

# How to measure test repeatability when stability and constant variance are not observed?

Gary G. Jing\*

Edina, MN, USA

Received: 8 November 2017 / Accepted: 21 September 2018

**Abstract.** Passive intermodulation (PIM) is a critical measurement for radio frequency (RF) communication networks. Yet PIM measurement inherently has very poor repeatability, which makes product assessment unreliable. The RF industry struggles with the issue since there are no known solutions. With the increasing demand for low-PIM performance, there are pressing demands to address the challenge. Two fundamental problems make traditional gage R&R study invalid for PIM: (1) PIM in nature is unstable and unrepeatable; (2) PIM measurement has inherently inconstant variance at different PIM levels, which is primarily due to limited capability of PIM analyzer. This resulted in several less-known issues significantly impacting the estimation of PIM test repeatability, including sample selection, one-sided spec and differences between test R&R and gage R&R. The paper proposed two fundamental changes when studying R&R of PIM test or tests in general violating constant variance assumptions: (1) sample selection; (2) what measurement to use to better estimate and represent the test repeatability. Special sampling is proposed to minimize the impact of inconstant variance. A more direct R&R measurement, margin of error (MOE), also known as study variation, is proposed to replace traditional gage R&R metrics to more meaningfully represent PIM test R&R. Several statistically based techniques to improve the repeatability and reliability of PIM measurement are also discussed. The study and proposed solutions apply to not only PIM test but also tests in general violating constant variance assumptions.

**Keywords:** gage R&R / margin of error / measurement system analysis (MSA) / constant variance / PIM / repeatability and reproducibility / study variation / test R&R / reliability

## 1 Introduction

Passive intermodulation (PIM) is a critical performance metric for radio frequency (RF) communication networks. With the advancement of technology and high competition for more bandwidth and higher data rates, low-PIM performance has become more and more important. Yet, due to the nature of PIM, it has been very challenging to reliably measure it. This paper focuses on studying the repeatability and reproducibility (R&R) of PIM measurement, which is typically referred to as PIM test. RF industry struggles with poor R&R on PIM measurement, which makes PIM-based product assessment unreliable.

This paper will start with literature review and discussion on PIM test and traditional gage R&R study in the next two sections. They will be followed by the discussion on unique challenges when measuring repeatability of PIM test, which will be followed by proposals on how to handle these challenges. The paper will end with summary and conclusions, as well as implications and influences.

## 2 Background information about PIM

### 2.1 What is PIM

Passive intermodulation (PIM) is a form of intermodulation distortion that occurs in passive components such as antennas, cables, connectors, or duplexers with two or more high-power input signals. It is the generation of interfering signals caused by nonlinearities in the mechanical components of a wireless system. As illustrated in Figure 1, two signals (amplitude modulation  $F_A$ ,  $F_B$ ) mix together to produce sum and different signals and products within the same band, causing interference. PIM is measured as the relative difference between the amplitude of intermodulation product and the amplitude of the carrier. PIM in the transmission path degrades the quality of the wireless communication system.

### 2.2 Characteristics of PIM

Figure 2 displayed five rounds of continuous monitoring of PIM of a device in 30 s duration. As is easily seen, PIM is not stable in nature.

\* Corresponding author: [Gary\\_jing@hotmail.com](mailto:Gary_jing@hotmail.com)

**2.3 How is PIM measured**

PIM measurement is typically done through sampling the continuous data. A commonly used practice in the industry is to, for example,

- take five sweeps each with a different frequency;
- take five readings (data points) from each sweep;
- take the peak PIM and standard deviation (STV) out of five data points from each sweep as the PIM measurements for each sweep;
- take the highest peak PIM and the highest STV from five sweeps as the PIM measurements for this round of testing;
- fail the product if the peak PIM > -153 dBc or STV > 3 dB (-153 dBc and 3 dB are specs for illustration).

The reason peak is used is that the industry typically wants to see the worst-case performance. Yet peak in nature is an unstable statistic for data, not to mention that the data themselves are not stable.

**2.4 PIM test system**

Holes' discussion [1] illustrated a typical PIM test system setup as shown in Figure 3. DUT is the device (product) under testing. Analyzer is the tester (gage) extracting the PIM reading from DUT.

**2.5 A widely known problem or limitation in the RF industry**

Figure 4 displayed the typical measurement error of analyzer published by IEC 62037-1:2012 [2]. The horizontal axis represents the delta between DUT PIM (True PIM) and noise floor (System PIM). The vertical axis represents the possible measurement error embedded in the analyzer, which can be called margin of error (MOE) of the analyzer. It represents the measurement capability of the analyzer. For example, if the delta is 10, the analyzer may produce a reading within the range of -3.1 to +2.3 dB (or simplified as +/-3 dB) from the true PIM.

Noise floor (also known as residual PIM, system PIM) represents the lowest PIM level the analyzer can detect. Usually it is below -125 dBm (or -168 dBc if the signal power is 43 dBm). As shown in the chart, once approaching the noise floor, the analyzer will no longer be able to give reliable readings, and readings can be unrealistically low. In general, readings below noise floor are considered noise effect, unreal. The industry typically expects to have a margin of at least 10 dB between the system residual PIM and PIM spec for the DUT.

