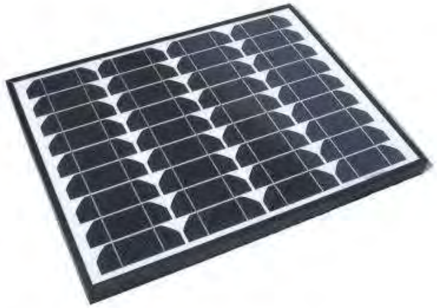
Design for High Throughput Experimentation in Industrial Research

James N. Cawse

Cawse and Effect LLC

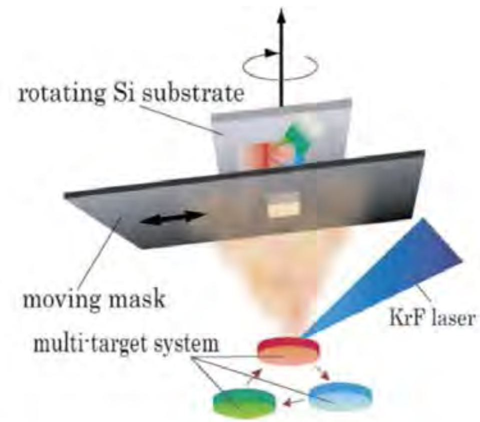
Designs feed into “real” problems

- Catalysts
  - Homogeneous
  - Heterogeneous
- Electronic materials
  - Photoelectric
  - Magnetic
- Coatings
  - Abrasion and UV resistant
- Energy
  - Solar cells
  - H2 Storage
  - Fuel cells
- Personal Care
  - Topical drugs

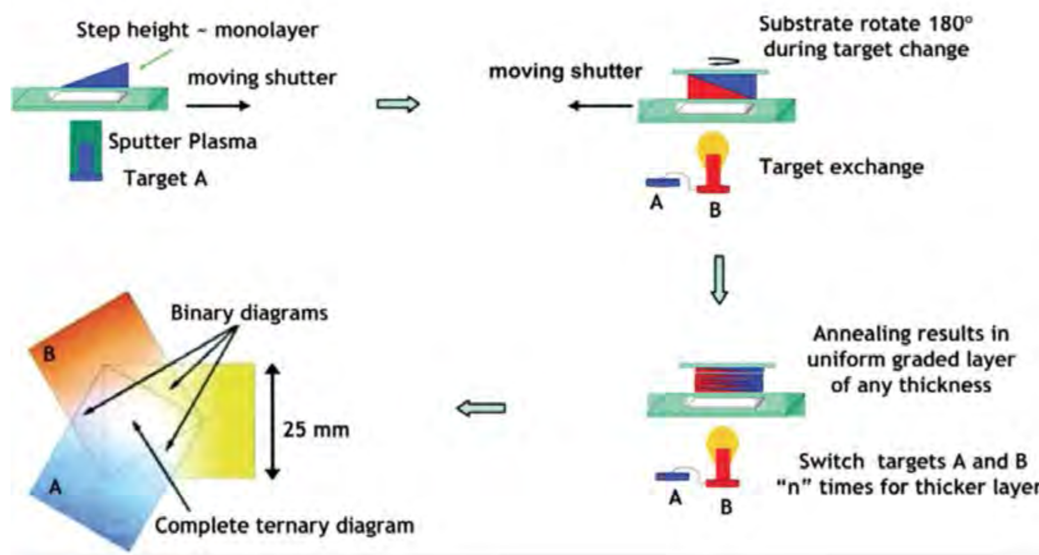


I. Planar Experiments

- Electronic, magnetic and similar materials
- Key elements
  - ”resolution”: pixel size of sensor
  - Ingenious methods of overlaying materials



