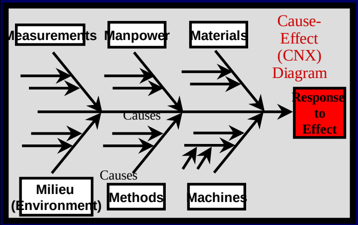
National Defense Industrial Association

Promoting National Security Since 1919

Process

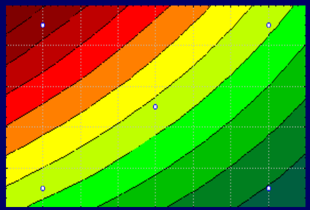


Plan

		In Front		In Back	
		Face East	Face West	Face East	Face West
Eyes Open	Left Hand	0.43	0.58	0.52	0.40
	Right Hand	0.62	0.29	0.28	0.36
Eyes Closed	Left Hand	0.62	0.57	0.47	0.40
	Right Hand	0.42	0.26	0.42	0.47

Produce

Ponder



Raising the Bar: Equipping DoD Testers to Excel with DOE

presented to:

NDIA T&E Conference

1 Mar 2010

Greg Hutto
Wing Ops Analyst, 46th Test Wing
gregory.hutto@eglin.af.mil

Bottom Line Up Front

- **Test design is not an art…it is a *science***
  - **Talented scientists in T&E Enterprise however…limited knowledge in test design…alpha, beta, sigma, delta, p, & n**
- **Our decisions are too important to be left to professional opinion alone…our decisions should be based on *mathematical fact***
- **53d Wg, AFOTEC, AFFTC, and 46 TW/AAC experience**
  - **Teaching DOE as a sound test strategy not enough**
  - **Leadership from senior executives (SPO & Test) is key**
- **Purpose: DOD adopts experimental design as the default approach to test, wherever it makes sense**
  - **Exceptions include demos, lack of trained testers, no control**

Background -- Greg Hutto

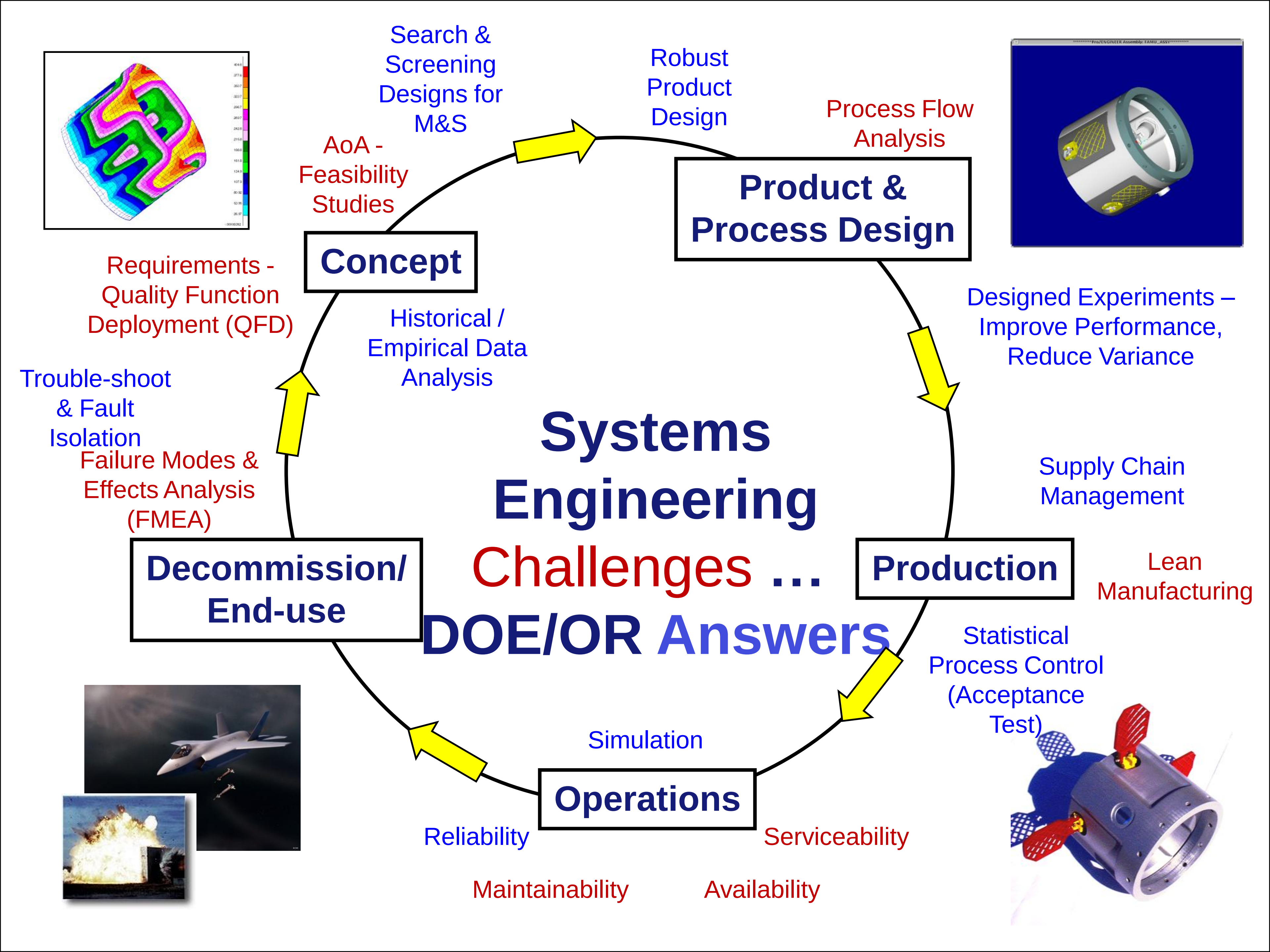
- B.S. US Naval Academy, Engineering - Operations Analysis
- M.S. Stanford University, Operations Research
- USAF Officer -- TAWC Green Flag, AFOTEC Lead Analyst
- Consultant -- Booz Allen & Hamilton, Sverdrup Technology
- Mathematics Chief Scientist -- Sverdrup Technology
- Wing OA and DOE Champion – 53rd Wing, now 46 Test Wing
- USAF Reserves – Special Assistant for Test Methods (AFFTC/CT) and Master Instructor in DOE for USAF TPS

Practitioner, Design of Experiments -- 19 Years

Selected T&E Project Experience – 18+ Years

- Green Flag EW Exercises '79
- F-16C IOT&E '83
- AMRAAM, JTIDS ' 84
- NEXRAD, CSOC, Enforcer '85
- Peacekeeper '86
- B-1B, SRAM, '87
- MILSTAR '88
- MSOW, CCM '89
- Joint CCD T&E '90

- SCUD Hunting '91
- AGM-65 IIR '93
- MK-82 Ballistics '94
- Contact Lens Mfr '95
- Penetrator Design '96
- 30mm Ammo '97
- 60 ESM/ECM projects '98--'00
- B-1B SA OUE, Maverick IR+, F-15E Suite 4E+,100's more '01-'06


Search & Screening Designs for M&S
Robust Product Design
Process Flow Analysis
AoA - Feasibility Studies
Product & Process Design
Concept
Requirements - Quality Function Deployment (QFD)
Designed Experiments – Improve Performance, Reduce Variance
Historical / Empirical Data Analysis
Trouble-shoot & Fault Isolation
Systems Engineering Challenges …
DOE/OR Answers
Supply Chain Management
Failure Modes & Effects Analysis (FMEA)
Decommission/ End-use
Production
Lean Manufacturing
Statistical Process Control (Acceptance Test)
Simulation
Operations
Reliability
Serviceability
Maintainability
Availability

Systems Engineering Employ Many Simulations of Reality

<table>
<tr><td></td><td></td><td colspan="6">Simulation of Reality</td></tr>
<tr><td colspan="2">Acq Phase</td><td colspan="2">M&S</td><td colspan="2">Hardware</td><td colspan="2">System/Flight Test</td></tr>
<tr><td>Reqt'ts Dev</td><td></td><td rowspan="7">Warfare</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>AoA</td><td rowspan="6">Physics</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Concepts</td><td></td><td rowspan="5">HWIL/SIL</td><td rowspan="5">Captive</td><td rowspan="5">Subsystem</td><td rowspan="2">Prototype</td></tr>
<tr><td></td><td>Risk Reduction</td></tr>
<tr><td>EMD</td><td></td><td rowspan="2">Prod Rep</td></tr>
<tr><td></td><td>Prod & Mfr</td></tr>
<tr><td>Sustain</td><td></td><td>Production</td></tr>
</table>

- At each stage of development, we conduct experiments
  - Ultimately – how will this device function in service (combat)?
  - Simulations of combat differ in fidelity and cost
  - Differing goals (screen, optimize, characterize, reduce variance, robust design, trouble-shoot)
  - Same problems – distinguish truth from fiction: What matters? What doesn’t?

DOE in Action: Six Sigma at General Electric

- Fortune Global 2000 #1 in 2009 - based on Sales, Profit, Assets and Market Value ($180B, $17B, $800B, $90B)
- Climate Brand Index #1
- Global Most Admired Company - Fortune 2005 - #1 or #2 last 10 years
- Global Most Respected – Barron's – top 15 consistently
- Recognized globally as one of the best for for innovation, investor relations, reporting, working mothers, ethics, accountability
- Products and Services
  - Aircraft engines, appliances, financial services, aircraft leasing, equity, credit services, global exchange services, NBC, industrial systems, lighting, medical systems, mortgage insurance, plastics

GE's (Re)volution In Improved Product/Process

- Began in late 1980's facing foreign competition
- 1998, Six Sigma Quality becomes one of three company-wide initiatives
- '98 Invest $0.5B in training; Reap $1.5B in benefits!
- "Six Sigma is embedding quality thinking - process thinking - across every level and in every operation of our Company around the globe"[1]
- "Six Sigma is now the way we work – in everything we do and in every product we design" [1]

1 Jack Welch - General Electric website at ge.com

DOE in Historical Stats Timeline[1]

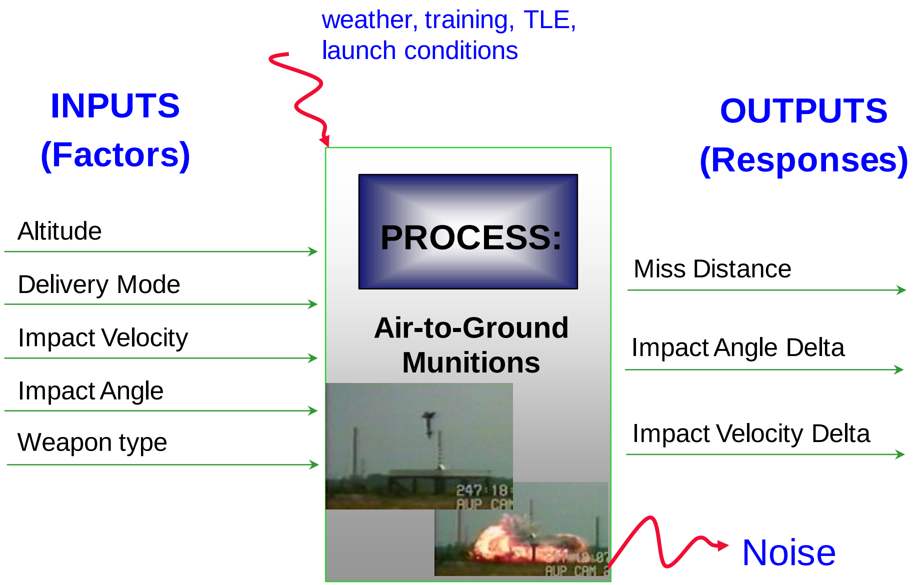
1800	Least Squares (Gauss, Legendre)
1850	
1900	Regression Concepts (Pearson, Galton)
1910	t-test (Gosset)
1920	**Factorial Experiments and ANOVA (Fisher)**
1930	Formalized Hypothesis Tests and Power (Neyman, Pearson)
1940	2^k Factorial Designs (Yates) Fractional Factorial Designs (Finney, Rao)
1950	Taguchi develops his methods Central Composite, Response Surface Methods (RSM) Designs (Box, Wilson)
1960	Optimal Designs (Kiefer, Wolfowitz) Box-Behnken Designs (Box, Behnken)
1970	2^{k-p} Fractional Factorial Resolution (Box, Hunter)
1980	Algorithm for D-optimal designs (Mitchell, Nachtsheim)
1990	Detecting dispersion effects ratio of variances (Montgomery)
2000	53 Wing Starts using DOE

1. Source: Appendix K -- *Understanding Industrial Designed Experiments*, Schmidt and Launsby, 1998

Overview

- **Four Challenges – The JPADS Drop Test**
- **What's the *BEST* strategy of test?**
- **DOE Fables for the T&E Enterprise**
- **How can we deploy DOE to revolutionize Test?**
- **Summary**

What are Statistically Designed Experiments?



- Purposeful, systematic changes in the inputs in order to observe corresponding changes in the outputs
- Results in a mathematical model that predicts system responses for specified factor settings

$$\text{Responses} = f\left(\text{Factors}\right) + \varepsilon$$

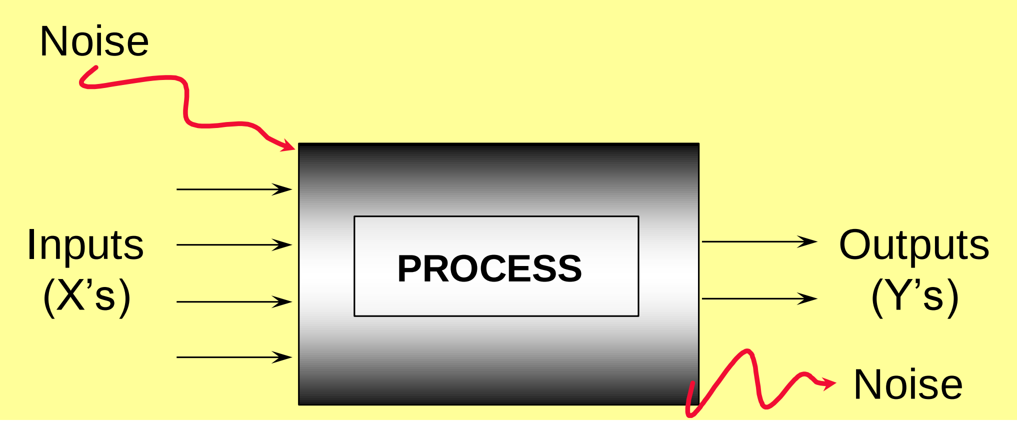
Why DOE? One Slide… DOE Gives Scientific Answers to Four Fundamental Test Challenges

Four Challenges faced by any test

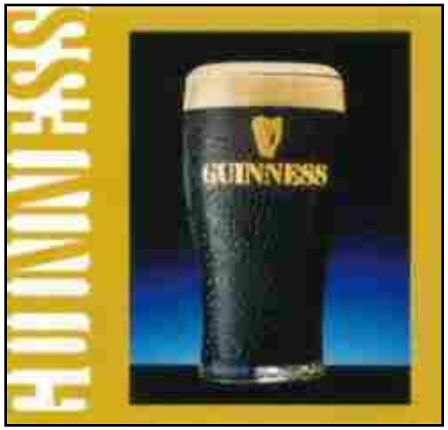
1. *How many? Depth of Test* – effect of test size on uncertainty
2. *Which Points? Breadth of Testing* – spanning the vast employment battlespace
3. *How Execute? Order of Testing* – insurance against “unknown-unknowns”
4. *What Conclusions? Test Analysis* – drawing objective, scientific conclusions while controlling noise

DOE effectively addresses all these challenges!

In our short time today, address primarily #1 and #2.



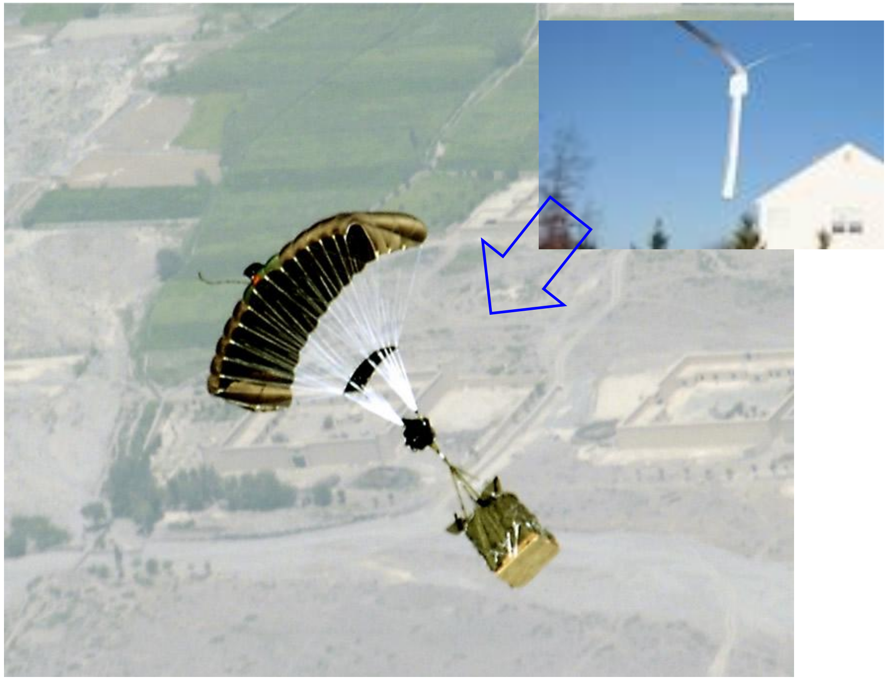
A beer and a blemish …



- 1906 – W.T. Gossett, a Guinness chemist
- Draw a yeast culture sample
- Yeast in this culture?
- Guess too little – incomplete fermentation; too much -- bitter beer
- He wanted to get it *right*

- 1998 – Mike Kelly, an engineer at contact lens company
- Draw sample from 15K lot
- How many defective lenses?
- Guess too little – mad customers; too much -- destroy good product
- He wanted to get it *right*

Today's Example – Precision Air Drop System



The dilemma for airdropping supplies has always been a stark one. High-altitude airdrops often go badly astray and become useless or even counter-productive. Low-level paradrops face significant dangers from enemy fire, and reduce delivery range. Can this dilemma be broken?

A new advanced concept technology demonstration shows promise, and is being pursued by U.S. Joint Forces Command (USJFCOM), the U.S. Army Soldier Systems Center at Natick, the U.S. Air Force Air Mobility Command (USAF AMC), the U.S. Army Project Manager Force Sustainment and Support, and industry. The idea? Use the same GPS-guidance that enables precision strikes from JDAM bombs, coupled with software that acts as a flight control system for parachutes. JPADS (the Joint Precision Air-Drop System) has been combat-tested successfully in Iraq and Afghanistan, and appears to be moving beyond the test stage in the USA… and elsewhere.

Capability:

Assured SOF re-supply of material

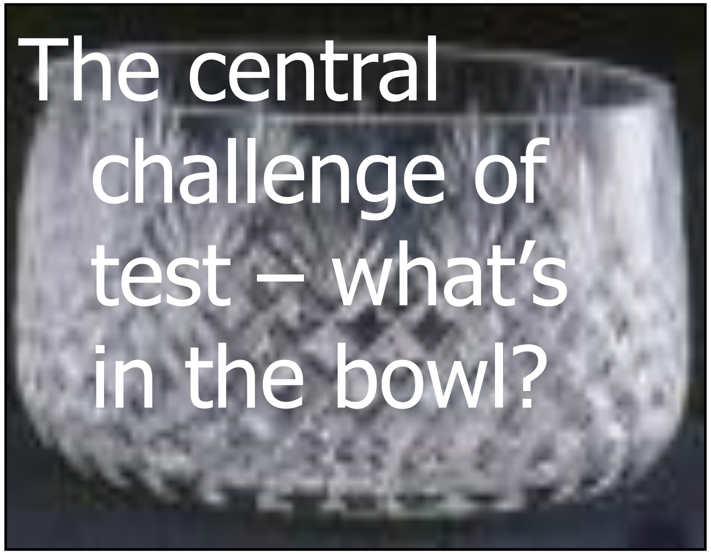
Requirements:

Probability of Arrival
Unit Cost $XXXX
Damage to payload
Payload
Accuracy
Time on target
Reliability …

- **Just when you think of a good class example – they are already building it!**
- **46 TS – 46 TW Testing JPADS**

Challenge #1 … How Many?

- In all our testing – we reach into the bowl (reality) and draw a sample of JPADS performance
- Consider an "80% JPADS"
  - Suppose a required 80% P(Arrival)
  - Is the Concept version acceptable?
- We don't know in advance which bowl God hands us …
  - The one where the system *works* or,
  - The one where the system *doesn't*



Start -- Blank Sheet of Paper: How Many?

- Let’s draw a sample of _*n*_ drops
- How many is enough to get it *right*?
  - 3 – because that’s how much $/time we have
  - 8 – because I’m an 8-guy
  - 10 – because I’m challenged by fractions
  - 30 – because something good happens at 30!
- Let’s start with 10 and see …

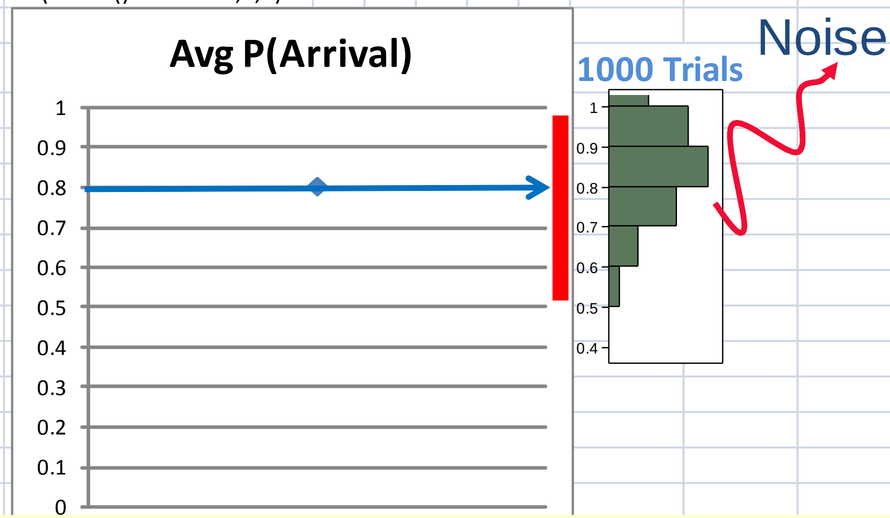
=> Switch to Excel File – JPADS Pancake.xls

Embedded Excel Simulation to Address "How Many?"

