Journal of Electronic Testing (2023) 39:57–69
https://doi.org/10.1007/s10836-023-06047-w

A Weighted-Bin Difference Method for Issue Site Identification in Analog and Mixed-Signal Multi-Site Testing

Isaac Bruce[1] · Praise O. Farayola[1] · Shravan K. Chaganti[1] · Abalhassan Sheikh[1] · Srivaths Ravi[1] · Degang Chen[1]

Received: 21 December 2022 / Accepted: 30 January 2023 / Published online: 21 February 2023

Abstract
Higher complexity in recent chip designs, module integration, and increasing test quality requirements have expanded measurement needs and further increased chip test costs. Multi-site testing (parallel measurement) solves this issue by taking test measurements from multiple chips simultaneously, massively increasing throughput, and significantly reducing the test time per chip. Massive multi-site testing system, a setup with significant measurement site count, further improves throughput and maximizes gains. However, it unavoidably amplifies site-to-site variations in the measured specifications. This problem is particularly magnified in analog and mixed-signal chips. Some measurement sites now exhibit pronounced induced errors, and their measurements no longer reflect the actual performance of the device under test (DUT). This problem presents a solid need to identify sites that suffer from extreme site-to-site variations (issue sites). We propose an automated method to investigate site-to-site variations in volume multi-site data and identify issue sites that may not be obvious via human inspection or basic statistical methods. Assuming that all measurement sites have the same accuracy and precision, we consider an issue site to be one whose weighted-bin difference score is greater than an analytically derived upper bound. We apply the proposed method to simulation data and volume test data obtained from an industrial analog and mixed-signal system on chips (SoCs) that were tested using multi-site testing hardware and show that the technique can effectively identify issue sites in the testing system. We compare the proposed algorithm to existing methods and demonstrate its superior performance.

Keywords Multi-site/Parallel Measurement Testing · Hardware Systematic Errors · Site-to-Site Variations · Weighted-Bin Difference · Automatic Test Equipment (ATE) · Analog and mixed-signal

1 Introduction

Increasing design complexities and levels of integration in modern SoCs have resulted in growing data acquisition and measurement time for device characterization due to increased measurement accuracy requirements, raising production costs. Coupled with the high demand for chips today, a huge need and market for reduced measurement time and low-cost testing exist [1].

Different methods and algorithms have been proposed to reduce test time and cost for SoCs. Some methods focus on more efficient algorithms, significantly reducing test time [2–4]. Built-in-self-test (BIST) techniques embed in each DUT a mechanism to obtain test measurements from itself, thereby reducing test time and cost [5, 6]. Authors in [2, 3] present algorithms that relax the stringent requirements of some automatic test equipment (ATE) parts, thereby reducing their cost and, subsequently, measurement/test cost. However, multi-site/parallel measurement is arguably one of the best methods for combating increasing test time and chip cost [7–10].

Multi-site testing reduces test time by simultaneously obtaining test measurements from several DUTs. It uses the ATE channel-sharing method, taking advantage of the ATE infrastructure concerning data storage and test interface facilities [11]. It was primarily introduced in the digital domain, especially in memory testing; however, it has now gained ground and acceptance in analog and mixed-signal testing [12]. The method has been proven to provide high-yield capacity and additional cost advantages [13, 14].

Responsible Editor: T. Xia

✉ Isaac Bruce
ibruce@iastate.edu

1 Iowa State University of Science and Technology, Iowa State University, Ames, IA, USA

One challenge of the method, usually pronounced in the massive multi-site testing approach in the analog and mixed-signal domain and central to this paper, is multi-site variations in chip measurement data. Site-to-site variations in a multi-site testing system may arise from various sources. Some of these variations, such as process variations and design marginalities, correspond to actual problems with the device being tested and should lead to scrapped devices or changes made to the design of the DUT [14]. However, in other cases (which call for caution), the measurement equipment could introduce unwanted systematic errors, affecting test results.

Site-to-site variations may arise due to difficulty maintaining equal resource management and sharing system resources, e.g., voltage and current sources, arbitrary waveform generators, and digitizers, amongst test sites. Another reason could be the undesired interactions among analog, digital, and power sub-blocks within the tester hardware. It is well-known that these occurrences can significantly affect analog signals. These variations may necessitate loosening device specifications to guarantee device performance, leading to a reduction in the unit costs of DUTs and, hence, a loss in revenue.

Figure 1 is the boxplot visualization for the measured analog-to-digital converter (ADC) offset data for multiple chips from an industrial multi-site testing system. We expect the mean and standard deviation of the offset measurements to vary from measurement site to site due to wafer and lot variations. However, an inspection of Site A shows an unwanted shift exhibited by that site, clearly demonstrating one case of how site variations manifest.

Despite the difficulties, the demand is to have more sites in production to satisfy the high volume demand for SoCs

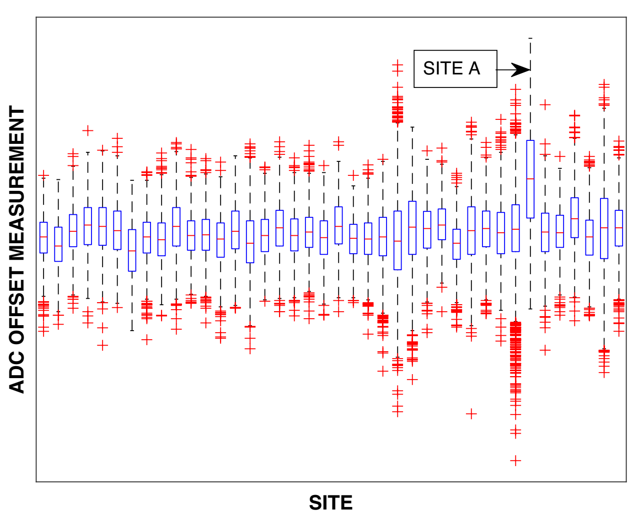

Fig. 1 ADC Offset measurements across multiple test sites showing Site A as an issue site

without the accompanying increase in cost [14, 15]. Hence, developing algorithms to identify sites that introduce significant systematic errors and lead to erroneous test results (issue sites) is vital.

Visual inspection of measurement data is usually the first step taken by the test engineer. For example, it is visually evident in Fig. 1 that Site A is an issue site. However, this method is crude and relies on humans, who are prone to error. Comparing the mean, standard deviation, and quartiles of measurement data at each site could be automatic. However, it relies on summary statistics that do not always accurately represent data. This method is primarily affected by other sources of variation, like wafer to wafer and others. Different methodologies for detecting issue sites and why they are not suitable are further discussed in Section 2.

In [16], issue sites were flagged by comparing the distribution of the measurement sites to a reference. However, the technique tends to generate a lot of false positives. It does not present a well-defined boundary for flagging issue sites and is more suited for Gaussian and symmetric distributions. Hence the need for a new flagging approach.

This paper presents a weighted-bin difference (WBD) algorithm to identify issue sites in a multi-site testing system. The algorithm first estimates the expected distribution of the measured parameter by selecting a subset of high-confidence good sites called Site0. It then uses it as a reference frame for issue site identification. Scores are computed for each site and compared to an analytically derived upper bound, which serves as a means to identify issue sites. The proposed method has the following advantages.

1. The method is, at no cost, automated and robust. An additional and complementary tool for the test engineer.
2. The method does not rely on single test statistics for issue site identification. Instead, it compares the distributions of each measurement site to a reference distribution.
3. The method performs an excellent job in estimating reference distribution compared to other methods.
4. The method proposed a well-defined boundary condition for identifying issue sites compared to other methods.
5. The method works for all types of distributions.

The rest of this paper is organized as follows. The second section provides a background for this paper and a summary of prior work. The third section emphasizes the need for a carefully estimated reference distribution. The proposed method is presented in the fourth section, and the guard boundaries are derived. The fifth section provides experimental results to demonstrate the proposed method's effectiveness and robustness using simulation data and actual multi-site ATE data from the semiconductor industry. The results are compared to those obtained by applying the Quantile-Quantile issue site detection algorithm proposed in [16]. Future directions concerning this work are discussed in the sixth section before finally concluding the paper in the seventh section.

