Testing the Limits of Designed Experiments

Dr. Paul Basore, Laboratory Director
Advanced Research and Development Laboratory (ARDL)
Hanwha Solar America, Santa Clara, California
18 July 2012

Hanwha

The Value of Replication in Experimentation

- A single good result may look nice in your thesis, but.....
  - Was it a lucky coincidence that you can never repeat?
  - Will other labs confirm the result, or will they prove that it's misleading?
  - Will it have any impact on high-volume manufacturing?
- Some rules for estimating whether a result is "real" or just a fluke:
  - It takes three *independent* replicates to give a trustworthy average
  - It takes ten *independent* replicates to give a trustworthy standard deviation
- *Independent* data capture the sources of uncontrollable variation
  - Cell-to-cell
  - Wafer-to-wafer
  - Day-to-day
  - Tool-to-tool? Person-to-person? Lab-to-lab?
- Example: 8 small cells on each of 10 wafers processed in 2 lots of 5
  - For multi-Si, typically every cell is independent (80 replicates)
  - For mono-Si, probably only every lot is independent (2 replicates!)

Statistical Significance: A-B Comparison

Hanwha

- Common definition of "statistically significant difference":
  - There is less than a 5% chance that the observed difference between the two things being compared is just due to random noise (p value).
  - If $p<0.05$, the better process will likely be the better process in the long run.
  - If $p>0.05$, you may need more data, or the difference may not be important.
- Conclusions that are based on results that are not statistically significant:
  - Usually lead to a great deal of wasted time and effort!
  - Commonly occur when the first result is the desired result.
  - People like to see what they expect. People don't like repetition.

One-way ANOVA: Reflectance(%) versus Process (3 runs each)

```
Source    DF      SS      MS      F      P
Process    1   5.134   5.134  10.00  0.034
Error      4   2.053   0.513
Total      5   7.187

S = 0.7164   R-Sq = 71.43%   R-Sq(adj) = 64.29%

                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level  N    Mean  StDev   ---------+---------+---------+---------+
New    3  13.307  0.374   (---------*--------)
Old    3  15.157  0.942                   (--------*---------)
```

Mean of each process lies beyond the 95% confidence limit for the mean of the other process -> Stat Sig!

D-I-Y 95% Confidence Intervals

Hanwha

df	$t_{.025}$	df	$t_{.025}$	df	$t_{.025}$	df	$t_{.025}$
1	12.71	6	2.45	12	2.18	25	2.06
2	4.30	7	2.36	14	2.14	30	2.04
3	3.18	8	2.31	16	2.12	40	2.02
4	2.78	9	2.26	18	2.10	60	2.00
5	2.57	10	2.23	20	2.09	120	1.98

With 95% confidence, the *true mean* lies within the range given by this expression →

$$m \pm \frac{t_{.025} s_p}{\sqrt{n}}$$

Where:
- m is the measured average (mean) value of the data
- $t_{.025}$ is from the above table (Student's-t)
- S_p is the pooled standard deviation of the available data*
- n is the number of data points averaged to obtain m

For A-B comparison example:
df = 4**, so $t_{.025}$=2.78
s_{old} = .94, s_{new}= .37, s_p=.72
n_{old}=3, n_{new}=3
m_{old}=15.2, m_{new}=13.3

*Can be based on more data than just what was used to calculate m (if all data should have same Stdev)
**df = total number of data points used minus the number of different mean values represented

Choosing Experimental Factors

- Think of all of the factors that might affect your result
  - It's better to be all-inclusive to avoid overlooking something important
  - Decide which factors are readily controllable and which are not
  - Rank controllable factors by significance (performance, cost, repeatability)
  - Design experiment to uniformly and randomly sample uncontrollable factors
- Pick several (e.g. four to six) of the most significant controllable factors
  - For each factor, pick a nominal *midpoint* value (often the current best practice)
  - Pick a range for each, within which you are confident the process won't fail
  - Adjust the midpoint to either the arithmetic mean or the geometric mean
- Example: Alkaline Texturing
  - Controllable and important: KOH concentration, Time, Temperature
  - Controllable but less important: Additive concentration, Number of wafers
  - Uncontrollable: Silicate concentration, Wafer orientation, Surface roughness
- Screen the factors to find the most important ones
  - Vary one factor at a time – OK on limited budget, but not optimal
  - Full factorial (all high/low combinations) – Thorough, but time consuming
  - Fractional factorial (resolution IV) - Usually the most efficient approach

Screening Experimental Factors

Hanwha

O Designing a "resolution-IV" fractional-factorial experiment

- Tests a fraction of the corners of the multi-dimensional "cube" of factor space
- This design finds the effect of each factor, undistorted by factor interactions
- Add three replicates of the center point at beginning, middle and end
- Table below shows this design for a four-factor screening experiment
  - ✓ The range for each factor is normalized to -1, 0, +1 for the low, mid, high values

Factor A	Factor B	Factor C	Factor D
0	0	0	0
-1	1	-1	1
1	1	-1	-1
1	1	1	1
-1	1	1	-1
0	0	0	0
-1	-1	1	1
1	-1	-1	1
-1	-1	-1	-1
1	-1	1	-1
0	0	0	0