True P(Arrival)	80%
Trial/Rep	Miss?
1	0
2	1
3	1
4	1
5	1
6	1
7	1
8	0
9	1
10	1
AvgMiss	0.8

0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

=IF(RAND()<"TRUTH",1,0)



Definitions:

α - false positive error rate - concluding a difference exists (good or bad) when the difference is noise.

Confidence is 1-α.

β - false negative error rate - failing to detect a difference when a difference is causally-based

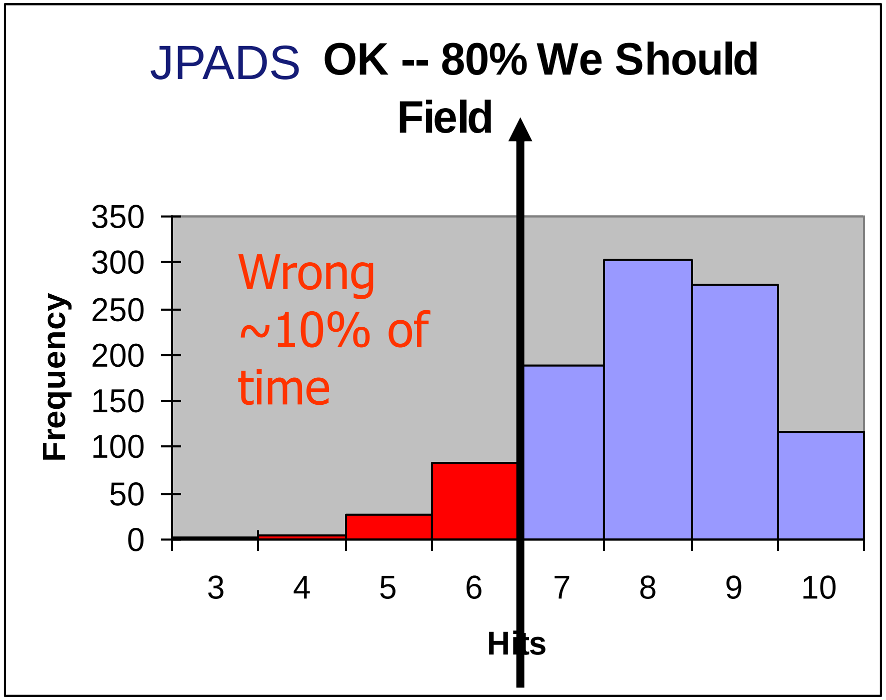
Power is 1- β.

We replicate to overcome sampling error but fail to quantify the *uncertainty* in our estimates.

A false positive – declaring JPADS is degraded (when it's not) -- α

- **Suppose we *fail* JPADS when it has 4 or more misses**
- **We'll be wrong (on average) about 10% of the time**
- **We can tighten the criteria (fail on 7) by failing to field more good systems**
- **We can loosen the criteria (fail on 5) by missing real degradations**
- **Let's see how often we miss such degradations …**

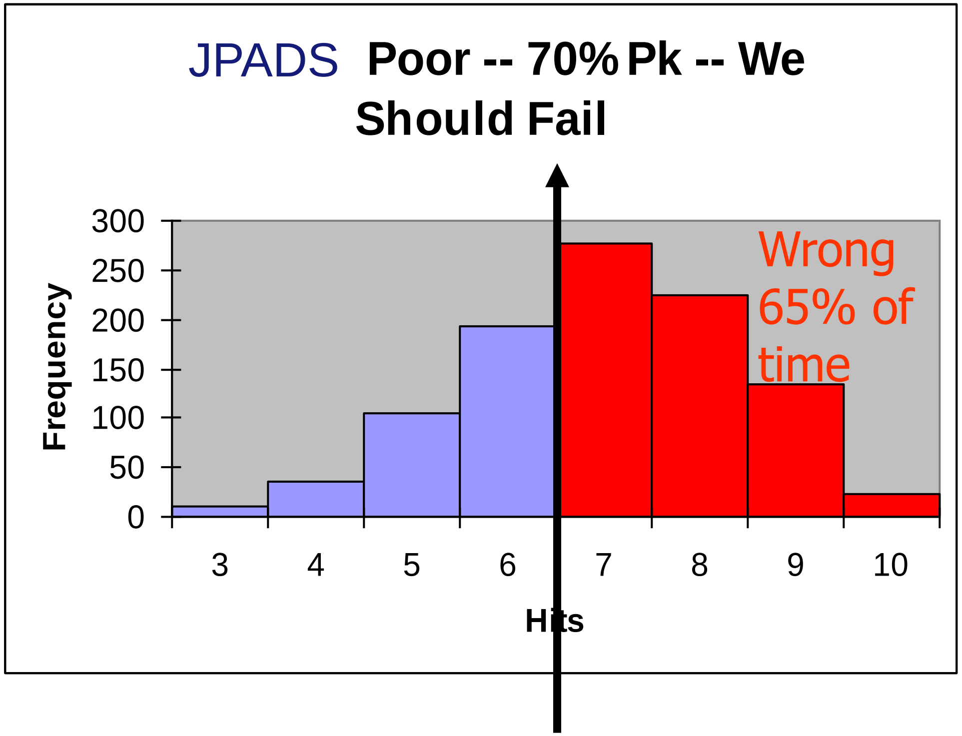
In this bowl – JPADS performance is acceptable



A false negative – we field JPADS (when it's degraded) -- β

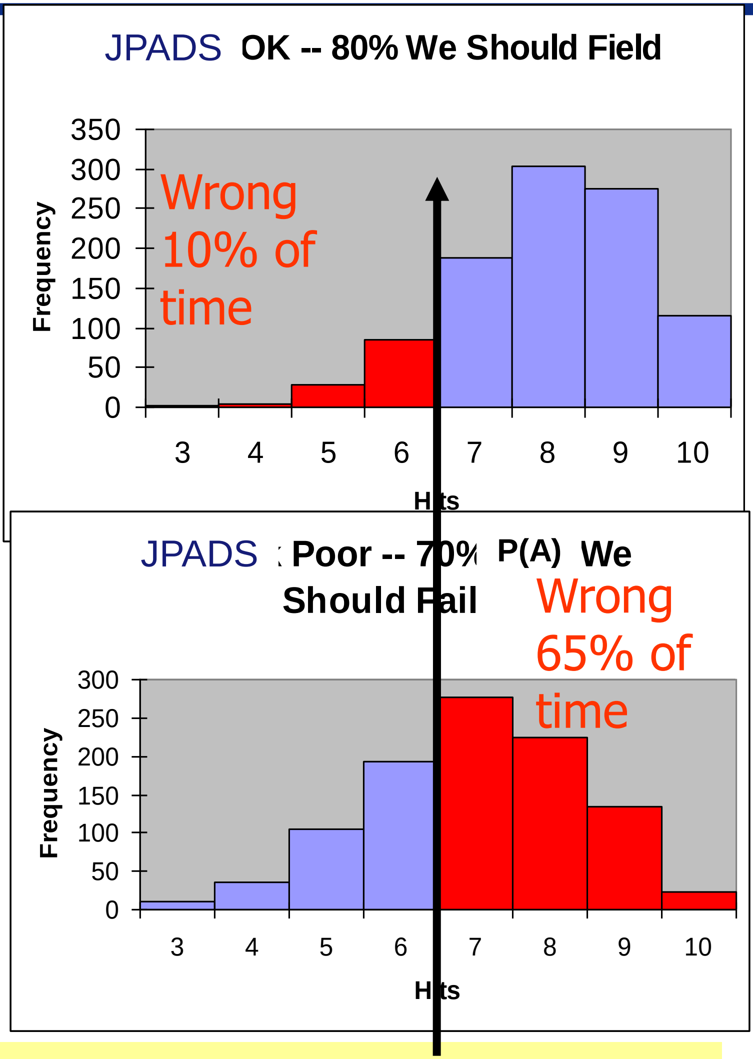
- **Use the failure criteria from the previous slide**
- **If we *field* JPADS with 6 or fewer hits, we fail to detect the degradation**
- **If JPADS has degraded, with n=10 shots, we're wrong about 65% of the time**
- **We can, again, tighten or loosen our criteria, but at the cost of increasing the other error**

In this bowl – JPADS P(A) ***decreased*** 10% -- it is degraded



We seek to balance our chance of errors

- **Combining, we can trade one error for other (α for β)**
- **We can also increase sample size to decrease our risks in testing**
- **These statements not opinion –<u>mathematical fact</u> and an inescapable challenge in testing**
- **There are two *other* ways out … factorial designs and real-valued MOPs**



Enough to Get It Right: ***Confidence*** in stating results; ***Power*** to find small differences

Challenge 2: Which Points? How Designed Experiments Solve This

Designed Experiment *(n). Purposeful control of the inputs (factors) in such a way as to deduce their relationships (if any) with the output (responses).*

Statistician G.E.P Box said …

"All math models are false …but some are useful."

"All experiments are designed … most, poorly."

Battlespace Conditions for JPADS Case

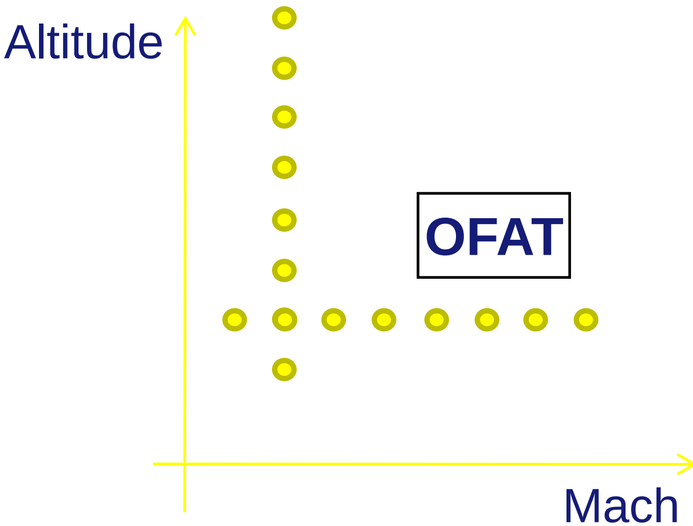
- **Systems Engineering Question: Does JPADS perform at required capability level across the planned battlespace?**

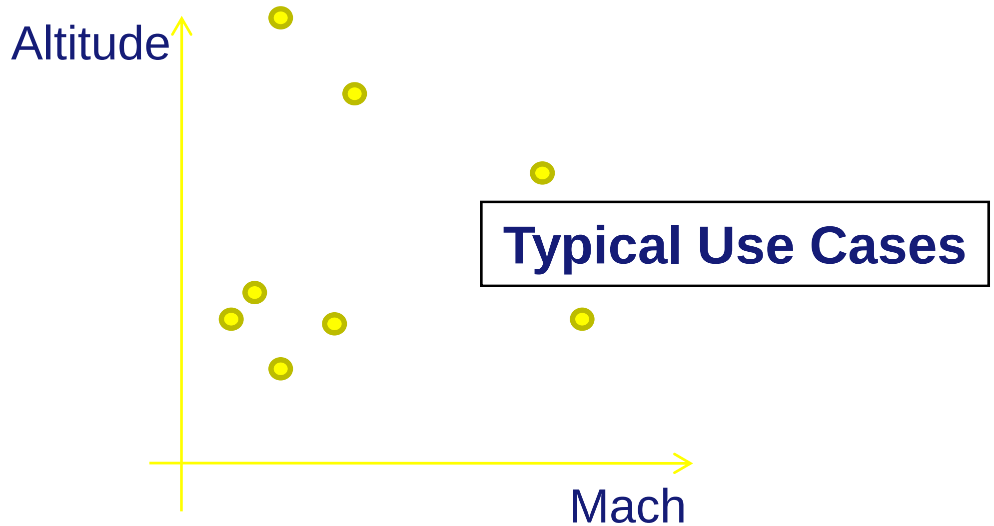
Type	Measure of Performance
Objective	Target acquisition range
	Target Standoff (altitude)
	launch range
	mean radial arrival distance
	probability of damage
	reliability
Subjective	Interoperability
	human factors
	tech data
	support equipment
	tactics

Conditions	Settings	# Levels
JPADS Variant:	A, B, C, D	4
Launch Platform:	C-130, C-17, C-5	3
Launch Opening	Ramp, Door	2
Target:	Plains, Mountain	2
Time of Day:	Dawn/Dusk, Mid-Day	3
Environment:	Forest, Desert, Snow	3
Weather:	Clear (+7nm), Haze (3-7nm), Low Ceiling/Visibility (<3000/3nm)	3
Humidity:	Low (<30%), Medium (31-79%), High (>80%)	3
Attack Azimuth:	Sun at back, Sun at beam, Sun on nose	3
Attack Altitude:	Low (<5000’), High (>5000’)	2
Attack Airspeed:	Low (Mach .5), Medium (Mach .72), High (Mach .8)	3
JPADS Mode:	Autonomous, Laser Guidance	2
	Combinations	**139,968**

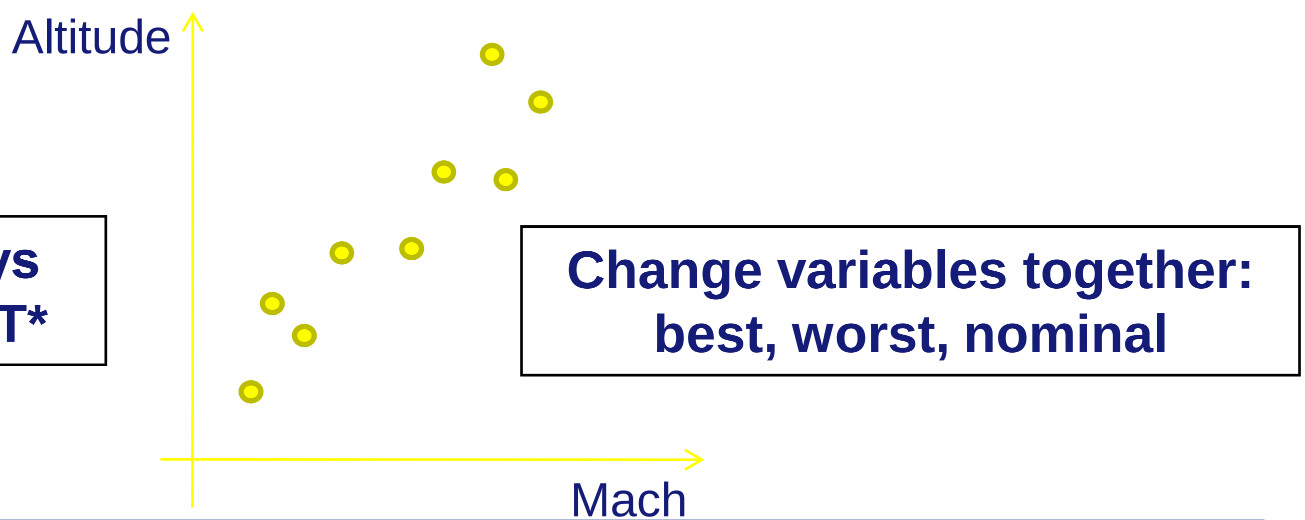
12 Dimensions - Obviously a large test envelope … how to search it?