**2.6 Two key properties of PIM measurement**

More discussion on PIM test details can be seen in references [1,3]. In summary, there are two key properties in PIM measurement:

- PIM in nature is unstable and unrepeatable.
- PIM measurement has inherently inconstant variance at different PIM levels. It is primarily due to the limited capability of PIM analyzer.

**3 Traditional gage R&R study**

Traditional gage R&R (GRR) study focuses on studying the repeatability and reproducibility of measurement

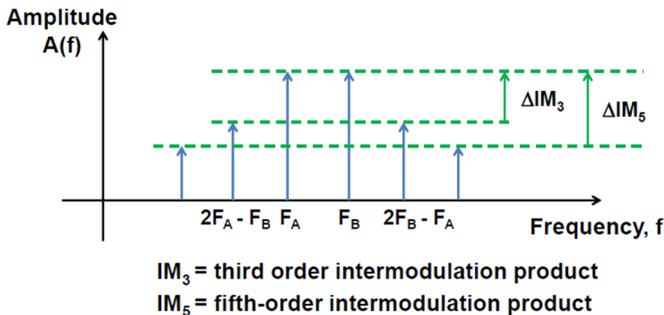


Fig. 1. Passive intermodulation.

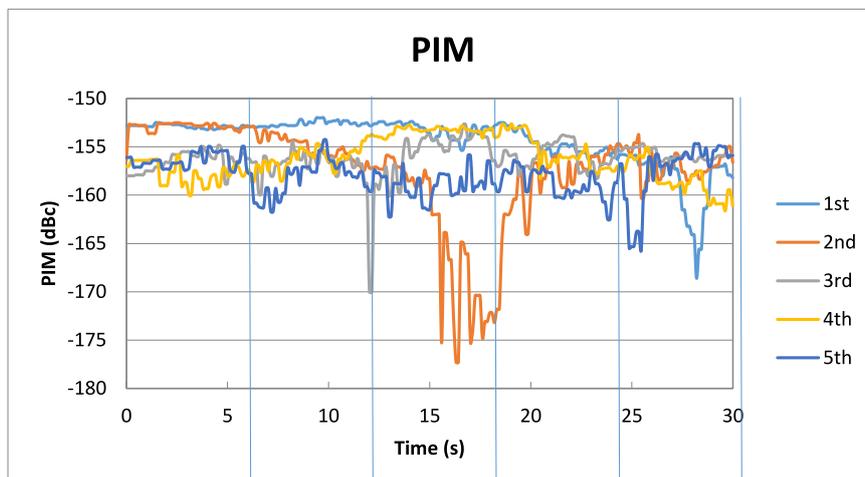


Fig. 2. Five rounds of continuous monitoring of PIM of same device in 30s duration.

system. In many cases, it is interchangeably used with measurement system analysis (MSA), since it is the most critical aspect of MSA.

Figure 5 illustrates the typical breakdown of variance components for gage R&R. Variance components have relationship shown in equations (1)–(3), where  $\sigma^2$  represents the variance of a variance component. The subscripts “ms”, “rpt”, “rpd”, “op” and “opXpart” each represents measurement system, repeatability, reproducibility, operator and operator-to-part interaction, as a variance component. This is based on the Variance Sum Law, a fundamental property

of variance and one of the cornerstones of statistics. For general discussion of traditional gage R&R, please refer to references [4,5]:

$$\sigma^2_{\text{total}} = \sigma^2_{\text{part}} + \sigma^2_{\text{ms}}, \tag{1}$$

$$\sigma^2_{\text{ms}} = \sigma^2_{\text{rpt}} + \sigma^2_{\text{rpd}}, \tag{2}$$

$$\sigma^2_{\text{rpd}} = \sigma^2_{\text{op}} + \sigma^2_{\text{opXpart}}. \tag{3}$$

### 3.1 Typical gage R&R metrics

Table 1 displayed the typical gage R&R metrics for variable measurement and expectations established by Automotive Industry Action Group (AIAG) [6]. There are some variations in expectations in the industry, but AIAG guideline is most commonly followed as a golden standard.

In Table 1, VarComp refers to variance component; TV refers to total variation (of data) in the form of  $6 \times \text{STV}$ ; P or SV, also called study variation, refers to gage variation in the form of  $6 \times \text{STV}$ . Study variation may not be widely known by people. Another term, MOE is more commonly used in daily life. MOE is typically used to describe the estimation error of survey result. Since it is widely used for polling results, typically using  $\pm 3\%$ , most people are familiar with the concept. This concept or term can also be used to describe study variation, using  $\pm 3 \times \text{STV}$  (or  $\pm \text{half SV}$ ) of estimated gage variation to call it “gage MOE”.

