GRR & TCS

Index

1. Why need for GRR or TCS?
2. Estimating Measurement Uncertainty
3. Test-capability Study
4. Gauge Reproducibility and Repeatability (GRR) Study
5. Test Guardbands and Defect Rates
6. Validating the Test Set-up and Correlation Lockout
7. Systematic Error , Repeatability Error and Parameter Distribution

Why need GRR or TCS

Variation

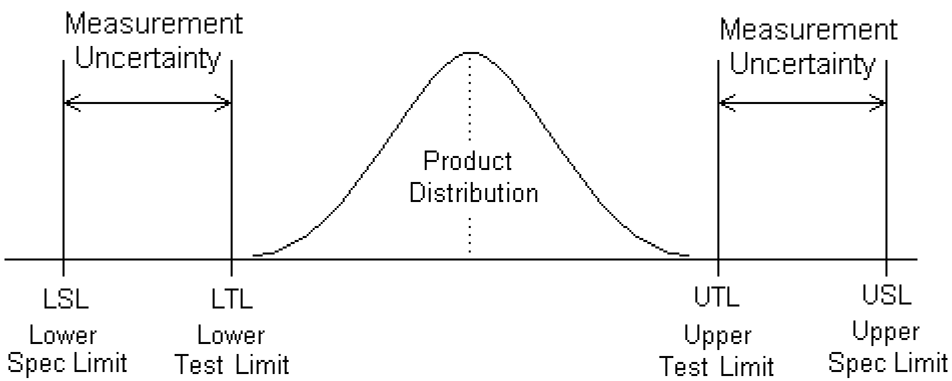
- Variation exists in the IC manufacturing process. Variation is the source of defects, but can be adjusted and reduced.
  - We make measurements at the important steps in the overall manufacturing process to identify and help minimize manufacturing variation.
- To reduce process variation, it is important to understand that “true” process variation is comprised of multiple sources of variation.
  - One of these sources of variation being the variation of the measurement tool
  - All measurement systems are subject to variation.
    - We have to determine the amount of variation.
- To assess measurement tool variation, the analysis tool known as Gauge (Gage) Repeatability and Reproducibility, or “GRR”, is typically used.

GRR – Where did it come from?

- GRR got its start in work done by General Motors in the 1960s
- The GRR method was formally transferred to the Automotive Industry Action Group in 1989
  - It is documented in the “Measurement System Analysis” handbook published by the AIAG starting in 1990
- The original GRR method was defined in terms of Average and Range. Later it was redefined in terms of Average and Variance using ANOVA methods
- GRR has spread to many industries, including the semiconductor industry since the late ‘80s

Why GRR or TCS in the first place?

- In production testing (of anything) we are checking to see if the product meets specifications
- There is guaranteed to be some error in the measurement (how much is the question, experiment needed to quantify)
- Therefore we guardband against that error to protect our customer from getting a bad part (oh, and to maintain yield… remember a bad system can *pass bad parts* **and** *fail good parts*!)

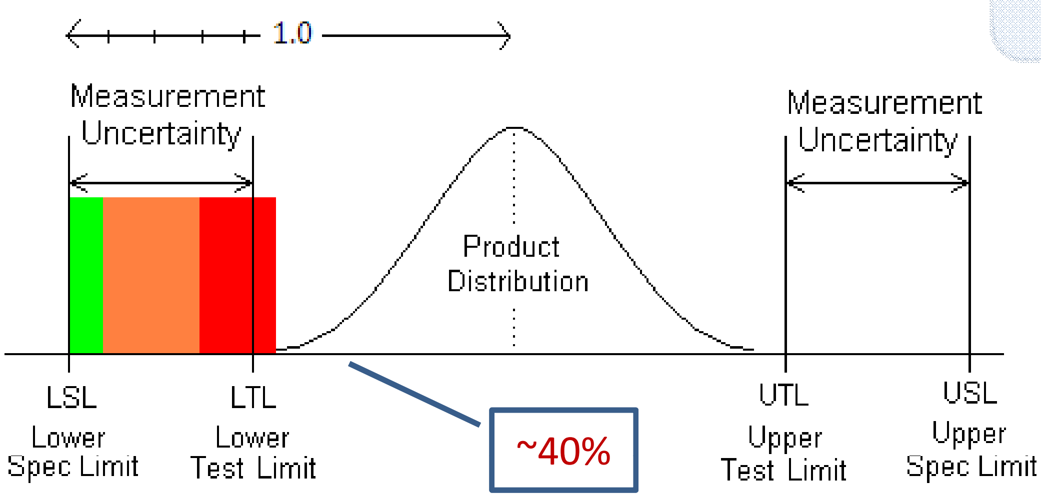


Why GRR or TCS in the first place?

- Since we've established that the test system is allowed some error in measurement, how much?
  - A fair metric is to <u>compare the error to the spec limits</u>
    - An error < 10% of the spec limit is acceptable
    - An error > 10% and < 30% should be reviewed
    - An error > 30% is generally unacceptable

$$\%TCS = \frac{100 * 2 * 3 * \sigma_M}{USL - LSL}$$

For a 2 sided limit



Why GRR or TCS in the first place?

- Why does this metric matter?
- If your error is
  - *100%+*, you are not even measuring anything?!?!
  - *50%*, your spec must be much worse than part performance
  - *25%*, your spec is wider than you'd like or yield loss occurs
  - *10%*, a reasonable allowance to give a test engineer/equipment
  - *5%*, good job test engineer, you beat your budget!
  - *1%*, why are you spending so much test time test engineer?!?!