2 Background and Prior Work

A flagging system was proposed in [17]. This technique removes human visualization and provides automation by flagging sites based on specific criteria. Four significant criteria are used for flagging sites: comparison of location using the mean or median; comparison of dispersion using the standard deviation of each site and IQR; comparison of skewness; and comparison of outliers. Test sites that were flagged across multiple criteria are given more weight than sites that were only flagged once. While this method worked better than visual inspection, many flags were raised, and these flags had to be analyzed by the test engineer. The method relied on single-test statistics like the mean or median, which may not accurately present and reflect the distribution of some measurement data [17]. Also, the method was primarily affected by other sources of variation, like intrinsic variations inherent in test data due to wafer variations.

Well-established inferential statistical tests like Analysis of Variance (ANOVA) were considered. However, this test requires some assumptions that were not true in our case. It assumes that measurement data was from a normal distribution and had equal variance, which was not true in our case. Several other inferential statistical tests that do not make either or both assumptions, like the Kruskal-Wallis test, the Levene test, etc., were also investigated. However, these tests gave different and conflicting results, were significantly affected by test site data size, and only provided information about whether an assumption was being met to some probability level or not.

Another approach that could be used is Principal Component Analysis [18]. This method considers issue sites to be those sites whose data points cannot be faithfully reconstructed using the principal vectors. While this may be a valid approach, PCA is a linear technique and may fail to provide meaningful solutions for cases where the error introduced by the site is nonlinear in nature.

If the sites suffering from extreme site-to-site variations are considered outliers, then outlier detection techniques can be used to identify issue sites. However, this still begs the question of what parameters would be used for the detection. The mean and standard deviation do not give an accurate description of the data, hence the idea of comparing the distribution of each site to a reference distribution. Multi-site variations will affect the distribution of the measured parameters at each site, with issue sites expected to suffer more. If we have an expected and true distribution, comparing each site's distribution to that true distribution is a valid way to identify issue sites.

Comparing histograms of the measurements at each site is a way to compare the distribution of each site. One method employed in the comparison of histograms is the total variation distance [19, 20]. If we let S_0, Q_x be the discrete distributions for Site0 and site x over probability space χ respectively, then the total variation distance is

$$d_{TV}\left(S_0, Q_x\right) = \frac{1}{2}\sum_{A\in\chi}\left|S_0(A) - Q_x(A)\right| \tag{1a}$$

Measurement sites with large d_{TV} are identified as issues sites. The method, however, fails to provide a boundary condition for identifying issue sites.

Another distance metric is the Chi-Square distance which is defined as:

$$D_{\chi^2} = \sum_{A\in\chi}\frac{\left(S_0(A) - Q_x(A)\right)^2}{S_0(A)} \tag{1b}$$

and the Kullback-Leibler divergence,D_{KL}, which is a measure of how much information is lost when the distribution Q_x is approximated as S_0. One major issue with the Chi-Square distance and Kullback-Leibler divergence metric is the case where $S_0(A) = 0$. This case is possible because bin intervals used in the construction of the normalized histograms typically cover a wide range, which implies that some bins will be zero-valued. Theoretically speaking, this would lead to a value of infinity, even if this only occurs at one point. Such zero-valued bins may indicate issue sites, but given the nature of multi-site testing, such zero-valued bins may be tolerable and expected. If the bin intervals are defined in such a way as to cater to outliers, then there will be bins for which S_0 will have a zero bin count while Q_x would not. This would lead to infinite Kullback-Leibler divergence scores and Chi-square distances, even though the variations in the distribution may be tolerable for the particular parameter being tested.

Any algorithm for identifying issue sites must account for cases where the histograms have zero-valued bins, as depicted in Eq. (2). It must also provide a boundary condition and, most importantly, be robust to what is considered "small/tolerable" variations due to process variations in the wafer. It must also account for small shifts in measurement data from site to site.

In this paper, we introduce a way to compare the normalized histogram by averaging the widths and lengths of the normalized histograms. The Weighted Bin Difference (WBD) assigns scores to each site on a parameter-by-parameter basis. We derive a boundary condition that, when violated, indicates that the site under consideration is an issue site. Before introducing the proposed method in

Section 4, we briefly review the technique used to generate the reference distribution.

3 Reference Distribution Estimation (Site0)

Every measured parameter has a true and expected distribution devoid of multi-site variation. This reference distribution is essential as we compare each site's distribution to it.

Several pre-silicon circuit simulations can be used to estimate die distributions. However, they do not contain manufacturing uncertainties, making the use of silicon measurements important. It is possible to obtain the theoretical distribution of parameters from simulations using Integrated Circuit (IC) layout schematics and fabrication parameters. However, this could be time-consuming, especially for a large number of measured specifications, and inaccurate due to model inaccuracies.

An assumption that most of the measurement sites are good and only a few are issue sites is used in [16]. In the paper, all the measurements by test sites are lumped together to form a reference distribution. However, this is not good enough. While this may work if only a few sites have issues, it will be a problem if more than one site has issues. Given that each site has been affected by small amounts of nonlinearities and noise, this approach can lead to a quick build-up of error.

For example, in scenarios like [21], when only one test site has issues, it does not affect the 3IQR limit for outlier detection. However, when the number of issue test sites is significant, it makes a lot of difference as the 3IQR limit is affected. This bolsters the point that lumping the measurements of all sites together will not be accurate as a reference distribution. In Fig. 2, there are three good sites and one issue site. We see that if all the sites are lumped together, the resulting normalized histogram does not accurately represent the normalized histograms of the good sites. However, suppose by some algorithm, only the good sites are used in the generation of the reference distribution. In that case, the resulting estimate is a good representation of the distribution of the good sites. With such a reference, it will be easier to identify issue sites, while in the other case, issue sites may not be easily distinguishable.

The generation of the reference distribution in this paper involves first using an algorithm to identify high-confidence good sites. High-confidence good sites refer to sites that suffer from minimal site-to-site variations. The resulting set of sites is termed Site0 [16]. In this paper, we employ an algorithm that uses ordinal optimization [22–24] and K-Means Clustering [25] for an effective and accurate selection of sites to form Site0. Details of the steps are presented in Algorithm 1.

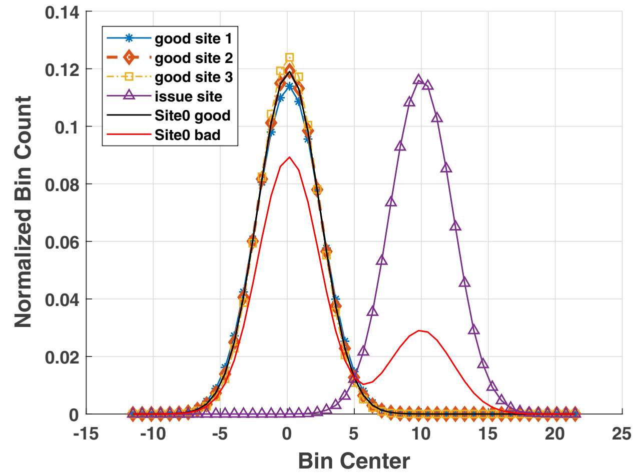

Fig. 2 Normalized histogram plot of three good sites, one bad site, bad Site0 and a good Site0. We show that lumping all the measurements of all dies together is not an accurate estimation of the reference distribution

In Algorithm 1, the result is Site0, which is the set of good sites. Before lumping all the measurements from the good sites together, we implement a data transformation method to correct for systematic errors introduced by each test site.