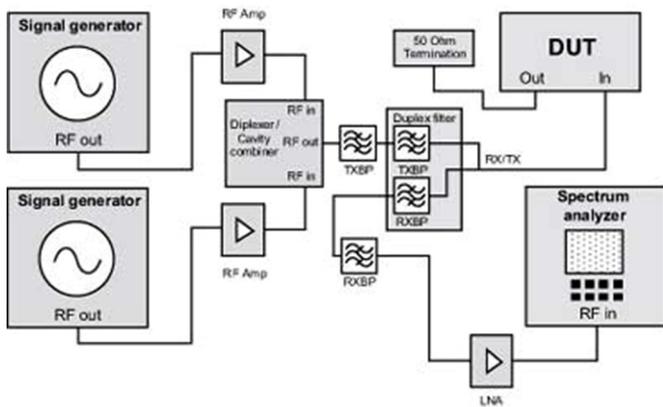


Fig. 3. System diagram of a typical PIM test setup.

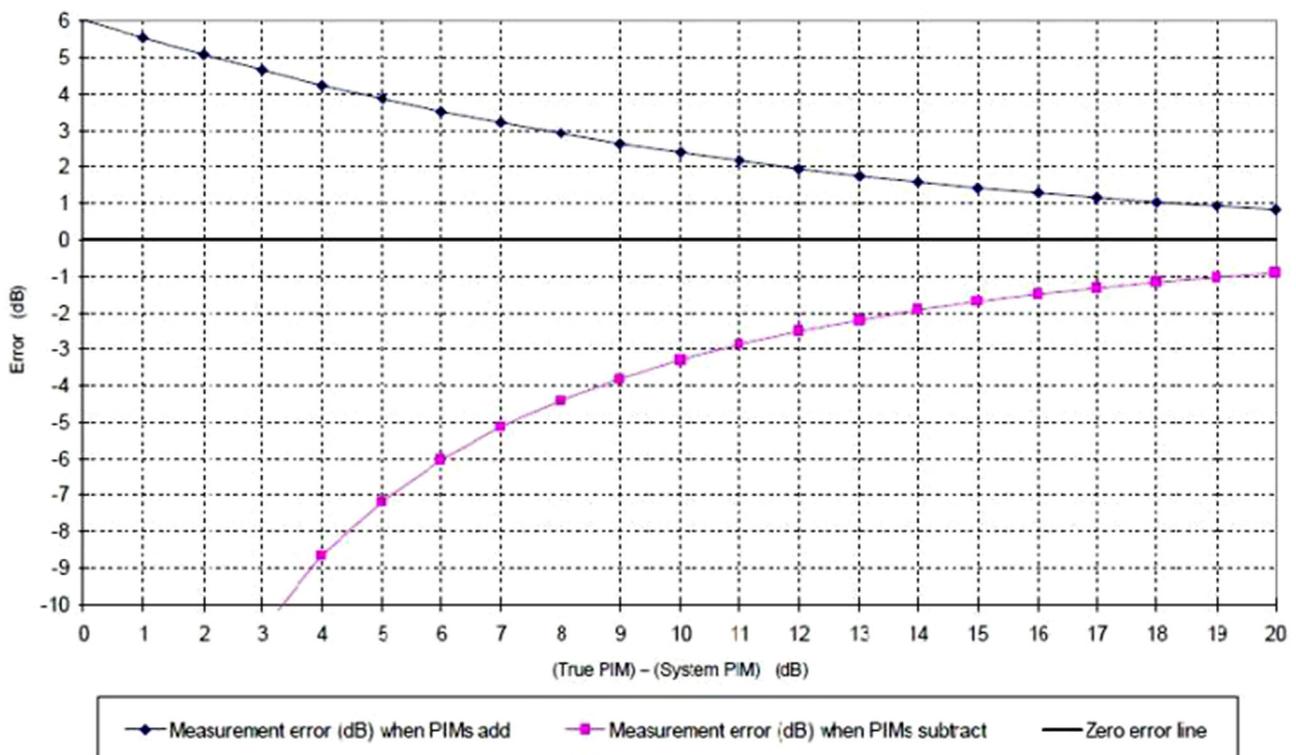
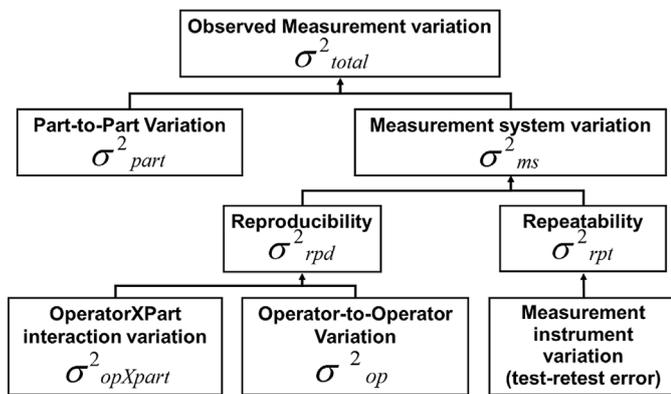


Fig. 4. Measurement error range associated with the delta between DUT PIM and residual PIM.