Spanning the Battlespace – Traditional Test Designs



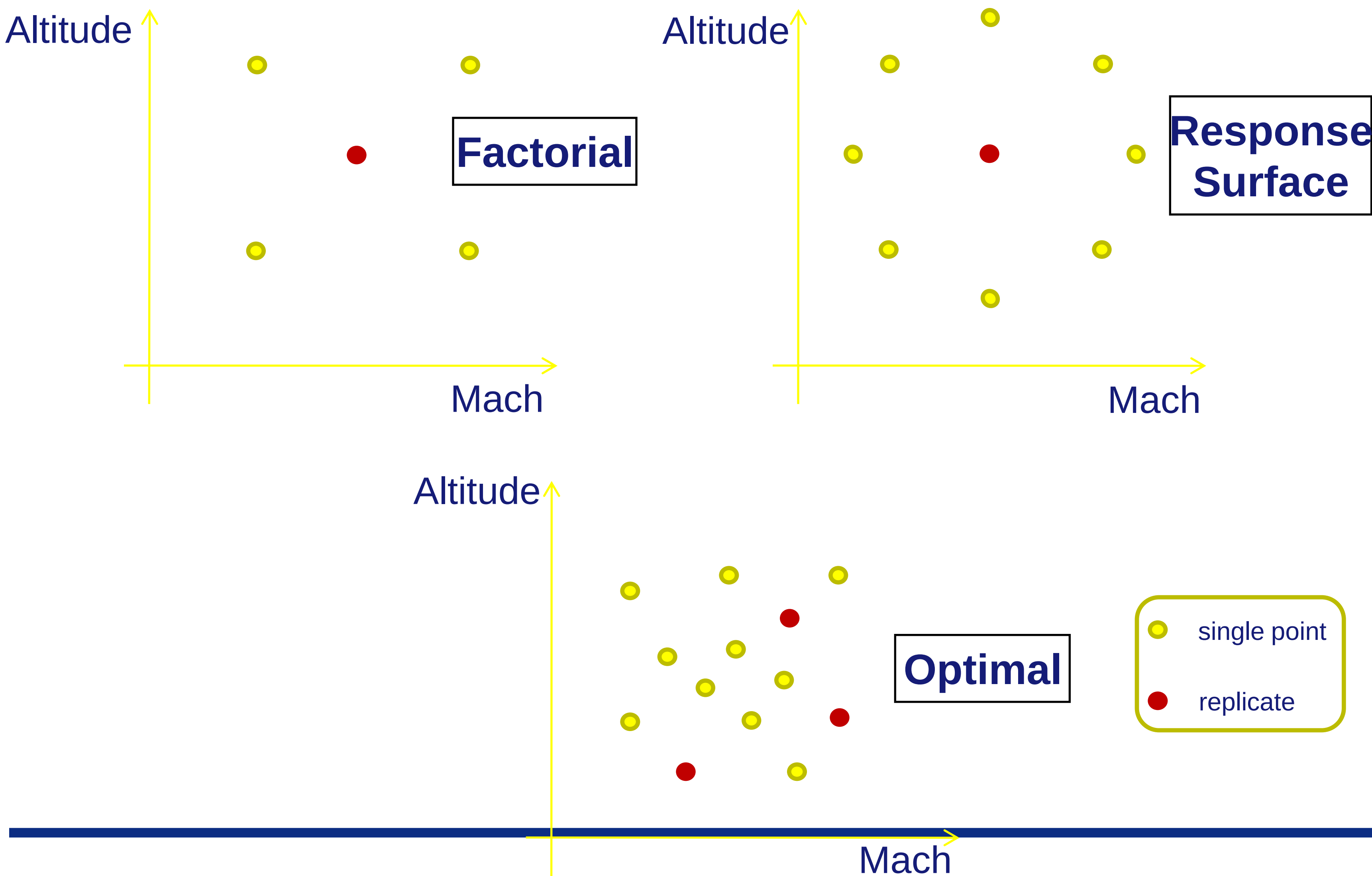


And … the always popular DWWDLT*

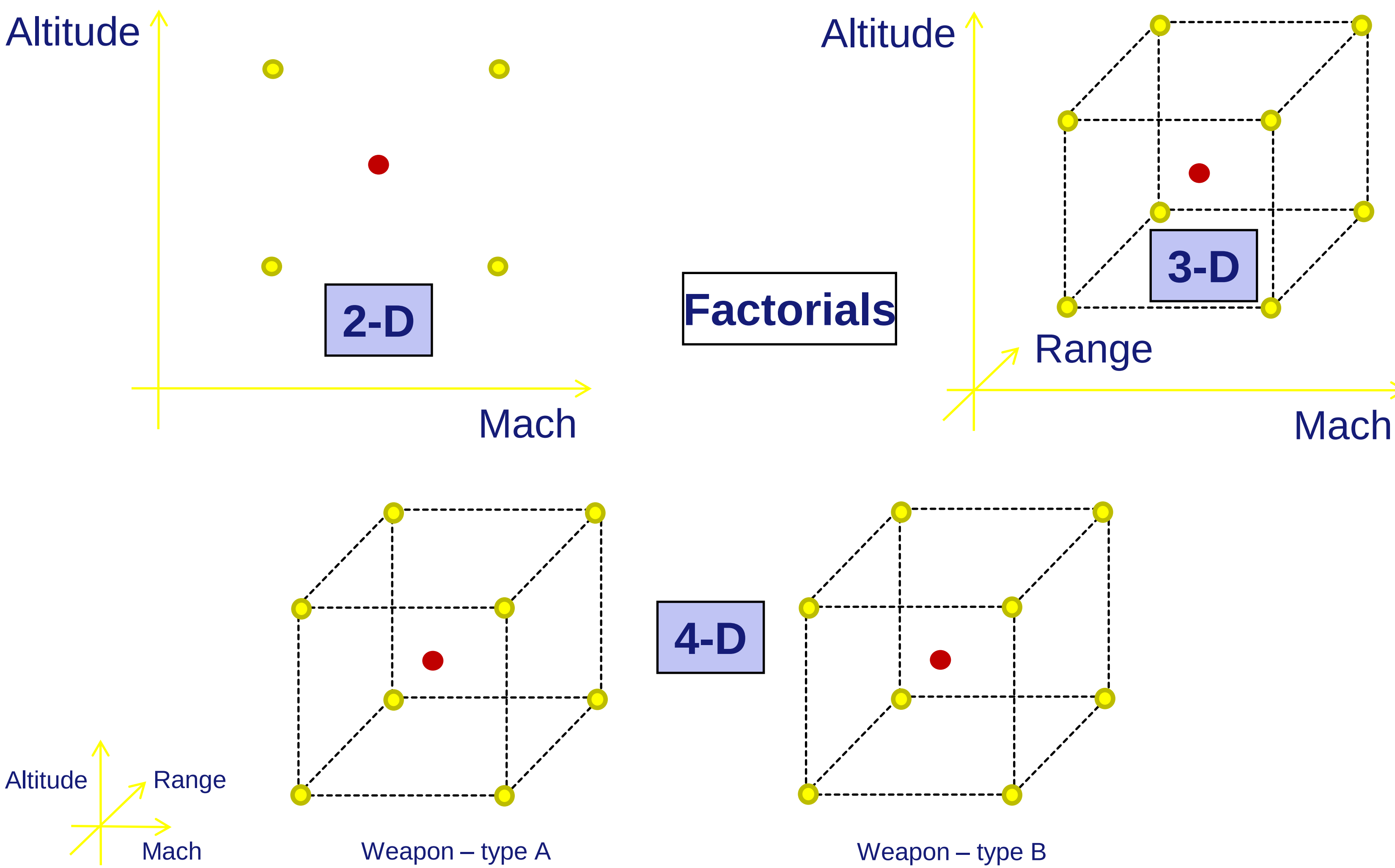


*** Do What We Did Last Time**

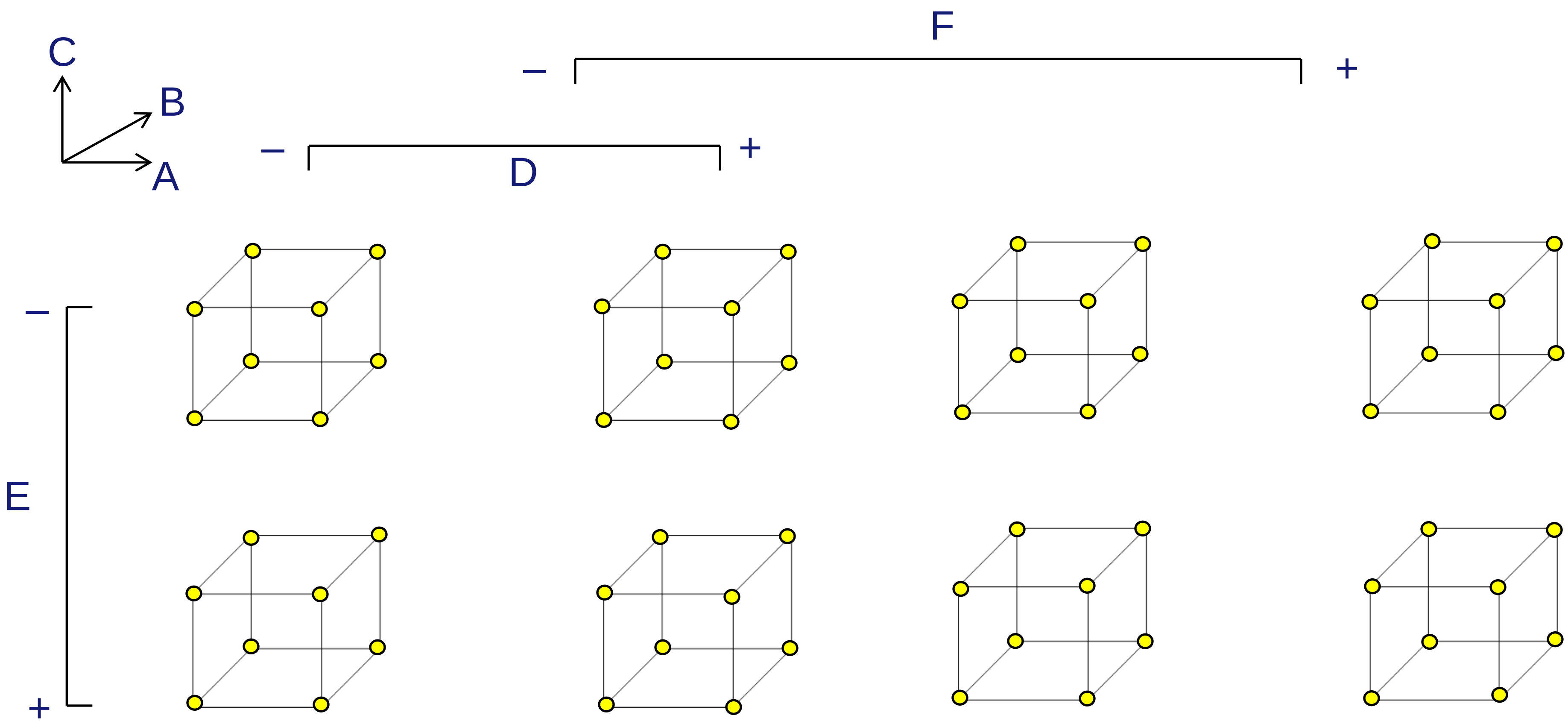
Spanning the Battlespace - DOE



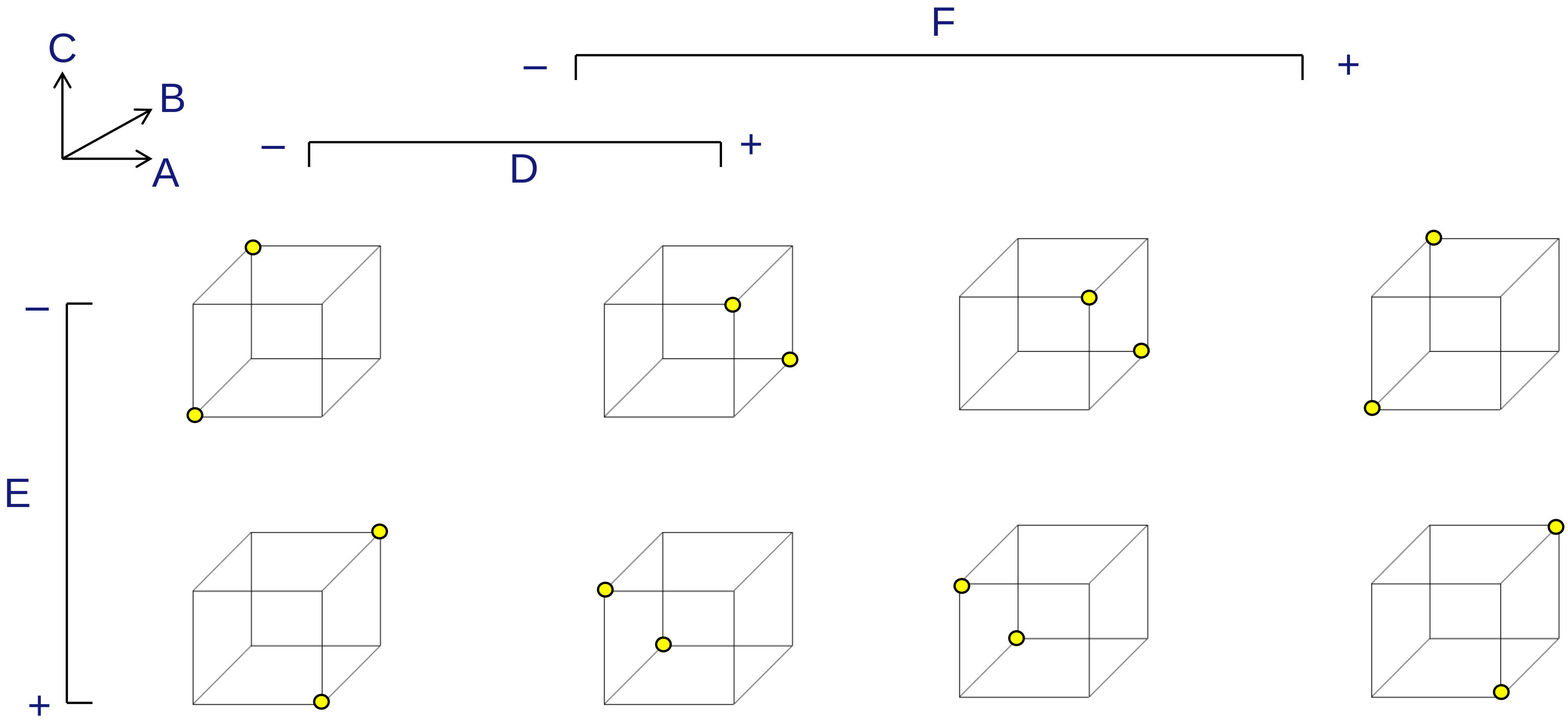
More Variables – DOE Factorials



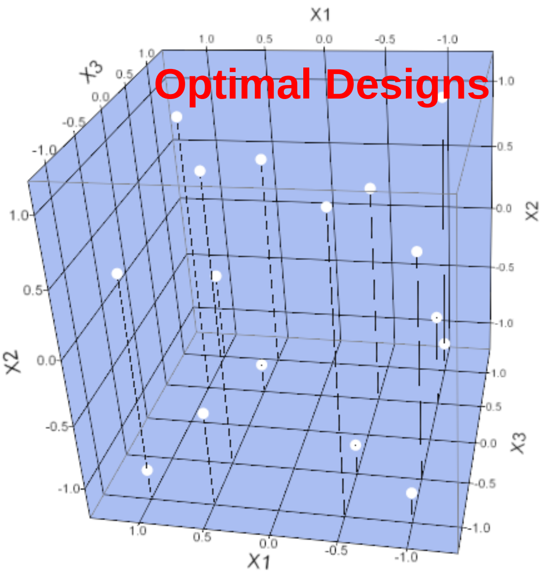
Even More Variables (here – 6)

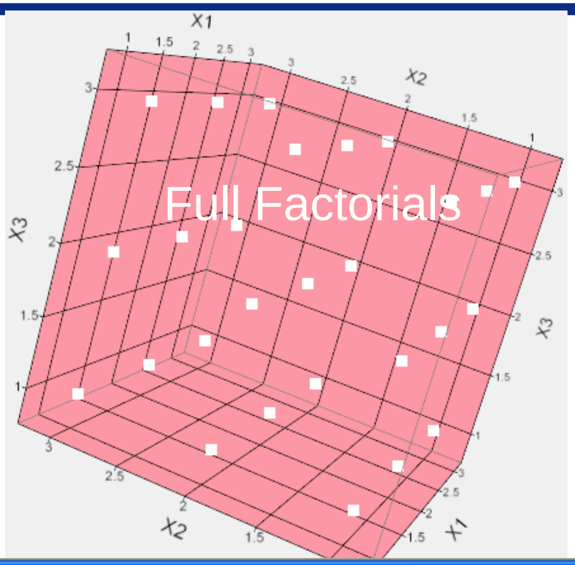


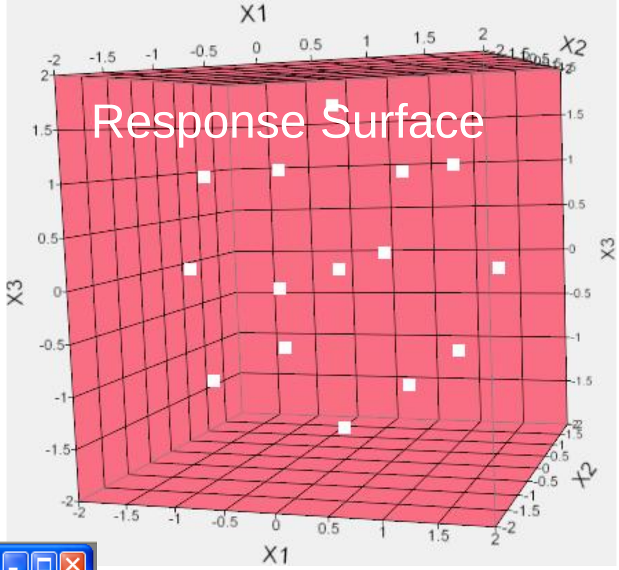
Efficiencies in Test - Fractions

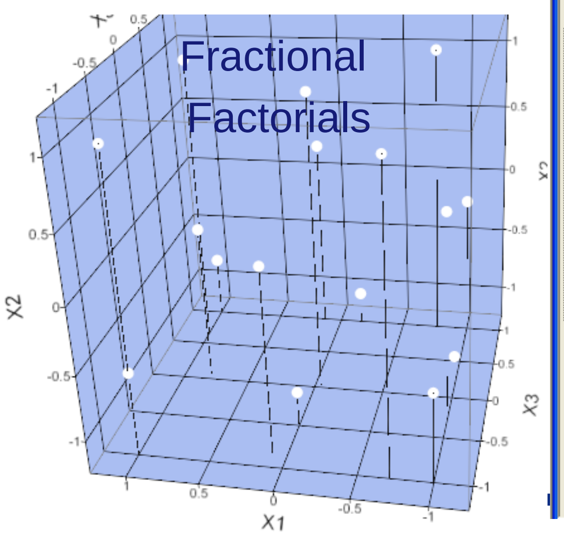


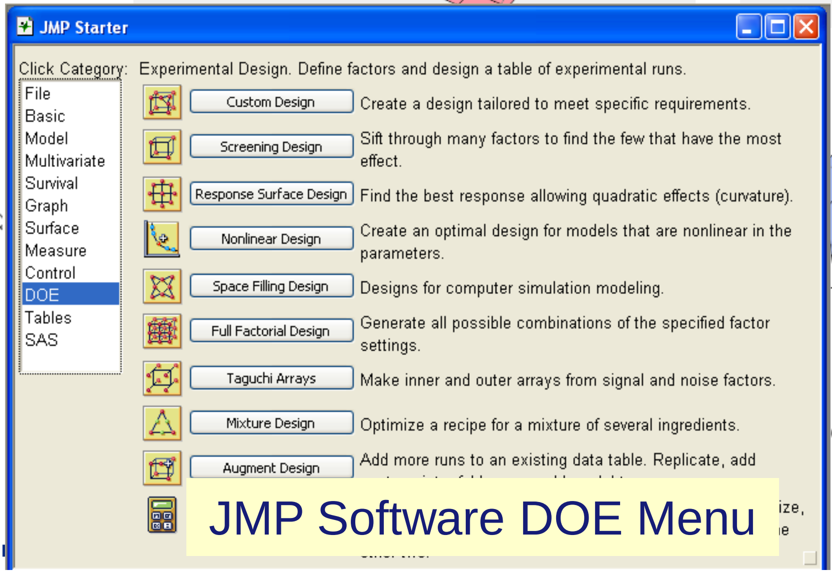
We have a wide menu of design choices with DOE

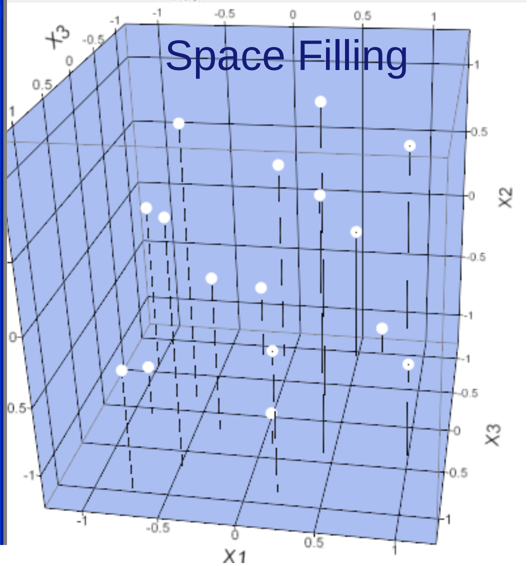




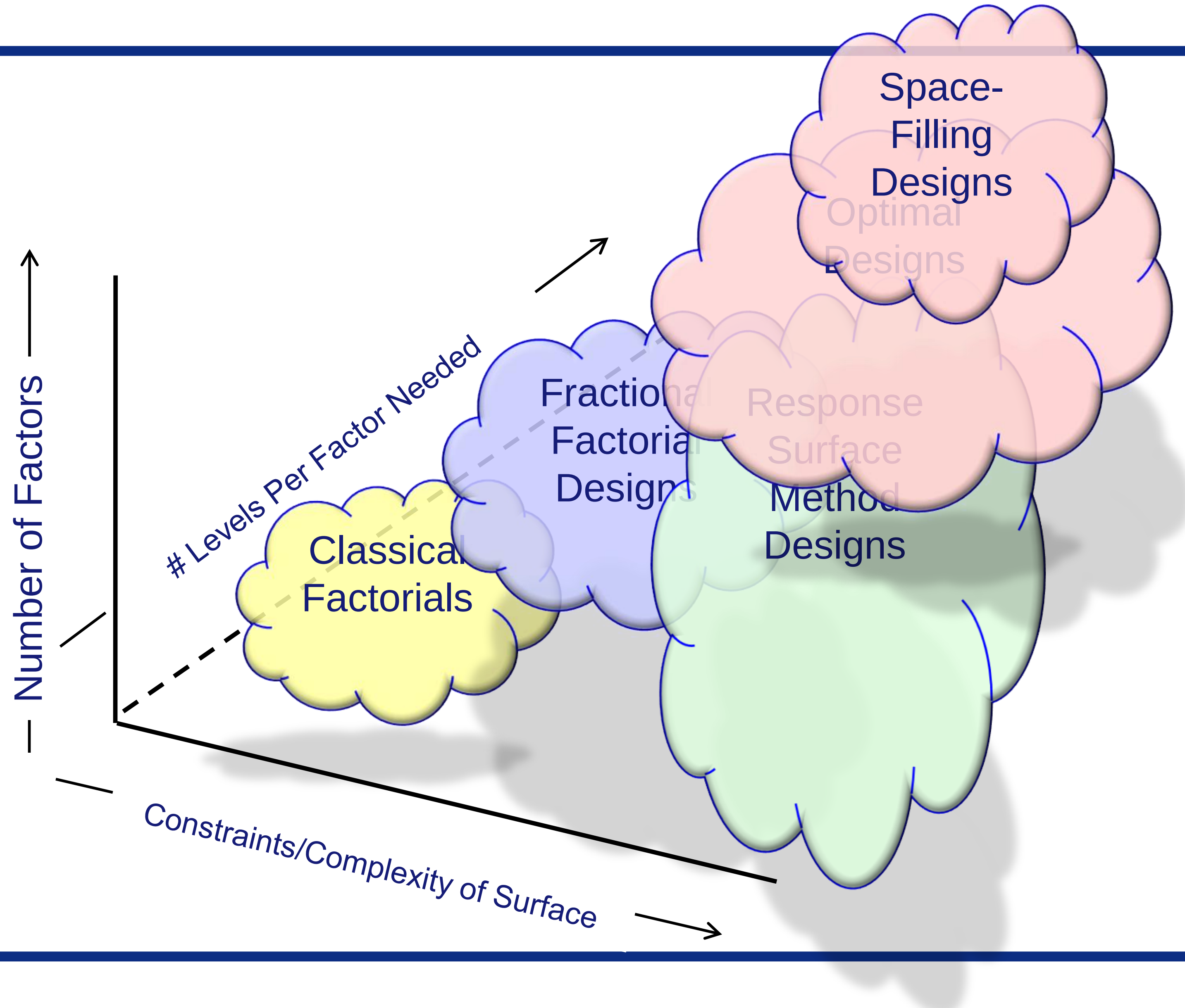








Problem context guides choice of designs



Pros and Cons of Design Classes

Design Class	Pros	Cons
Full Factorials	Easy to construct, analyze Robust to missing points	Get large fast with variables and levels
Fractional Factorials	Smaller while spanning space	More fragile, may require extra runs to determine causes
Response Surface	Moderate in size, more robust, model nonlinearities	Not many
Optimal Designs	Very flexible for complex problems	Construct & analyze with care - more technical depth needed
Space-Filling Designs	Excellent spanning properties	Difficulties in cause and effect linking

- **Simplified exposition: The best grad schools teach a 3-semester series in DOE**

Challenge 2: Which Points to Span the Relevant Battlespace?

4 reps 1 var

JPADS A	JPADS B
4	4

2 reps 2 vars

	JPADS A	JPADS B
Ammo	2	2
Food	2	2

1 reps 3 vars

		JPADS A	JPADS B
Eglin (Low)	Ammo	1	1
	Food	1	1
Nellis (High)	Ammo	1	1
	Food	1	1

½ rep 4 vars

			JPADS A	JPADS B
Dawn (low light)	Eglin (Low)	Ammo	1	
		Food		1
	Nellis (High)	Ammo		1
		Food	1	
Midday (bright)	Eglin (Low)	Ammo		1
		Food	1	
	Nellis (High)	Ammo	1	
		Food		1

- **Factorial (crossed) designs let us *learn more* from the same number of assets**
- **We can also use Factorials to *reduce assets* while maintaining confidence and power**
- **Or we can *combine* the two**

All four Designs share the same ***power*** and ***confidence***

- **How to support such an amazing claim?**

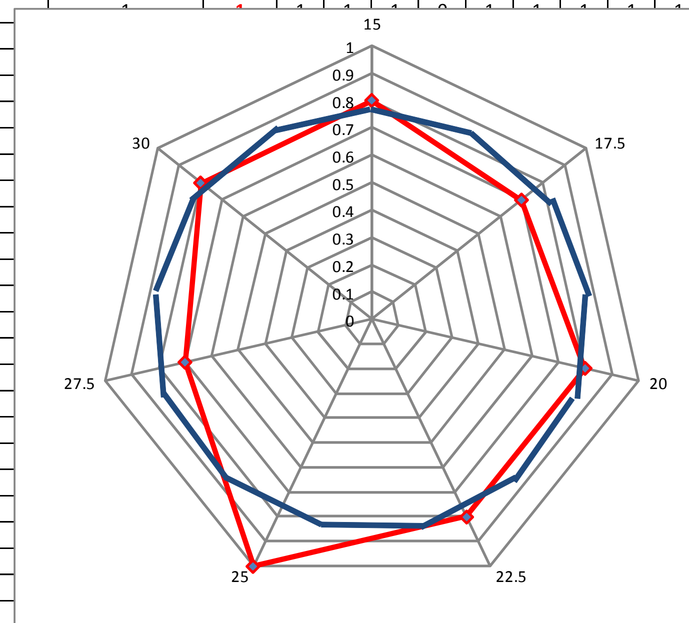
=> Switch to Excel File – JPADS Pancake.xls

Equal Power? A preposterous claim … how to justify it?

