these 2 outliers as latent failures or hardware bins. RETE was then used to display the parameters of gyroscope x-, y-, and z-axis values across all test indices.

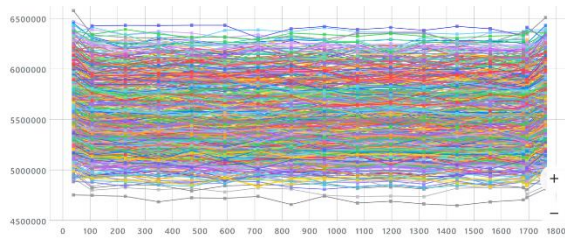


Fig. 5. Time series of gyroscope parameters for passing MEMS accelerometer devices.

Figure 5 illustrates the time series gyroscope parametric results in x-, y- and z- axis for the remaining 286 devices that formally passed the standard ELFR screening. The dense, almost parallel band indicates the majority of the population remains stable over the 24 hours stress. The graph also highlights some variations and small deviations in particular devices. Visual inspection is not enough for such analysis and therefore, RETE applies mahalanobis distance outlier analysis taking acceleration and gyroscope parameters into account to screen out potential outliers and analyses the drift of the weak devices.

Many outliers out of 286 devices were flagged using these algorithms but to avoid an extensive listing, the case study will represent the results for three representative devices:

- Site number 275 (late drift device)
- Site number 140 (unstable device)
- Site number 1 (early life failure)

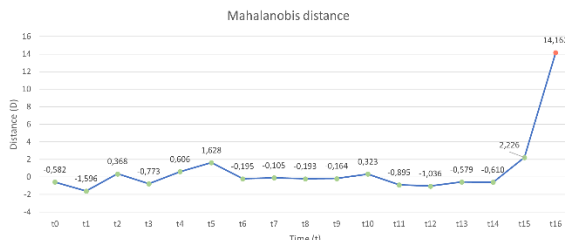


Fig. 6. Mahalanobis distance of gyroscope parameters across 17 test indices for site number 275

Figure 6 highlight the mahalanobis distance evolution for site number 275 showing late emerging drift in the gyroscope parameters. The device remains stable over the 24-hour stress period and behaves like the normal population until the last test index where the distance rises sharply to 14.162 and RETE classifies this as a late emerging outlier as the threshold value is 11.345 for 99% pass criteria. This device would not be visible in simple pass/fail check but RETE classifies the device as a drift outlier and potential latent failure with late emerging degradation.

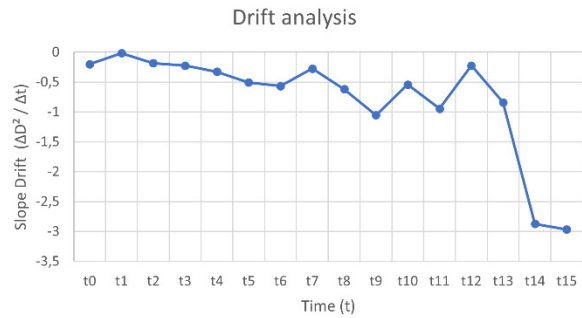


Fig. 7. Mahalanobis distance drift of gyroscope parameters across 17 test indices for site number 140

Figure 7 represents the RETE evaluation for time dependent mahalanobis distance drift for site number 140 across 17 test indices. The drift slope shows a negative drift and varies noticeably from interval to interval. The drift becomes completely unstable and strongly negative in the last stretch of the test indicating accelerated changes in the parameters. The mahalanobis distance remains under the specified limits despite showing variance. Therefore, the device is not classified as a hard outlier and RETE flags site number 140 as potential drift outlier to be reviewed by reliability engineers for further analysis.

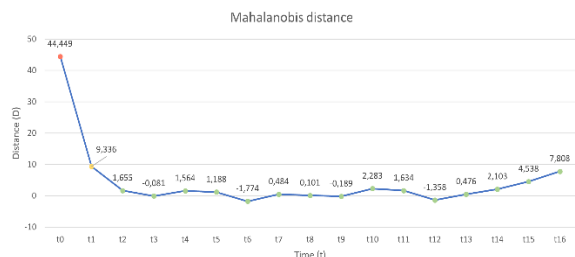


Fig. 8. Mahalanobis distance drift of gyroscope parameters across 17 test indices for site number 1

The graph in figure 8 represents the mahalanobis distance outlier analysis for site number 1 across 17 test indices. For site number 1, the distance already crosses the threshold at high values of $D = 44.449$ at t_0 and $D = 9.336$ at t_1 . These values cross the threshold and indicate that this device is a hard outlier and RETE flags this as an early life failure due to the tendency of deviation from normal population at the start of the test.

In this case study signifying early life failure screening of 288 accelerometer devices, RETE demonstrated the capability of screening out additional outliers beyond the two devices rejected by convention pass/fail criteria. RETE highlighted additional weak devices with advanced outlier algorithms showcasing the potential to strengthen infant mortality screening in semiconductor test industry without increasing the stress duration of the test.

7. Impact and Conclusion

Comprehensive summary of RETE workflow:

- Prototype toolchain availability as a desktop application.
- Test Data File Handling:
 - Multiple test data file import
 - File integrity and consistency checks
 - Data validation and interpretation
- Advanced outlier and failure detection tools
- Visualization engine through interactive plotter
- HTOL/ELFR standard analysis configurations
- Automated reporting

RETE displays how an unified data structure and advanced data analytics can improve burn-in test data from screening steps to predictive reliability tools. Through the combination of silicon telemetry data and structured analysis workflows with PAT, univariate/multivariate outlier detection, drift monitoring and device traceability, RETE exposes subtle degradation signatures that are difficult to uncover through traditional data analytics.

Key findings of RETE:

- Improved correlations and finding between stress conditions, failure mechanisms and equipment data
- Enable optimisation of burn-in and HTOL conditions
- AI/ML derived insights can be integrated into design and manufacturing process to eliminate major failure mechanisms

The two industrial case studies highlight this capability as RETE detected catastrophic and latent thermal outliers in the screening for ADAS SoCs, while also screening out functional degradation signatures during burn-in test of accelerometers. Detection and characterization of these outliers supports process optimization which results in improved overall device reliability. Therefore, the approach of RETE intends to evolve standard reliability testing platforms into a fast, predictive, “Zero-Defect” test infrastructure for semiconductor devices.

References

- [1] RETE: Design and Test for Reliability. Data Analysis for Zero Defects. Luca Moriconi,

Alessandro Maseri, Giovanni Taormina: ARTs23, 2023.

- [2] JEDEC Solid State Technology Association, “Stress-Test-Driven Qualification of Integrated Circuits,” JEDEC Standard JESD47M, 2025.
- [3] JEDEC Solid State Technology Association, “Temperature, Bias, and Operating Life,” JEDEC Standard JESD22-A108.
- [4] Automotive Electronics Council, “Guideline for Part Average Testing (PAT) of Integrated Circuits”, AEC-Q001, Rev. D.
- [5] B. Iglewicz, D. C. Hoaglin, How to Detect and Handle Outliers, ASQC Quality Press, Milwaukee, WI (1993)
- [6] P. C. Mahalanobis, “On the generalized distance in statistics,” *Proceedings of the National Institute of Sciences of India* 2, 49–55 (1936).
- [7] C. C. Aggarwal, Outlier Analysis, 2nd ed., Springer, New York (2017); doi: 10.1007/978-3-319-47578-3.