**Table 1.** Commonly used gage R&R metrics and expectations.

Metric	Formula	Excellent	Good	Marginally acceptable
% Contribution (%R&R, % VarComp)	$\frac{\sigma_{ms}^2}{\sigma_{total}^2} 100\%$	<1%	<4%	<9%
% Study variation (P/TV, % process)	$\frac{\sigma_{ms}}{\sigma_{total}} 100\%$	<10%	<20%	<30%
% Tolerance (P/T, SV/tolerance)	$\frac{6\sigma_{ms}}{\text{Tolerance}}$	<10%	<20%	<30%
Distinct categories (gage resolution)	$\text{round}\left(\frac{1.41\sigma_{part}}{\sigma_{ms}}\right)$	>10	>7	>5

**Fig. 5.** Gage R&R component breakdown.

Note that each gage R&R metric has a different purpose, thus they are not equivalent to and do not represent each other. All metrics need to be good to consider a gage overall good. It is possible that a gage is acceptable in some areas but not in others.

% Tolerance (P/T) is an external looking metric, focusing on gage variation against a given tolerance. It visually depicts the % that the tolerance is consumed or occupied by gage variation. If the tolerance is relaxed, an incapable gage may become capable. On the other hand, if tolerance is tightened, a capable gage may become unacceptable. P/T performance is equally dependent on the independent external tolerance as much as on gage's own merit.

The remaining three are internal looking metrics, focusing on gage's ability to differentiate parts. They are inherently connected, equivalent and representable to each other. Yet they each have a different focus and situation to use. P/TV enables visual comparison between gage variation and total variation, thus is liked by many people. Distinct category gives practical meaning of gage resolution: how much a gage can differentiate parts, by sorting them into how many groups. % Contribution directly presents the fundamental components of gage R&R, the variance breakdown and their contribution to the total variance. Among various metrics and the associated terms,

%R&R is a term often loosely used and needs special attention. Different people may refer to different metrics with this term. With the original intent of GRR analysis to breakdown variance components, the abbreviation of % repeatability and reproducibility, %R&R, should primarily refer to % contribution. But in daily use, it is also widely used to refer to P/TV. Without clarification, a reader would not know for sure which metric is being referred to, and the expectations for the two metrics are totally different. So when using %R&R, it is important to use other references to clarify which metric it is referring to.

### 3.2 Sample selection impact on gage R&R metrics

It is worth clarifying and highlighting that the gage's ability to differentiate parts depends on the selection of samples used in gage R&R study. The key here is if the total variation represents the intended use or population or not. If it does, the internal looking metrics represent the actual performance for the intended use, otherwise not. Thus, in theory, the samples should cover the whole range of intended measurement, ideally representing the population distribution. In practice, people usually rely on one of two ways to achieve

- randomly select samples from the population,
- evenly spread out samples across the intended measurement range.

Yet even with these practices, gage R&R samples may not represent the intended population, and at the beginning of the production, population distribution may not be available. In this case, the GRR samples are likely to be skewed from the population, which may distort the gage R&R metrics. If the selection of sample is too narrow, the total variation will be arbitrarily reduced and the internal looking metrics will be inflated and look worse than what they actually are. On the other hand, if the selection of samples is broader than the intended or normal use, the total variation will be arbitrarily inflated and those metrics will be deflated and look better. The sample selection is a frequently overlooked aspect which can make gage R&R estimation less reliable. This may not be widely known by people.

### 3.3 Special considerations for one-sided spec

A special note to the one-sided spec: The tolerance used in the formula in Table 1 is corresponding to two-sided spec. In many cases, there is only a one-sided spec, which can potentially cause misleading effect. For one-sided spec, a finite tolerance does not exist, thus in theory P/T does not exist.

One way that has been used to handle the one-sided situation is to do one-sided calculation. It is very similar to the traditional capability metrics:  $C_p$  vs.  $C_{pk}$ . The difference between  $C_p$  and P/T is primarily that  $C_p$  compares tolerance against process variation, while P/T compares gage variation against tolerance.  $C_p$  is two-sided analysis ignoring the location (center or mean) of the process.  $C_{pk}$  looks at each side from the center (mean) of the process, thus the location of the process is factored in  $C_{pk}$  but not in  $C_p$ . Similar approach is applied to P/T in one-sided situation. It uses half of gage variation to compare an assumed tolerance, the space between spec and center (mean) of samples. In this case, sample selection becomes a very important factor to P/T metric, which can inflate (if too close to the spec) or deflate (if too far from the spec) the metric. In two-sided situation, P/T is relatively independent from sample selection, due to the assumption of constant variance. Yet skewed sample selection will have significant impact on one-sided P/T metric.

To minimize the impact of sample selection (or location) on one-sided P/T metric, one alternative is to use a hypothetical or arbitrary tolerance. In this case, any mentioning of P/T also needs to mention the corresponding tolerance being used. I personally prefer using an arbitrary tolerance to handle one-sided spec to avoid the impact of sample selection.

Another alternative to minimize the impact of sample selection (location) is to abandon the traditional metrics and report study variation (P) directly, which is a more direct measurement of gage variation (or precision) when tolerance is not available. It can be used to compare against any given tolerance at a later time.