- **Consider again our JPADS problem across 2 dimensions**
- **13 wind speeds x 5 altitudes = 65 cases x 10 reps each = 650 trials**
- **Surely this will solve our problem with noise?**

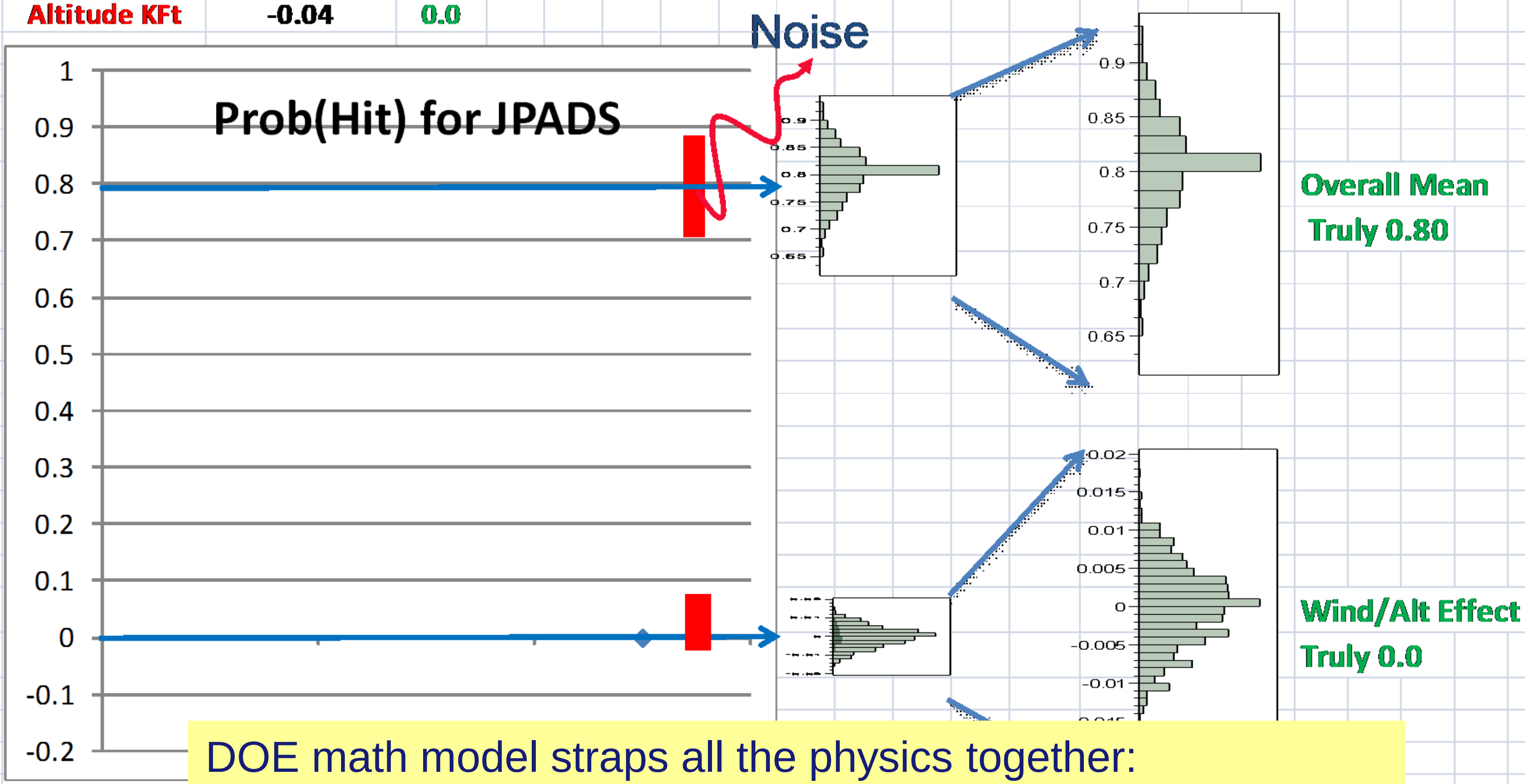
It will **not** … we have 65 separate 10-sample trials

		True P(Arrival)	80%										
			Replicates per Shot Condition										
Case	Wind	Altitude KFt	1	2	3	4	5	6	7	8	9	10	Phit
1	0	1	1	1	0	1	1	0	1	0	1	1	0.7
2	2.5	1	1	1	1	1	1	1	1	1	1	1	1
3	5	1	1	1	1	0	0	1	1	1	1	1	0.8
4	7.5	1	1	1	1	1	1	1	1	1	1	0	0.9
5	10	1	1	1	1	1	1	1	1	1	1	1	1
6	12.5												0.9
7	15												0.9
8	17.5												1
9	20												0.9
10	22.5												0.6
11	25												0.8
12	27.5												0.7
13	30												0.9
14	0												0.7
15	2.5												0.7
16	5												0.7
17	7.5												0.6
18	10												0.4
19	12.5												0.7
20	15												0.9
21	17.5												0.7
22	20												0.9
23	22.5												1
24	25												0.7
25	27.5												0.8
26	30												0.8
27	0												0.9
28	2.5												0.9
29	5												0.9
30	7.5	3	1	0	1	1	1	1	1	1	1	1	0.9
59	15	5	0	0	1	1	1	1	1	1	1	1	0.8
60	17.5	5	1	1	0	1	1	1	0	1	1	0	0.7
61	20	5	1	1	1	0	1	1	0	1	1	1	0.8
62	22.5	5	1	1	1	0	1	1	1	0	1	1	0.8
63	25	5	1	1	1	1	1	1	1	1	1	1	1
64	27.5	5	0	1	1	1	1	1	0	0	1	1	0.7
65	30	5	1	1	1	1	1	0	1	0	1	1	0.8



But, discard 9/10th of trials … strap 1/10th into a math model

	Estimate	Truth
Overall Mean	0.78	0.8
Wind	0.00	0.0
Altitude KFt	-0.04	0.0



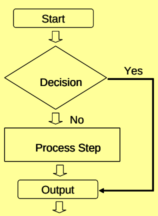
DOE math model straps all the physics together:

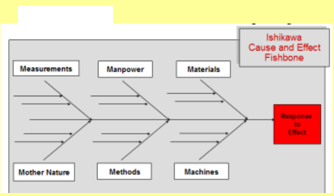
- *reducing* samples per condition by 90% while
- *increasing* our prediction accuracy 50%

Note: this speaks to the method of analysis (Challenge #4.)

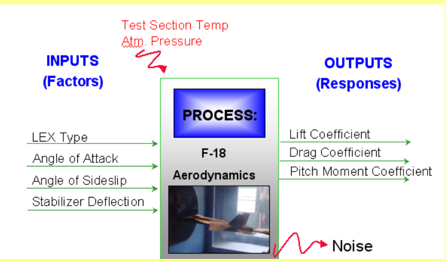
Test as Science vs. Art: Experimental Design Test Process is Well-Defined

Planning: Factors Desirable and Nuisance

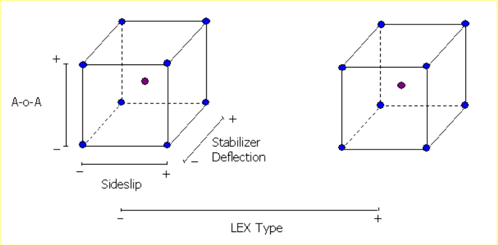




Desired Factors and Responses



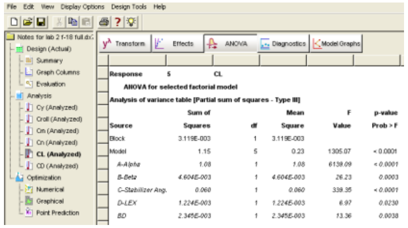
Design Points



Test Matrix

A-o-A	Sideslip	Stabilizer	LEX Type
2	0	5	-1
10	0	-5	1
10	8	5	-1
2	8	5	-1
2	8	-5	-1
2	0	-5	-1
10	8	-5	1
2	0	5	1
2	8	5	1
10	8	5	1
10	8	-5	-1
10	0	5	-1
10	0	-5	-1
2	8	-5	1
10	0	5	1
2	0	-5	1

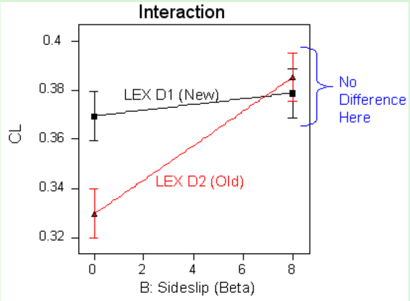
Randomize & Block -> Results and Analysis



Model Build

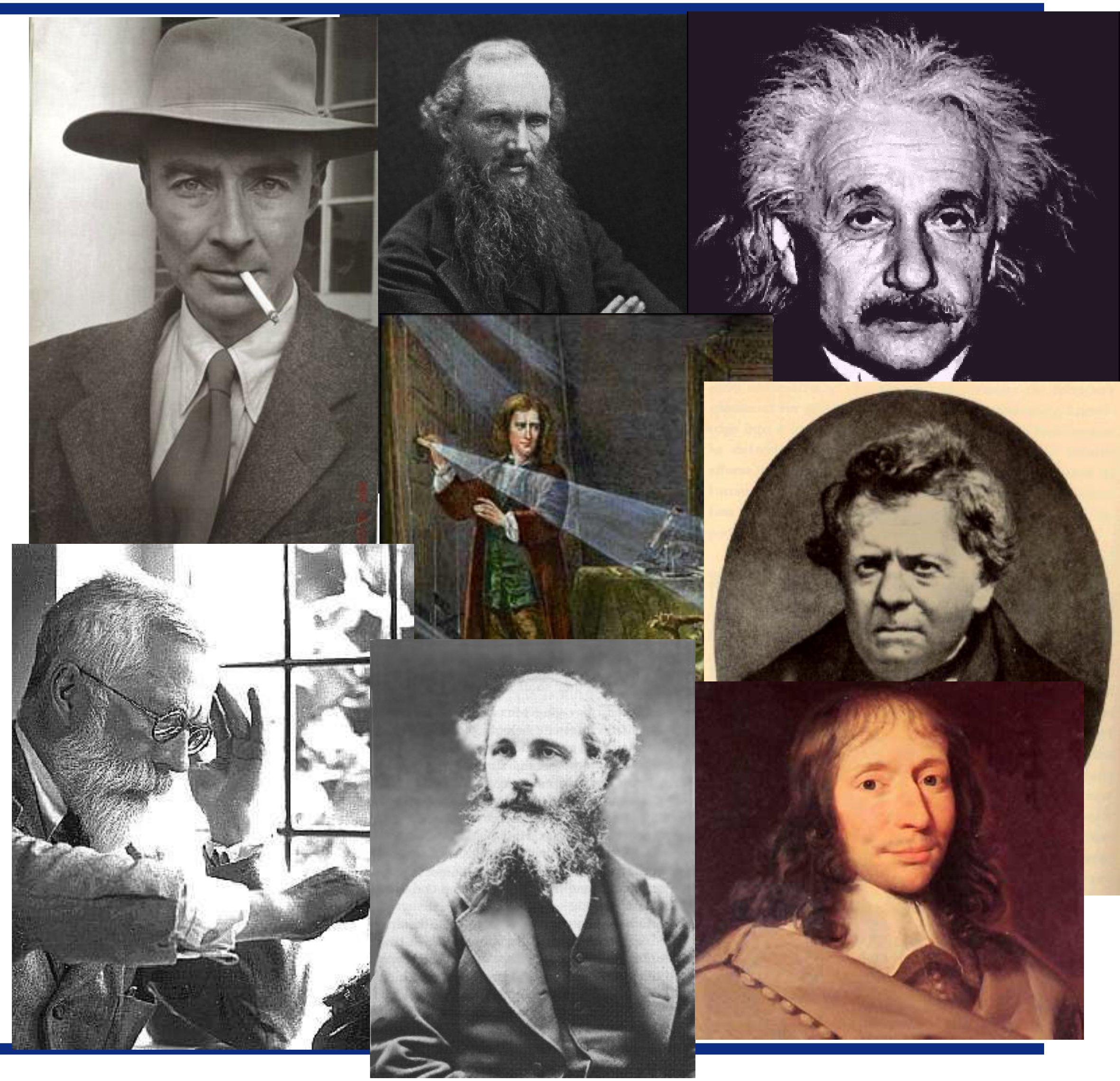
$$C_L = +0.38 + 0.26 \times \text{A-o-A} + 0.017 \times \text{Sideslip} + 0.061 \times \text{Stabilizer Deflection} - .00875 \times \text{LEX Type} + 0.012 \times \text{Sideslip} \times \text{LEX Type}$$

Discovery, Understanding Prediction, Re-design



Caveat – we *need* good science!
Good news – we *have* good science

- **We understand operations, aero, mechanics, materials, physics, electro-magnetics …**
- **To our good science, DOE adds the *Science of Test***



Bonus: Match faces to names – Ohm, Oppenheimer, Einstein, Maxwell, Pascal, Fisher, Kelvin

It applies to <u>our</u> tests: DOE in 50+ operations over 20 years

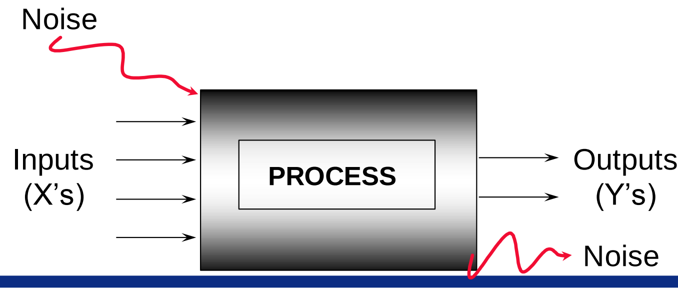
- IR Sensor Predictions
- Ballistics 6 DOF Initial Conditions
- Wind Tunnel fuze characteristics
- Camouflaged Target JT&E ($30M)
- AC-130 40/105mm gunfire CEP evals
- AMRAAM HWIL test facility validation
- 60+ ECM development + RWR tests
- GWEF Maverick sensor upgrades
- 30mm Ammo over-age LAT testing
- Contact lens plastic injection molding
- 30mm gun DU/HEI accuracy (A-10C)
- GWEF ManPad Hit-point prediction
- AIM-9X Simulation Validation
- Link 16 and VHF/UHF/HF Comm tests
- TF radar flight control system gain opt
- New FCS software to cut C-17 PIO
- AIM-9X+JHMCS Tactics Development
- MAU 169/209 LGB fly-off and eval
- Characterizing Seek Eagle Ejector Racks
- SFW altimeter false alarm trouble-shoot
- TMD safety lanyard flight envelope
- Penetrator & reactive frag design
- F-15C/F-15E Suite 4 + Suite 5 OFPs
- PLAID Performance Characterization
- JDAM, LGB weapons accuracy testing
- Best Autonomous seeker algorithm
- SAM Validation versus Flight Test
- ECM development ground mounts (10’s)
- AGM-130 Improved Data Link HF Test
- TPS A-G WiFi characterization
- MC/EC-130 flare decoy characterization
- SAM simulation validation vs. live-fly
- Targeting Pod TLE estimates
- Chem CCA process characterization
- Medical Oxy Concentration T&E
- Multi-MDS Link 16 and Rover video test

Recap - Scientific Answers to Four Fundamental Test Challenges

Four Challenges faced by any test

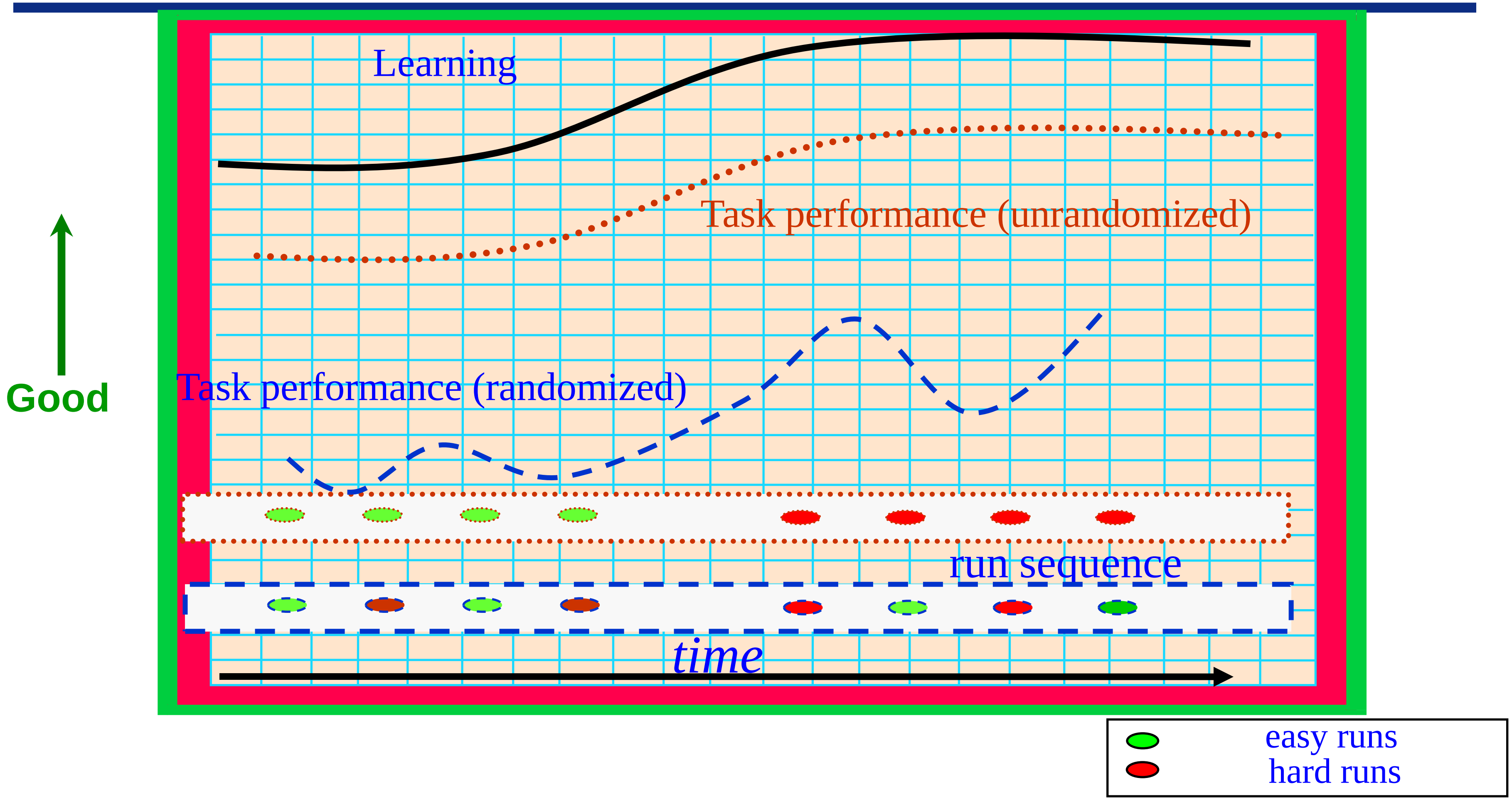
1. *How many?* A: sufficient samples to control our twin errors – false positives & negatives
2. *Which Points?* A: span/populate the battle-space while linking cause and effect with orthogonal run matrices
3. *How Execute?* A: Randomize and Block runs to exclude effects of the unknown-unknowns
4. *What Conclusions?* A: build math-models* of input/output relations, quantifying noise, controlling error

DOE effectively addresses all these challenges!