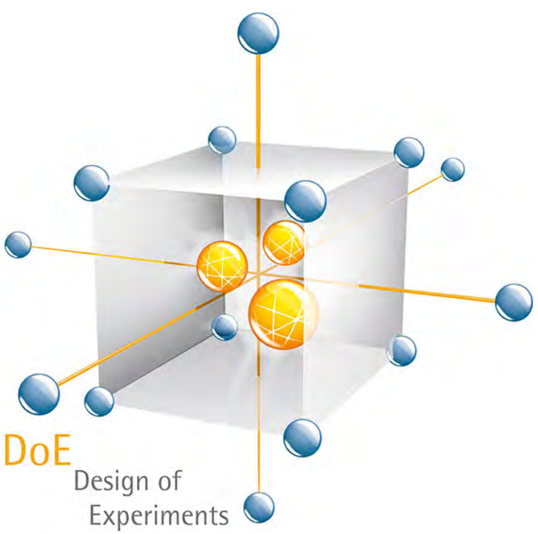
(a)



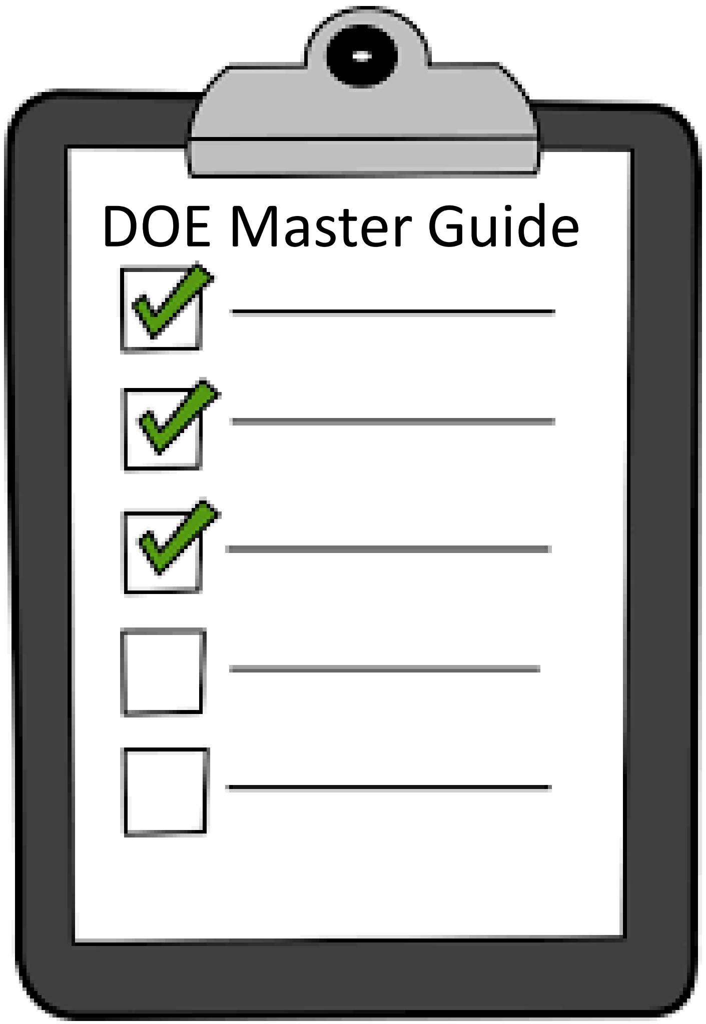
(b)

II. Physical Mixing

- Most other applications
  - Chemspeed
  - HTE
  - Unchained Labs
    - Freeslate
      - Symyx



Strategy



- **Project Team**
- **Project Title**
- **Objective(s)**
- **Knowledge**
- **Resources**

Management!

- Factors
- Responses
- Constraints
- Design

Science!

Objectives

Business

- Who are your customers?
- What is the flow of customer needs from the ultimate customer to you?
- What are the goals for each specific customer need?
- What is their priority?
- How are those goals measured?
- What is the specification for success?