Algorithm 1: Site0 Selection

Data: ;
X = Measurement Dataset, N = Number of sites;
Q = Number of Top Sites, P = Number of parameters
Result: ;
$Site0$ = Set of good sites for each parameter
$V \in R^{N \times P}$;
$Rank \in Z^{N \times P}$;
$G \in R^{N \times 4}$ **for** ($j = 1$ **to** P) {
  for ($i = 1$ **to** N) {
    X_{ij} = Site i, Parameter j measurements ;
    V_{ij} = Standard deviation of X_{ij} ;
  }
}
$Rank$ = Sort each column of V in ascending order;
$Rank_{top}$ = Pick first $ceil(Q \times N)$ rows of $Rank$;
$S_{seed} = \cap_{i=1}^{P} Rank_{top_i}$;
for ($j = 1$ **to** P) {
  for ($i = 1$ **to** N) {
    X_{ij} = Site i, Parameter j measurements ;
    G_{i1} = Mean of X_{ij} ;
    G_{i2} = Standard deviation of X_{ij} ;
    G_{i3} = Median of X_{ij} ;
    G_{i4} = IQR of X_{ij} ;
  }
  Cluster G into 2 groups ;
  W = Cluster containing most of S_{seed};
  $C = covariance(G(S_{seed} \in W))$;
  $\mu = mean(G(S_{seed} \in W))$;
  Compute the Mahalanobis distance for each Site $i \in W$;
  $d_i = \sqrt{(G_{i(1:4)} - \mu)^T C^{-1} (G_{i(1:4)} - \mu)}$;
  σ =Standard deviation(d_i);
  $Site0_j = \{i : d_i \leq 0.5\sigma\}$;
}

It should be emphasized that Algorithm 1 is not the only way of identifying Site0. A simple comparison of the statistical properties of the sites can be used to find the most similar sites. However, our proposed method provides better guarantees of not including potential issue sites.

Let D_x^j be the volume test measurement data for site x, parameter j, and site x belongs to Site0. This data is transformed as follows:

$$\overline{D_x^j} = \frac{\sigma_0}{\sigma_x}\left(D_x^j - \mu_x\right) + \mu_0 \quad (2)$$

where σ_0 and σ_x are the standard deviations for Site0 and site x, respectively, for parameter j. Similarly, μ_x and μ_0 are the mean for Site0 and site x for parameter j. Since we do not know the true mean and variance of the various sites, we use unbiased estimates, which are given as:

$$\mu_x = \frac{1}{N_x}\sum_{j=1}^{N_x} y_j \quad (3)$$

$$\sigma_x = \sqrt{\frac{1}{N_x - 1}\sum_{j=1}^{N_x}\left(y_j - \mu_x\right)^2} \quad (4)$$

The standard deviation and mean for Site0 are computed as weighted averages of the same values for each site in Site0. i.e.

$$\mu_0 = \frac{\sum_{x=1}^{N_0} N_x \mu_x}{\sum_{x=1}^{N_0} N_x} \quad (5)$$

$$\sigma_0^2 = \frac{\sum_{x=1}^{N_0}\left(N_x - 1\right)\sigma_x^2}{\left(\sum_{x=1}^{N_0} N_x\right) - 1} \quad (6)$$

where N_0 is the total number of sites in Site0 of parameter j, y_j is the data samples and N_x is the die count for each site x.

The proposed transformation method is applied to ADC Offset error measurements obtained from an actual multi-site testing system. The results are presented In Fig. 3.

Figure 3 shows the boxplots for the offset error test measurements from a subset of sites. The figure also shows the Site0 boxplot obtained by just lumping the data together and the boxplot obtained after applying the data transformation method. The Site0 boxplot of the transformed data does not indicate a similar high variance, has the same width as most sites, and is properly centered. This suggests that the transformation yields a reference distribution much closer to the ideal distribution.

In addition to the visual inspection, we adopted a statistical method to demonstrate the efficacy of the proposed methodology. The Kruskal-Wallis test is a non-parametric method for testing if random samples belong to the same distribution [26]. Theoretically speaking, all the Site0 sites die test measurements after transformation must come from the same underlying distribution. The Kruskal Wallis test tests for the null hypothesis that all the samples are from the same distribution and, depending on the probability value, rejects or accepts it. The hypothesis is accepted if the probability value is greater than a set significance level. A similar test is performed on the Site0 sites die measurements without the transformation. Another test is also performed on the die measurements of all sites, which represents the case when all sites are lumped together, and the results are displayed in Table 1 and Fig. 4.

From Table 1, we see that the proposed approach yields the highest probability and passes the test, whiles the other two approaches fail. This result demonstrates the superiority of our algorithm. It must be emphasized that the method works only if the assumption that the majority of the sites are good is valid. During test board design, engineers conduct rigorous checks to ensure that all test sites are good. Changes in the test site hardware may occur during use; hence there are periodic checks during tests to ensure that the test sites are behaving as expected.

4 Proposed Weighted-Bin Difference Method

The Site0 distribution serves as the reference distribution for the measured parameter j. We proceed to construct histograms for all test sites and references. We define the actual bin center value and the normalized bin count for bin m from a Site x to be b_m^x and h_m^x respectively. For Site0

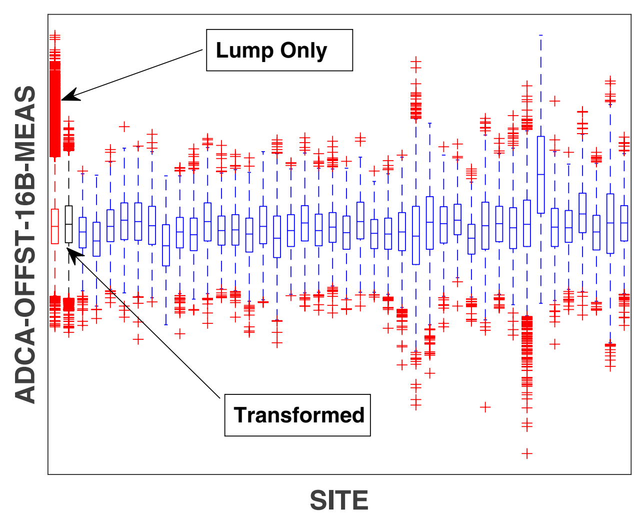

Fig. 3 ADC Offset boxplot for selected sites (blue). The conventional approach of just lumping data together results in lots of outliers which leads to high variance (red) whiles our approach results in a much better reference distribution (black) when compared to the N sites

Table 1 Table showing the Kruskal Wallis test that they are from the same distribution

Reference	probability value
Proposed Method	0.9999
Proposed Method without transformation	0
Lump all sites	0

this corresponds to b_m^0, h_m^0. The same bin intervals are used for all sites, including Site0, therefore $b_m^0 = b_m^x = y_m$.

The idea behind the WBD method is to give a score to each site based on the "difference" between its normalized histogram and that of Site0.

$$WBD\ score\ for\ Site\ x = \sum_{m=1}^{B} f\left(\frac{y_m - \mu_0}{\sigma_0}\right) g\left(h_m^0 - h_m^x\right) \tag{9}$$

where B = number of bins, μ_0 = mean of Site0 and σ_0 = standard deviation of Site0.

The choice for *f(x)* and *g(x)* is unlimited; however, the aim is to choose functions that make differences in the normalized histogram more evident and make it easier to identify issue sites. This point is illustrated In Fig. 5, which shows the distribution of Site0 and site x. Since these distributions are symmetric, the mean corresponds to the center of the curves. In this figure, two different points have been marked, A and B. The distance between point A and the center of the Site0 distribution is smaller when compared to that of point B. Any function, $f(x)$, chosen must make these differences more pronounced, i.e., the small distance between A and the mean of Site0 must be made smaller, whiles the more considerable distance observed for point B must be increased further. The function $f(x) = x^2$ satisfies the requirements. This choice also precludes the possibility of artificially generated low scores since the square of a number is always non-negative.

Also, the function, *g(x),* must differentiate between cases where distances to the mean are the same but normalized histogram counts are different. This scenario is seen In Fig. 5, where points C and D are the same distance from the mean of Site0; however, their normalized histogram counts are different. Any function, *g(x),* must account for this and circumvent any chance of the cumulative sum of the differences "artificially," resulting in a zero value. i.e., $2 * (3-2) + 2 * (2-3) = 0$, which would falsely imply that there are no differences between Site0 and site x. A good choice for $g(x)$ is x^2 for the same reasons quoted above. Another viable choice is $g(x) = |x|$. From empirical observations, $g(x) = |x|$ and $f(x) = x^2$ are suitable choices for the WBD definition given that they satisfy the requirements. It should, however, be noted that these are not the only applicable functions, but these choices make the derivation of a boundary condition more mathematically tractable.