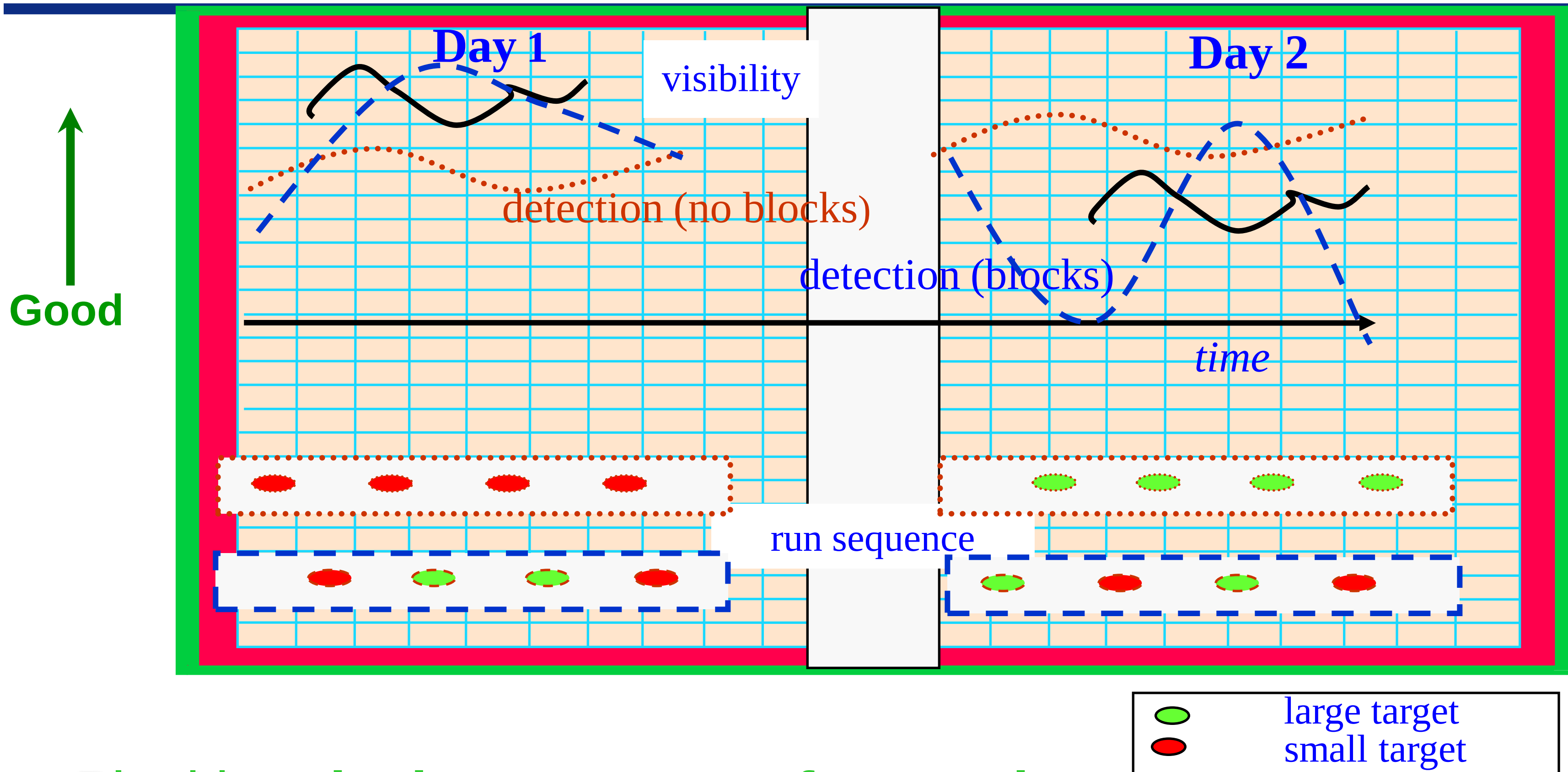
* Many model choices: regression, ANOVA, mixed, Gen Lin Model, CART, Kriging, MARS

Challenge 3: What Order? Guarding against "Unknown-Unknowns"



- Randomizing <u>runs</u> protects from unknown background changes *within* an experimental period (due to Fisher)

Blocks Protect Against Day-to-Day Variation



- Blocking designs protects from unknown background changes *between* experimental periods (also due to Fisher)

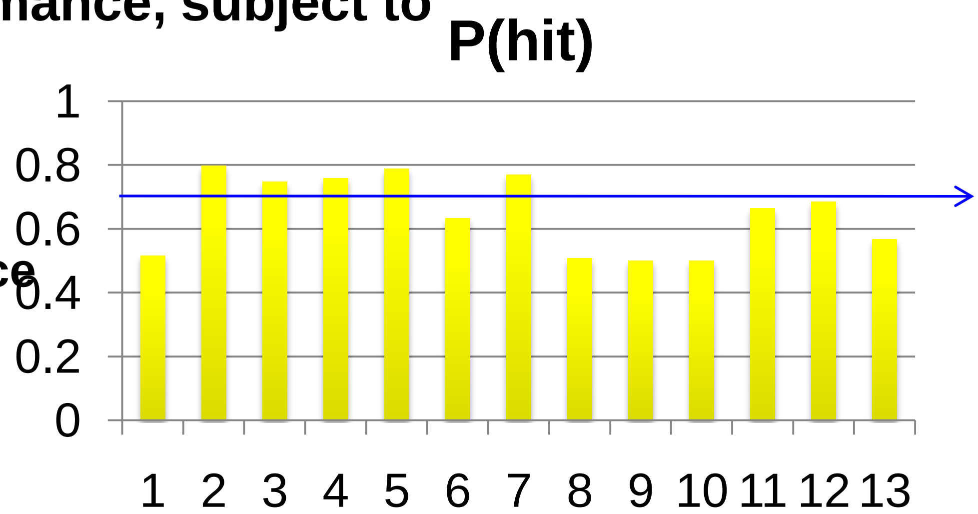
Challenge 4: What Conclusions - Traditional "Analysis"

- Cases or Scenario settings and findings: table of summary stats

Sortie	Alt	Mach	MDS	Range	Tgt Aspect	OBA	Tgt Velocity	Target Type	Result
1	10K	0.7	F-16	4	0	0	0	truck	Hit
1	10K	0.9	F-16	7	180	0	0	bldg	Hit
2	20K	1.1	F-15	3	180	0	10	tank	Miss

- Graphical summary of performance, subject to
  - Change in scale
  - Subset the domain to show desired trend
  - Tell stories about performance
- Seldom quantify uncertainty
- Seldom test if result could be

By chance alone …



How Factorial Matrices Work -- a peek at the linear algebra under hood

- We set the X settings, observe Y

Simple 2-level, 2 X-factor design
Where y is the grand mean,
A,B are effects of variables alone
and AB measures variables working together

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

$$\begin{bmatrix} y_{(1)} \\ y_a \\ y_b \\ y_{ab} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \bar{\bar{y}} \\ A \\ B \\ AB \end{bmatrix} + \begin{bmatrix} e_{(1)} \\ e_a \\ e_b \\ e_{ab} \end{bmatrix}$$

- Solve for b s.t. the error (e) is minimized

$$\mathbf{e} = \mathbf{y} - \mathbf{Xb}$$

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$$

$$\frac{d(\mathbf{e}'\mathbf{e})}{d\mathbf{b}} = \frac{d((\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}))}{d\mathbf{b}} = \mathbf{0}$$

How Factorial Matrices Work II

$$\frac{d((\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}))}{d\mathbf{b}} = \mathbf{0}$$

$$\frac{d((\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb}))}{d\mathbf{b}} = -2\mathbf{X}'(\mathbf{y}-\mathbf{Xb}) = \mathbf{0}$$

$$-2\mathbf{X}'(\mathbf{y}-\mathbf{Xb}) = \mathbf{0}$$

$$\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{Xb} = \mathbf{0}$$

$$\mathbf{X}'\mathbf{Xb} = \mathbf{X}'\mathbf{y}$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

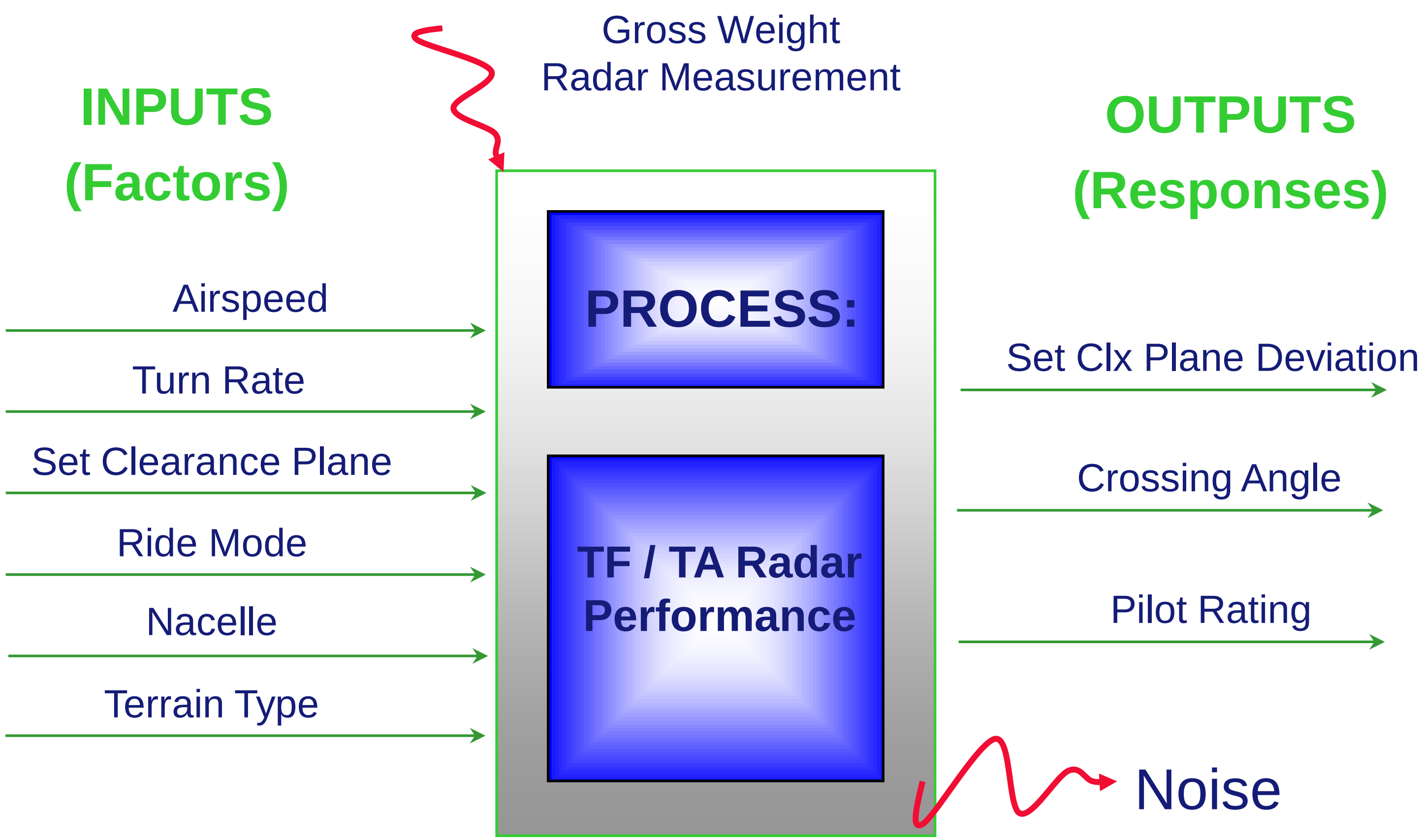
- To solve for the unknown **b**'s, **X** must be <u>invertible</u>
- Factorial design matrices are generally ***orthogonal***, and therefore invertible ***by design***

With DOE, we fit an empirical (or physics-based) model

$$y = \beta_0 + \sum_i \beta_i x_i + \sum_i \beta_{ii} x_i^{\,2} + \sum\sum_{i \neq j} \beta_{ij} x_i x_j + \sum_i \beta_{iii} x_i^{\,3} + \varepsilon \qquad i = 1, 2, \ldots, k$$

- **Very simple to fit models of the form above (among others) with polynomial terms and interactions**
- **Models fit with ANOVA or multiple regression software**
- **Models are easily interpretable in terms of the physics (magnitude and direction of effects)**
- **Models very suitable for "predict-confirm" challenges in regions of unexpected or very nonlinear behavior**
- **Run to run noise can be explicitly captured and examined for structure**

Analysis using DOE: CV-22 TF Flight



Analysis Using DOE

Gross Weight	SCP	Turn Rate	Ride	Airspeed	SCP Dev	Pilot Ratings
55	300.00	0.00	Medium	160.00	5.6	4.5
47.5	500.00	0.00	Medium	160.00	0.5	4.8
47.5	300.00	4.00	Medium	160.00	7.5	4.2
55	500.00	4.00	Medium	160.00	2.3	4.8
47.5	300.00	0.00	Hard	160.00	5.2	4.2
55	500.00	0.00	Hard	160.00	1.2	4.6
55	300.00	4.00	Hard	160.00	12.0	3.2
47.5	500.00	4.00	Hard	160.00	6.7	3.4
47.5	300.00	0.00	Medium	230.00	4.0	4.8
55	500.00	0.00	Medium	230.00	0.2	5.4
55	300.00	4.00	Medium	230.00	15.0	2.8
47.5	500.00	4.00	Medium	230.00	8.3	3.2
55	300.00	0.00	Hard	230.00	5.8	4.5
47.5	500.00	0.00	Hard	230.00	1.9	5.0
47.5	300.00	4.00	Hard	230.00	16.0	2.0
55	500.00	4.00	Hard	230.00	12.0	2.5
47.5	400.00	2.00	Medium	195.00	4.0	4.2
47.5	400.00	2.00	Hard	195.00	7.2	3.7
55	400.00	2.00	Medium	195.00	6.6	4.4
55	400.00	2.00	Hard	195.00	7.7	3.8

SCP Deviation Interpretation

Design-Expert® Software

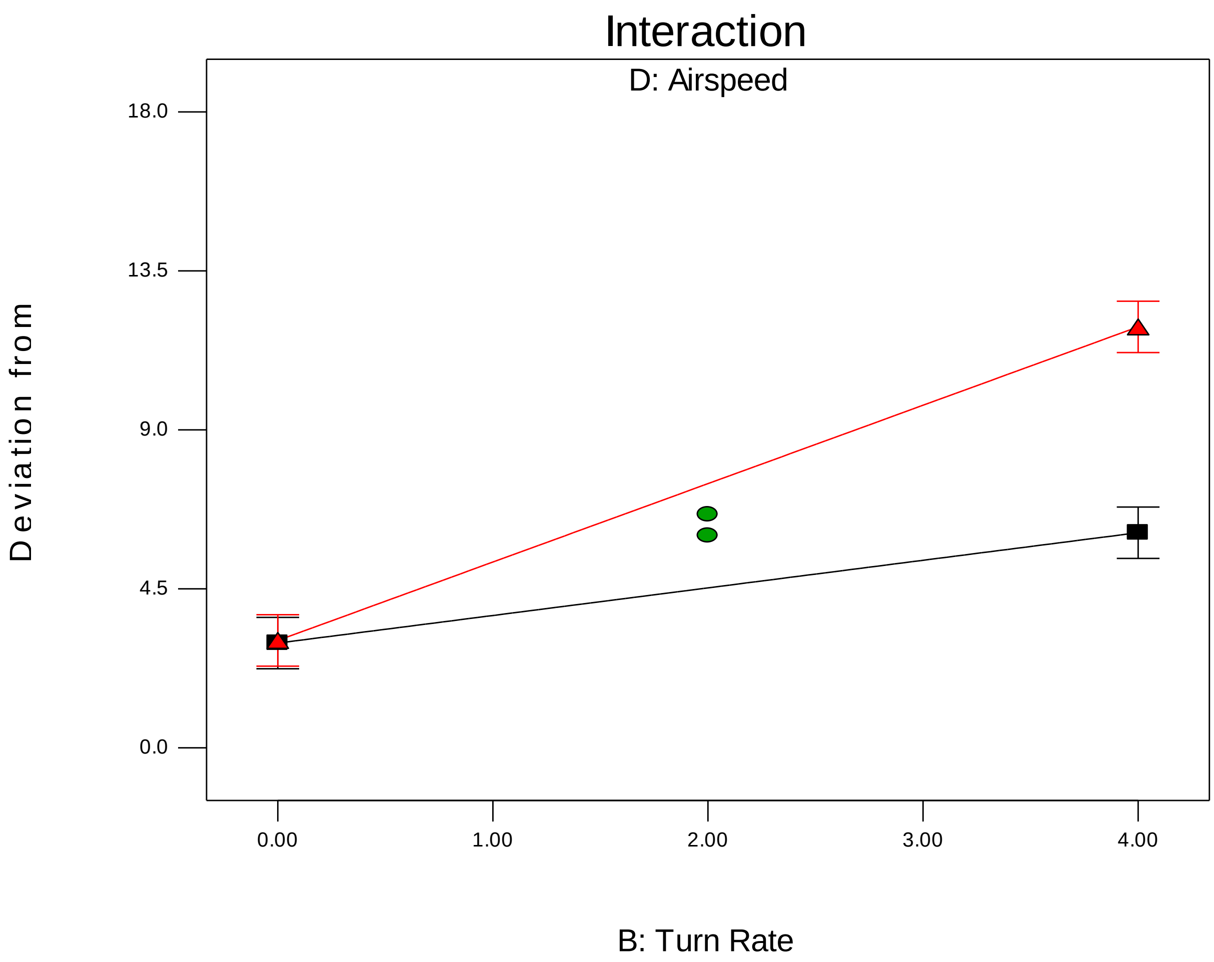
Deviation from SCP

● Design Points

■ D- 160.000
▲ D+ 230.000

X1 = B: Turn Rate
X2 = D: Airspeed

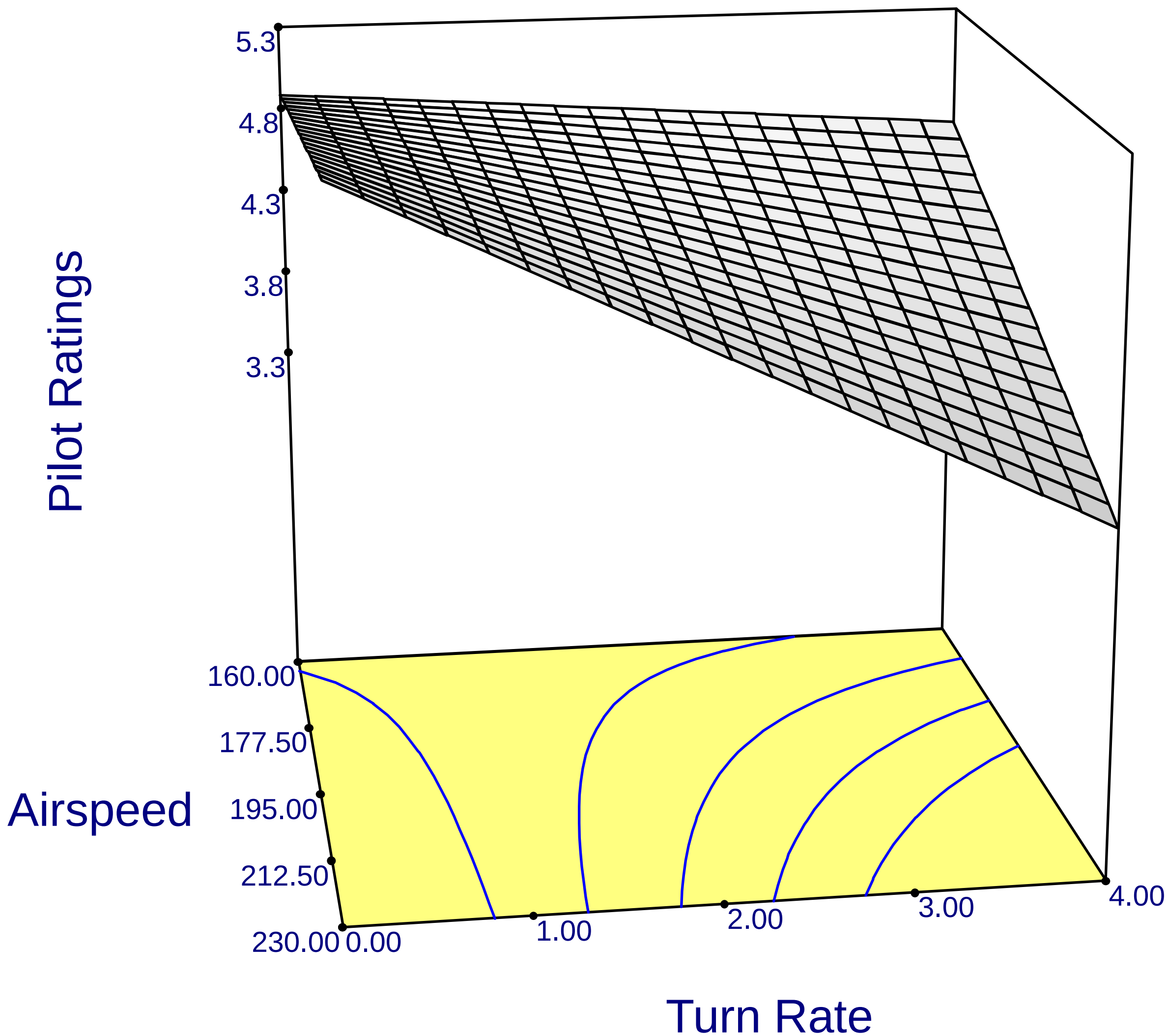
Actual Factors
A: SCP = 400.00
C: Ride = Medium



Pilot Ratings Response Surface Plot

Actual Factors:
X = Airspeed
Y = Turn Rate

Actual Constants:
SCP = 500.00
Ride = Medium



Radar Performance Results

SCP Deviation Estimating Equation

Prediction Model - Coded Units (Low=-1, High=+1)

Deviation from SCP =		
	+6.51	
	-2.38	* SCP
	+3.46	* Turn Rate
	+1.08	* Ride
	+1.39	* Airspeed
	+0.61	* Turn * Ride
	+1.46	* Turn * Airspeed

Note magnitude and direction of effect s and what is not there – Gross Weight

Performance Predictions

Name	Setting	Low Level	High Level
SCP	**460.00**	300.00	500.00
Turn Rate	**2.80**	0.00	4.00
Ride	**Hard**	Medium	Hard
Airspeed	**180.00**	160.00	230.00

	Prediction	95% PI low	95% PI high
Deviation from SCP	**6.96**	4.93	8.98
Pilot Ratings	**3.62**	3.34	3.90

Confirm or validate our model -- predict points that have not been tested and compare predictions to performance

Actual Maverick H/K Test – A great success

- Extensive captive carry
  - 22 sorties
  - Approx 100 sim shots
  - Old/new seekers on each wing to equalize Wx
  - 3 platforms: F-16, F-15E, A-10
  - Eglin & Nellis
- Results – approx 2x acq/trk range
- 9 shots comparable to current performance
- Full Disclosure – Error Type III

Mission: MAV-1

Eglin Range   Range Time: 0700-0900 (Dawn)   Target: Point (Tank)   Weather: As Scheduled

Launch Airspeed: 400 KIAS

F-16 - #1   Left Wing: CATM-65K on LAU-88   Right Wing: TGM-65B on LAU-88

Run#	Target Type	Altitude	Sun Angle	Missile Type	Cueing	Comments*
1	Tank/Truck	18,000’	Sun at 6	H/K	Visual	
2	Tank/Truck	18,000’	Sun at 6	B	“	
3	Tank/Truck	1500’	Sun at 3/9	H/K	NAV/GPS	
4	Tank/Truck	1500’	Sun at 3/9	B	“	
5	Tank/Truck	1500’	Sun at 6	H/K	Radar	
6	Tank/Truck	1500’	Sun at 6	B	“	
7	Tank/Truck	18,000’	Sun at 3/9	H/K	LANTIRN	
8	Tank/Truck					

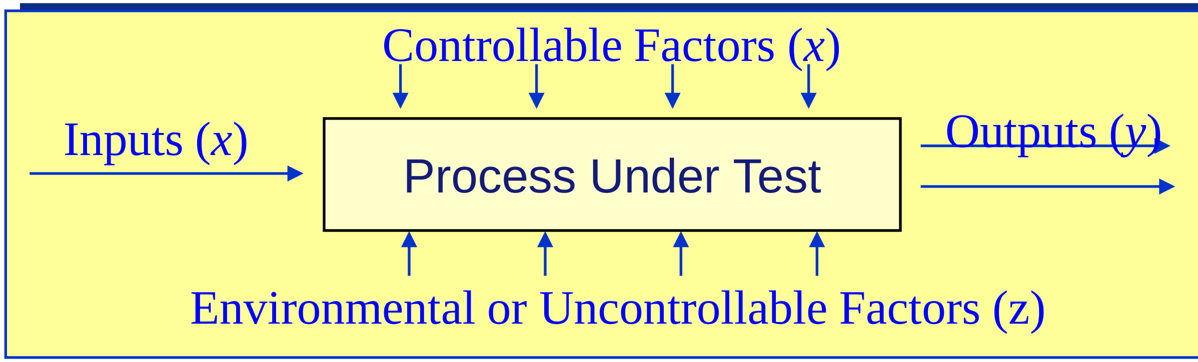
Typical Mav H/K F-16 Run Card

F-16 - #2   Left Wing: CATM-65K on LAU-88   Right Wing: TGM-65D on LAU-88

Run#	Target Type	Altitude	Sun Angle	Missile Type	Cueing	Comments*
1	Tank/Truck	18,000’	Sun at 6	H/K	Visual	
2	Tank/Truck	18,000’	Sun at 6	D	“	
3	Tank/Truck	1500’	Sun at 3/9	H/K	NAV/GPS	
4	Tank/Truck	1500’	Sun at 3/9	D	“	
5	Tank/Truck	1500’	Sun at 6	H/K	Radar	
6	Tank/Truck	1500’	Sun at 6	D	“	
7	Tank/Truck	18,000’	Sun at 3/9	H/K	LANTIRN	
8	Tank/Truck	18,000’	Sun at 3/9	D	“	

* After simulated pickle, simulate missile flyout by overflying target and recording seeker video.