Test System Quality Levels
(Implementation Can Be Progressive)

Test System Quality Levels				
Level	Test Capability Study Done (Improvements Made As Needed)	Guardband Used	Correlation Lock-out Used	Comprehensive Calibration (Includes Calibration of User Interface Adapters & Other User Developed Hardware)
1	YES	No	No	No
2	YES	YES	No	No
3	YES	YES	YES	No
4	YES (And Test Capability Studies For Worst Systems Are Re-Done Every Year *)	YES	YES	No
5	YES (And Test Capability Studies For Worst Systems Are Re-Done Every Year *)	YES	YES	YES

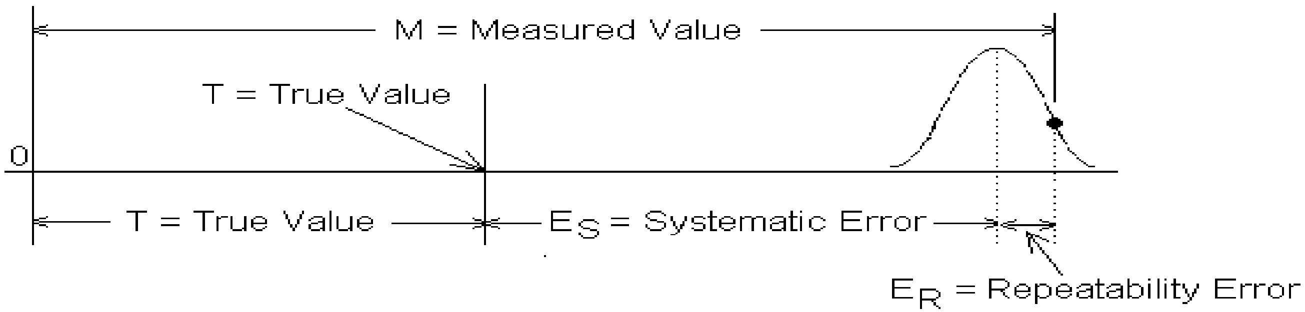
* Some Test Capability Studies are re-done every year based on the previous year's test production problems. For example, if there are 100 different products every year, test capability studies could be re-done and improvements made for the 20 products which had the worst test production problems throughout the year.

Estimating Measurement Uncertainty

Measured Value, True Value, Systematic Error and Repeatability Error

Measured Value, True Value, Systematic Error, and Repeatability Error

Measured Value = True Value + Systematic Error + Repeatability Error



THE MEASURED VALUE:
The Measured Value, M, is the sum of three terms:

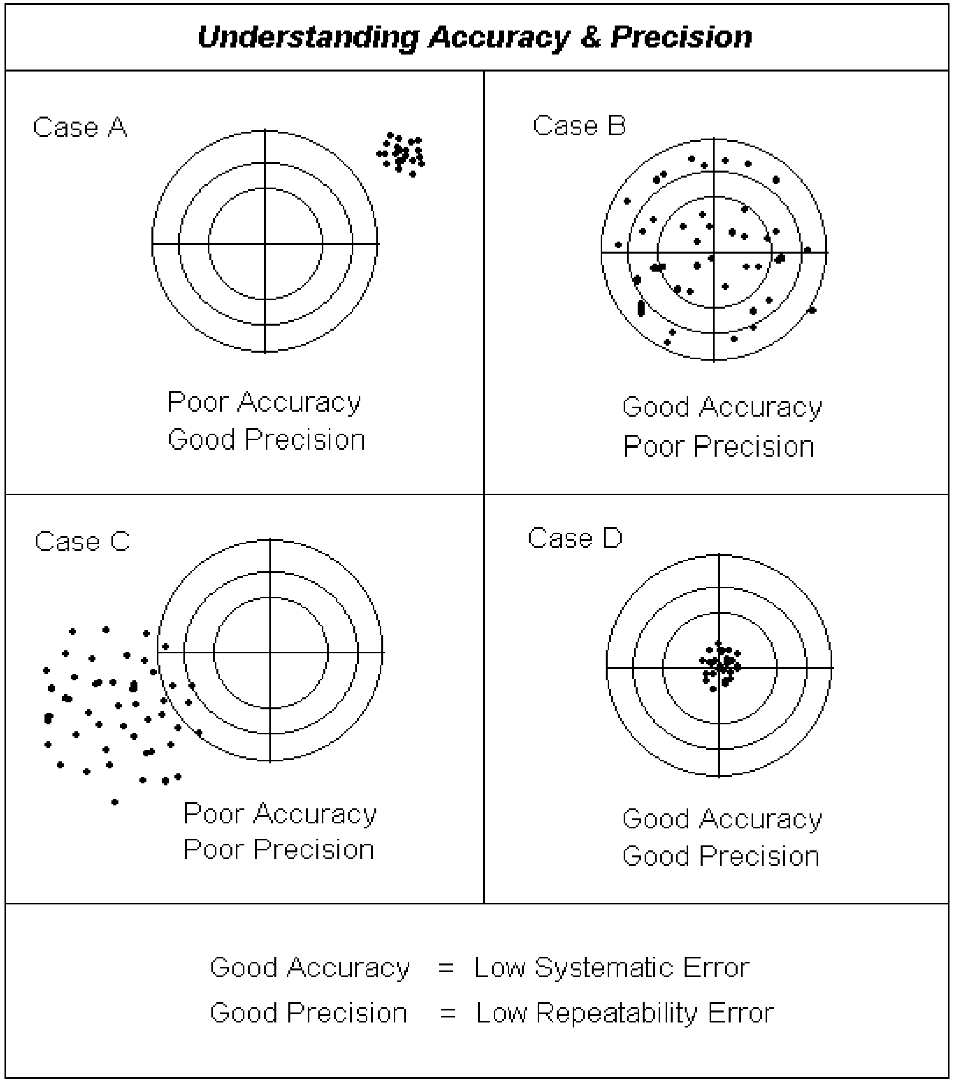
1. The TRUE VALUE,
2. The SYSTEMATIC ERROR of the test system, and
3. The REPEATABILITY ERROR of the test system.