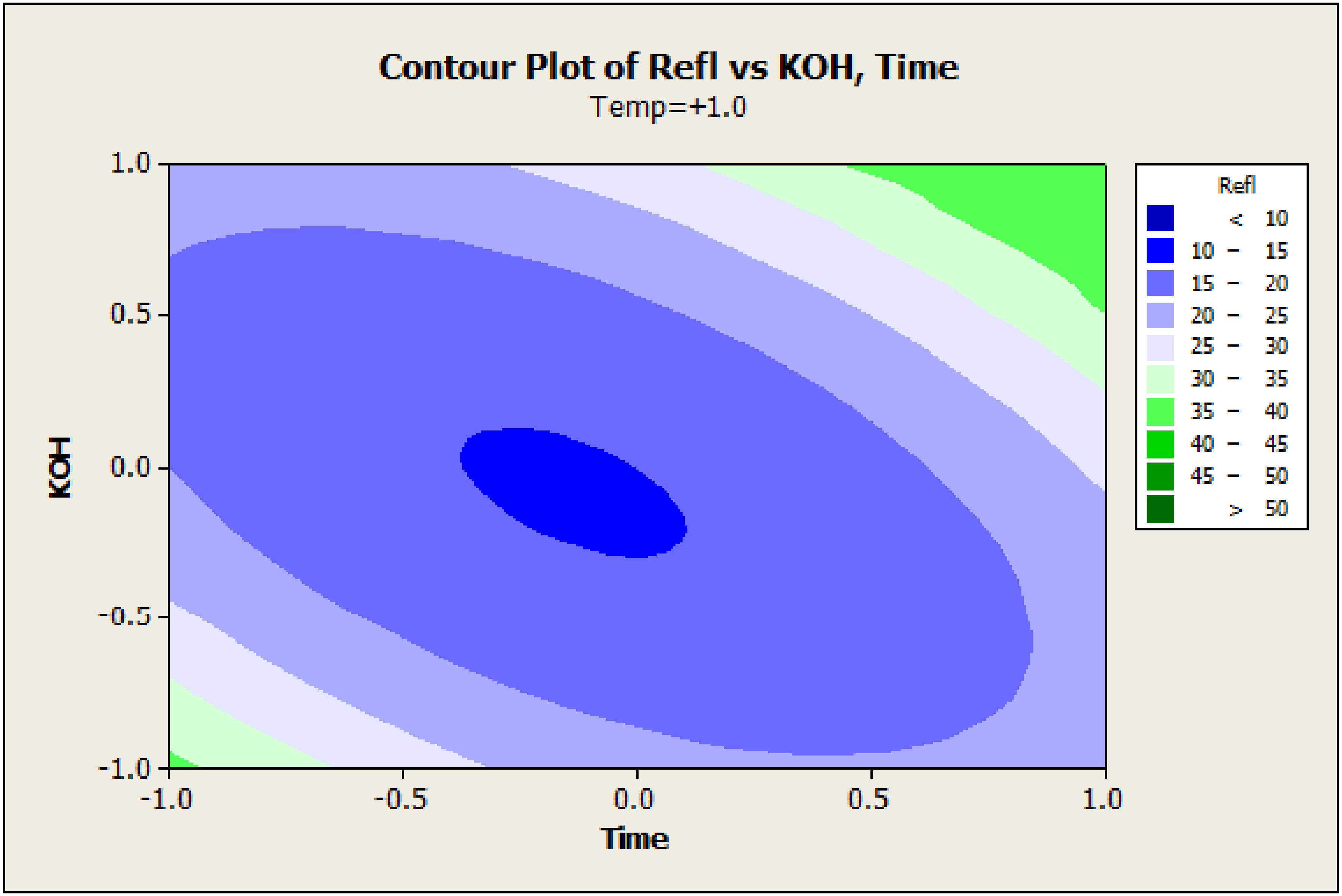
For this design, 11 runs are needed to screen these four factors to determine:

- Which factors have the most effect
- Consistency of the midpoint process
- Deviation of midpoint from linear model