### 3.4 Typical gage R&R setting

A typical setting for gage R&R study for variable measurement is  $10 \times 3 \times 3$ , meaning 10 parts measured by three operators three times each. There are some details that if handled inappropriately may significantly affect, distort or even invalidate the result.

- *Part selection*: As mentioned earlier, samples should cover the whole range of intended measurement, ideally representing the population distribution. Frequently, sample parts are selected from a narrower range, which will inflate %R&R and make gage performance look worse.
- *Operator*: The traditional gage R&R study was originally designed for manual measurement of mechanical parts. Part property is typically stable. Major sources of variation are typically gage repeatability and operators' interaction with gage and part (reproducibility). Thus, operator is singled out to almost exclusively represent

reproducibility. In modern measurement, automation and automated measurement are prevalent. In this case, variation induced by human operator can be well controlled or minimized, thus no longer be a major source of measurement variation. Other sources may become more dominate. The convention of using "operator" to represent sources for reproducibility is preserved, yet the meaning of "operator" is much broader than human now. It becomes the symbol to represent anything that changes in the measurement. Frequently seen examples are different testing equipment or systems, testing stations or even facilities, fixtures or jigs that are used to perform the same test.

- *Repeat*: Gage R&R in nature is a form of DOE  $10 \times 3 (3)$ . DOE rules apply. Randomization is a critical technique to minimize the impact of noises. For gage R&R, the ideal setting is to replicate instead of repeat. In each replicate, all operators take one round of measurement of all parts in random order; replicate the activity three times.

### 3.5 Typical analyses

There are essentially two types of analyses to estimate the variation components in measurement:

- Average and Range (Xbar-R) method,
- ANOVA method.

There are variations with the Xbar-R method, but they essentially all use range (R) to estimate measurement variation for the same part, using empirical statistics and calculation similar to those for Xbar-R chart. There are many publications comparing the two methods; references [7,8] are some examples. In general, Xbar-R method is less reliable since it uses only two extreme data points to estimate R and empirical statistics (a constant as multiplier) to estimate variation based on R. Comparatively, ANOVA method is more solid, involving all data points in calculation, able to provide more breakdowns to variation components such as interactions. Our further discussion will be based on ANOVA method.

### 3.6 Default assumptions and conditions

No matter which method is used, they all have following assumptions and conditions to make the analyses valid:

- (1) *Parts are stable*: Meaning the parameter being measured does not change by itself or over time.
- (2) *Constant variance*: Meaning the variation of both parts and gage stays at the same level across the whole range of intended measurement, and over time.
- (3) *No environmental effect*: Meaning the environment of measurement is kept stable, and the impact of noise variables is minimized and nominal.

If these assumptions and conditions are violated, the result of gage R&R study becomes unreliable and questionable. Assumptions (1) and (3) are no brainers. Assumption (2) is critical for ANOVA to be valid, otherwise the result will be heavily influenced by data points that have significantly bigger variance. Same impact applies to Xbar-R method as well.

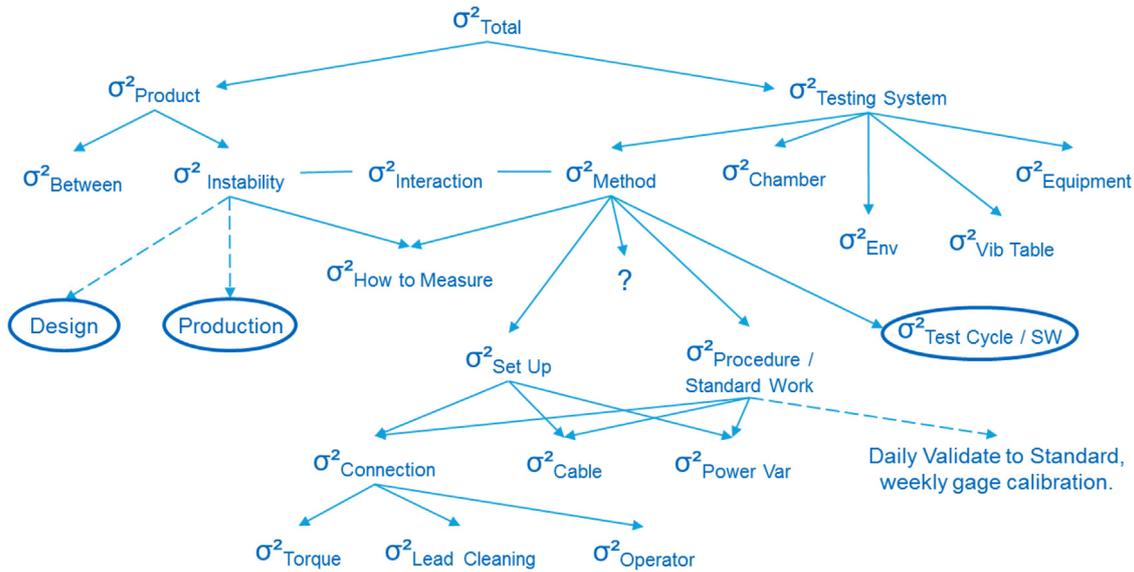


Fig. 6. PIM test R&R component breakdown example.