This is represented as:

$$M = T + E_S + E_R$$

M = Measured Value
T = True Value
E_S = Systematic Error
E_R = Repeatability Error

Understanding Accuracy &Precision

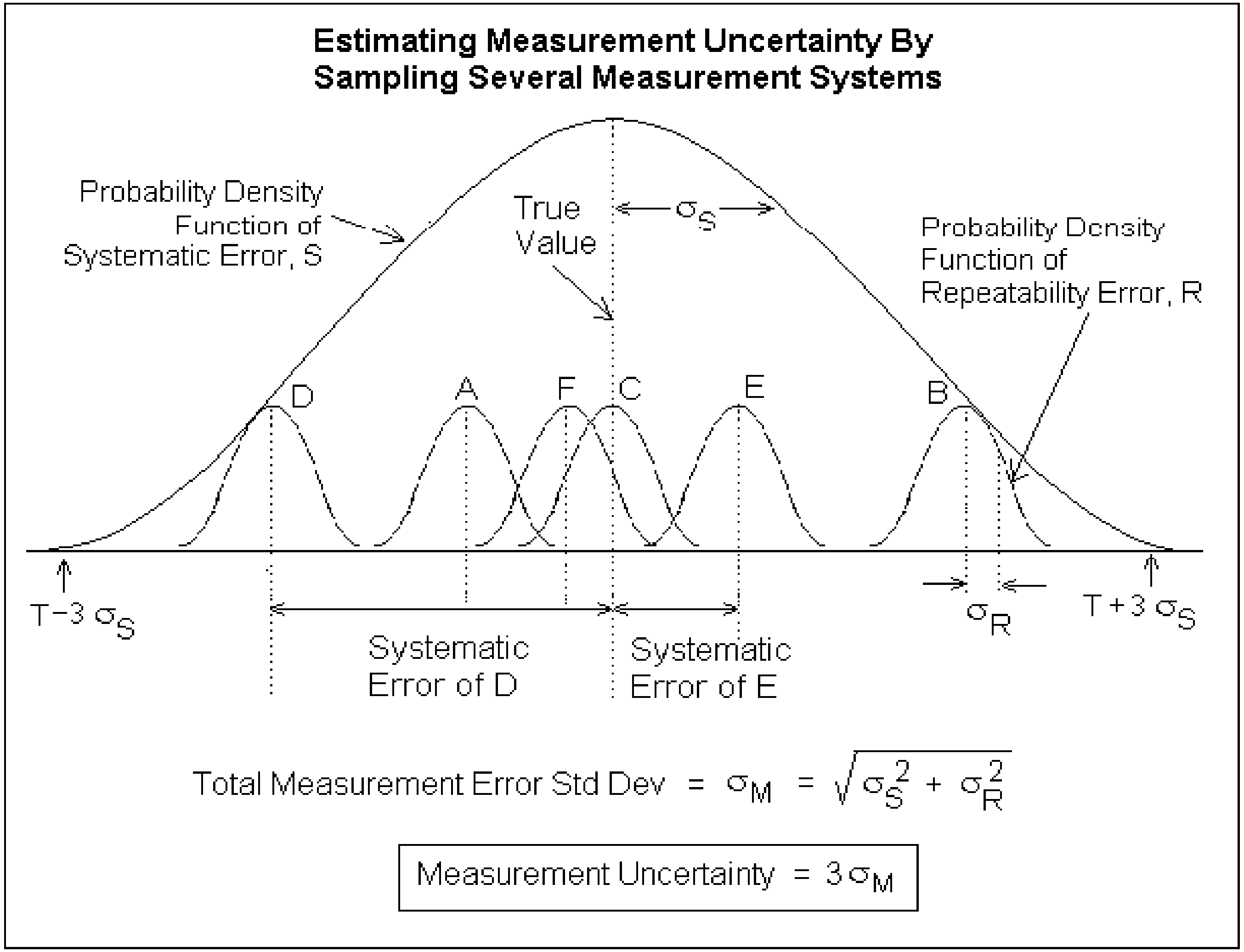


Good **accuracy** means readings centered on the true value (case B or D), i.e., low systematic errors.

Good **precision** means readings grouped closely together (case A or D), i.e., low repeatability errors.

Statistics for Test

Estimating Measurement Uncertainty



Test-capability Study

Equations for the Test Capability Study

$$s_i = \sqrt{\frac{1}{T-1}\sum_{j=1}^{T}(x_{ij} - \bar{x}_i)^2}$$

Sample std. dev. for device *i*

$$\hat{\sigma}_m = \frac{1}{c_4 D}\sum_{i=1}^{D} s_i$$

Average of s_i across all devices corrected for bias in s_i

x_{ij} = Measurement of device *i* on test setup *j*

$\bar{x}_i$ = Mean of measurements for device *i* over all test setups

Example TCS with Three Test Set-ups

D = device, T = test set-up bias, R = repeatability error

DUT	Test Setup A	Test Setup B	Test Setup C	Sample Variance
1	D1 + TA + RA1	D1 + TB + RB1	D1 + TC + RC1	Var(TA, TB, TC) + Var(RA1, RB1, RC1)
2	D2 + TA + RA2	D2 + TB + RB2	D2 + TC + RC2	Var(TA, TB, TC) + Var(RA2, RB2, RC2)
3	D3 + TA + RA3	D3 + TB + RB3	D3 + TC + RC3	Var(TA, TB, TC) + Var(RA3, RB3, RC3)
:	:	:	:	:
N	DN + TA + RAN	DN + TB + RBN	DN + TC + RCN	Var(TA, TB, TC) + Var(RAN, RBN, RCN)

σ_m^2 = E[Var(TA, TB, TC) + Var(R)]

Example

UNIT#:	CFG A:	CFG B:	CFG C:	CFG D:	CFG E:	STD DEV
1	19.00	23.00	20.00	19.00	15.00	2.86
2	32.00	38.00	33.00	42.00	37.00	4.04
3	7.00	12.00	20.00	23.00	9.00	6.98
4	33.00	50.00	51.00	54.00	46.00	8.23
5	-72.00	-77.00	-70.00	-76.00	-72.00	2.97
6	-28.00	-31.00	-37.00	-33.00	-27.00	4.02
7	40.00	50.00	55.00	47.00	45.00	5.59
8	-32.00	-31.00	-37.00	-37.00	-39.00	3.49
9	5.00	12.00	12.00	11.00	11.00	2.95
10	-22.00	-19.00	-15.00	-17.00	-14.00	3.21
11	-32.00	-27.00	-37.00	-33.00	-32.00	3.56
12	-9.00	-4.00	-8.00	-9.00	-10.00	2.35
13	25.00	28.00	34.00	27.00	28.00	3.36
14	-6.00	-3.00	-9.00	-7.00	-8.00	2.30
15	-7.00	-12.00	-11.00	-15.00	-16.00	3.56
AVG	-3.13	0.60	0.07	-0.27	-1.80	3.97
STD DEV	30.81	35.00	36.17	36.44	33.54	1.69

The average of the device std. devs. is the estimate of the total measurement-error std. dev.