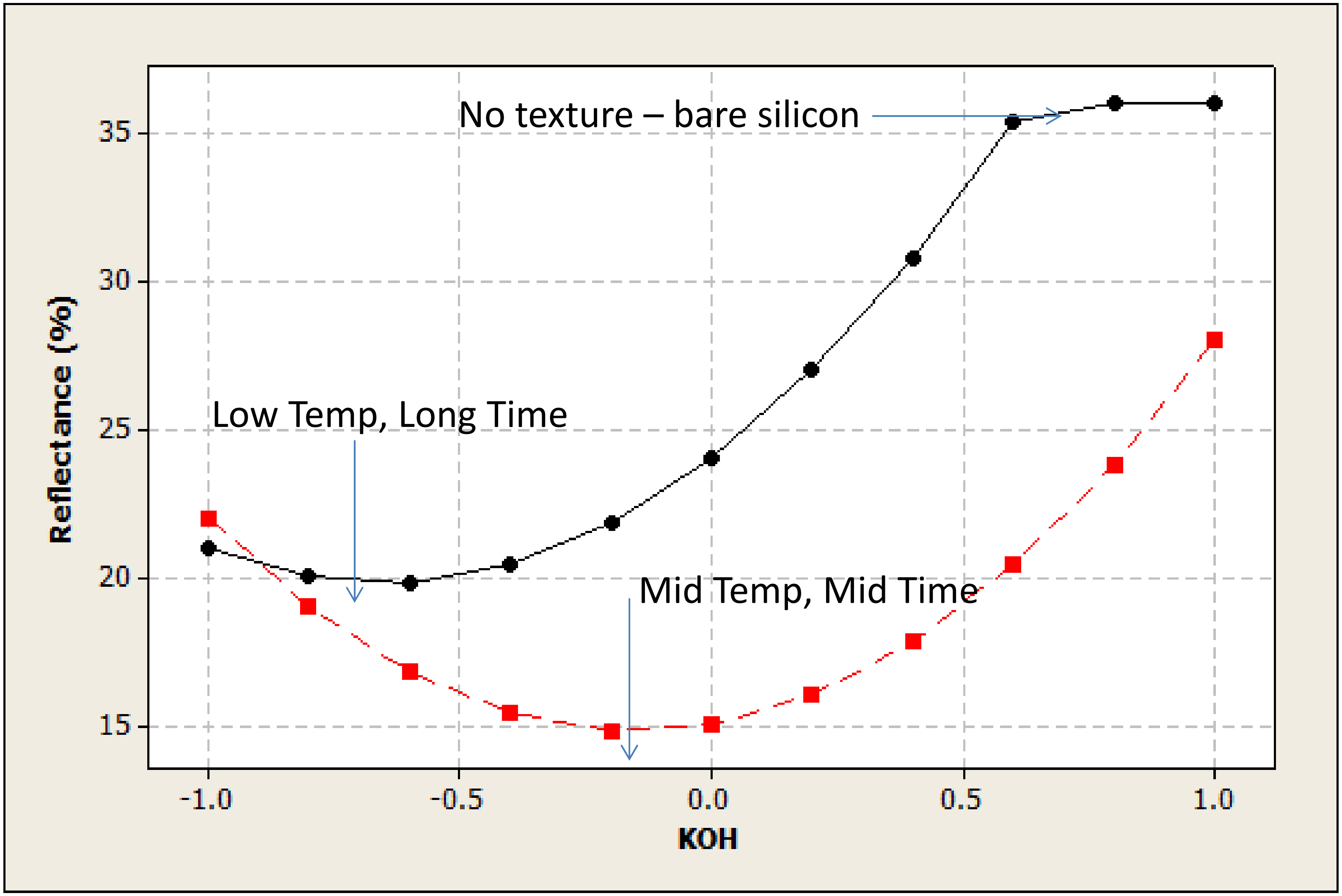
↓

If the average at the midpoint is significantly different than the average of the corners, then expect a maximum/minimum within this space. Non-linear (squared) terms must be included in the model if you want to locate this optimum.

Alkaline Texture Process (3 factors)



One Factor at a Time (KOH)

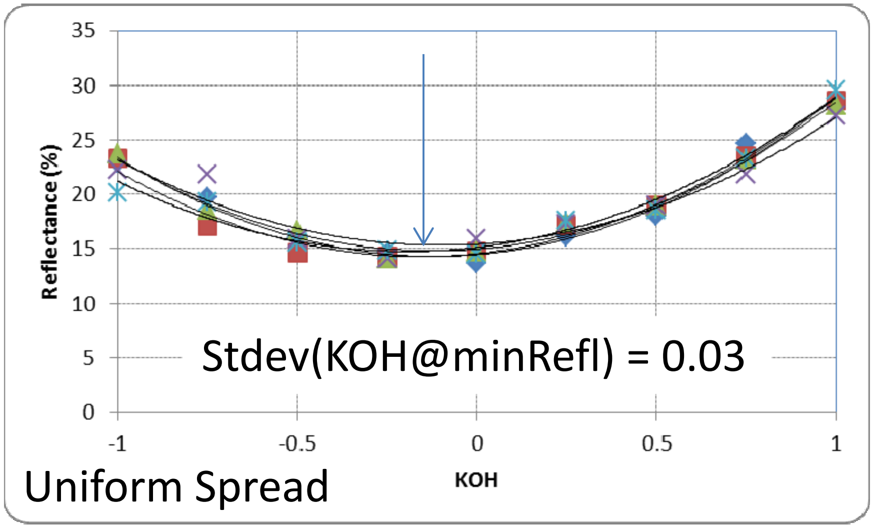


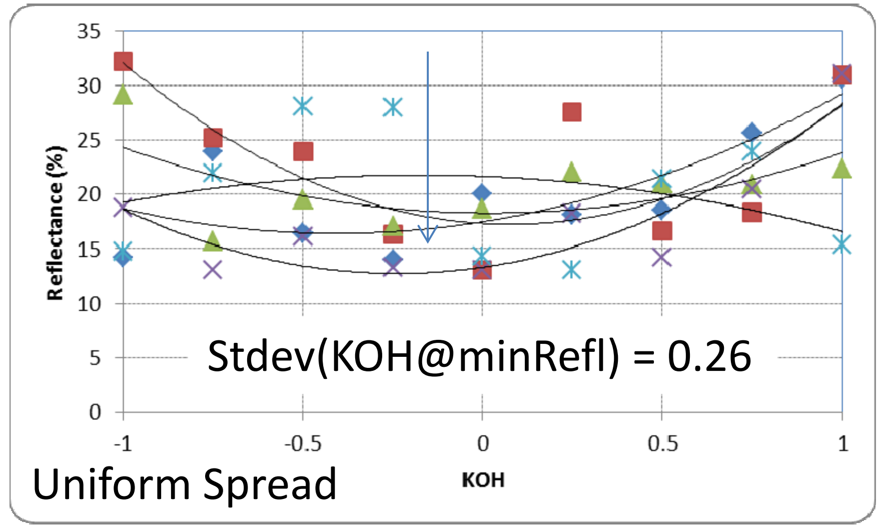
Now add experimental noise: y' = MIN(36,MAX(13,y+StDev*NORM.INV(RAND(),0,1)))