Technical

- Unbiased
  - Don't pre-solve problem
  - Diverse input
  - Success clearly defined
- Specific
- Measurable
- Practical consequence

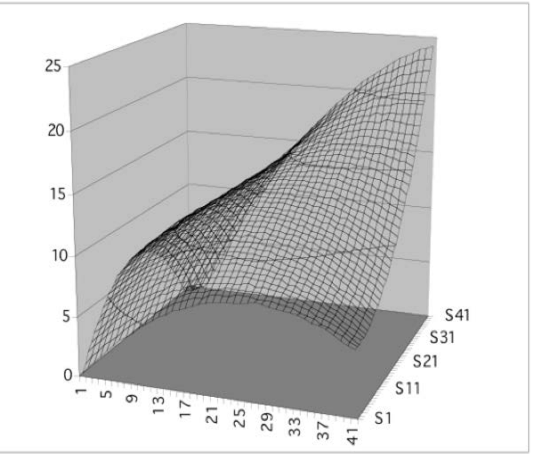
Management Discussion → RE$OURCE$

Design Elements

- Factors
  - Types
    - Quantitative
    - Qualitative
    - Formulation
  - Normal Levels
  - Setting Error
  - Proposed settings
- Responses
  - Types
    - Quantitative
    - Ordinal
    - Count
    - Binary
  - Expected ranges
  - Precision and Accuracy
  - Relationship to objective

Conventional Experimental Spaces

- Space is relatively smooth
- Factors are ordered
- Simple interactions
- Not too many dimensions.



- Curvature is moderate; Equations are continuous
- Preferably binary or real
- 2-way and quadratic
- Factor reduction early in program by simple screening experiments

High Throughput Spaces

- Phase diagram space
- Combinatorial Space
- High-Interaction space
- High dimensional spaces
- Split-Plot Space
- Organic Chemical Space

- Space isn’t smooth (phase transitions)
- Factors are not ordered (qualitative)
- 3-way and higher interactions
- Cannot eliminate factors early
- Process/Formulation Interactions
- The Pharmaceutical world

Approaches to high throughput spaces

“Standard”DOE

- DOE of DOE’s
- Principal Components
- Split Plots

Pruned Combi

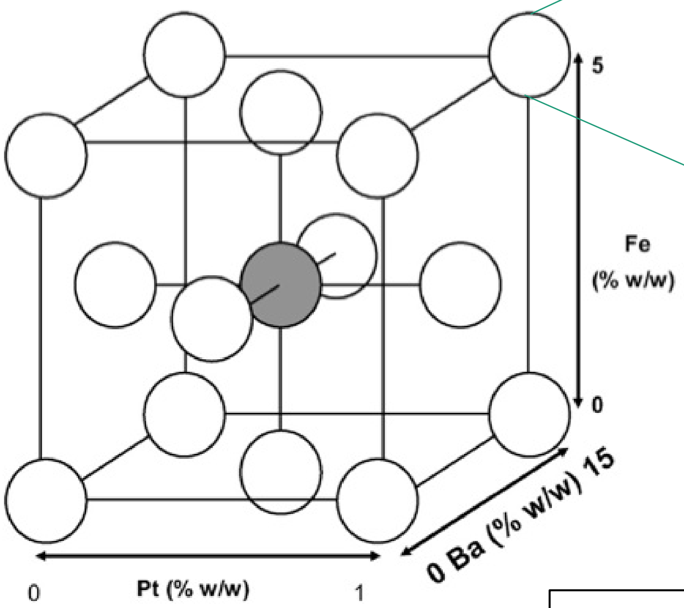
- Brainstorm
- Deconstruct
- Prune
- Design

Machine Learning

- Random Initiation
- Artificial Intelligence
  - Genetic Algorithm
  - Neural Net
  - Nondominated Sort
  - Pareto Tools
  - Clustering
  - Bayesian Network
  - Kriging

DOE of DOE’s: 576 total runs

Catalyst formulation:
3 factor 18 run RSM design



Process Conditions:
32 run 6 factor RSM
Design

Process model as f (conditions)

Model process coefficients as f (formulation)

Hendershot, Lauterbach et. al. (2007)