The resulting WBD definition is consequently given as:

$$\begin{aligned} WBD(x) &= \sum_{m=1}^{B} \left(\frac{y_m - \mu_0}{\sigma_0}\right)^2 \left|h_m^0 - h_m^x\right| \\ &= \sum_{m=1}^{B} \frac{\left(y_m - \mu_0\right)^2}{\sigma_0^2} \left|h_m^0 - h_m^x\right| \end{aligned} \tag{10}$$

From Eq. (10), an upper bound can be computed for the WBD score for each parameter *j,* and any site whose score

A) Proposed Method

Kruskal-Wallis ANOVA Table

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Groups	7.75462e+07	14	5.53901e+06	2.2	0.9999
Error	7.24806e+11	20551	3.52686e+07		
Total	7.24883e+11	20565			

B) Proposed Method without transformation

Kruskal-Wallis ANOVA Table

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Groups	8.07934e+10	14	5.77096e+09	2292.27	0
Error	6.44042e+11	20551	3.13387e+07		
Total	7.24836e+11	20565			

C) All sites

Kruskal-Wallis ANOVA Table

Source	SS	df	MS	Chi-sq	Prob>Chi-sq
Groups	1.34946e+12	39	3.46016e+10	4149.43	0
Error	1.89671e+13	62432	3.03805e+08		
Total	2.03166e+13	62471			

Fig. 4 Summary of results from Kruskal Wallis test for the three approaches. The proposed method (p – value Site0) has the least Chi square error and the highest probability showing it passes the test even at high significance levels

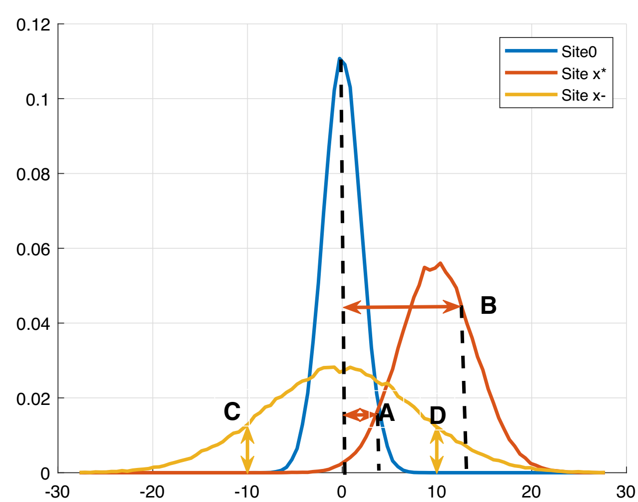

Fig. 5 Normalized histogram of Site0 in blue and normalized histogram of Site x in orange

is greater than this upper bound is identified as an issue site. The derivation of this upper bound is as follows:

$$
\begin{aligned}
WBD(x) &= \sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}\left|h_m^x - h_m^0\right| \\
WBD(x) &\le \max\left(\sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}h_m^x, \sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}h_m^0\right) \\
WBD(x) &\le max(Z, Q)
\end{aligned}
\tag{11}
$$

where $Z = \sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}h_m^x$ and $Q = \sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}h_m^0$.

As $B \to \infty$ and bin intervals get smaller, the summation term in Eq. (11) can be treated as an integral, and the Z and Q can be computed as follows.

$$
\begin{aligned}
Q &= \sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}h_m^0 \\
Q &= \int \frac{(y_m-\mu_0)^2}{\sigma_0^2}p_{y_m}\,dy_m = \frac{\sigma_0^2}{\sigma_0^2} = 1
\end{aligned}
\tag{12}
$$

A similar analysis can be done for Z.

$$
\begin{aligned}
Z &= \sum_{m=1}^{B} \frac{(y_m-\mu_0)^2}{\sigma_0^2}h_m^x = \sum_{m=1}^{B} \frac{(y_m-\mu_x+\mu_x-\mu_0)^2 h_m^x}{\sigma_0^2} \\
Z &= \sum_{m=1}^{B} \frac{(y_m-\mu_x)^2 h_m^x}{\sigma_0^2} + \sum_{m=1}^{B} \frac{(\mu_0-\mu_x)^2 h_m^x}{\sigma_0^2} \\
&\quad + \sum_{m=1}^{B} \frac{2(\mu_0-\mu_x)(y_m-\mu_x)h_m^x}{\sigma_0^2} \\
Z &= \int \frac{(y_m-\mu_x)^2 p_{y_m}dy_m}{\sigma_0^2} + \int \frac{(\mu_0-\mu_x)^2 p_{y_m}dy_m}{\sigma_0^2} \\
&\quad + \int \frac{2(\mu_0-\mu_x)(y_m-\mu_x)p_{y_m}dy_m}{\sigma_0^2} \\
Z &= \frac{\sigma_x^2}{\sigma_0^2} + \frac{(\mu_0-\mu_x)^2}{\sigma_0^2}
\end{aligned}
\tag{13}
$$

where μ_x, σ_x^2 are the true mean and variance for each site x measurements of parameter j. It should be noted that μ_0 and σ_0 are unbiased estimates which are computed using Eqs. (5) and (6). Even if all the sites follow the Site0 distribution, μ_x and σ_x will not necessarily equal μ_0 and σ_0 respectively and this is because of the finite number of DUTs measured at each site. The deviations from the desired values are, however, expected to fall within tolerable intervals centered around μ_0 and σ_0. In other words, for each good Site x, we expect that

$$\mu_x \in \left[\mu_0 - \delta, \mu_0 + \delta\right] \tag{14}$$

From the theory of hypothesis testing [27], we can then determine the required δ for a given significance level α. i.e.

$$P\left(\left|\mu_x - \mu_0\right| > \delta\right) = \alpha \tag{15}$$

Given that the maximum number of samples/ test measurements from the sites in Site0 is N_T, then we can approximate δ to be

$$\delta = \frac{\sigma_0 k}{\sqrt{N_T}}, k = norminv(1 - \alpha/2) \tag{16}$$

where *norminv* refers to the inverse of the cumulative distribution function of the Gaussian distribution [27].

A similar approach can be adopted for σ_x; the standard deviation for each site x. Each σ_x^2 is treated as an estimate of σ_0^2. This means that for

$$0 \le l \le u, P\left(\frac{N_T - 1}{u}\sigma_0^2 \le \sigma_x^2 \le \frac{N_T - 1}{l}\sigma_0^2\right) = \alpha \tag{17}$$

We can then find l and u as $u = chi2inv\left(1 - \alpha/2, N_T - 1\right)$ and $l = chi2inv\left(\alpha/2, N_T - 1\right)$ where *chi2inv* refers to the inverse cumulative distribution function of the Chi-square random variable [27]. The smaller α is, the tighter the bounds. Inherent in this approach is the assumption that the underlying distribution is Gaussian. Whiles this is not necessarily true, it is sufficient for our purposes and works no matter the distribution of the parameter of interest.

Consequently, we have

$$\overline{\mu_x} = \mu_0 + \delta \tag{18}$$

$$\overline{\sigma_x^2} = \frac{N_T - 1}{l}\sigma_0^2 \tag{19}$$

$$Upper\ Bound = \max\left(\frac{\overline{\sigma_x^2}}{\sigma_0^2} + \frac{\left(\mu_0 - \overline{\mu_x}\right)^2}{\sigma_0^2}, 1\right) \tag{20}$$

For any parameter j, we specify an α and compute the corresponding μ_0, σ_0 and consequently the upper bound. Any site whose WBD score is greater than this upper bound is identified as an issue site. It should be noted in this case that the bound is capped at 1. We cannot go below that which means reducing α beyond a specific limit provides no further benefit.

5 Results

This section demonstrates the effectiveness of the proposed algorithm. We apply the proposed method to simulation data in Part A and real ATE volume test data in Part B.