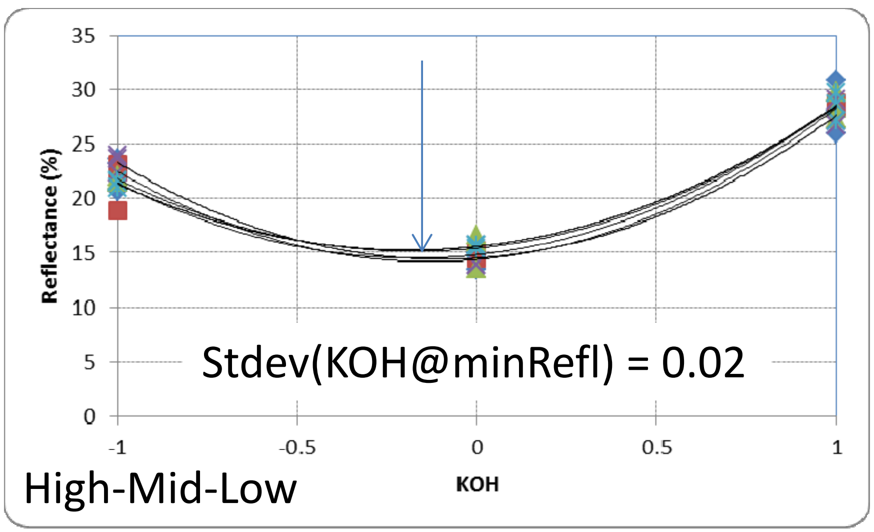
Uniform Spread, or High-Mid-Low?

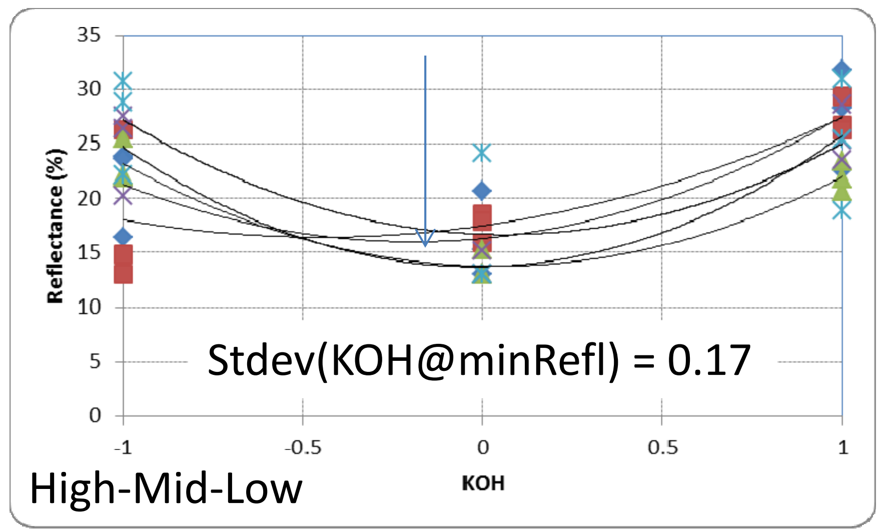
Stdev(Refl) = 1 percentage point

Stdev(Refl) = 5 percentage points



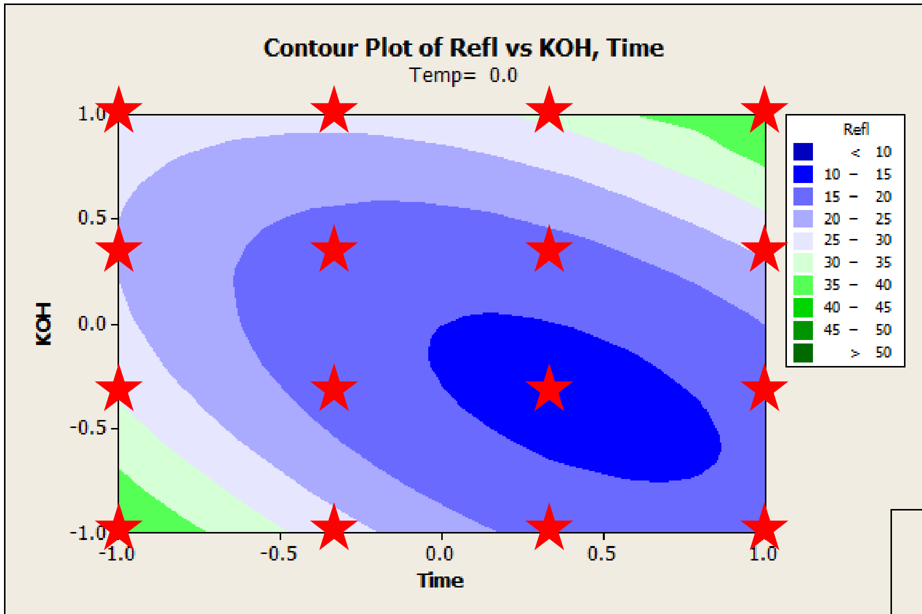






☞ High-Mid-Low better than Uniform Spread for normal distribution, smooth variation