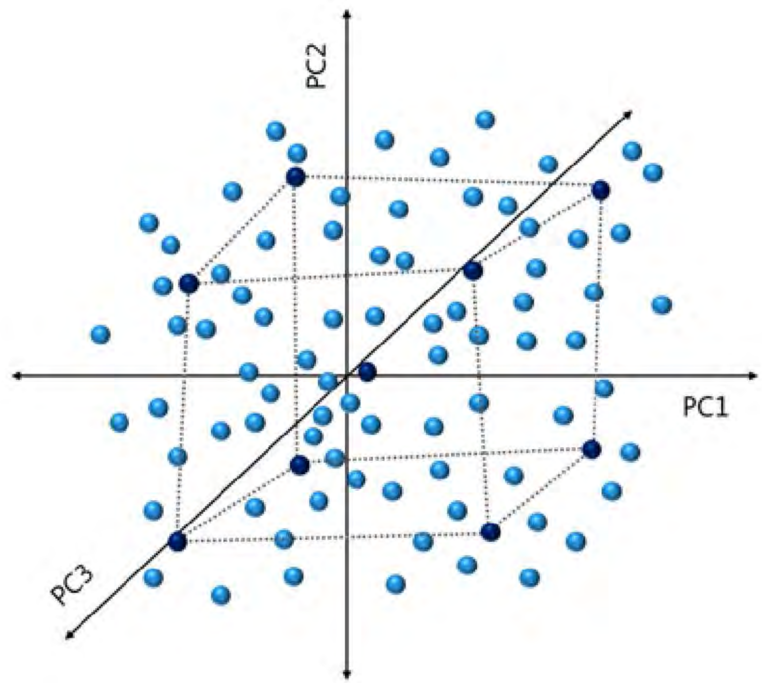
Principal Components in Chemical Space

”Descriptors”

- Physical factors
  - bp, density, bond length lipophilicity, polarizability, charge, flexibility, rigidity…
- Calculated factors
  - electron density, solubility parameters, solvent polarity parameters…

Design

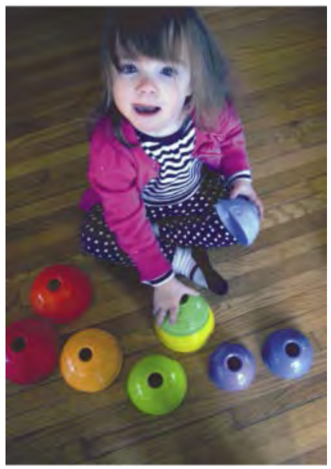
- PCA: reduce dimensionality
- Select representative compounds for “DOE”



“Pruned” Combi: Solid State Lighting

Business Opportunity

sunlight



standard LED

– Phosphors are critical to markets ranging from lighting to medicine applications.

– Conventional phosphors: difficult to produce & handle, environmentally unstable due to heat & moisture.

– Need to develop new phosphors for natural spectrum.

– Challenges: new matrices, host materials, deposition & environmental considerations.

– 2015 Lighting Market $250B; Encapsulant/Phosphor $230M

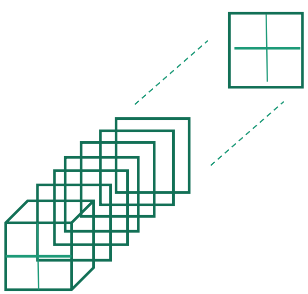
Initial Experimental Factors

Factor	Type	Qual	Quant
Ligand:Eu Salts	Qualitative	Various Salts	mol% range
Primary Metals, M1	Qualitative	Ti, Zr, Al, …	mol% range
Secondary Metals, M2	Qualitative	Lanthanides (Ln), Non-Ln	mol% range
Siloxanes	Qualitative	MDTQ Resins	mol% range
Ligands	Qualitative	Hundreds (?)	mol% range
Solvents	Qualitative	Different Solvents	
Ligand/Eu ratio	Quantitative		
Ln/M1 ratio	Quantitative		
Total Metal (Ln+M1+M2)	Quantitative		
Water	Quantitative		
Temperature	Quantitative		
Pressure	Quantitative		

Experimental Factor Deconstruction

Qualitative Formulation

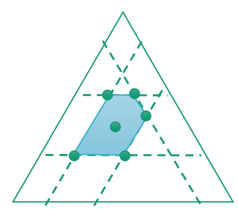
Factors	Levels
Ln Salts	3
~~M2~~	1
M1	3
Siloxanes	2
Ligands	100