The Basic Test Problem

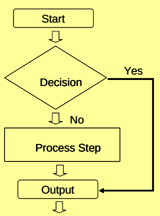


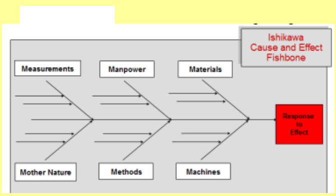
Answers to these questions determine your **strategy of experimentation**.

- **Questions that must be answered for *any* test:**
  - **Which *x*'s, *z*'s should be changed, through what range, pattern?**
  - **How should the runs be ordered?**
  - **How many runs should we/can we make?**
  - **What method of analysis should be employed?**
- **Possible Objectives:**
  - **Characterizing where y's usually fall depending on x settings**
  - **Determining which variables have an influence on outputs**
  - **Determining where to set influential variables so outputs are near goal**
  - **Determining how to set inputs so variability of y is small**
  - **Determining where to set x's so that effects of z's are small**

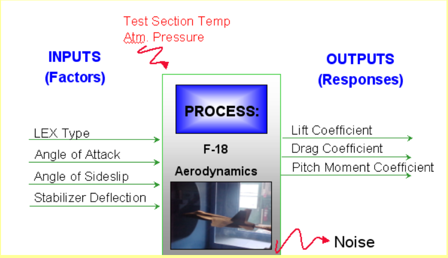
Design of Experiments Test Process is Well-Defined

Planning: Factors Desirable and Nuisance

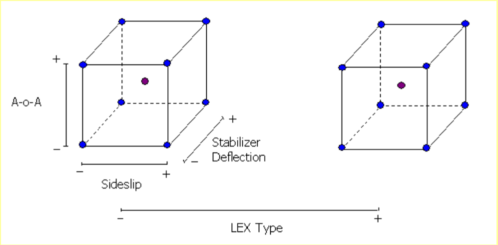




Desired Factors and Responses



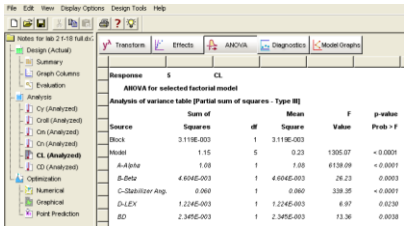
Design Points



Test Matrix

A-o-A	Sideslip	Stabilizer	LEX Type
2	0	5	-1
10	0	-5	1
10	8	5	-1
2	8	5	-1
2	8	-5	-1
2	0	-5	-1
10	8	-5	1
2	0	5	1
2	8	5	1
10	8	5	1
10	8	-5	-1
10	0	5	-1
10	0	-5	-1
2	8	-5	1
10	0	5	1
2	0	-5	1

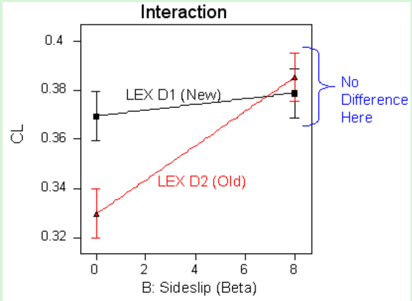
Results and Analysis



Model Build

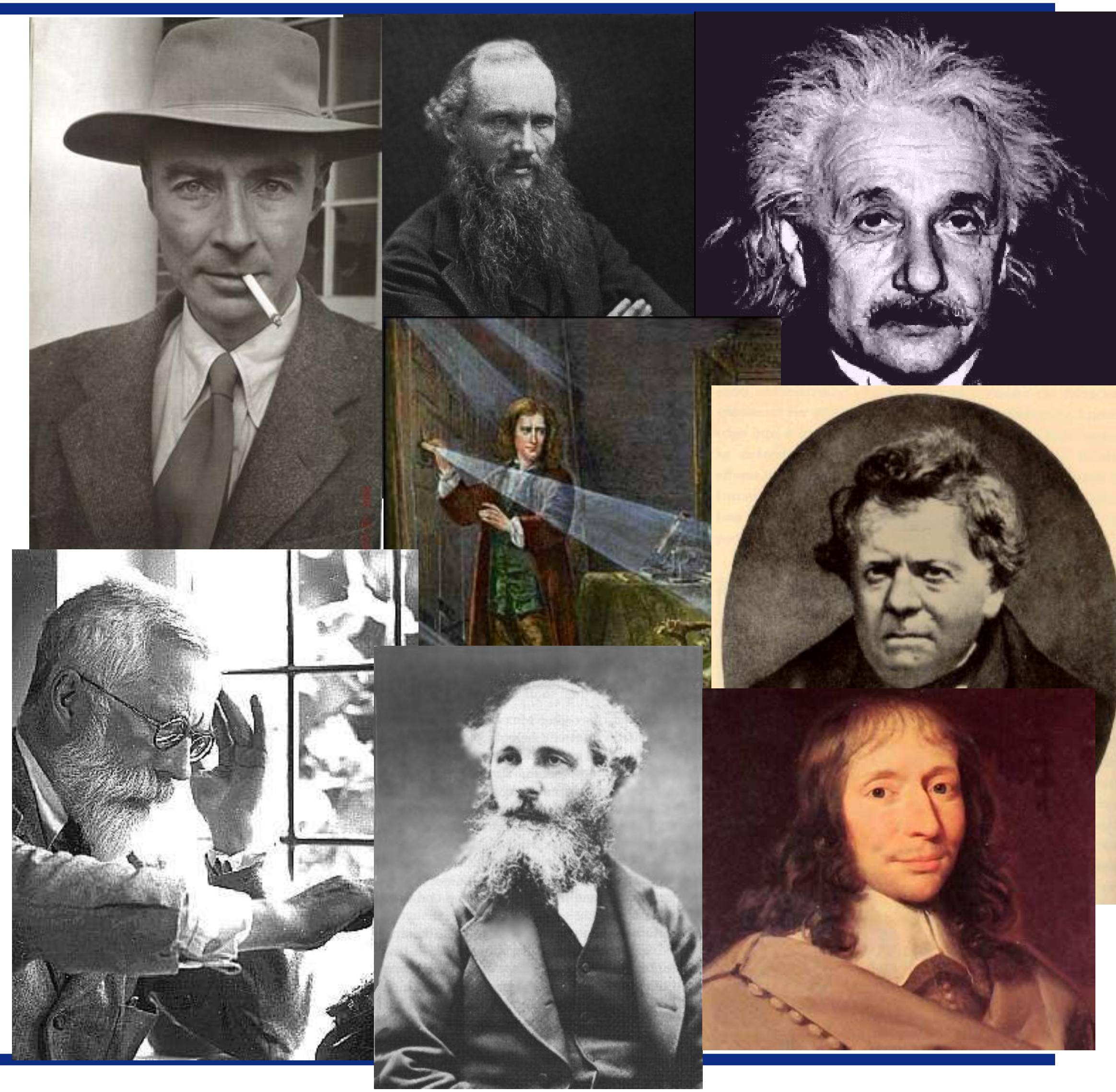
C_L = +0.38
+0.26 x A-o-A
+0.017 x Sideslip
+0.061 x Stabilizer Deflection
-.00875 x LEX Type
+0.012 x Sideslip x LEX Type

Discovery, Prediction



Caveat – we ***need*** good science!

- **We understand operations, aero, mechanics, materials, physics, electro-magnetics …**
- **To our good science, DOE introduces the *Science of Test***



Bonus: Match faces to names – Ohm, Oppenheimer, Einstein, Maxwell, Pascal, Fisher, Kelvin

Many Strategies of Test have been tried .. Which is ***best***?

Consider how to *characterize B-1B Radar Target Location Error* as a function of several factors.

Altitude
Target RCS
Angle off nose
Aircraft tail number
Target Range
Calibration date of radar
Time of last doppler update
Target Elevation
Operator Skill Level

Inputs (test conditions)

Target location error

Outputs (MOPs)

- Intuition
- One factor at a time (OFAT)
- Scenario (best guess or case)
- +2 (DWWDLT, Rent-a-TIS)

Remember – Large Test Space: 2*2*2*…2 = 2^9 = 512 combos

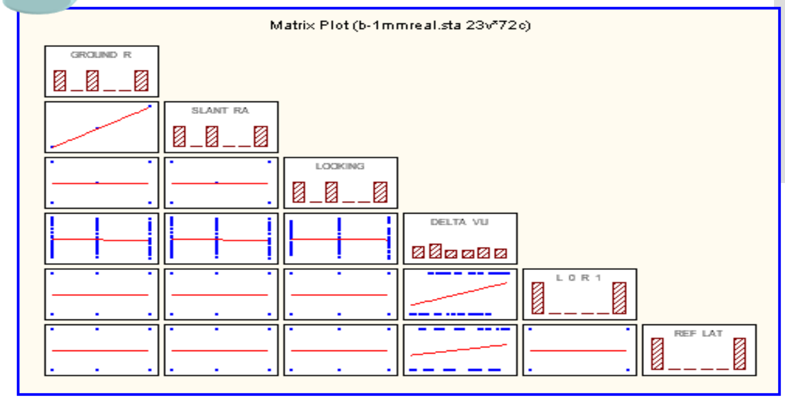
Pros and Cons of Various Test Strategies

Test Strategy	Description	Pros	Cons
DWWDLT	Do What We (They) Did Last Time	Conservative, preserves heritage	If not well thought-out, perpetuates mediocrity. No proof of goodness
"Rent A TIS"	Ask the Hardware Contractor	Good system understanding, good science in test. Track record.	As above, and no proof of goodness.
Subject Expertise/Intuition	Test Expert-Selected points	Expert assistance on where to test	Discourages new knowledge, unexpected behavior. Subjective, sparse coverage.
OFAT	Change one factor at a time	Know how changing one factor affects the system under test.	Ignores interactions (two factors change response). Sparse coverage of factor space.
Scenario/Cases	Use themes or war/use scenarios to pick test conditions	Good understanding of typical use performance.	Fails to answer why good-bad. Leads to "mythological tactics." Sparse coverage.
DOE	Decompose process into suspected cause-effect, replicate for power.	Most efficient and effective test strategy. Test becomes a science with rules.	Requires substantial training and experience, but can be taught and learned.

DOE – the only way we know to test broadly and deeply with power!

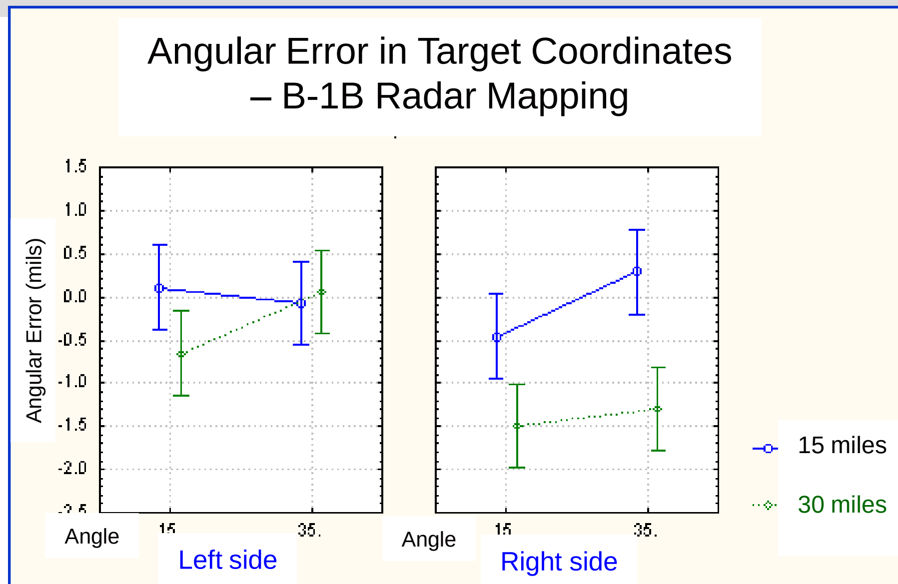
B-1 OT Radar Mapping in One Mission (53d Wg)

Problem: Characterize B-1 radar coordinate accuracy for variety of operational setups.



Design: Responses: absolute error

Conditions: angle, side of nose, tail number, target, and range to target. 4 replicates.



Results: Similar accuracy across scan volume, target type, tail number.

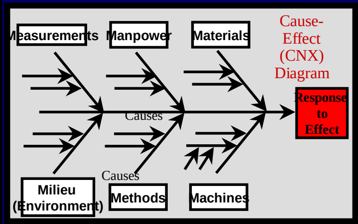
Result: Single two-aircraft mission answered operational accuracy questions raised by 7 previous missions using conventional test methods.

It applies to our tests: DOE in 36+ operations over 20 years

- IR Sensor Predictions
- Ballistics 6 DOF Initial Conditions
- Wind Tunnel fuze characteristics
- Camouflaged Target JT&E ($30M)
- AC-130 40/105mm gunfire CEP evals
- AMRAAM HWIL test facility validation
- 60+ ECM development + RWR tests
- GWEF Maverick sensor upgrades
- 30mm Ammo over-age LAT testing
- Contact lens plastic injection molding
- 30mm gun DU/HEI accuracy (A-10C)
- GWEF ManPad Hit-point prediction
- AIM-9X Simulation Validation
- Link 16 and VHF/UHF/HF Comm tests
- TF radar flight control system gain opt
- New FCS software to cut C-17 PIO
- AIM-9X+JHMCS Tactics Development
- MAU 169/209 LGB fly-off and eval
- Characterizing Seek Eagle Ejector Racks
- SFW altimeter false alarm trouble-shoot
- TMD safety lanyard flight envelope
- Penetrator & reactive frag design
- F-15C/F-15E Suite 4 + Suite 5 OFPs
- PLAID Performance Characterization
- JDAM, LGB weapons accuracy testing
- Best Autonomous seeker algorithm
- SAM Validation versus Flight Test
- ECM development ground mounts (10's)
- AGM-130 Improved Data Link HF Test
- TPS A-G WiFi characterization
- MC/EC-130 flare decoy characterization
- SAM simulation validation vs. live-fly
- Targeting Pod TLE estimates
- Chem CCA process characterization
- Medical Oxy Concentration T&E
- Multi-MDS Link 16 and Rover video test

Seven DOE Stories for T&E

Process

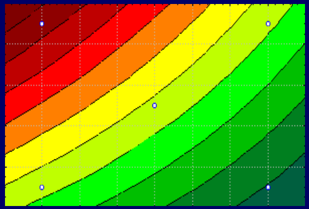


Plan

		In Front		In Back	
		Face East	Face West	Face East	Face West
Eyes Open	Left Hand	0.43	0.58	0.52	0.40
	Right Hand	0.62	0.29	0.28	0.36
Eyes Closed	Left Hand	0.62	0.57	0.47	0.40
	Right Hand	0.42	0.26	0.42	0.47

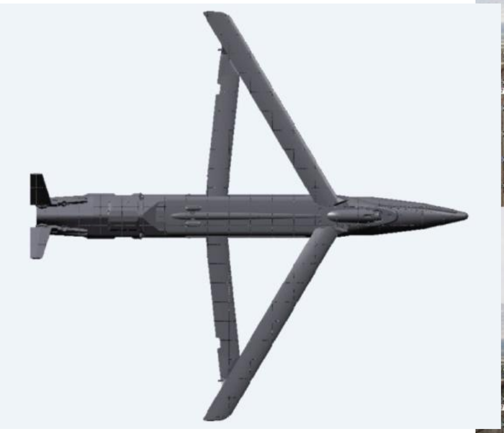
Produce

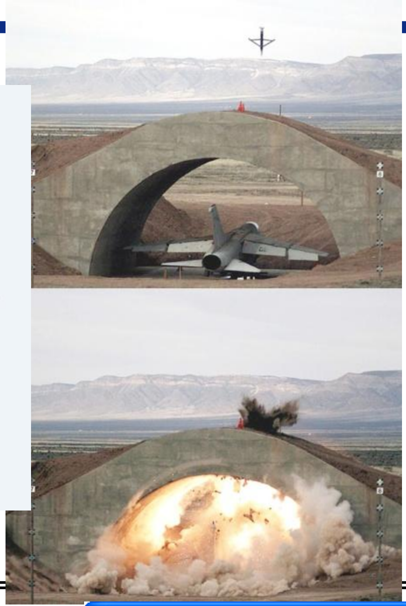
Ponder



We've selected these from 1000's to show T&E Transformation

Testing to Diverse Requirements: SDB II Shot Design



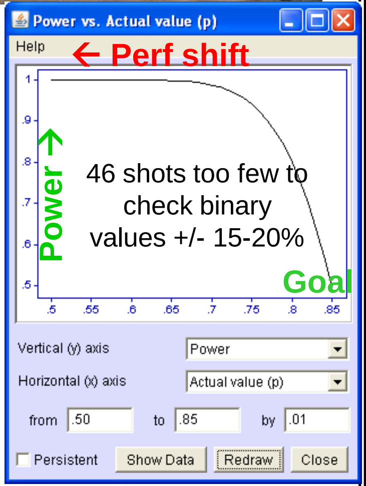


Test Objective:

- SPO requests help – 46 shots right N?
- Power analysis – what can we learn?
- Consider Integrated Test with AFOTEC
- What are the variables? We do not know yet …
- How can we plan?
- What "management reserve"

DOE Approach:

- Partition performance questions: Pacq/Rel + laser + coords + "normal mode"
- Consider total test pgm: HWIL+Captive+Live
- Build 3x custom, "right-size" designs to meet objectives/risks



Results:

- Binary P_{acq} N=200+
- Demo laser + coords N=4 ea
- Prove normal mode N=32

	Performance Shift from KPP in Units of 1 Standard Deviation						
Shots	0.25	0.5	0.75	1	1.25	1.5	2
4	0.176	0.387	0.65	0.855	0.958	0.992	0.999
6	0.219	0.52	0.815	0.959	0.995	0.999	0.999
8	0.259	0.628	0.905	0.989	0.999	0.999	0.999
10	0.298	0.714	0.952	0.997	0.999	0.999	
12	0.335	0.783	0.977	0.999	0.999	0.999	
14	0.371	0.837	0.989	0.999	0.999	0.999	
16	0.406	0.878	0.994	0.999	0.999		
18	0.44	0.909	0.997	0.999	0.999		
20	0.472	0.933	0.998	0.999	0.999		
24	0.532	0.964	0.999	0.999			
28	0.587	0.981	0.999	0.999			
32	0.636	0.99	0.999	0.999			
36	0.681	0.995	0.999				
40	0.721	0.997	0.999				
50	0.802	0.999	0.999				
60	0.862	0.999					
70	0.904	0.999					
80	0.934	0.999					
100	0.97	0.999					

32-shot factorial screens 4-8 variables to 0.5 std dev shift from KPP

- Integrate 20 AFOTEC shots for "Mgt Reserve"