5.1 Results on Multi-site Measurement Testing Simulation Data

A 14-bit SAR ADC is modeled in MATLAB®, and each ADC has been modeled to have capacitor mismatches which account for the unavoidable intrinsic errors one would expect due to process variations in the wafer. The simulation setup consists of 60 sites with 200 ADCs per site. Measurement errors have been modeled to stem from noise on voltage supplies and small nonlinearities present in the voltage supplies. Out of the 60 sites, some of the sites were randomly selected to have higher noise variance and nonlinearity coefficients which models the typical site-to-site variations observed in reality. We consider two parameters: Offset error and Gain error. We apply the algorithm with a significance level, α, of 0.05, which is standard in most statistical studies. We focus on the sites that violate the flagging criteria and sites with a relatively higher WBD score (marginally bad sites).

Test measurement data are usually visualized via boxplots, and we follow the same approach here to display the results of the algorithm. We begin by looking at Offset Error.

From Fig. 6A, we see that some of the identified issue sites (sites in red) have a wider bar than those of the non-issue sites. Some also display a median shift when compared to Site0. In Fig. 6B, we see that these issue sites have a WBD score greater than that of the non-issue sites and bigger than the flagging or upper bound of 1. The marginally bad sites also display similar characteristics but without violating the bounds.

We apply the QQ [16] algorithm to the same measurement data, and the results are displayed In Fig. 7 for a flagging boundary of 3. The boundary is then changed to 2, and the results are displayed In Fig. 8. In addition to the flagging boundary not being obvious, it is observed that the number of sites identified as issue sites can change drastically with just a small change in the boundary. This could lead to test engineers wrongfully pausing production tests over an issue that may not be severe and probably due to wafer-to-wafer variations.

From Fig. 9A, we realize that the issue sites, in this case, have the same median as the non-issue sites, but there is a difference in the length of the boxplot of each issue site compared to that of the non-issue sites. This is indicative of a higher spread/variance in the issue sites. The algorithm successfully identifies these sites, as shown In Fig. 9B. The sites that have been identified as marginally bad also display such long bars in comparison to Site0. The QQ scores are displayed In Fig. 10. Note that changing the boundary from 3 to 2 does not increase the number of issue sites identified in this case.

These two examples demonstrate our algorithm's strength in identifying two possible forms of variations; median shifts and an increase in the variance. These are not necessarily the only variations the algorithm can capture. It also demonstrates the robustness and consistency of the flagging boundary introduced in this paper.

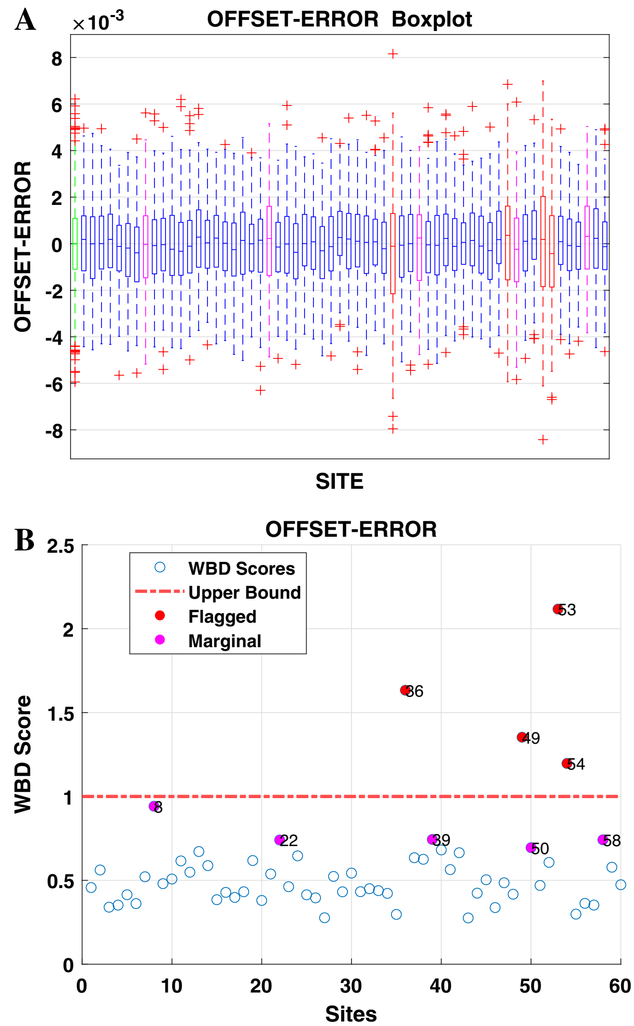

Fig. 6 **A** Boxplot for Offset Error with Issue sites in red, non-issue sites in blue and marginally bad sites in magenta. Site0 is in green. **B** Scatter plot of WBD scores for each site based on Offset Error measurements taken by the sites

5.2 Results on Industry Multi-site Volume Silicon Measurement Test Data

The algorithm is applied to ATE volume test measurement data obtained from Texas Instruments. The parameter considered is the ADC minimum integral nonlinearity, which has an unsymmetrical distribution. For confidentiality reasons, we do not reveal the actual number of sites, but we present selected measurement/test sites together with Site0. We also cannot display the range of the measurement values. Just as was done in the case of the simulation data, we display the results via boxplots. The algorithm is applied to this dataset with α = 0.05, and the results are displayed In

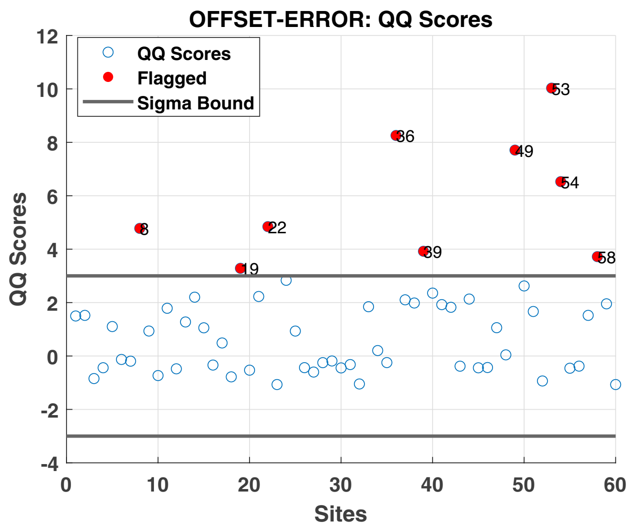

Fig. 7 Scatter plot of QQ scores for each site based on Offset Error measurements taken by the sites. Flagging boundary is ± 3

Fig. 11A with issue sites in red, non-issue sites in blue, and marginal issue sites in magenta.

From Fig. 11B, we observe that the sites flagged by the algorithm display the most shift in the median. The marginal issue sites do not display any obvious variations and are only marginal because they have the largest scores amongst sites that do not violate the flagging boundary. Test engineers are free to further examine if these sites need extra attention but the results, which is further corroborated by the boxplot, shows there is no need for concern.

The QQ algorithm is also applied to this same set of measurements, and the results are displayed in Figs. 12 and 13 for a boundary of 3 and 2, respectively.

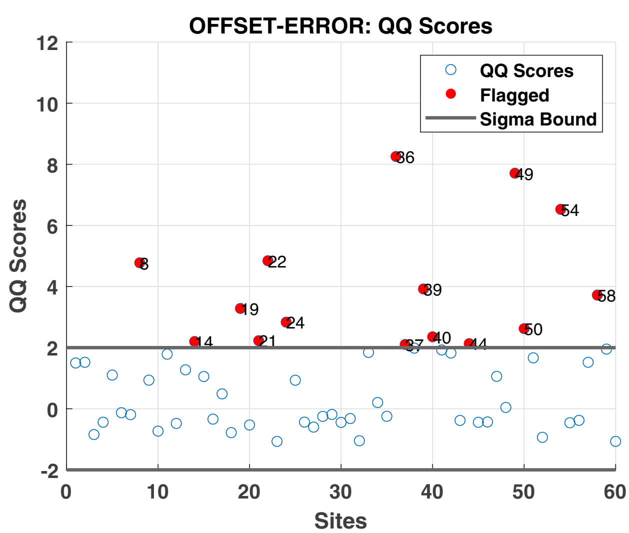

Fig. 8 Scatter plot of QQ scores for each site based on Offset Error measurements taken by the sites. Flagging boundary is ± 2