All Combinations: 1800

Quantitative Formulation

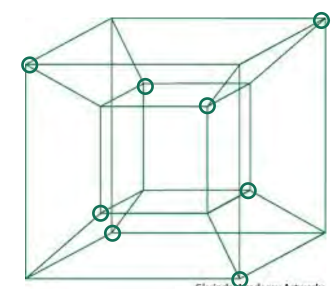
Factors	Levels
Ln Salts	mol% range
~~M2~~	~~1~~
M1	mol% range
Siloxanes	mol% range
Ligands	mol% range



11 run mixture

Process

Factors	Levels
Water	% range
Temp	range
Pressure	range
Dwell	range
Heat Rate	range



16 run Fractional Factorial

1800 x 11 x 16 = 316,800 possible runs

Experimental Factor Reduction

Qualitative Formulation

Factors	Levels
Ln Salts	3
~~M2~~	~~1~~
M1	3
Siloxanes	~~2~~ 1
Ligands	100
~~Solvents~~	~~1~~

All Combinations: 900

Quantitative Formulation

Factors	Levels
Ln Salts	mol% range
~~M2~~	~~1~~
M1	mol% range
Siloxanes	mol% range
Ligands	mol% range

9 run mixture

Process

Factors	Levels
Water	% range
~~Resin~~	~~% range~~
Temp	range
~~Pressure~~	~~range~~
Dwell	range
~~Heat Rate~~	~~range~~

5 run Fractional Factorial

900 x 9 x 5 = 40,500 possible runs

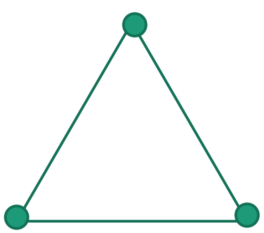
Experimental Design Space Reduction

Qualitative Formulation

Factors	Levels
Ln Salts	3
M1	3
Ligands	100

900 combinations

Quantitative Formulation



Effect Type

Effect	Type
Ln Salts	Main
M1	Main
Ligands	Main
Ln * M1	2-way
Ln* Ligands	2-way
M1*Ligands	2-way
Ln*M1*Ligands	3-way

Optimal Design Size

104 to 300

504 to 600

900

Process

504 x 3 x 1 = 1,512 possible runs

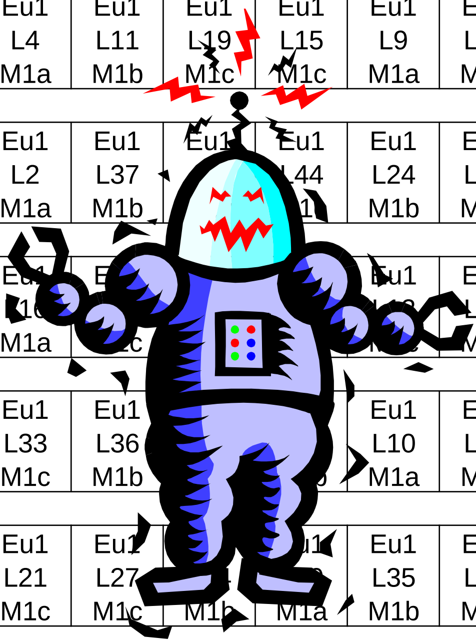
Is the Experimental Design Robot-Efficient?