Acquisition: F-15E Strike Eagle Suite 4E+ (circa 2001-02)

Test Objectives:

- **Qualify new OFP Suite for Strikes with new radar modes, smart weapons, link 16, etc.**
- **Test must address dumb weapons, smart weapons, comm, sensors, nav, air-to-air, CAS, Interdiction, Strike, ferry, refueling…**
- **Suite 3 test required 600+ sorties**

DOE Approach:

- **Build multiple designs spanning:**
  - **EW and survivability**
  - **BVR and WVR air to air engagements**
  - **Smart weapons captive and live**
  - **Dumb weapons regression**
  - **Sensor performance (SAR and TP)**

Results:

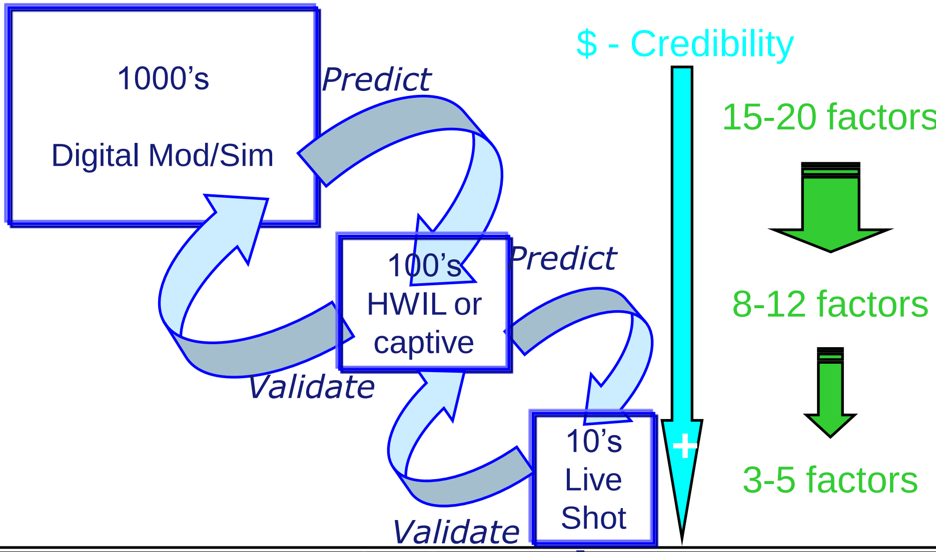
- **Vast majority of capabilities passed**
- **Wrung out sensors and weapons deliveries**
- **Dramatic reductions in usual trials while spanning many more test points**
- **Moderate success with teaming with Boeing on design points (maturing)**

Source: F-15E Secure SATCOM Test, Ms. Cynthia Zessin, Gregory Hutto, 2007 F-15 OFP CTF 53d Wing / 46 Test Wing, Eglin AFB, Florida

Strike Weapons Delivery a *Scenario* Design Improved with DOE

	Cases Design	Same N - Factorial	-50% N Factorial	Comparison
Cases	6	32	16	2.5 to 5x more (530%)
n	10	2	2	
N	60	64	32	-50% to +6.7% more
Sensitivity (δ/σ)	1	1	1	
α	5%	5%	5%	Same
β	20%	2.5%	20%	$1/10^{th}$ error rate
Power (1-β)	80%	97.50%	80%	detect shifts equal or

Case: Integration of Sim-HWIL-Captive-Live Fire Events



Test Objective:

- **Most test programs face this – AIM-9X, AMRAAM, JSF, SDB II, etc…**
- **Multiple simulations of reality with increasing credibility but increasing cost**
- **Multiple test conditions to screen for most vital to performance**
- **How to strap together these simulations with prediction and validation?**

DOE Approach:

- In digital sims screen 15-20 variables with fractional factorials and predict performance
- In HWIL, confirm digital prediction (validate model) and further screen 8-12 factors; predict
- In live fly, confirm prediction (validate) and test 3-5 most vital variables
- Prediction Discrepancies offer chance to improve sims

Results:

- Approach successfully used in 53d Wing EW Group
- SIL labs at Eglin/PRIMES > HWIL on MSTE Ground Mounts > live fly (MSTE/NTTR) for jammers and receivers
- Trimmed live fly sorties from 40-60 to 10-20 (typical) today
- AIM-9X, AMRAAM, ATIRCM: ***90***% sim reduction

Case: Secure SATCOM for F-15E Strike Eagle

Test Objective:

- To achieve secure A-G comms in Iraq and Afghanistan, install ARC-210 in fighters
- Characterize P(comms) across geometry, range, freq, radios, bands, modes
- Other ARC-210 fighter installs show problems – caution needed here!

DOE Approach:

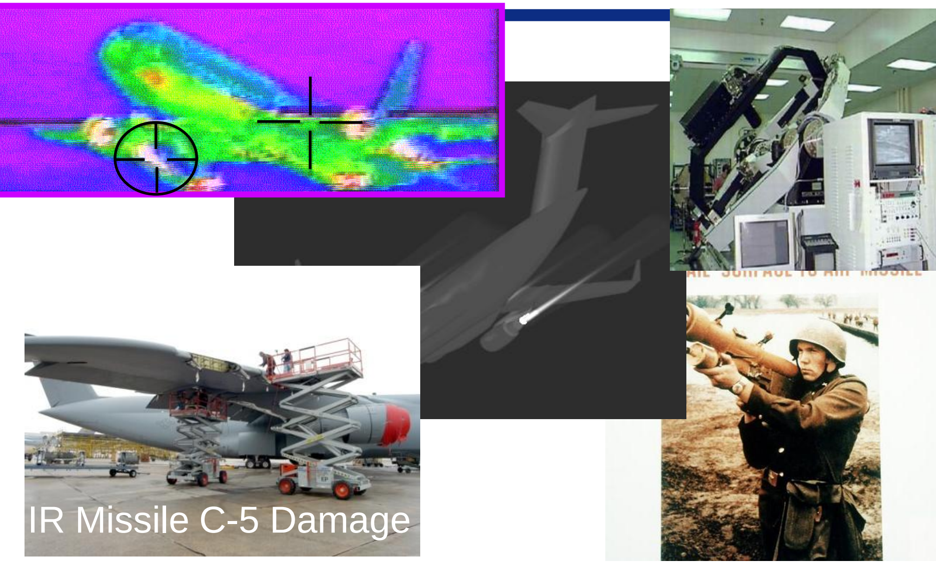
- Use Embedded Face-Centered CCD design (created for X-31 project, last slide)
- Gives 5-levels of geometric variables across radios & modes
- Speak 5 randomly-generated words and score number correct
- Each Xmit/Rcv a test event – 4 missions planned

Results:

- For higher-power radio – all good
- For lower power radio, range problems
- Despite urgent timeframe and canceled missions, enough proof to field

Source: F-15E Secure SATCOM Test, Ms. Cynthia Zessin, Gregory Hutto, 2007 F-15 OFP CTF 53d Wing / 46 Test Wing, Eglin AFB, Florida

Case: GWEF Large Aircraft IR Hit Point Prediction



Test Objective:

- IR man-portable SAMs pose threat to large aircraft in current AOR
- Dept Homeland Security desired Hit point prediction for a range of threats needed to assess vulnerabilities
- Solution was HWIL study at GWEF (ongoing)

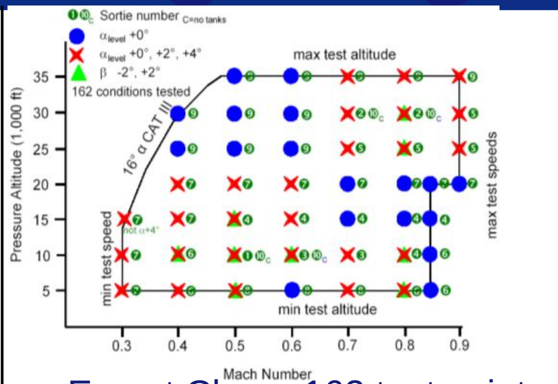
DOE Approach:

- Aspect – 0-180 degees, 7each
- Elevation – Lo,Mid,Hi, 3 each
- Profiles – Takeoff, Landing, 2 each
- Altitudes – 800, 1200, 2 each
- Including threat – 588 cases
- With usual reps nearly 10,000 runs
- DOE controls replication to min needed

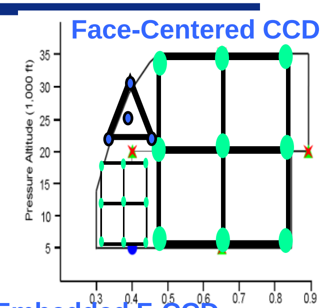
Results:

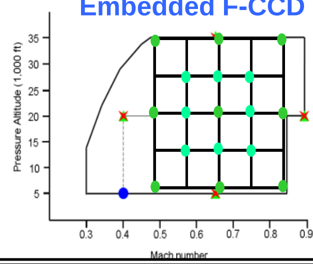
- Revealed unexpected hit point behavior
- Process highly interactive (rare 4-way)
- Process quite nonlinear w/ 3rd order curves
- Reduced runs required 80% over past
- Possible reduction of another order of magnitude to 500-800 runs

Case: Reduce F-16 Ventral Fin Fatigue from Targeting Pod



Expert Chose 162 test points





DOE Approach:

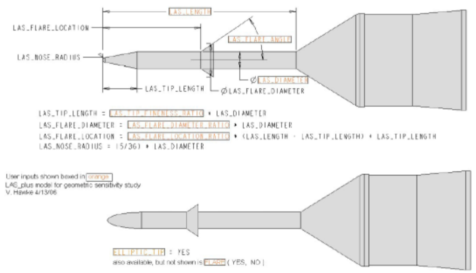
- Many alternate designs for this 5-dimensional space (α, β , Mach , alt, Jets on/off)

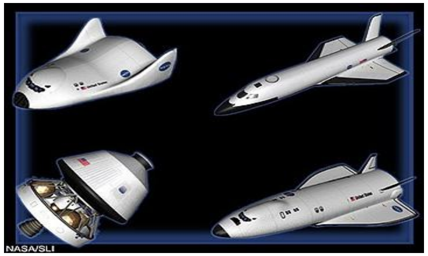
	Choice	Design Name	Test Points	Percent of Baseline	Model
	Base	Subject-expert	324	100%	none - inspection
Designed Experiments	1	CCD	58	18%	quadratic + interactions
	2	FCD	58	18%	quadratic + interactions
	3	Fractional FCD	38	12%	quadratic + interactions
	4	1/3rd fraction 3-level	54	17%	quadratic interactions
	5	Box-Behnken	54	17%	quadratic + interactions
	6	S-L Embedded FCD	76	23%	Up to cubic model

Results:

- New design invented circa 2005 capable of efficient flight envelope search
- Suitable for loads, flutter, integration, acoustic, vibration – full range of flight test
- Experimental designs can increase knowledge while dramatically decreasing required runs
- A full DOE toolbox enables more flexible testing

Case: CFD for NASA CEV





Test Objective:

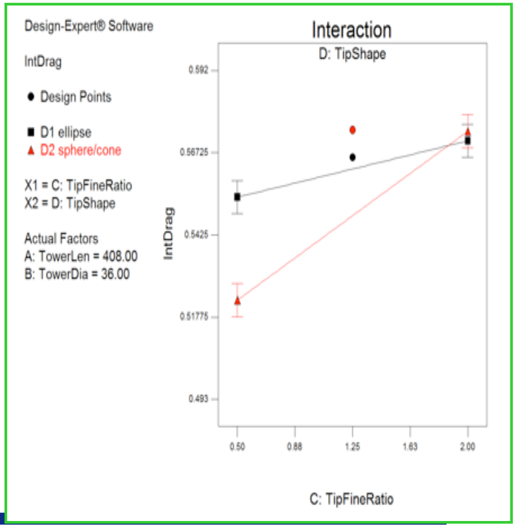
- **Select geometries to minimize total drag in ascent to orbit for NASA's new Crew Exploration Vehicle (CEV)**
- **Experts identified 7 geometric factors to explore including nose shape**
- **Down-selected parameters further refined in following wind tunnel experiments**

DOE Approach:

- Two designs – with 5 and 7 factors to vary
- Covered elliptic and conic nose to understand factor contributions
- Both designs were first order polynomials with ability to detect nonlinearities
- Designs also included additional confirmation points to confirm the empirical math model in the test envelope

Results:

- Original CFD study envisioned 1556 runs
- DOE optimized parameters in 84 runs – 95%!
- ID'd key interaction driving drag



Source: **A Parametric Geometry CFD Study Utilizing DOE** *Ray D. Rhew, Peter A. Parker, NASA Langley Research Center,* **AIAA 2007 1616**

A Strategy to be the *Best* … Using Design of Experiments

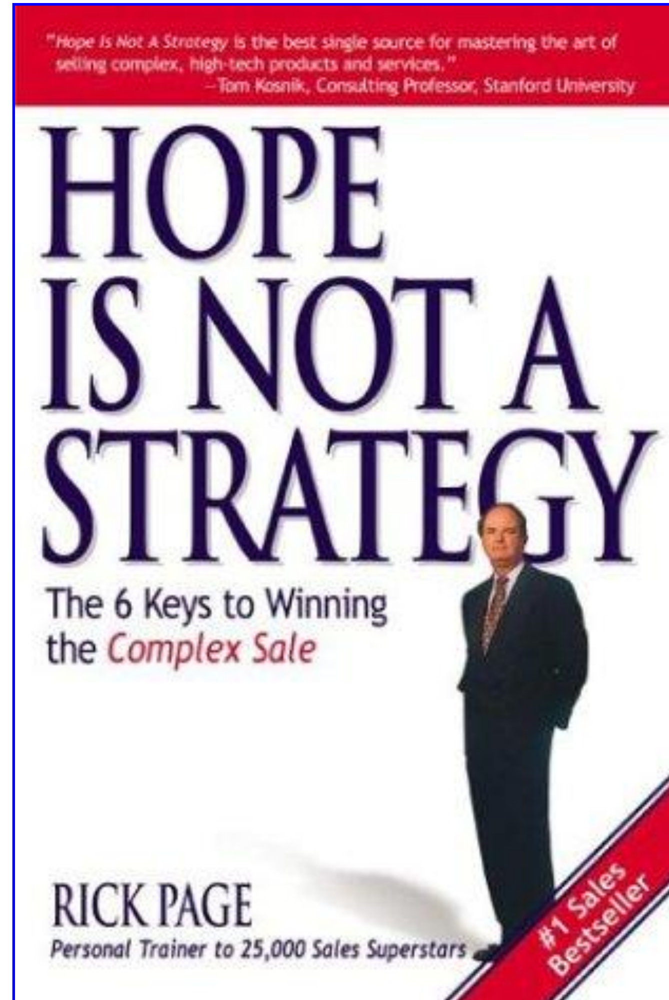
- Inform OT Leadership of Statistical Thinking for Test
- Adopt most powerful test strategy (DOE)
- Train & mentor total Ops Test teams
- Combo of AFIT, Center, & University
- Revise AF T&E policy, procedures
- Share these test improvements

<u>Adopting DOE</u>

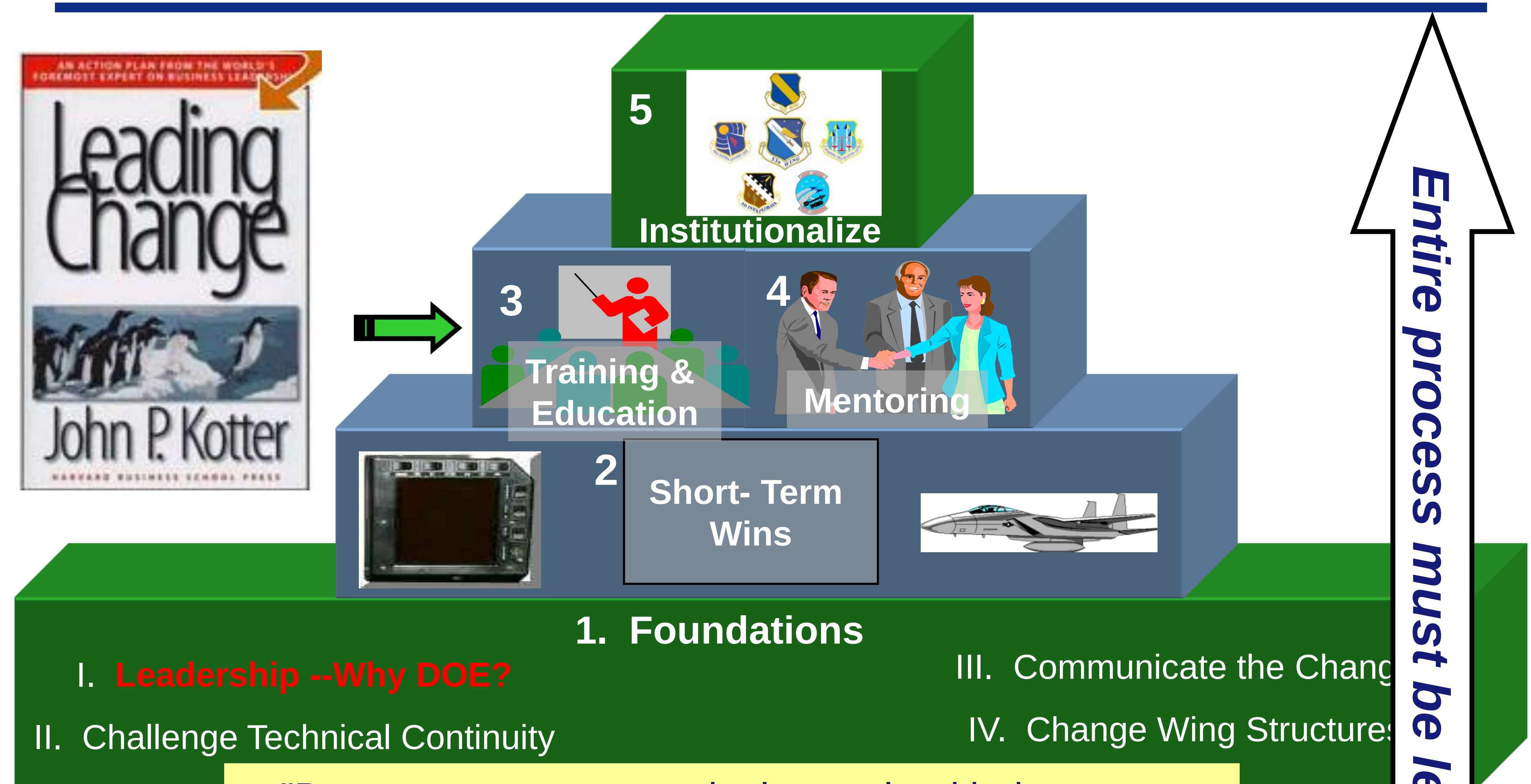
- 53d Wing & AFOTEC
- 18 FTS (AFSOC)
- RAF AWC
- DoD OTA: DOT&E, AFOTEC, ATEC, OPTEVFOR & MCOTEA
- AFFTC & TPS
- 46 TW & AAC
- AEDC

<u>Targets</u>

- HQ AFMC
- ASC & ESC
- Service DT&E



Five Steps to Teach Testers to Fish...



"Because management deals mostly with the **status quo** and leadership deals mostly with **change**, in the next century we are going to have to try to become much more skilled at creating leaders." -- Dr. John Kotter

53d Wing Policy Model: Test Deeply & Broadly with Power & Confidence

- **From 53d Wing Test Manager's Handbook*:**

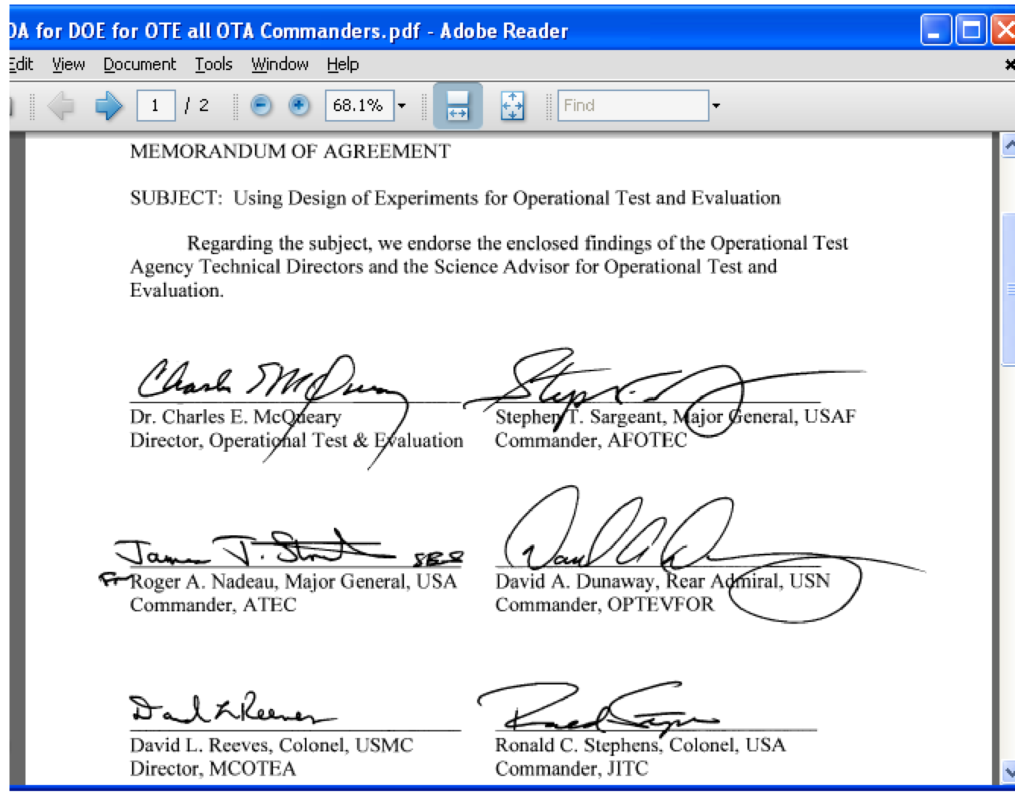
  "While this [*list of test strategies*] is not an all-inclusive list, these are well suited to operational testing. The test design policy in the 53d Wing supplement to AFI 99-103 mandates that we achieve confidence and power across a broad range of combat conditions. After a thorough examination of alternatives, the DOE methodology using factorial designs should be used whenever possible to meet the intent of this policy."