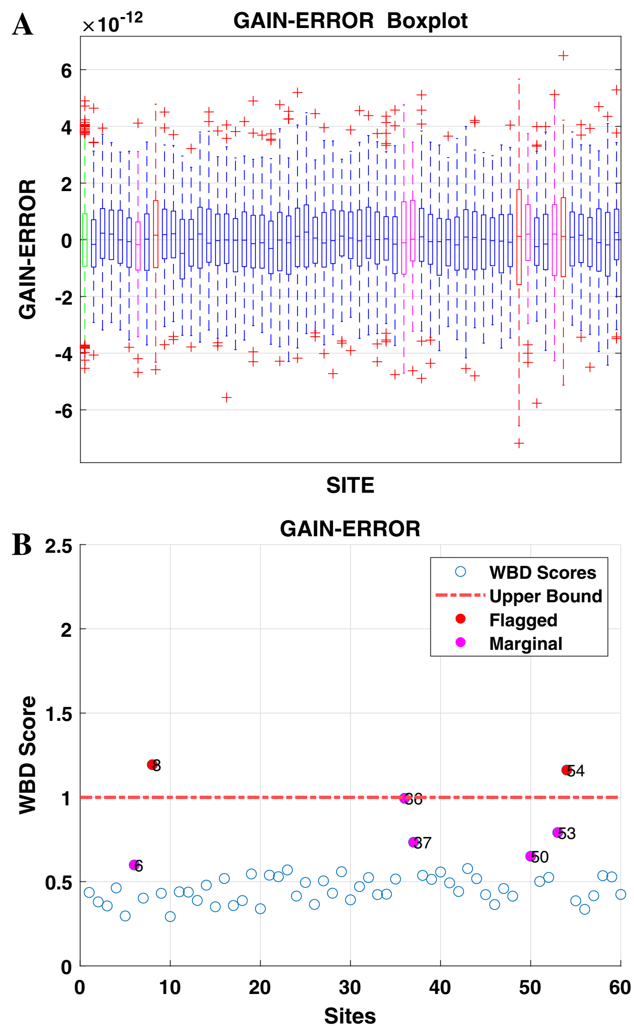

Fig. 9 **A** Boxplot for Gain Error with Issue sites in red, non-issue sites in blue and marginally bad sites in magenta. Site0 is in green. **B** Scatter plot of WBD scores for each site based on Gain Error measurements taken by the sites

For a boundary of 3, the QQ algorithm fails to identify sites 32 and 47, which are obvious issue sites, consequently implying them to be good enough. The boxplot In Fig. 11A clearly contradicts this inference. The boundary must be lowered to ± 2 before identifying any issue sites. The QQ algorithm flags the first site even though, from the boxplot, it is not an issue site. This, however, leads to a number of false positives. This is clearly seen by visually comparing site 1 to the reference.

6 Future Work

While important, identifying issue sites in the multi-site testing system for analog and mixed-signal devices is only the first step toward increasing the number of test

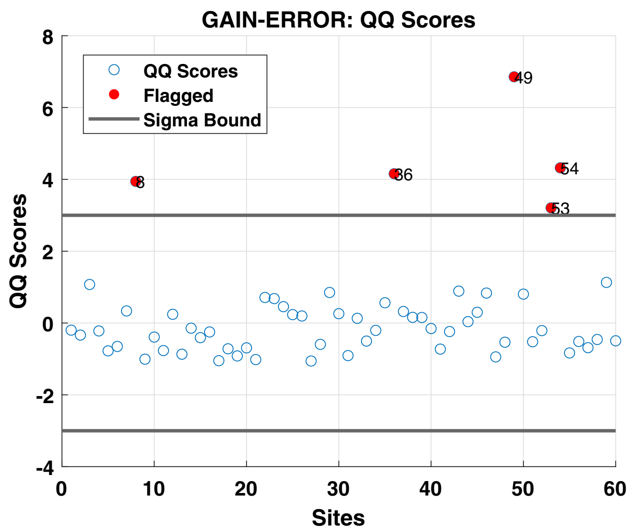

Fig. 10 Scatter plot of QQ scores for each site based on Gain Error measurements taken by the sites. Flagging boundary is ±3

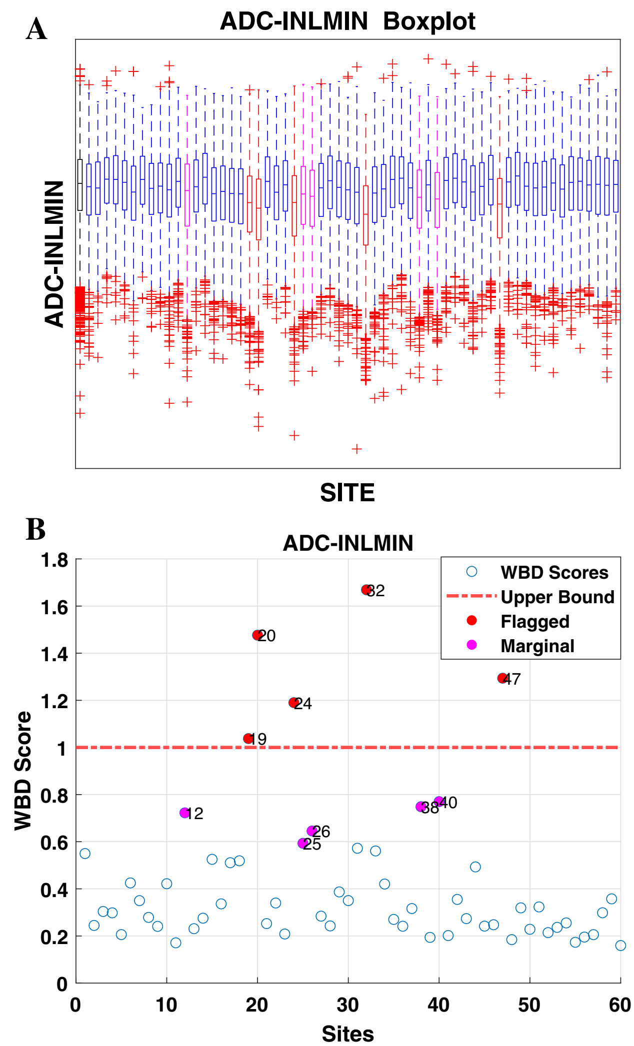

Fig. 11 **A** Boxplot for ADC minimum Integral non-linearity error with Issue sites in red, non-issue sites in blue and marginally bad sites in magenta. Site0 is in green. **B** Scatter plot of WBD scores for each site based on ADC minimum Integral non-linearity error measurements taken by the sites

sites. The identified issue sites must be worked on to remove the site-induced errors. To the best of our knowledge, only [23, 28, 29] propose a method to calibrate for the errors introduced by test sites during multi-site testing. The proposed method is run offline and has been shown to identify systematic errors introduced by test sites successfully. There is room for more novel calibration methods that can be run online. In other words, there is a need for methods that identify the induced systematic error whiles the test is being done.

Another significant problem has to do with identifying the root causes of site-to-site variations. Once issue sites are identified, it would be beneficial to know which hardware variations or possibly testing schemes contributed to the errors introduced by these issue sites. This can help test engineers design better test programs and better inform them on how best test boards can be designed to minimize site-to-site variations.

All the methods discussed so far assume a smooth wafer pattern and no cross-wafer variation. However, it has been observed in [30, 31] that wafers exhibit a mixture of smooth and systematic stepper patterns. The effects of these non-smooth patterns on site-to-site variations need to be investigated. Die outlier screening is also a well-known problem in the semiconductor field. Site-to-site variations can cause false labeling of dies which could lead to yield loss and test escapes. There is, therefore, a need for an issue site-aware outlier identification algorithm.