Test-capability Study

From the example,

Average of the Standard Deviations = $\bar{S} = 3.97$

Standard Deviation of the Standard Deviations = $\sigma_S = 1.69$

To cull outliers in the standard deviations calculated for each device, discard any device standard deviation beyond $\bar{S} \pm 3\sigma_S$.

The overall measurement-error standard deviation is $\sigma_M = \bar{S}/c_4$.

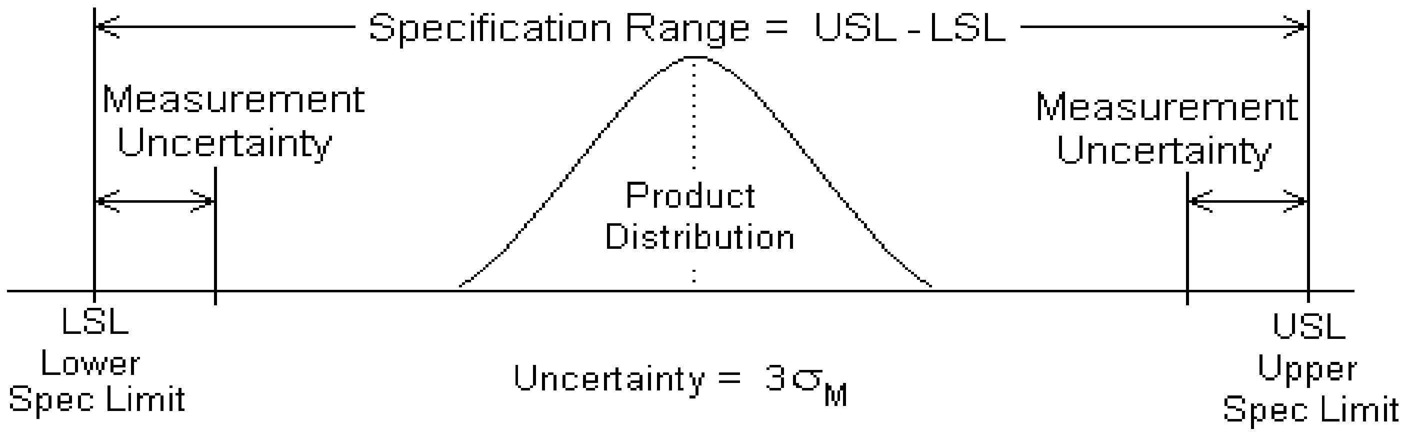
Recall that correction factor c_4 is needed to obtain an unbiased estimate of σ_M .

For our case, we have 5 test configurations; hence, $c_4 = 0.94$

and $\sigma_M = \bar{S} / c_4 = 3.97 / 0.94 = 4.223$.

Measurement Uncertainty = $3\sigma_M = 12.669$.

Measurement Error as a Percent of Specification Range



$$\text{Measurement Error (\% of Spec Range)} = \begin{cases} 100\% * \dfrac{6\,\sigma_M}{USL - LSL} & \text{For two sided limits} \\[2ex] 100\% * \dfrac{3\,\sigma_M}{LSL} \ \text{or} \ \dfrac{3\,\sigma_M}{USL} & \text{For one sided limit} \end{cases}$$

Effective Systematic & Repeatability Variance Sample Sizes:

DUT*	Test Setup A	Test Setup B	Test Setup C	Sample Std Dev
1	T1 + SA + RA1	T1 + SB + RB1	T1 + SC + RC1	$[\mathrm{Var}(SA,SB,SC) + \mathrm{Var}(RA1,RB1,RC1)]^{1/2}$
2	T2 + SA + RA2	T2 + SB + RB2	T2 + SC + RC2	$[\mathrm{Var}(SA,SB,SC) + \mathrm{Var}(RA2,RB2,RC2)]^{1/2}$
3	T3 + SA + RA3	T3 + SB + RB3	T3 + SC + RC3	$[\mathrm{Var}(SA,SB,SC) + \mathrm{Var}(RA3,RB3,RC3)]^{1/2}$
:	:	:	:	:
N	TN + SA + RAN	TN + SB + RBN	TN + SC + RCN	$[\mathrm{Var}(SA,SB,SC) + \mathrm{Var}(RAN,RBN,RCN)]^{1/2}$
				Avg: $[\mathrm{Var}(SA,SB,SC) + \mathrm{Var}(R)]^{1/2}$

Let

N_{SETUP} = number of test set-ups (configurations)

N_{DEVICE} = number of devices

Then

Effective Sample Size for Repeatability-error Variance = $N_{SETUP} * N_{DEVICE} >> 30$.

Effective Sample Size for Systematic-error Variance = N_{SETUP}

To accurately estimate systematic error N_{SETUP} must be as large as practical.

Gauge Reproducibility and Repeatability (GRR) Study

Gauge Repeatability and Reproducibility Study (GRR)

The calculations in the GRR study are more complex than in the test capability study due to this separate treatment of repeatability errors. The following are the calculations of the GRR study[1]:

Let

D = number of devices,

T = number of test setups,

R = number of repeated measurements for each test setup,

$i = 1, 2, \ldots, T$ = index for the test setup,

$j = 1, 2, \ldots, D$ = index for the device, and

$k = 1, 2, \ldots, R$ = index for the repeated measurements.