* Original Wing Commander Policy April 2002

March 2009: OTA Commanders Endorse DOE for both OT & DT

"Experimental design further provides a valuable tool to identify and mitigate risk in all test activities. It offers a framework from which test agencies may make well-informed decisions on resource allocation and scope of testing required for an adequate test. A DOE-based test approach will not necessarily reduce the scope of resources for adequate testing.

Successful use of DOE will require a cadre of personnel within each OTA organization with the professional knowledge and expertise in applying these methodologies to military test activities. Utilizing the discipline of DOE in all phases of program testing from initial developmental efforts through initial and follow-on operational test endeavors affords the opportunity for rigorous systematic improvement in test processes."



MEMORANDUM OF AGREEMENT

SUBJECT: Using Design of Experiments for Operational Test and Evaluation

Regarding the subject, we endorse the enclosed findings of the Operational Test Agency Technical Directors and the Science Advisor for Operational Test and Evaluation.

Dr. Charles E. McQueary
Director, Operational Test & Evaluation

Stephen T. Sargeant, Major General, USAF
Commander, AFOTEC

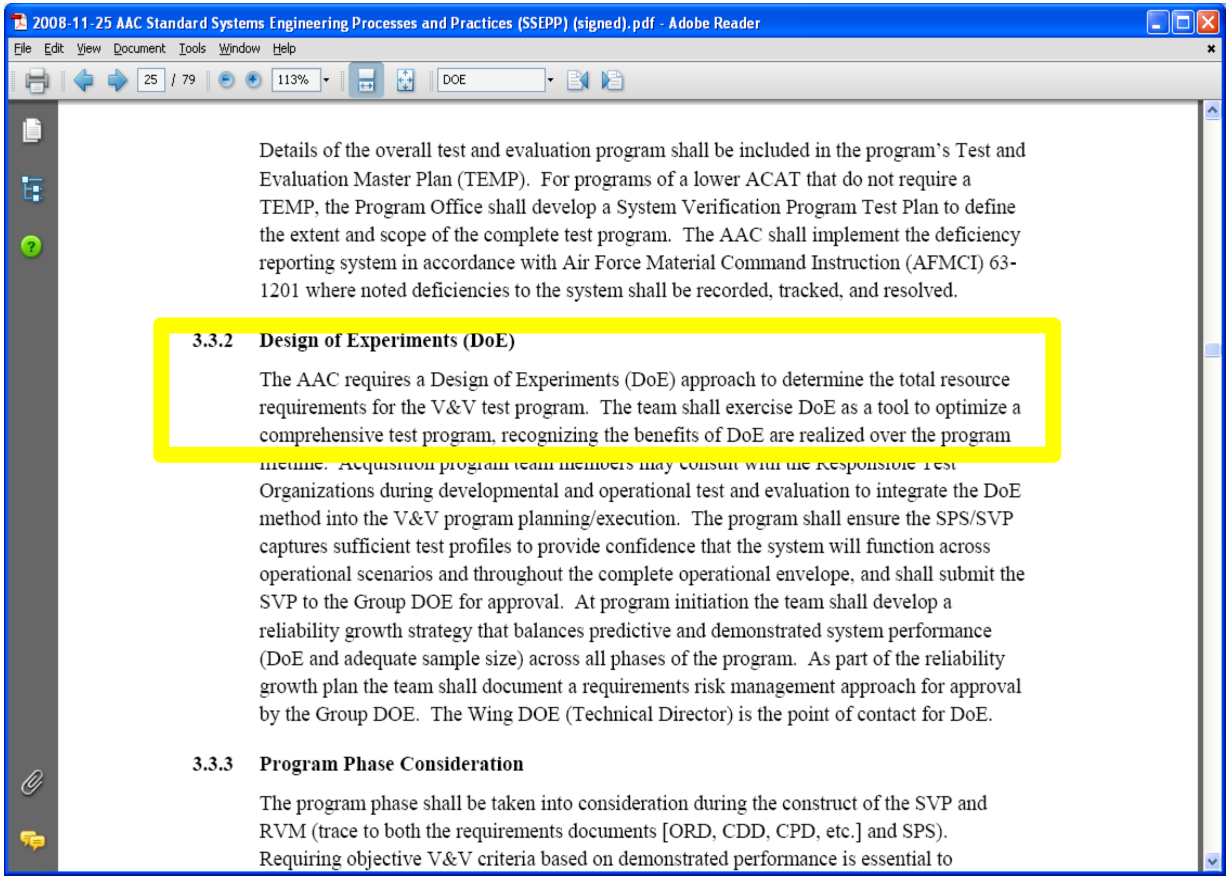
Roger A. Nadeau, Major General, USA
Commander, ATEC

David A. Dunaway, Rear Admiral, USN
Commander, OPTEVFOR

David L. Reeves, Colonel, USMC
Director, MCOTEA

Ronald C. Stephens, Colonel, USA
Commander, JITC

Nov 2008: AAC Endorses DOE for RDT&E Systems Engineering



2008-11-25 AAC Standard Systems Engineering Processes and Practices (SSEPP) (signed).pdf - Adobe Reader

Details of the overall test and evaluation program shall be included in the program's Test and Evaluation Master Plan (TEMP). For programs of a lower ACAT that do not require a TEMP, the Program Office shall develop a System Verification Program Test Plan to define the extent and scope of the complete test program. The AAC shall implement the deficiency reporting system in accordance with Air Force Material Command Instruction (AFMCI) 63-1201 where noted deficiencies to the system shall be recorded, tracked, and resolved.

3.3.2 Design of Experiments (DoE)

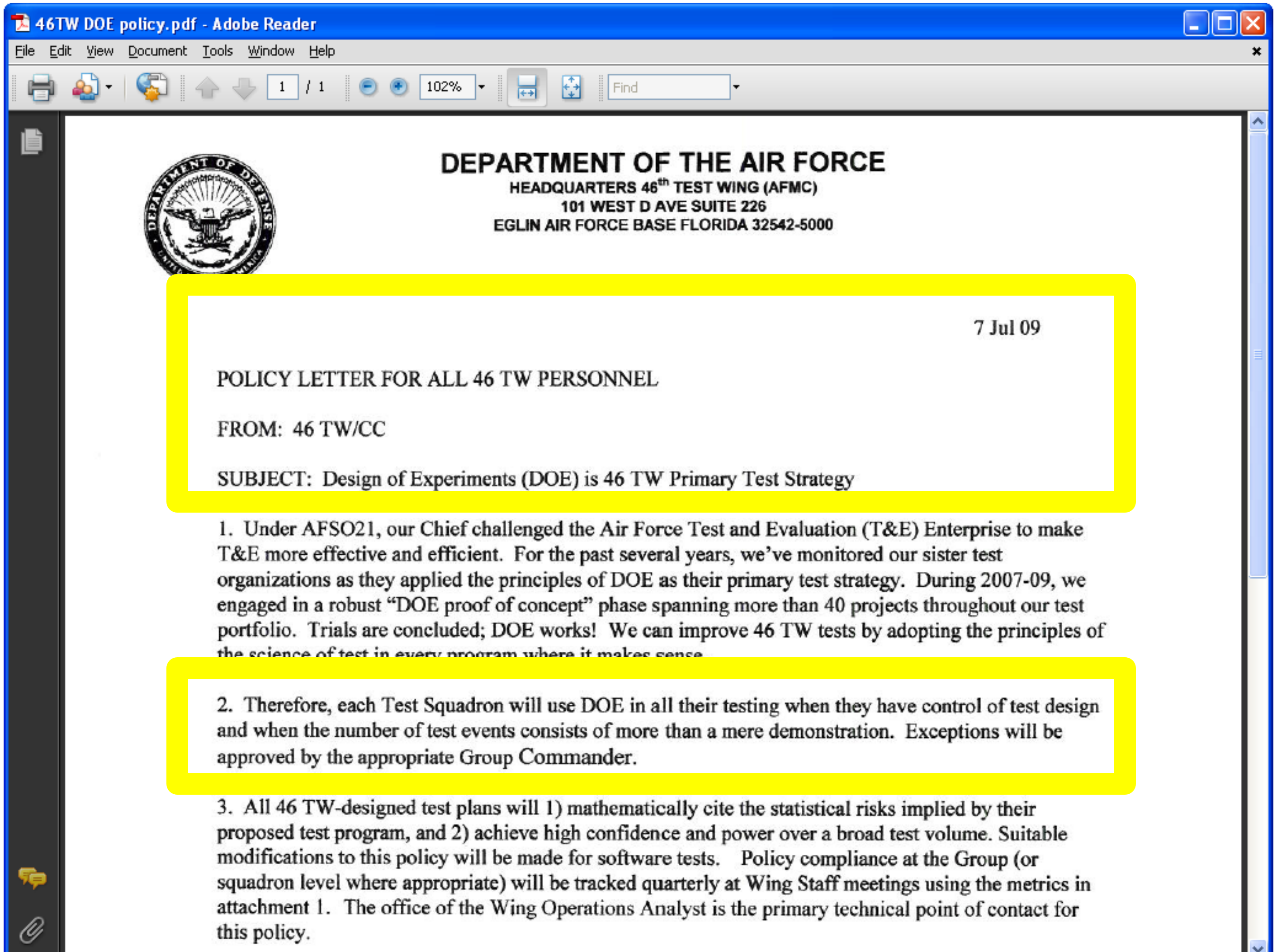
The AAC requires a Design of Experiments (DoE) approach to determine the total resource requirements for the V&V test program. The team shall exercise DoE as a tool to optimize a comprehensive test program, recognizing the benefits of DoE are realized over the program lifetime. Acquisition program team members may consult with the Responsible Test Organizations during developmental and operational test and evaluation to integrate the DoE method into the V&V program planning/execution. The program shall ensure the SPS/SVP captures sufficient test profiles to provide confidence that the system will function across operational scenarios and throughout the complete operational envelope, and shall submit the SVP to the Group DOE for approval. At program initiation the team shall develop a reliability growth strategy that balances predictive and demonstrated system performance (DoE and adequate sample size) across all phases of the program. As part of the reliability growth plan the team shall document a requirements risk management approach for approval by the Group DOE. The Wing DOE (Technical Director) is the point of contact for DoE.

3.3.3 Program Phase Consideration

The program phase shall be taken into consideration during the construct of the SVP and RVM (trace to both the requirements documents [ORD, CDD, CPD, etc.] and SPS). Requiring objective V&V criteria based on demonstrated performance is essential to

- **AAC Standard Systems Engineering Processes and Practices**

July 2009: 46 TW Adopts DOE as default method of test



DEPARTMENT OF THE AIR FORCE
HEADQUARTERS 46th TEST WING (AFMC)
101 WEST D AVE SUITE 226
EGLIN AIR FORCE BASE FLORIDA 32542-5000

7 Jul 09

POLICY LETTER FOR ALL 46 TW PERSONNEL

FROM: 46 TW/CC

SUBJECT: Design of Experiments (DOE) is 46 TW Primary Test Strategy

1. Under AFSO21, our Chief challenged the Air Force Test and Evaluation (T&E) Enterprise to make T&E more effective and efficient. For the past several years, we've monitored our sister test organizations as they applied the principles of DOE as their primary test strategy. During 2007-09, we engaged in a robust "DOE proof of concept" phase spanning more than 40 projects throughout our test portfolio. Trials are concluded; DOE works! We can improve 46 TW tests by adopting the principles of the science of test in every program where it makes sense.

2. Therefore, each Test Squadron will use DOE in all their testing when they have control of test design and when the number of test events consists of more than a mere demonstration. Exceptions will be approved by the appropriate Group Commander.

3. All 46 TW-designed test plans will 1) mathematically cite the statistical risks implied by their proposed test program, and 2) achieve high confidence and power over a broad test volume. Suitable modifications to this policy will be made for software tests. Policy compliance at the Group (or squadron level where appropriate) will be tracked quarterly at Wing Staff meetings using the metrics in attachment 1. The office of the Wing Operations Analyst is the primary technical point of contact for this policy.

We Train the Total Test Team … but first, our Leaders!

Leadership Series

- ***DOE Orientation* (1 hour)**
- ***DOE for Leaders*** (half day)
- ***Introduction to Designed Experiments*** (PMs- 2 days)

OA/TE Practitioner Series

- **10 sessions and 1 week each**
- **Reading--Lecture--Seatwork**
- *Basic Statistics Review* (1 week)
  - **Random Variables and Distributions**
  - **Descriptive & Inferential Statistics**
  - **Thorough treatment of t Test**
- *Applied DOE I and II* (1 week each)
  - **Advanced Undergraduate treatment**
  - **Graduates know both how and why**

Journeymen Testers

Continuation Training

- Weekly seminars online
- Topics wide ranging
  - New methods
  - New applications
  - Problem decomposition
  - Analysis challenges
  - Reviewing the basics
  - Case studies
  - Advanced techniques
- DoD web conferencing
- Q&A session following

Operations Analyst Forum/DOE Continuation Training for 04 May 09:

Location/Date/Time: B1, CR220, Monday, 04 May 09, 1400-1500 (CST)

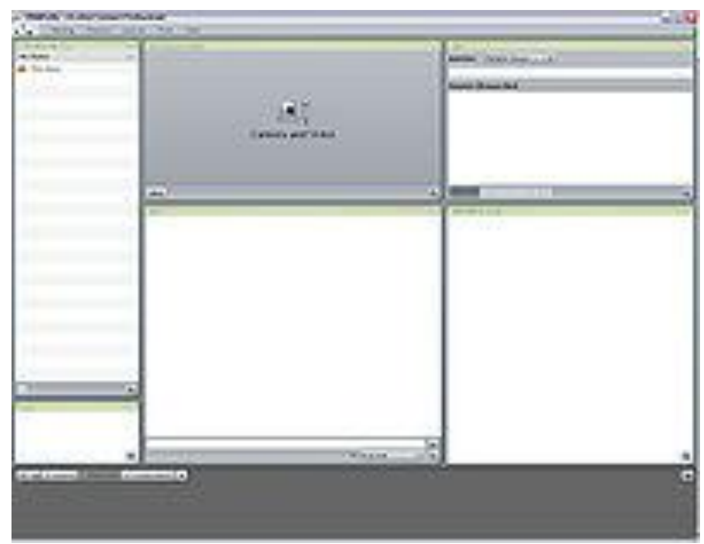
Purpose: OPS ANALYST FORUM, DOE CONTINUATION TRAINING, USAF T&E COLLABORATION

Live: Defense Connect Online: https://connect.dco.dod.mil/eglindoe

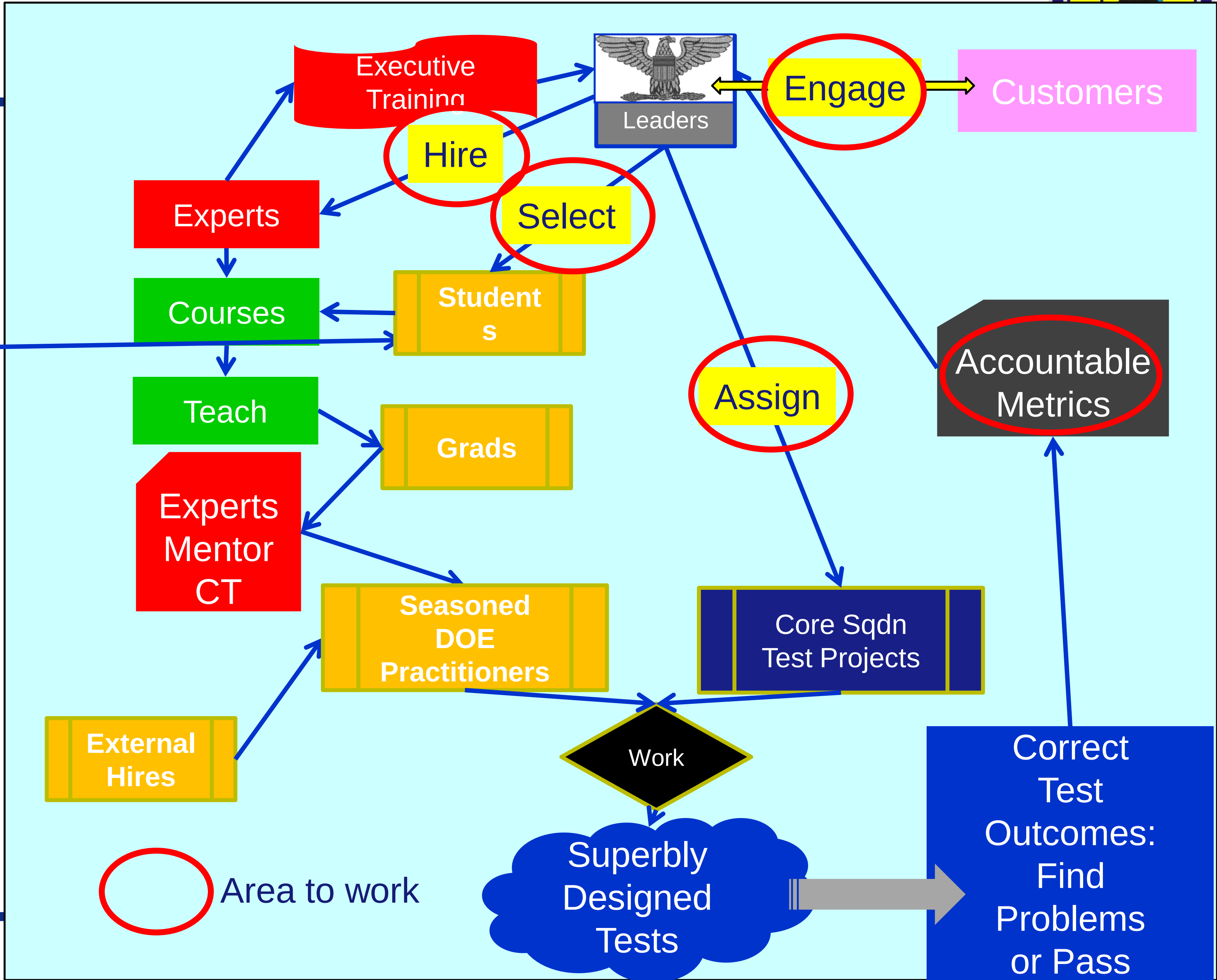
Dial-In: (VoIP not available yet) 850-882-6003/DSN 872-6003

Data:

- USAF DOE CoP
- https://afkm.wpafb.af.mil/SiteConsentBanner.aspx?ReturnUrl=%2fASPs%2fCoP%2fOpenCoP.asp%3fFilter%3dOO-TE-MC-79&Filter=OO-TE-MC-79 (please let me know if this link works for you)



Our Goal



TW210-Bootcamp-TW105 Intro Cse:

- Awareness
- Helpers
- Feeder

Kotter's 8 Steps for Making Change Stick

John Kotter's Recipe for Change and Institutionalization	
1	Create a sense of urgency
2	Sculpt and Polish a Vision (and credible path to achieve)
3	Assemble a ***Powerful*** Coalition
4	Over-Communicate Vision x10 x100!
5	Empower People (ID, direct, permit, train & equip)
6	Get the ***quick, core*** wins (projects)
7	Do **not** declare victory & retire early
8	Anchor it: Change systems, organization & People
	-- from ***Leading Change***

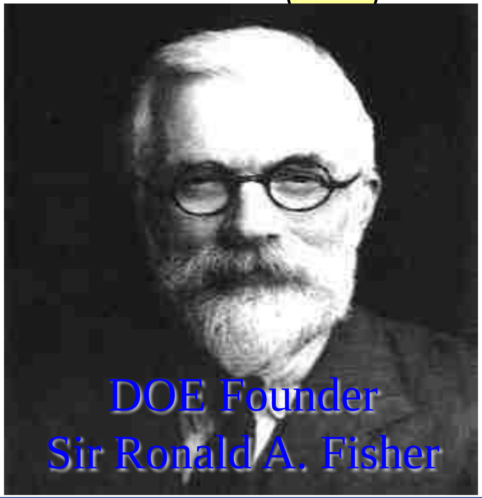
DOE Initiative Status

- ✓ Gain AAC Commander endorsement and policy announcement
- ✓ Train and align leadership to support initiative
- ✓ Commence training technical testers
- ✓ Launch multiple quick-win pilot projects to show applicability
- ✓ Communicate the change at every opportunity
- ✓ Gain AAC leadership endorsement and align client SPOs
- ❑ Influence hardware contractors with demonstrations and suitable contract language
- ❑ Keep pushing the wheel to build momentum
- ❑ Confront nay-sayers and murmurers
- ❑ Institutionalize with promotions, policies, Wing structures and practices
- ❑ Roll out to USAF at large and our Army & Navy brethren

But Why Should *T&E* Pursue DOE?

- Why:
  - **It's the scientific, structured, objective way to build better ops tests**
  - **DOE is faster, less expensive, and more informative than alternative methods**
  - **Uniquely answers deep and broad challenge: Confidence & Power across a broad battlespace**
  - **Our less-experienced testers can reliably succeed**
  - **Better chance of *getting it right*!**
- What I Need ...
  - **Practitioners – students!**
  - **People – a field grade civilian ea Sqdn**
  - **Projects – give me the biggest and baddest**
  - **Accountability – "if not DOE, why not??"**

"To call in the statistician after the experiment is ... asking him to perform a postmortem examination: he may be able to say what the experiment died of."
Address to Indian Statistical Congress, 1938.



What's *Your* Method of Test?

DOE: The *Science* of Test

Questions?