Statistically “ideal” random plate

	A	B	C	D	E	F	G	H
1	Eu1 L22 M1c	Eu1 L4 M1a	Eu1 L11 M1b	Eu1 L19 M1c	Eu1 L15 M1c	Eu1 L9 M1a	Eu1 L42 M1b	Eu1 L28 M1b
2	Eu1 L32 M1a	Eu1 L2 M1a	Eu1 L37 M1b	Eu1 L6 M1b	Eu1 L44 M1c	Eu1 L24 M1b	Eu1 L20 M1a	Eu1 L39 M1a
3	Eu1 L18 M1b	Eu1 L16 M1a	Eu1 L14 M1c	Eu1 L43 M1a	Eu1 L26 M1b	Eu1 L12 M1c	Eu1 L13 M1b	Eu1 L29 M1a
4	Eu1 L30 M1c	Eu1 L33 M1c	Eu1 L36 M1b	Eu1 L1 M1b	Eu1 L5 M1b	Eu1 L10 M1a	Eu1 L25 M1c	Eu1 L8 M1a
5	Eu1 L23 M1b	Eu1 L21 M1c	Eu1 L27 M1c	Eu1 L34 M1b	Eu1 L40 M1a	Eu1 L35 M1b	Eu1 L41 M1c	Eu1 L31 M1c
6	Eu1 L3 M1c	Eu1 L17 M1a	Eu1 L7 M1c	Eu1 L38 M1a	Eu1 L45 M1a	STD	STD	STD

A Robot-Efficient Experimental Design

Statistically "Ideal" Random Plate



	A	B	C	D	E	F	G	H
1	Eu1 L22 M1c	Eu1 L4 M1a	Eu1 L11 M1b	Eu1 L19 M1c	Eu1 L15 M1c	Eu1 L9 M1a	Eu1 L42 M1b	Eu1 L28 M1b
2	Eu1 L32 M1a	Eu1 L2 M1a	Eu1 L37 M1b	Eu1 [illegible]	Eu1 L44 [illegible]	Eu1 L24 M1b	Eu1 L20 M1a	Eu1 L39 M1a
3	Eu1 L18 M1b	Eu1 L16 M1a	[illegible]	[illegible]	[illegible]	Eu1 [illegible]	Eu1 L13 M1b	Eu1 L29 M1a
4	Eu1 L30 M1c	Eu1 L33 M1c	Eu1 L36 M1b	[illegible]	[illegible]	Eu1 L10 M1a	Eu1 L25 M1c	Eu1 L8 M1a
5	Eu1 L23 M1b	Eu1 L21 M1c	Eu1 L27 M1c	[illegible] M1b	[illegible] M1a	Eu1 L35 M1b	Eu1 L41 M1c	Eu1 L31 M1c
6	Eu1 L3 M1c	Eu1 L17 M1a	Eu1 L7 M1c	Eu1 L38 M1a	Eu1 L45 M1a	STD	STD	STD

Blocked by Plates and Rows

	A	B	C	D	E	F	G	H
1	Eu1 L22 M1a	Eu1 L4 M1a	Eu1 L11 M1a	Eu1 L19 M1a	Eu1 L15 M1a	Eu1 L9 M1a	Eu1 L42 M1a	Eu1 L28 M1a
2	Eu1 L32 M1b	Eu1 L2 M1b	Eu1 L37 M1b	Eu1 L6 M1b	Eu1 L44 M1b	Eu1 L24 M1b	Eu1 L20 M1b	Eu1 L39 M1b
3	Eu1 L18 M1c	Eu1 L16 M1c	Eu1 L14 M1c	Eu1 L43 M1c	Eu1 L26 M1c	Eu1 L12 M1c	Eu1 L13 M1c	Eu1 L29 M1c
4	Eu1 L30 M1c	Eu1 L33 M1c	Eu1 L36 M1c	Eu1 L1 M1c	Eu1 L5 M1c	Eu1 L10 M1c	Eu1 L25 M1c	Eu1 L8 M1c
5	Eu1 L23 M1b	Eu1 L21 M1b	Eu1 L27 M1b	Eu1 L34 M1b	Eu1 L40 M1b	Eu1 L35 M1b	Eu1 L41 M1b	Eu1 L31 M1b
6	Eu1 L3 M1a	Eu1 L17 M1a	Eu1 L7 M1a	Eu1 L38 M1a	Eu1 L45 M1a	STD	STD	STD

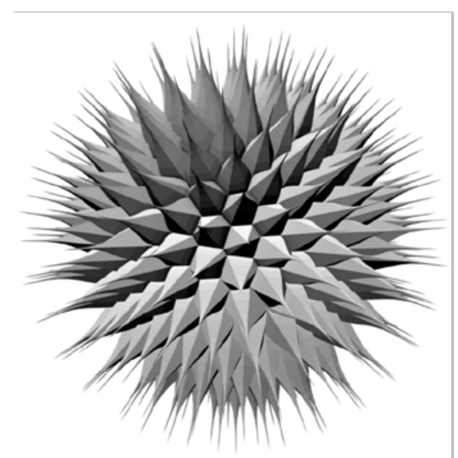
Not Ideal… but it will get done!