#### 4 Some unique challenges when measuring repeatability of PIM test

It is the case that the above three assumptions and conditions are not held in PIM test. As a result,

- the traditional gage R&R analysis is potentially inaccurate (or invalid) for PIM test;
- PIM test R&R is hard to meet the traditional gage R&R expectations.

As stated in Section 2.6, assumptions (1) and (2) in Section 3.6 are clearly violated in PIM test. Assumption (3) is also in question. To better understand the situation for assumption (3), we need to study the differences between (PIM) Test R&R and traditional gage R&R.

##### 4.1 The differences between test R&R and gage R&R

Test R&R has a broader scope than gage R&R. This is because a measurement system is usually a subset in a test system. Besides measurement system, a test system also includes and considers some other components as potential sources of variation. Products and environments are two typical examples. For simple tests where product or environment does not introduce additional variations, there may be no visible or practical differences between test R&R and gage R&R. But in situations such as PIM test, things are much more complicated. There are many more sources of variations to consider in test R&R than those in traditional gage R&R.

Figure 6 illustrated a PIM test R&R component breakdown for example. It considered many more variation components. In the traditional gage R&R, the testing system as a whole is considered the measurement system, and there is no more component breakdown other than operators and “equipment”. When the measurement capability is good, there is no much concern about specific sources of variations. Traditional gage R&R breakdown

can be sufficient. But when the test R&R is very poor, there will be strong interest to further drill down the major sources of variations. Further breakdown of components will be very much interested. For example, it is a common belief that cell phone signals near testing chambers may interfere with PIM test, and so may the connection of test cable to DUT significantly impact.

The components shown in Figure 6 are only one collection of interests. Items can be further broken down or regrouped if desired. Under Figure 6 breakdown structure, the gage part of the test system is narrowed down to equipment. Another way to look at this breakdown is that in traditional gage R&R’s view, most subcomponents under test system are considered having nominal or no contribution to variation and are thus ignored. When their contributions are no longer nominal, the traditional gage R&R will no longer be capable of capturing or reflecting them.

Besides the breakdown at the measurement system side, another potential source of significant variation considered by test R&R (but usually not by gage R&R) comes from products, specifically the instability of product, meaning the parameter being measured changes almost by itself each or over time. In addition, interactions between the product and test system can also be considered.

##### 4.2 Traditional gage R&R expectations are generally unrealistic for PIM test

All things combined, the traditional gage R&R measurements for PIM test are much worse than the traditional expectations. As previously highlighted, with only 10 dB margin between PIM spec and noise floor (effective tolerance of 10 dB) and the analyzer potentially taking 6 dB variation, it is very obvious that the traditional 30% P/T expectation is not realistic. Note that the traditional estimation of gage variation is essentially invalid due to inconstant variance. Based on the extensive repeatability

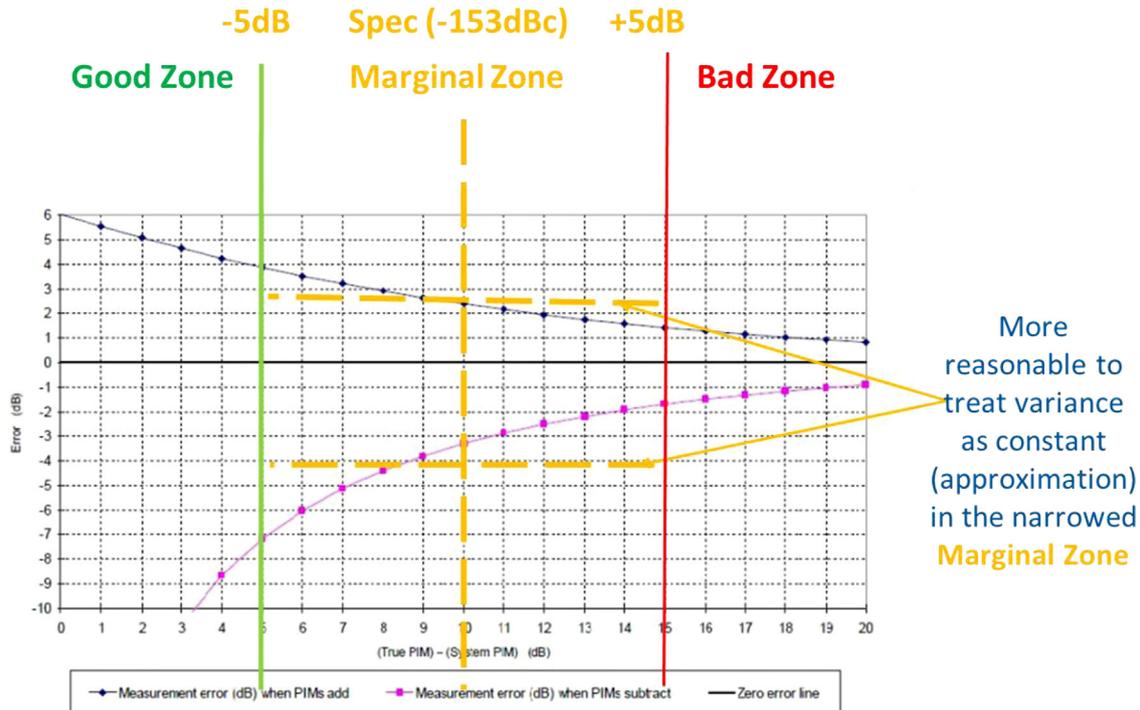


Fig. 7. Marginal zone.