$x_{ijk} = k^{th}$ parameter measurement on test setup i and device j

Then to estimate the repeatability variance:

$$\bar{x}_{ij} = \frac{1}{R}\sum_{k=1}^{R} x_{ijk}$$ = mean of the repeated measurements for test setup i and device j

$$s_{ij}^{2} = \frac{1}{R-1}\sum_{k=1}^{R} (x_{ijk} - \bar{x}_{ij})^{2}$$ = sample variance of the repeated measurements for test setup i and device j

$$s_{rj}^{2} = \frac{1}{T}\sum_{i=1}^{T} s_{ij}^{2}$$ = mean of s_{ij}^{2} across all test setups

$$s_{r}^{2} = \frac{1}{D}\sum_{j=1}^{D} s_{rj}^{2}$$ = mean of s_{rj}^{2} across all devices

s_r = estimated standard deviation of the repeatability error

Then to estimate the reproducibility variance:

$$\bar{\bar{x}}_{j} = \frac{1}{T}\sum_{i=1}^{T} \bar{x}_{ij}$$ = grand average for device j

$$s_{Tj}^{2} = \frac{1}{T-1}\sum_{i=1}^{T} (\bar{x}_{ij} - \bar{\bar{x}}_{j})^{2}$$ = sample variance of test setup means for device j

$$s_{Rj}^{2} = \max\left\{0, \left(s_{Tj}^{2} - \frac{s_{rj}^{2}}{R}\right)\right\}$$ = estimated reproducibility variance of device j

$$s_{R}^{2} = \frac{1}{D}\sum_{j=1}^{D} s_{Rj}^{2}$$ = mean of s_{Rj}^{2} across all devices

s_R = estimated standard deviation of the reproducibility error

Finally,

$s_x = \sqrt{s_r^{2} + s_R^{2}}$, where s_x = estimate of the measurement standard deviation , σ_t , for parameter x.

GRR versus Test Capability

The purpose of both methods is to estimate the total measurement-error variance for each parametric measurement. For most cases, the two methods produce estimates that are very close.

GRR separately estimates the repeatability variance and the reproducibility variance and then adds them to estimate the total measurement-error variance. Test capability estimates the total measurement-error variance directly.

The repeatability variance is independently sampled a large number of times in both methods.

> Test-capability repeatability samples = $T \bullet D$ (>120)
>
> GRR repeatability samples = $T \bullet D \bullet R$ (>240)
>
> T = number of test setups (>=4), D = devices (>=30), N = replicates (>=2)

The uncertainty in a standard-deviation estimate is slowly decreasing after 100 samples. Hence, the replicates in the GRR method add a lot of work for very little benefit.

Missing values for replicates must be estimated before ANOVA is applied.

Excel spreadsheet comparing the GRR and Test-capability studies:

GRR vs
Test-capability.xls

Test Guardbands and Defect Rates

Rate of Test Escapes

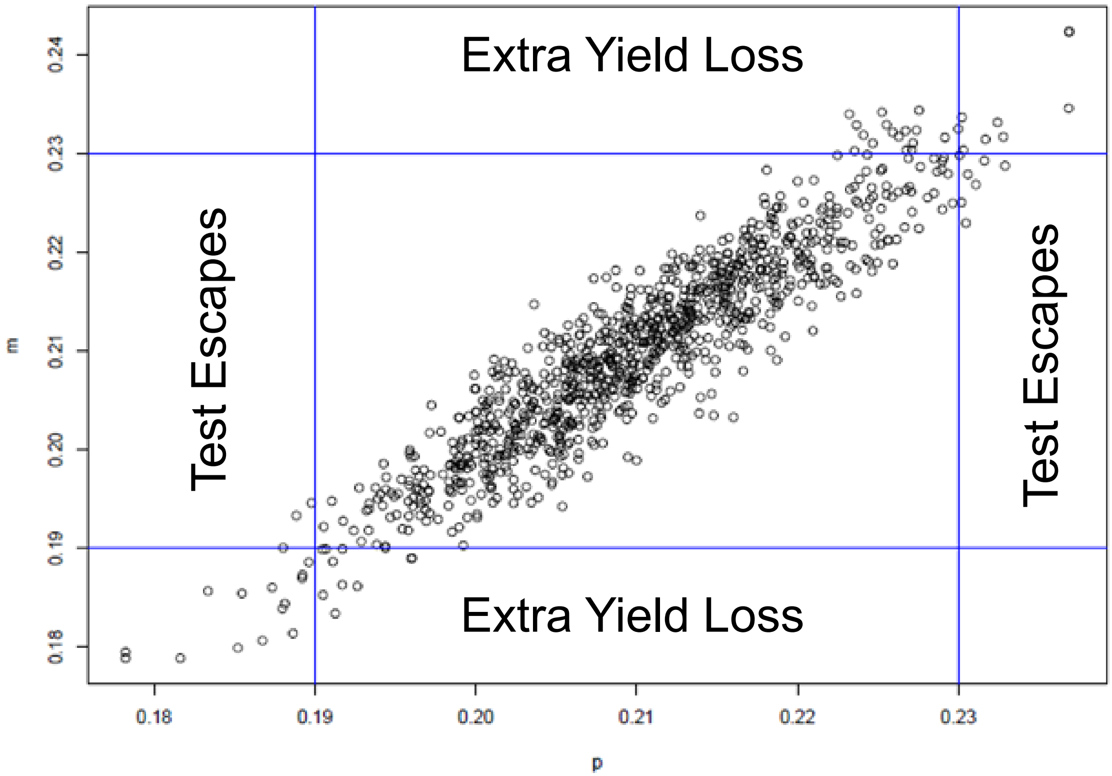
Measurement equation: M = P + E, where

P = true or population value, M = measured value

E = measurement error, and M and E are independent.

Construct a bivariate normal distribution without test guardbands for the following example:

LSL=0.19, USL=0.23, $\mu_p = 0.21$, $\sigma_p = 0.01$, $C_{pk} = 0.67$ (half of min. for robust process), and $\sigma_m = 0.004$.



Test guardbands and Defect Rates

Without test guardbands in this example, the probability of P being out of spec given that M is in spec is ~0.74% = 7400 ppm.

With test guardbands = $3\sigma_m$, where σ_m = the measurement-error standard deviation, the defect rate = 14 ppm.