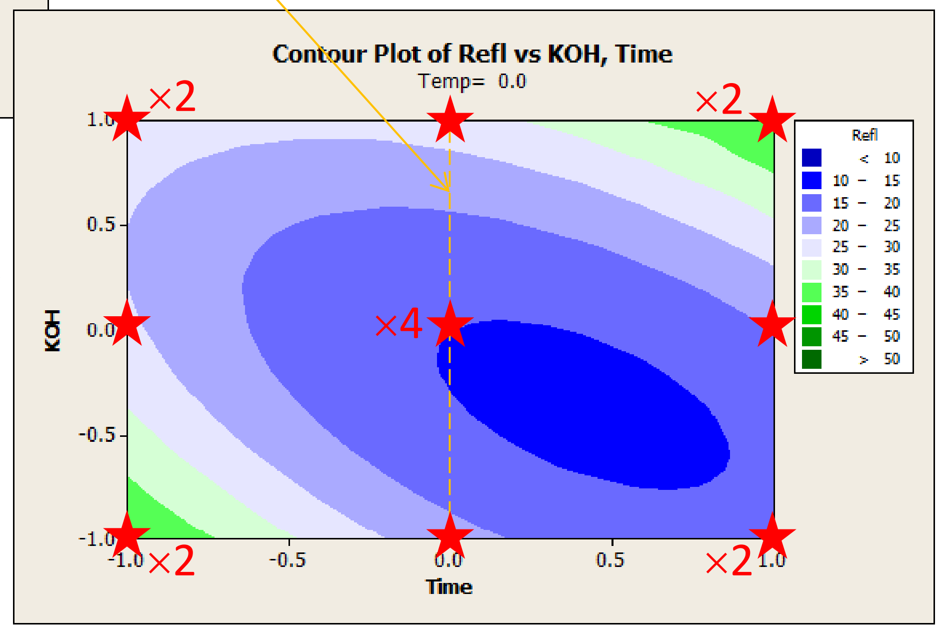
Picking Points in Two Dimensions



16 points distributed uniformly: "Uniform Spread"

Trace of previous 1D example

16 points concentrated around perimeter and at center: "Corner-Edge-Center"



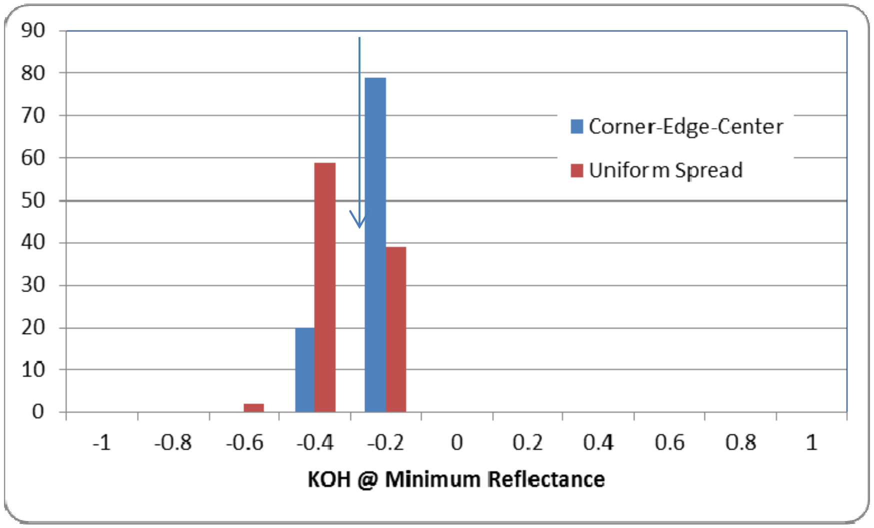
Note texture failure at two corners!