7 Conclusion

This paper investigates site-to-site variations inherent in massive multi-site testing systems and presents an algorithm to identify sites that suffer from extreme site-to-site variations. The proposed weighted bin difference method detects issue sites by comparing the weighted bin scores for each test site to an analytically derived boundary. A mathematical derivation of the upper bound of the WBD scores is provided. A method to estimate the reference distribution from multi-site test data is also reviewed and modified. The algorithm was applied to simulation data and real ATE volume measurement data, and its effectiveness was clearly demonstrated. Comparison of the

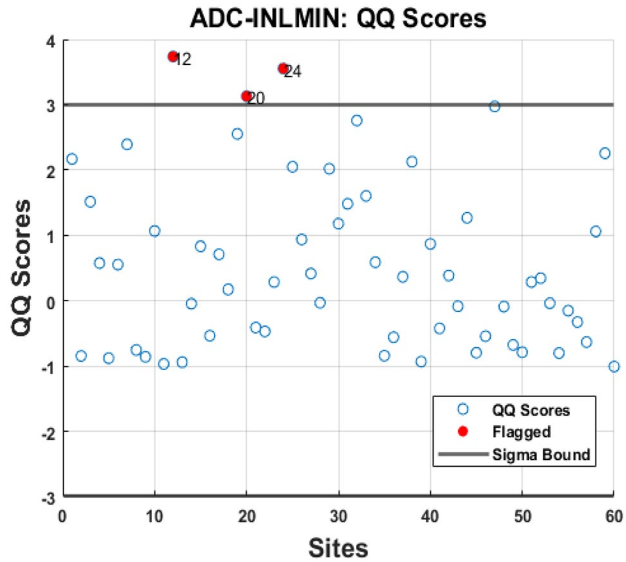

Fig. 12 Scatter plot of QQ scores for each site based on ADC minimum Integral non-linearity error measurements taken by the sites. Flagging Boundary of ± 3

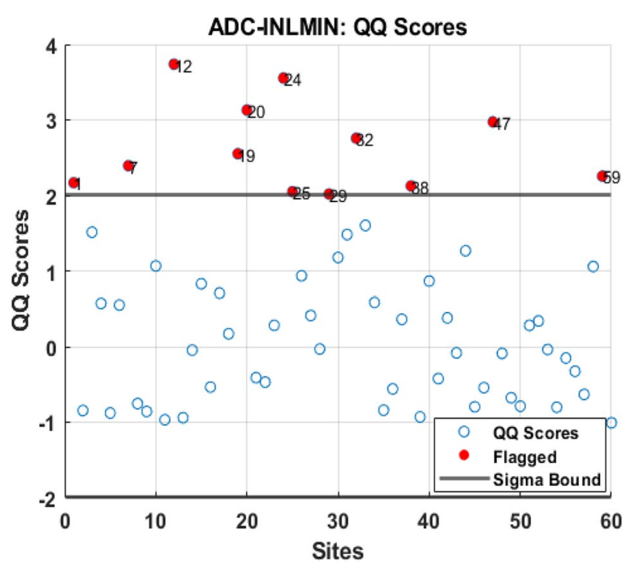

Fig. 13 Scatter plot of QQ scores for each site based on ADC minimum Integral non-linearity error measurements taken by the sites. Flagging Boundary of ± 2

proposed method was compared to the Quantile-Quantile algorithm and was shown to provide superior results.

Once issue sites have been identified, testing can be temporarily paused to identify the source of the errors. Mechanical hardware calibration can be applied. Other proposed approaches to calibrating the measurements at the issue sites could also be considered [28].

Acknowledgments This work was supported in part by Texas Instruments and in part by the Semiconductor Research Corporation.

Data Availability The data that support the findings of this study are available from Texas Instruments Inc, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however, available from the authors upon reasonable request and with permission of Texas Instruments Inc.

Declarations

Conflict of Interest I certify no actual or potential conflict of interest in this article.

Competing Interests The authors declare that they have no known competing financial interests or personal relationships that could influence the work reported in this paper.

References

1. Nourani M, Chin J (2002) Testing high-speed SoCs using low-speed ATEs. Proceedings 20th IEEE VLSI Test Symposium (VTS 2002), Monterey, CA, USA, pp 133–138. https://doi.org/10.1109/VTS.2002.1011124
2. Chaganti SK, Chen T, Zhuang Y, Chen D (2018) Low-cost and accurate DAC linearity test with ultrafast segmented model identification of linearity errors and removal of measurement errors (uSMILE-ROME). 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, pp 1–6. https://doi.org/10.1109/I2MTC.2018.8409877
3. Duan Y, Chen T, Chen D (2016) Low-cost dithering generator for accurate ADC linearity test. 2016 IEEE International Symposium on Circuits and Systems (ISCAS), Montréal, QC, Canada, pp 1474–1477. https://doi.org/10.1109/ISCAS.2016.7527536
4. Chen T, Chen D (2015) Ultrafast stimulus error removal algorithm for ADC linearity test. 2015 IEEE 33rd VLSI Test Symposium (VTS), Napa, CA, USA, pp 1–5. https://doi.org/10.1109/VTS.2015.7116249
5. Chen T et al (2020) A low-cost on-chip built-in self-test solution for adc linearity test. IEEE Trans Instrum Meas 69(6):3516–3526. https://doi.org/10.1109/TIM.2019.2936716
6. Manzone A, Bernardi P, Grosso M, Rebaudengo M, Sanchez E, Reorda MS (2005) Integrating BIST techniques for on-line SoC testing. 11th IEEE International On-Line Testing Symposium, French Riviera, France, pp 235–240. https://doi.org/10.1109/IOLTS.2005.38
7. Khoo VC (2014) A case study on the effectiveness of multi-sites test handler to improve of production output. IOSR J Eng 4(4):47–59. https://doi.org/10.9790/3021-04454759
8. Rivoir J (2003) Lowering cost of test: parallel test or low-cost ATE? Proceedings of the 7th International Conference on Properties and Applications of Dielectric Materials (Cat No 03CH37417) ATS-03, Xi'an, China, pp 360–363. https://doi.org/10.1109/ATS.2003.1250837
9. reducing_test. Reducing IC test costs through multisite and concurrent testing. https://www.techdesignforums.com/practice/technique/reduce-test-costs-multisite-concurrent-testing/. Accessed 3 Oct 2020
10. Kim H, Lee Y, Kang S (2015) A novel massively parallel testing method using multi-root for high reliability. IEEE Trans Reliab 64(1):486–496. https://doi.org/10.1109/TR.2014.2336395