Minimum Test Guardbands:

The defect rate is a function of the test guardbands and the fraction of the parameter distribution that exceeds the specification limits, $C_{pk} < 1.33$.

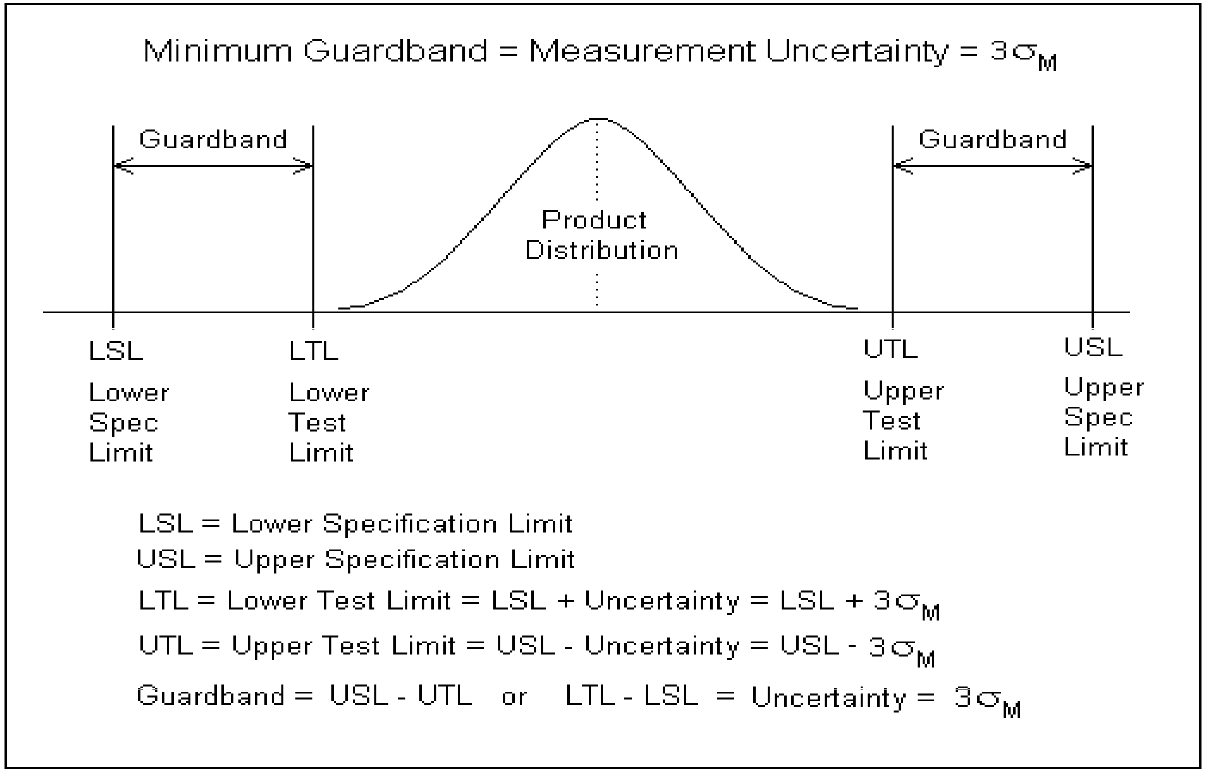
Test escapes occur when the true parameter value is slightly out of spec, the measurement error exceeds the test guardbands, and the error has the proper sign to create a measurement within the test limits.

Semiconductor fab process can create significant variability on critical parameters. In general, one must assume that $C_{pk} < 1.33$ in the product's lifetime.

The industry-standard defect rate is six-sigma quality, which is defined as 3.4 ppm. Experience and statistics show that a test guardband ≥ $3\sigma_m$ will create an average defect rate ≤ 3.4 ppm.

This average defect rate is a long-term average. Defect rates for a particular lot can be significantly higher.

Test Guardbands



Minimum Guardband = Measurement Uncertainty

The larger the measurement uncertainty, the larger the guardband will be and the smaller the limit window will be.

A large guardband may significantly reduce product yield, which is another reason to make the measurement uncertainty as small as possible.