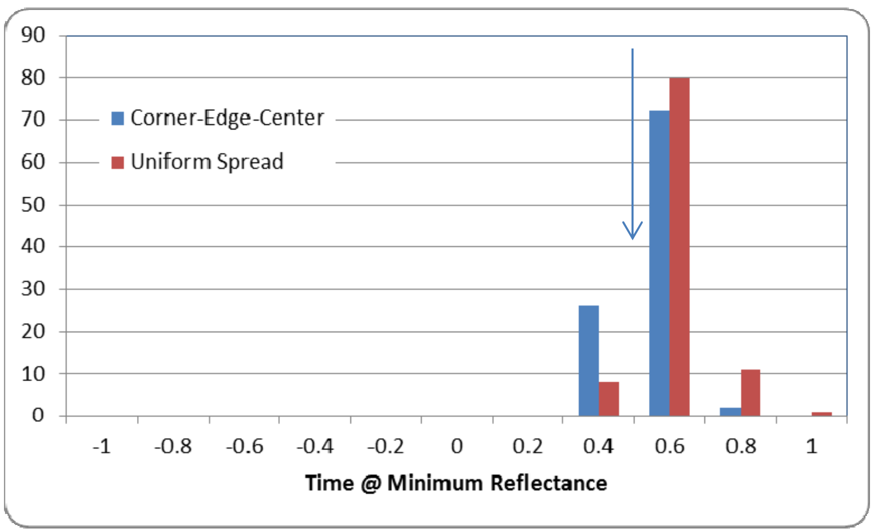
Optimum Process from Parabolic Fit

Results from 100 16-point experiments

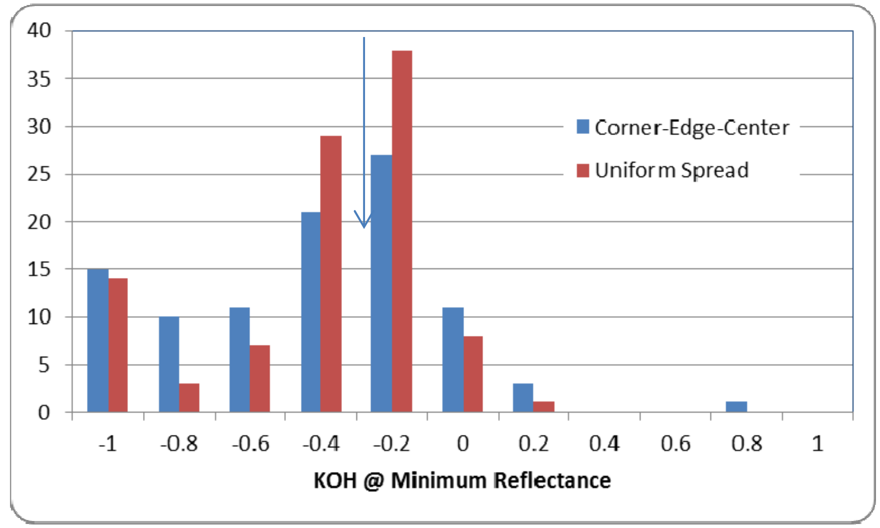
Hanwha

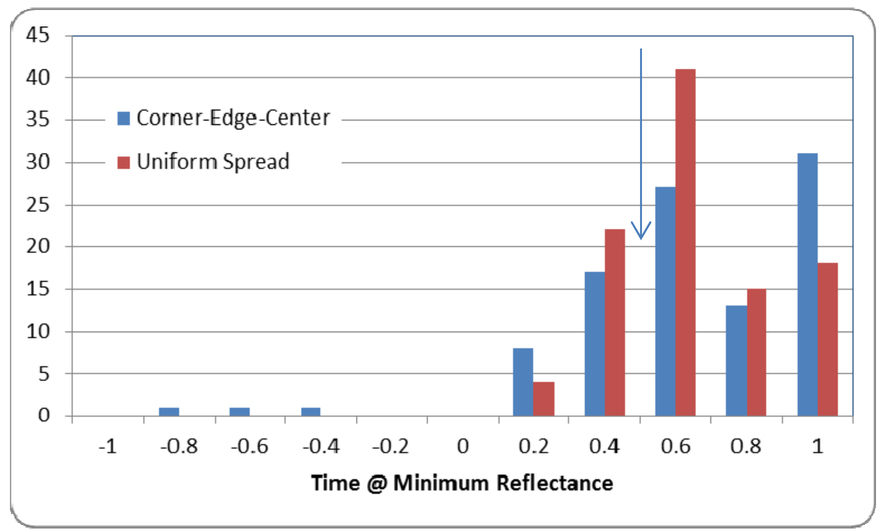
Stdev(Refl) = 1 percentage point





Stdev(Refl) = 5 percentage points





☞ Corner-Edge-Center still slightly better despite process failure in two corners

3D: Corners-Faces-Center (8+6+2=16 pts)

Response Surface Regression:
Reflectance versus Time, Temp, KOH

```
The analysis was done using coded units.

Estimated Regression Coefficients for Reflectance

Term          Coef   SE Coef       T      P
Constant    16.793     1.622  10.354  0.000
Time        -2.188     1.083  -2.020  0.090
Temp        -2.534     1.083  -2.339  0.058
KOH          2.432     1.083   2.245  0.066
Time*Time    5.765     2.110   2.732  0.034
Temp*Temp    1.836     2.110   0.870  0.418
KOH*KOH      6.486     2.110   3.074  0.022
Time*Temp    1.687     1.211   1.393  0.213
Time*KOH     5.285     1.211   4.363  0.005
Temp*KOH    -1.666     1.211  -1.376  0.218

S = 3.42584
```

Probability that each factor could actually be zero (P)

Maybe significant (Time, Temp, KOH)