11. Han D, Lee Y, Kang S (2014) A new multi-site test for system-on-chip using multi-site star test architecture. ETRI J 36(2):293–300. https://doi.org/10.4218/etrij.14.0113.0469
12. Lehner T, Kuhr A, Wahl M, Bruck R (2014) Site dependencies in a multisite testing environment. 2014 19th IEEE European Test Symposium (ETS), Paderborn, Germany, pp 1–6. https://doi.org/10.1109/ETS.2014.6847808
13. Kian LB (2006) Test cost saving and challenges in the implementation of /spl times/6 and /spl times/8 parallel testing on freescale 16-bit HCS12 microcontroller product family. Third IEEE International Workshop on Electronic Design, Test and Applications (DELTA'06), Kuala Lumpur, Malaysia, pp 7–82. https://doi.org/10.1109/DELTA.2006.85
14. Khasawneh Q, Dworak J, Gui P, Williams B, Elliott AC, Muthaiah A (2018) Real-time monitoring of test fallout data to quickly identify tester and yield issues in a multi-site environment. 2018 IEEE 36th VLSI Test Symposium (VTS), San Francisco, CA, pp 1–6. https://doi.org/10.1109/VTS.2018.8368661
15. Says AI (2021) Coping with parallel test site-to-site variation. Semicond Eng. https://semiengineering.com/coping-with-parallel-test-site-to-site-variation/. Accessed 5 May 2022
16. Farayola PO, Chaganti SK, Obaidi AO, Sheikh A, Ravi S, Chen D (2020) Quantile – quantile fitting approach to detect site to site variations in massive multi-site testing. 2020 IEEE 38th VLSI Test Symposium (VTS), pp 1–6. https://doi.org/10.1109/VTS48691.2020.9107616
17. Farayola PO, Chaganti SK, Obaidi AO, Sheikh A, Ravi S, Chen D (2021) Detection of site to site variations from volume measurement data in multisite semiconductor testing. IEEE Trans Instrum Meas 70:1–12. https://doi.org/10.1109/TIM.2021.3051666
18. Wold S, Esbensen K, Geladi P. Principal component analysis. p 16
19. Rosenberg D. 1 distances between probability measures. p 4
20. Acharya J, Luo M. An algorithmic and information-theoretic toolbox for massive data. p 5
21. Butler KM, Nahar A, Daasch WR (2016) What we know after twelve years developing and deploying test data analytics solutions. 2016 IEEE International Test Conference (ITC), pp 1–8. https://doi.org/10.1109/TEST.2016.7805844
22. Ho Y-C, Zhao Q-C, Jia Q-S (2007) Ordinal optimization: soft optimization for hard problems. Springer, New York
23. Farayola PO, Bruce I, Chaganti SK, Sheikh A, Ravi S, Chen D (2021) Massive multisite variability-aware die distribution estimation for analog/mixed-signal circuits test validation. 2021 16th International Conference on Design Technology of Integrated Systems in Nanoscale Era (DTIS), pp 1–6. https://doi.org/10.1109/DTIS53253.2021.9505144
24. Bruce I et al (2021) An ordinal optimization-based approach to die distribution estimation for massive multi-site testing validation: a case study. 2021 IEEE European Test Symposium (ETS), pp 1–4. https://doi.org/10.1109/ETS50041.2021.9465402
25. Steinley D (2006) K-means clustering: a half-century synthesis. Br J Math Stat Psychol 59(1):1–34. https://doi.org/10.1348/000711005X48266
26. Kruskal William H, Allen Wallis W (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47(260):583–621. https://doi.org/10.1080/01621459.1952.10483441
27. Mood AM (1920) Introduction to the theory of statistics, 3rd edn. http://archive.org/details/in.ernet.dli.2015.132521. Accessed 21 Feb 2021
28. Farayola PO et al (2021) Systematic hardware error identification and calibration for massive multisite testing. IEEE International Test Conference (ITC), Anaheim, CA, USA, pp 304–308. https://doi.org/10.1109/ITC50571.2021.00042
29. Farayola PO, Bruce I, Chaganti SK, Sheikh A, Ravi S, Chen D (2022) The least-squares approach to systematic error identification and calibration in semiconductor multisite testing. 2022 IEEE 40th VLSI Test Symposium (VTS), San Diego, CA, USA, pp 1–7. https://doi.org/10.1109/VTS52500.2021.9794216
30. Daasch WR, McNames J, Bockelman D, Cota K (2000) Variance reduction using wafer patterns in I/sub ddQ/ data. Proceedings International Test Conference 2000 (IEEE Cat. No.00CH37159), Atlantic City, NJ, USA, pp 189–198. https://doi.org/10.1109/TEST.2000.894206
31. Daasch WR, McNames J, Madge R, Cota K (2002) Neighborhood selection for I/sub DDQ/ outlier screening at wafer sort. IEEE Des Test Comput 19(5):74–81. https://doi.org/10.1109/MDT.2002.1033795

Isaac Bruce (Student Member, IEEE) received his B.Sc. degree in Electrical and Electronics Engineering from the Kwame Nkrumah University of Science and Technology. He then proceeded to obtain his Master's degree in Electrical Engineering from the University of Notre Dame. He is currently pursuing a doctorate degree in Electrical and Computer Engineering with emphasis on Analog and Mixed Signal Circuit design and testing at the Iowa State University under the tutelage of Dr. Degang Chen. He is currently focused on solving issues in multisite testing, a relatively novel technique in analog circuit testing that drastically reduces test time and consequently the test cost. His research interests include temperature sensors, low cost and low area data converters and statistical and machine learning algorithms for analog and mixed signal applications.

Praise O. Farayola (Student Member, IEEE) received the B.Sc. degree in electrical and electronics engineering from the University of Ibadan, Ibadan, Nigeria, in 2015. He is currently pursuing the Ph.D. degree in electrical engineering with Iowa State University, Ames, IA, USA. In 2020 (summer), he was an Analog Circuit Design Intern with Texas Instruments, Tucson, AZ, USA. He worked on current feedback amplifier design and verification with the High-Speed Operational Amplifier Team, Tucson, Arizona. His research interest includes multisite testing, test hardware error identification and calibration, low-cost high-precision ADC testing, analog circuit design, and fault coverage analysis for analog circuit test.

Shravan K. Chaganti (Member, IEEE) received the B.Tech. degree in electronics and communication engineering from IIT Roorkee, Roorkee, India, in 2014, and the Ph.D. degree in electrical engineering from Iowa State University, Ames, IA, USA in 2020. In 2016, he was an Intern with Skyworks Solutions, Cedar Rapids, IA, where he was involved in analog circuit design, verification, and test. In 2018, he interned at Texas Instruments, Sugar Land, TX, USA, where he worked on data converter linearity test time reduction. He is currently working with the C2000 MCU Team on new product development and test cost reduction. His research interests include low-cost high-precision SoC testing, analog circuit design-for-test, functional safety for automotive, and analog fault coverage simulation and improvement. Dr. Chaganti received the IEEE Ned Kornfield Best Paper Award in 2018.

Abalhassan Sheikh received the bachelor's and M.Sc. degrees in electrical engineering from the University of Houston, Houston, TX, USA, in 2003 and 2007, respectively. In 2007, he joined Texas Instruments, working on product and test engineering of SoC targeted for the Automotive markets. In 2010, he joined the C2000 32-bit MCU Team working on new product development of SoC comprising numerous IP's including ADC and DAC where his area of interests are in test cost reduction and data analysis. Mr. Sheikh was elected Member, Group of Technical Staff in 2016. He received the IEEE Ned Kornfield Best Paper Award in 2018.

Srivaths Ravi (Senior Member, IEEE) received the B.Tech. degree in electrical and electronics engineering from IIT Madras, Chennai, India, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, USA. He is the Hardware Development Manager for the C2000 microcontrollers business unit of Texas Instruments. He currently manages C2000's global RESEARCH AND DEVELOPMENT organization responsible for end-to-end silicon development— product strategy and roadmap, IP/SoC definition and architecture/specifications, new product silicon design, product/test engineering, as well as sustaining engineering activities. At TI, he has also played multiple leadership roles including managing SoC platform, design, design-for-test (DFT), and customer-facing system application teams. He has also been a Senior Member of TI's Technical Staff (SMTS). He was a Research Staff Member with NEC Laboratories America, Princeton, where he worked in the areas of low power EDA and embedded security. He co-architected and designed the mobile security processor MOSES used in NEC's cellphone chips, which was also awarded the NEC Laboratories Technology Commercialization Award in 2005. He also held a visiting research collaborator position at Princeton University. He has over hundred publications in the form of conferences/journal contributions and book chapters, and has received 7 best paper awards and 13 granted patents. He has served on the organizing/program committees of various leading conferences, including as the General Chair of VLSI Design Conference 2010, Program Chair of Asian Test Symposium 2011, Program Chair of VLSI Test Symposium 2016–2017, and General Chair of TI Technical Conference in 2018.

Degang Chen (Fellow, IEEE) received the B.S. degree in instrumentation and automation from Tsinghua University, Beijing, China, in 1984, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 1988 and 1992, respectively. His industry experience includes the Beijing Institute of Control Engineering, Beijing, China, from 1984 to 1986; Faculty Fellow with the Boeing Company, Seattle, WA, USA, in 1999; Maxim Integrated, Dallas, TX, USA, in 2001; and Texas Instruments (TI), Tucson, AZ, USA, in 2011, 2012, and 2014. His academia experience includes the John R. Pierce Instructor of electrical engineering with the California Institute of Technology, Pasadena, CA, in 1992. Since 1992, he has been with Iowa State University, Ames, IA, USA, where he is currently a Professor of Electrical and Computer Engineering and holds the Jerry Junkins Chair with the College of Engineering. He has authored and coauthored over 300 refereed journal articles, conference papers, and patents. His current research interests include analog and mixed-signal integrated circuit design and testing, integrated circuit sensor design for life-time reliability and functional safety, and high-accuracy test without requiring high-accuracy instrumentation. Dr. Chen and his students' work has received many best paper awards and other honors, including the IEEE Ned Kornfield Best Paper Awards in 2013, 2014, and 2019, and the IEEE EJ McCluskey Best Ph.D. Dissertation Awards in 2017 and 2019, in recent years.