Validating the Test Set-up and Correlation Lockout

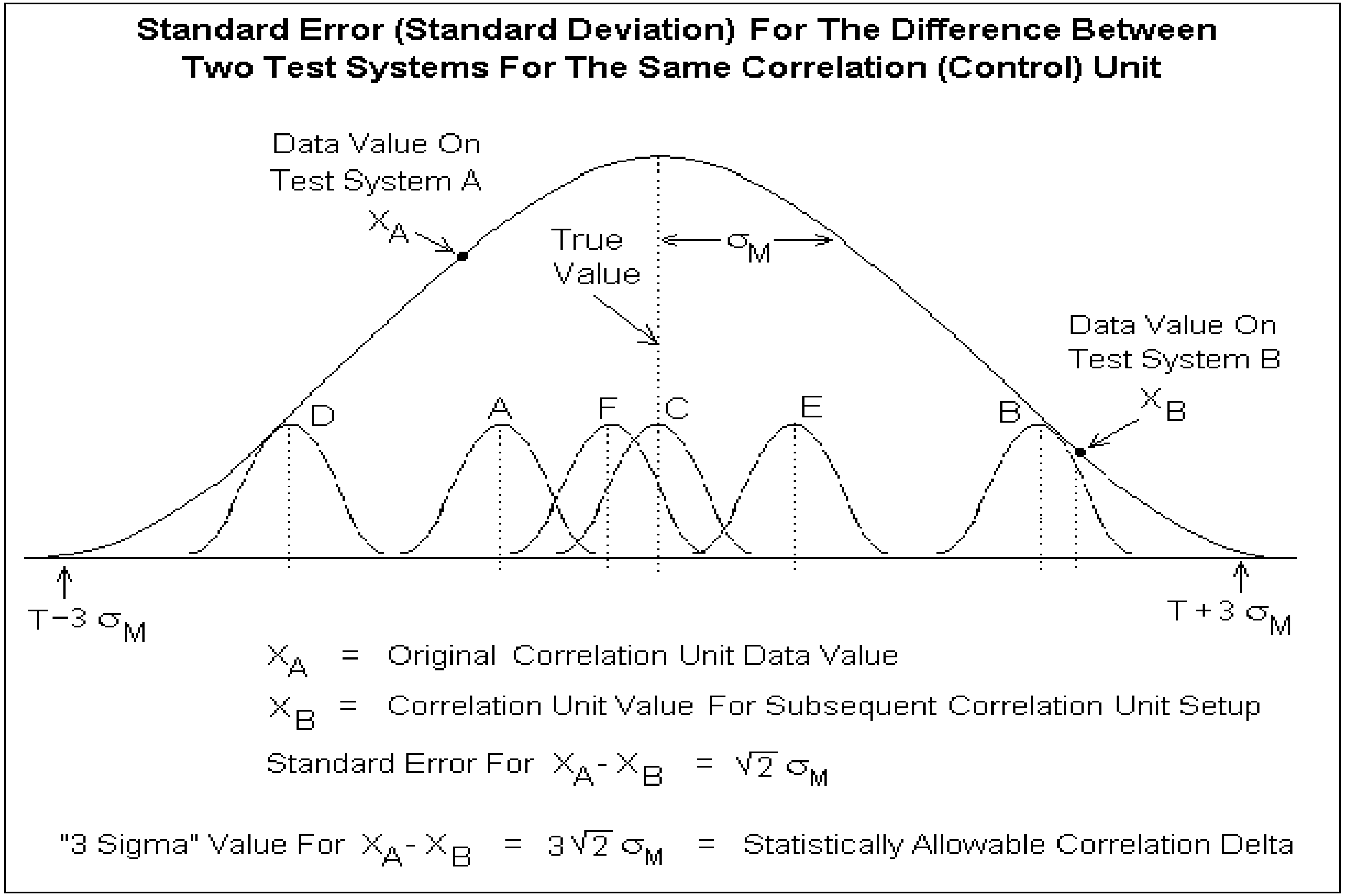
Control Units (Correlation Lockout):

1. Compares data from devices (control units) tested on a known-good test set-up to the same devices tested on the production set-up.
2. Proven, standard process-control technique.
3. Detects significant systematic errors prior to lot start.

Test Engineer's Actions to Implement Correlation Lockout:

1. Perform GRR or Test-capability study to estimate the standard deviations of measurement errors, σ_m.
2. Reduce σ_m to acceptable levels, e.g., < 0.1*(spec range).
3. Set test guardbands >= $3\sigma_m$
4. Set delta corr. limits <= $\sqrt{2}$ * test guardbands.
5. Generate and verify correlation units. Ship 20 correlations units to each production site.
6. Activate correlation lockout in the test application.

Standard Error For The Difference Between Two Test Systems



Standard Error For The Difference Between Two Test Systems

Why $3\sqrt{2}\,\sigma_M$ Should Be Used When Original Correlation Unit Data Is Based On Only <u>ONE</u> Test System

If A and B are statistically independent and if $\sigma_B^2 = \sigma_A^2$ then

$$\mathrm{Var}(A - B) = \mathrm{Var}(A) + \mathrm{Var}(B) \quad \longrightarrow \quad \sigma_{A-B}^2 = \sigma_A^2 + \sigma_B^2 = 2\,\sigma_A^2$$

This implies $\sigma_{A-B} = \sqrt{2}\,\sigma_A$

Since the standard measurement error for system A is $\sigma_A = \sigma_M$

where σ_M = Std Measurement Error (From Test Capability Study)

Then $\sigma_{A-B} = \sqrt{2}\,\sigma_M$

and so the allowable "3 sigma" variation for A-B is

$$\boxed{3\,\sigma_{A-B} = 3\sqrt{2}\,\sigma_M}$$

This is the statistically allowable difference between the original correlation unit value and subsequent correlation unit values taken on any arbitrary test system (test setup).

Why use test guardbands > $3\sigma_m$?

If a parameter distribution is far from the test limits with the minimum $3\sigma_m$ test guardbands, one should consider increasing the test guardbands to improve the probability of passing correlation lockout.

Correlation lockout suffers from the problem of multiple tests each with a small probability of a failure. All tests must pass correlation lockout for a site to be valid for production. In probe, all sites must pass to start production testing.

With $3\sigma_m$ test guardbands and $3\sqrt{2}\sigma_m$ correlation limits, the probability of a correlation failure on any test is 0.26%. With T such tests of a correlation unit, the probability of passing correlation lockout, $P_{PASS} = 0.9973^T$ For T = 100, P_{PASS} = 76.3%.

Because the probability of passing correlation lockout is calculated from the tails of a Gaussian distribution, a modest increase in the test guardbands and correlation limits will greatly improve P_{PASS}. For $4\sigma_m$ test guardbands and T = 100, P_{PASS} = 99.4%.

At all times the correlation limits must be $\pm\sqrt{2}$(test guardband).

Systematic Error , Repeatability Error and Parameter Distribution

σ_R σ_S σ_P *relationship*

Case Study:GRR%=10%, CPL=CPU=2.0

*$6\sigma_m$ /(USL-LSL)=10%. σ_P =(USL-LSL)/(6*2).*

So $\sigma_m < \sigma_P$, And in a normal parameter test （e.g. Vout,Vfb,Iq and so on) we know $\sigma_R < \sigma_S$