Machine Learning: GA's, NN's, etc.

- GA/NN: Baerns & Holena
- Bayesian Net: Poli
- Cluster Analysis: Bible
- Genetic Programing: Chakraborti

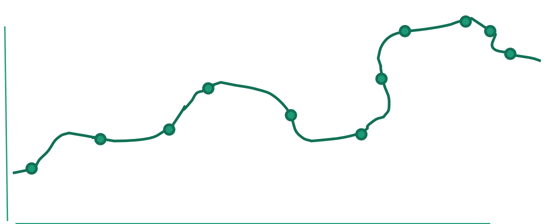
Advantages

- Many optimization paths in parallel
- Less attraction to local optima
- Deals with "sparse" sampling and high dimensions

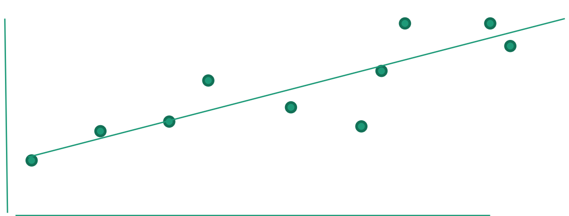


The first problems…

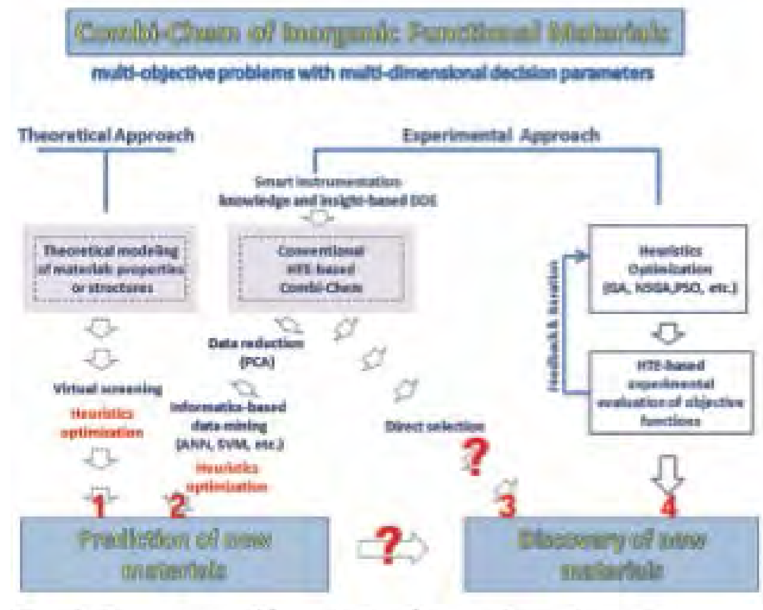
- Complexity
  - Mathematical sophistication
  - Statistical knowledge
  - Familiarity with MATLAB or similar
- Overfitting