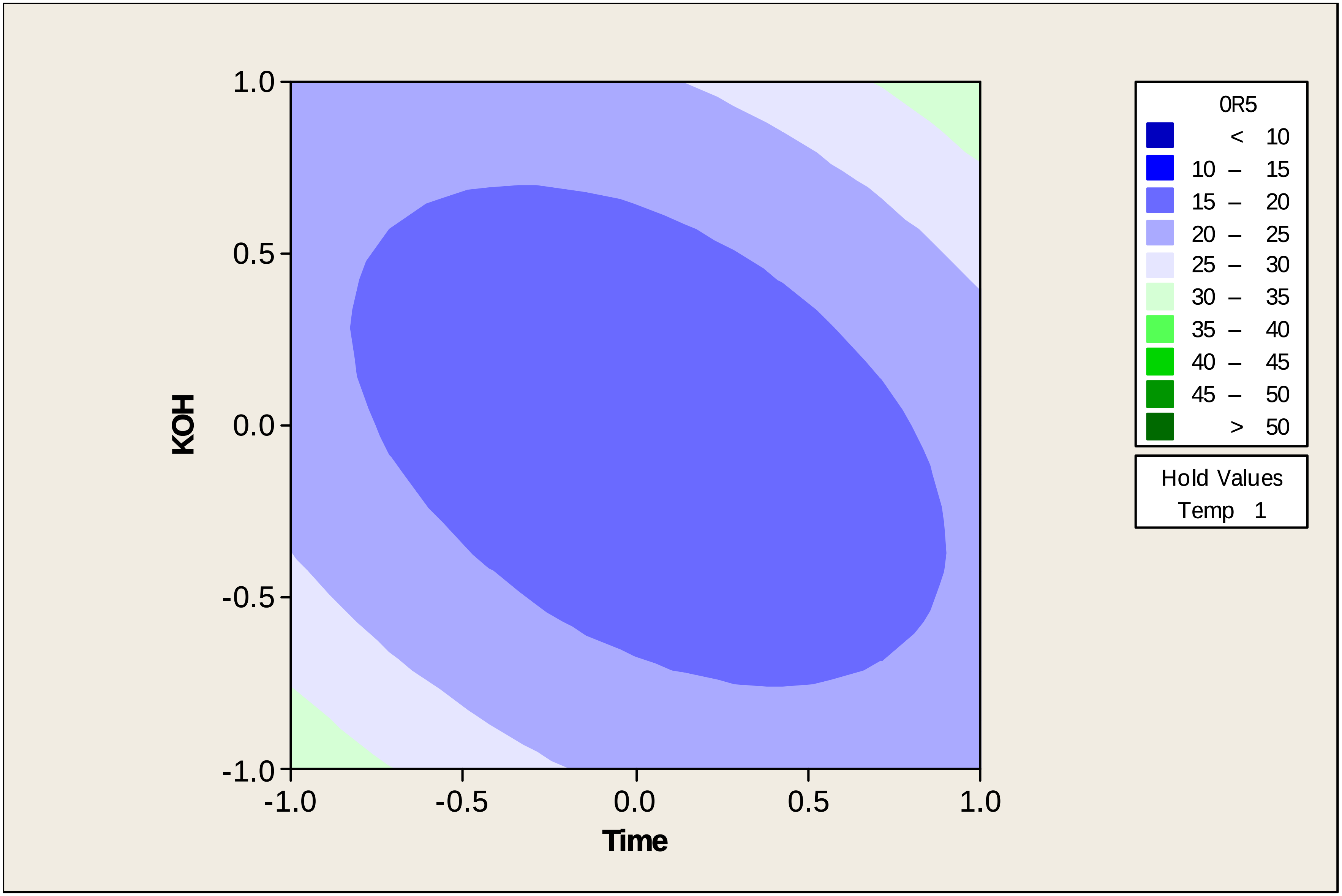
Statistically significant (Time*Time, KOH*KOH, Time*KOH)

Not significant (Temp*Temp, Time*Temp, Temp*KOH)

S: Stdev of the deviation of reflectance(%) from the model after accounting for the effect of the factors – due to both modeling error (3.3%) and random noise (1%).

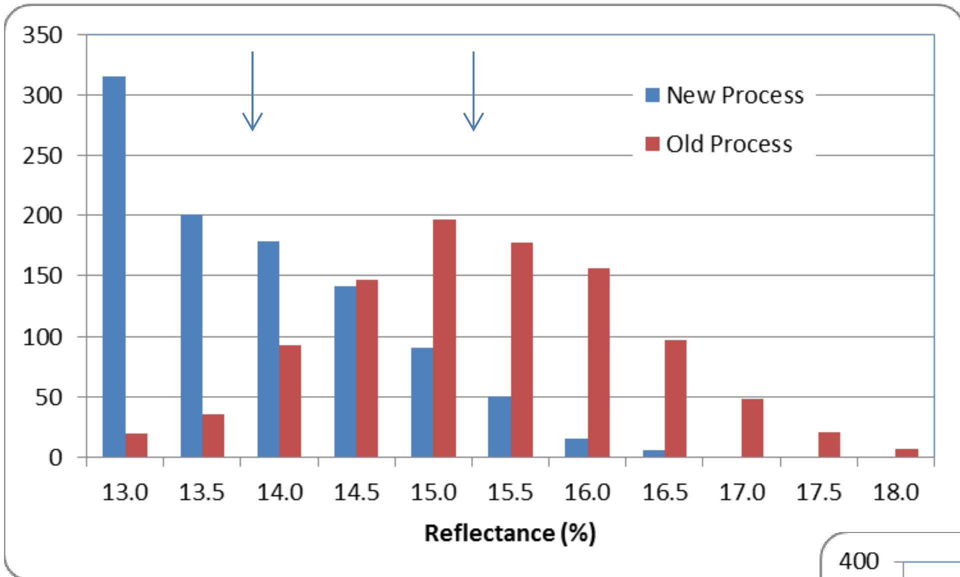
Here, due mostly to process failure in some regions of the parameter space