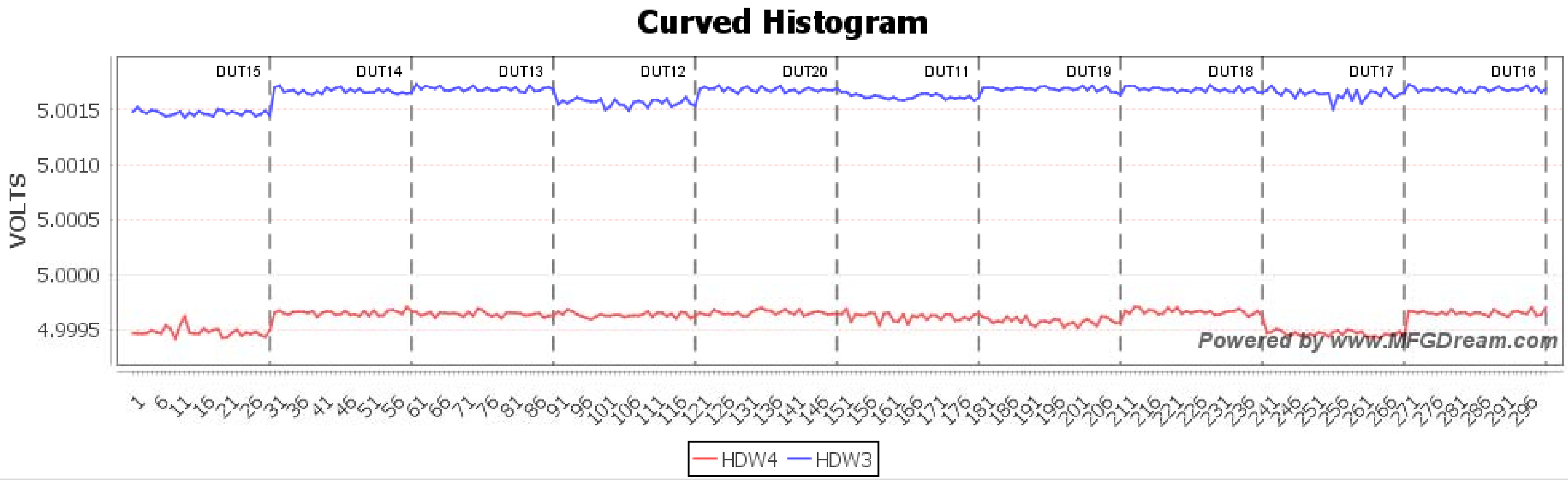
We can find some parameter which $\sigma_m < \sigma_P$ *and* $\sigma_R < \sigma_S$.

So the parameter $\sigma_R \ll \sigma_P$. *Actually most of parameter have this kind of characterization.*

Example:

200006:Vout_V_5V X

Number	TestItems	L L	U L	Units	R.PT	R.PD	O.E	New L L	New U L	%G&R&R	CPU	CPL	CPK	CP	EXECS	MEAN	MIN	MAX
200006	Vout_V_5V	4.9	5.1	VOLTS	0.00003	0.00144	0.00144	4.90144	5.09856	4.3%	32.49632	32.90298	32.49632	32.69965	600	5.00062	4.99942	5.00174



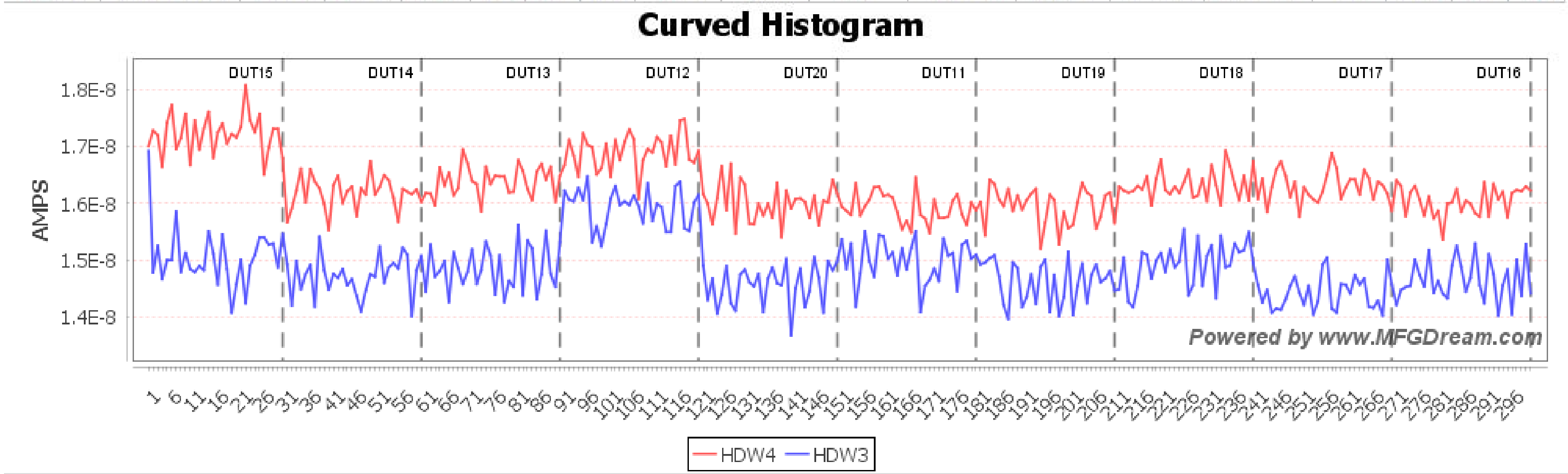
σ_R σ_S σ_P relationship

Not all parameter have $\sigma_R < \sigma_S$.   For example Ishdn,Ien which current is too small, and test system don't have so good accuracy.
For these parameter σ_R almost same as σ_S

Example:

1600005:Ileak_En_5V   X

Number	TestItems	L L	U L	Units	R.PT	R.PD	O.E	New L L	New U L	%G&R&R	CPU	CPL	CPK	CP	EXECS	MEAN	MIN	MAX
1600005	Ileak_En_5V	-1.0E-8	2.0E-7	AMPS	0.00	0.00	0.00	-0.00	0.00	3.1%	69.06278	9.58916	9.58916	39.32597	600	0.00	0.00	0.00



How to use this relationship.

- Good to debug when you get GRR or TCS check Fail.
- Known where the error come from.
- We can use $\sigma_R << \sigma_P$ *do something.*

Example: device stacked *together issue, we can dynamic to check* σ_P *in the program for each site, if* σ_P *results become smaller and smaller and almost equal to* σ_R *, that means the device stack happened.*

More information: http://www.kanwoda.com/
GRR or TCS Tool: http://www.mfgdream.com/

Question?

THANKS!