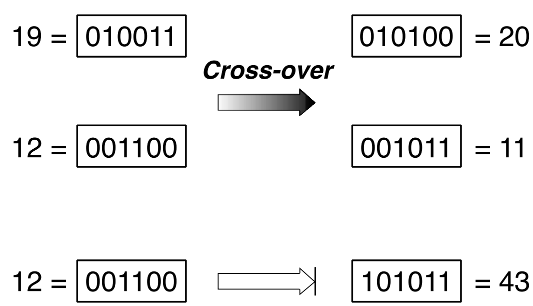
GA Model



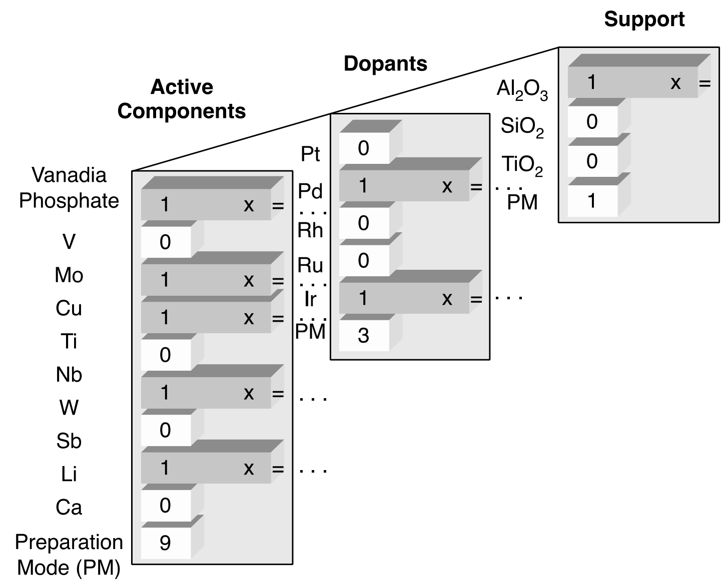
Regression Model



The Chemist’s GA problem…



Standard GA’s are based on “gene” structures with binary encoding



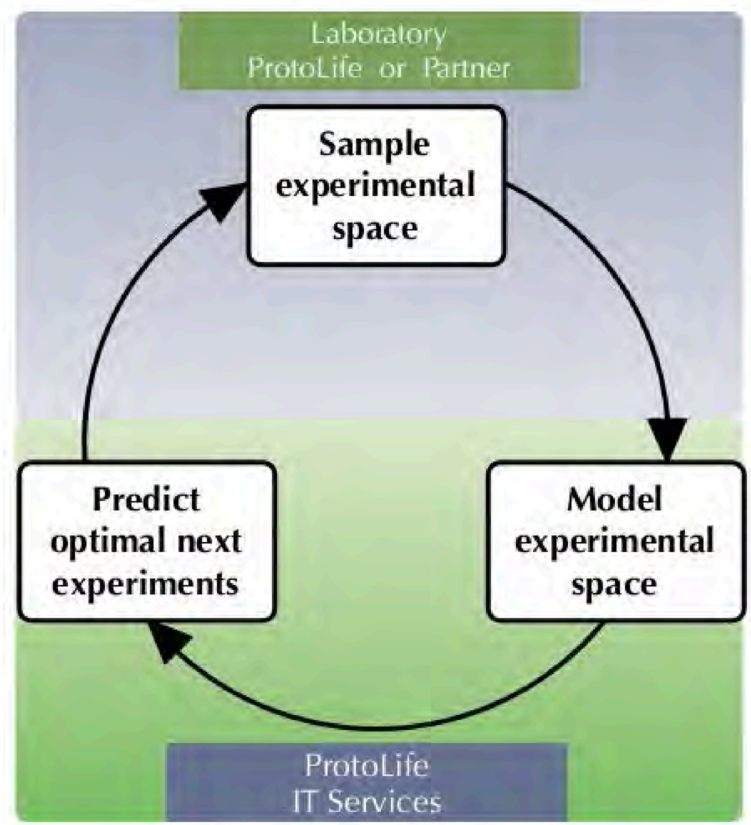
Chemistry must be described with “chromosome” structures and much more complex encoding

“Catalyst Description Language”

Berns, Holeňa, Cat. Sci. Series 7, 2009

Predictive Design Technology

- Web-based service
  - Predicts optimal experiments by modeling
  - Closed-loop iteration
  - Intelligent selection of machine learning options
  - Can exploit chemical information



www.protolife.com

Thank you!

And thanks to:
GE Global Research
Protolife Inc.
Dow Corning Silicones
Chemspeed

Cawse and Effect LLC

www.cawseandeffect.com