3D Response Surface from Regression Fit

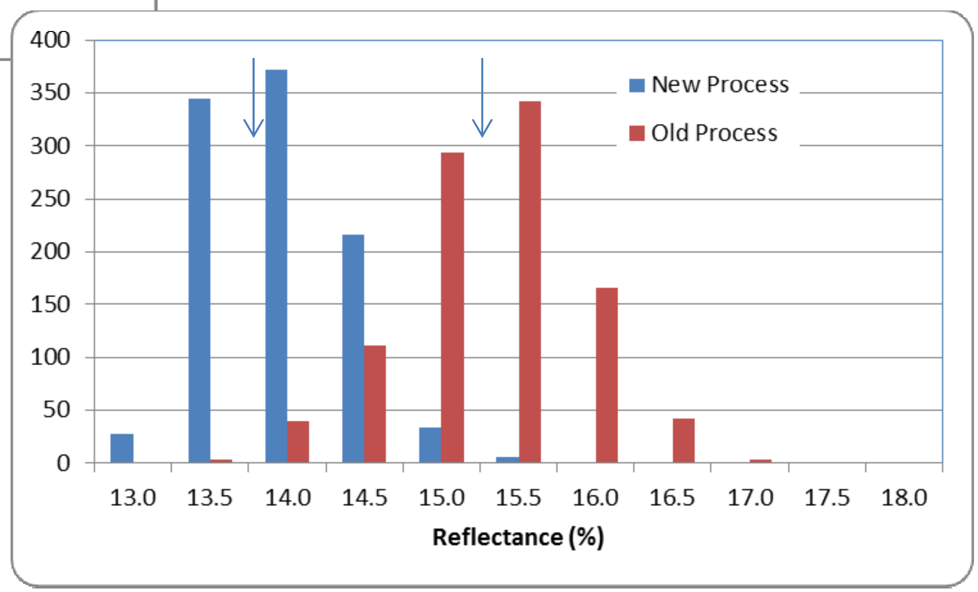


Old Process: Time=0, Temp=0, KOH=0 New Process: Time=0.25, Temp=0.31, KOH=-0.19

Validation



Do one run each of the Old Process and the New Process. This is the distribution of 1000 such attempts. The new process will appear better only 86% of the time.



Compare average of three runs each of Old Process and New Process. The new process will be better 97% of the time. Note the effect of Central Limit Theorem!

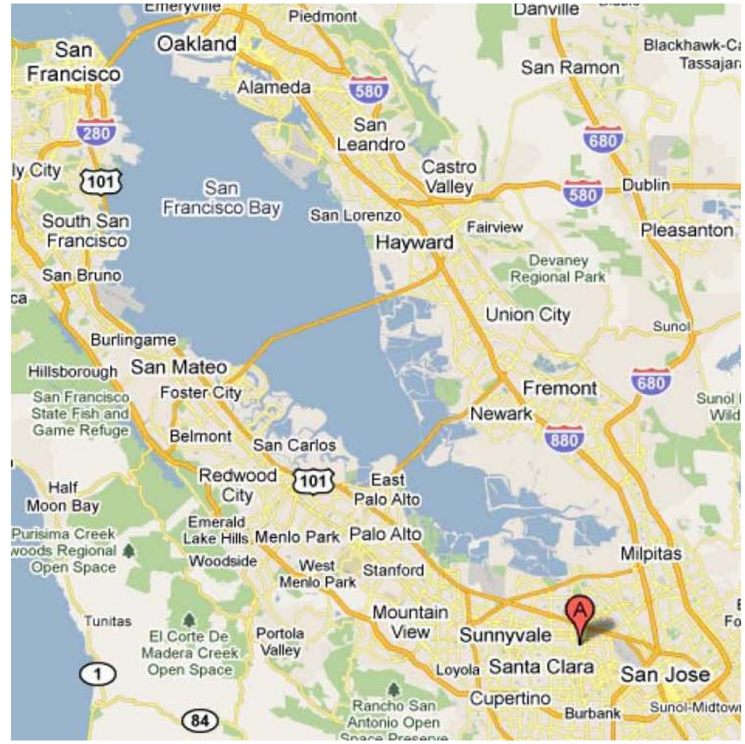
The above simulations used Stdev(Reflectance) of 1 percentage point.

Final Words of Caution

- If a result is statistically significant, still *heed the scope of the comparison*
  - If all data was collected on Monday, the conclusion may only apply on Monday
  - If all data was from a single tool, the conclusion may only apply to that tool
  - If all data used identical wafers, the conclusion may only apply to those wafers
  - If one person did all of the processing, the conclusion may only apply to them
- Rigorously keeping everything the same except the experimental factors?
  - Most likely to give a statistically significant result, *but with a narrow scope*
  - The result may not be repeatable under slightly difference circumstances!
    - ✓ Spread your replicates across multiple days
    - ✓ Randomize the run order to avoid misinterpretation of process drift over time
    - ✓ If wafer quality is the main source of variability, use an assortment of wafers
    - ✓ If you can, obtain replicates using different people, tools and process materials
- <u>The most valuable results are those that others can replicate</u>
  - In PV, unless it gets put into high-volume manufacturing, what's the point?
  - Collaboration between different R&D labs is a good thing!

In publications, show statistical significance and clarify the scope of your results

When in the Bay Area, Visit the ARDL!



2424 Walsh Avenue, Santa Clara, CA 95051 USA