analyses we have done in the industry, usually MOE for PIM test will not be less than  $\pm 5$  dB, with MOE of analyzer being the dominate source. There could be exceptions for simple products or components, such as input cables. Usually low MOE also corresponds to significantly lower noise floor, which is dependent on the complexity of the products and the testing system.

This situation is true across the whole industry. Since PIM is gradually becoming one of the most important performance measurements for RF-related communication networks, the poor R&R of PIM measurements present serious challenges to this industry.

## 5 Proposals for how to handle the unique challenges of PIM test or situations where constant variance is not observed

With the needed background information and issues thoroughly discussed above, we can now discuss the test strategy and techniques for how to deal with the unique challenges of PIM test (or measurement), as well as situations in general that constant variance is not observed. These proposals do not fundamentally or permanently fix the PIM test issues. They can improve the test performance and R&R.

### 5.1 For PIM test R&R assessment

These proposals apply to not only PIM test but also all situations that constant variance is not observed.

#### 5.1.1 Focusing on the marginal zone around the spec

As shown in Figure 7, the marginal zone is referring to the area around the spec covered by the MOE of test. The key here is that only the marginal zone matters to the test result. Readings are unreliable within the zone, yet “reliable” outside it. For example, if a result is below the marginal zone, it is safe to consider that it is reliably below the spec. On the other hand, if a result is above the marginal zone, it is safe to consider that it is reliably above the spec. If a result falls within the marginal zone, by definition the “true” result can be anywhere inside the zone and no guarantee to stay at the same side of spec; results of repeated testing are very likely to change camp.

With analyzer having  $\pm 3$  dB MOE around the spec, historical data from the industry show that the marginal zone for PIM spec will usually not be smaller than  $\pm 5$  dB. It does vary from product to product, largely related to product complexity. More sophisticated products tend to have a bigger marginal zone, while zone for simple product can be smaller. To verify or refine the marginal zone for a specific product, select samples only from the initially assumed marginal zone ( $\pm 5$  dB around the spec) based on the initial test result, and then conduct traditional gage R&R study. The refined more precise marginal zone (MOE, study variation) can be estimated from the study. The key to this practice is that only within this narrowed initial zone, the variance can be reasonably treated or approximated as constant to make the traditional gage R&R analysis valid.

Yet to follow this sampling practice, the % study variation and % contribution will be inflated due to the clustered sampling from a narrowed marginal zone, instead

of the traditional whole range of possible measurement. In this case, the traditional gage R&R metrics, such as % study variation, % contribution and distinct categories, will not be valid to depict R&R capability for the whole range of measurement. % Tolerance is the only metric that can still be valid, yet for one-sided spec, there is no given tolerance but only arbitrary or hypothetical ones. With this, for PIM test, a more meaningful, new metric to better represent test R&R will add meaningful value.

### 5.1.2 Moving away from traditional gage R&R metrics, instead using study variation (or MOE) as more direct and meaningful measurement for PIM test R&R

As shown in Table 1, study variation is the common numerator for major traditional GRR metrics (% study variation, % tolerance). It describes the diameter of the marginal zone, a more direct and actual measurement of test R&R. MOE is another term describing the same, but typically shown in radius format,  $\pm$  radius of marginal zone.

### 5.1.3 Numerical examples

Extensive studies show that the typical MOE for PIM test is about  $\pm 5$  dB around the spec. Specific products or samples may vary around that. A specific R&R study is conducted following this proposed approach, to sample from the assumed marginal zone of  $\pm 5$  dB. The actual MOE calculated from this sample is  $\pm 4.63$  dB, thus the actual MOE is refined as  $\pm 4.63$  dB for this product. We recognize that a different sample is likely to give a different yet similar result. We have seen the situation that the calculated MOE from sampling within  $\pm 5$  dB zone is bigger than  $\pm 5$  dB,  $\pm 5.2$  dB to be specific. In this case, we consider the actual MOE is  $\pm 5.2$  dB for this product.

## 5.2 For more reliable PIM measurements

Knowing that PIM measurement R&R is very poor compared to the traditional expectations, there are some ways to improve the PIM measurement itself. There are ways to fundamentally change how PIM is captured, which is beyond the scope of this paper. This paper will highlight two possible ways to improve PIM measurement from statistical perspectives.

### 5.2.1 Repeated measurement or testing

Based on Central Limit Theorem, the variation (standard error) of mean is reduced from the variation of individual measurements by a factor of square root of sample size (number of repeats in this case). Four repeats can cut variation in half, thus double the precision.

Repeat is a commonly used technique, typically (but not only) used after exhausting all ways to improve equipment precision itself. It is widely used inside the electronic measurement systems. Most outputs of electronic measurement systems are some form of statistics

(likely the average) of multiple readings. The repeated measurement technique can be used again on measurement results to further improve the precision.

The repeat concept can be applied in various forms. One option to increase the reliability of results is to do repeated testing for results within the marginal zone. As mentioned earlier, results in the marginal zone are of typical concerns, since reliable conclusion cannot be made for them. Yet if two out of three repeats fail the spec, it is safer to consider it a true failure; if two out of three pass, it is safer to consider it a true pass. Drawing pass or fail conclusion based on two out of three repeats will dramatically increase the reliability of conclusion, although the actual reliability will be a function of how close the measurements are to the spec.

The downside of this repeat approach is that the testing time will be multiplied by the number of repeats. PIM test is a relatively long test, easily exceeding 10 min per test. Testing chambers require significant investment. PIM test is frequently a bottleneck in manufacturing. Repeated testing will add huge constraint to manufacturing and will not be preferred by the industry. Possible remedies to the added time include shortening the test cycle as a trade-off.

### 5.2.2 Consider using more stable parameters

An additional option to increase reliability of PIM measurement is to move to a more stable parameter to represent PIM. The current PIM measurement uses peak performance. A more stable parameter is average. Yet the peak was chosen to represent the worst-case scenario. The current STV of PIM uses the highest STV among five sweeps. A pooled STV from all five sweeps will be more stable.

Both practices (repeats and more stable parameters) have some drawbacks, thus both practices, especially the second, will face challenges to be accepted by the RF industry. With the increasing complexity of RF products, PIM test reliability issue will become more and more significant. Trade-offs will have to be made down the road. Besides the two statistical approaches, there are good signs that PIM measurement itself can be done differently to improve R&R, which is beyond the scope of this paper.

## 6 Summary and conclusions

The current PIM test has very poor R&R across the industry, comparing to the traditional gage R&R expectations. This is primarily due to the nature of PIM. Two fundamental problems make the traditional gage R&R study invalid for PIM test:

- *Parts are inherently unstable*: The parameter being measured changes by itself each or over time.
- *Inconstant variance*: The variation of gage (analyzer) is not constant across the whole range of the intended measurement.

Special sampling can be done to battle these problems and make more meaningful estimation of PIM test R&R:

- Select samples only from the marginal zone; suggest  $\pm 5$  dB around spec (for PIM) to begin with, refine it with real data.
- Can treat the variance as constant within the marginal zone (approximation).

Move away from traditional gage R&R metrics since they are invalid with or without the special sampling. Use a more direct R&R measurement, margin of error (study variation), instead for the following reasons:

- Study variation is the common numerator for major traditional GRR metrics (% study variation, % tolerance).
- % Study variation will be inflated due to clustered sampling from a narrow marginal zone instead of the traditionally whole range of intended measurement.
- Using one-sided spec to represent tolerance can be misleading and risky. Using an arbitrary hypothetical tolerance is more meaningful.

Repeated testing can be used to improve the repeatability (and reliability) of PIM measurement (or testing). The industry can also consider switching to more stable parameters to represent PIM. Current PIM is using peak reading which in nature is less stable.

## Implications and influences

RF industry has been operating with very poor R&R on the key performance measurement PIM. The industry has been passing or failing products based on unrepeatable PIM measurement. The results are very unreliable, yet no

known solutions are available. There is a pressing need to address the challenge. The proposed solutions do not fundamentally or permanently solve the problem but will better estimate and represent as well as improve the PIM test repeatability. The study and proposed solutions apply to not only PIM test but also tests in general violating constant variance assumptions.

## References

1. P. Holes, Locating passive intermodulation, PIM faults, 2012, <https://www.radio-electronics.com/info/rf-technology-design/>
2. IEC, IEC 62037-1:2012: Passive RF and Microwave Devices, Intermodulation Level Measurement – Part 1: General Requirements and Measuring Methods (2012)
3. C. Cook, J. Mosher, Gauge capability procedure for evaluating the stability of passive intermodulation measurement systems, in *Proceedings of the 63rd IWCS Conference* (IWCS, Shrewsbury, PA, 2014), pp. 815–821
4. D. Montgomery, *Introduction to Statistical Quality Control*, 2nd edn. (John Wiley & Sons, Inc., New York, 1991)
5. T. McCarty, L. Daniels, M. Bremer, P. Gupta, *The Six Sigma Black Belt Handbook* (McGraw-Hill, New York, 2005)
6. L.A. Brown, B.R. Daugherty, V.W. Lowe, *Measurement Systems Analysis*, 3rd edn. (Auto Industry Action Group, Southfield, Michigan, 2003)
7. D. Ermer, Improved gage R&R measurement studies, *Quality Progress*, 77–79 (2006)
8. D. Ermer, Appraiser variation in gage R&R measurement, *Quality Progress*, 75–78 (2006)

**Cite this article as:** Gary G. Jing, How to measure test repeatability when stability and constant variance are not observed, Int. J. Metrol. Qual. Eng. 9, 10 (2018)