Panel on Design of Experiment

Dennis K.J. Lin
University Distinguished Professor
Supply Chain & Information Systems
The Pennsylvania State University

at
QPRC & SRC Joint Conference
08 June, 2006

Just Returned from Greece

And Found the Most Difficult Statistical Problem Ever!

Dennis K. J. Lin

Καθηγητὴς Στατιστικῆς καὶ Ἐφοδιαστικῆς Ἀλυσίδος

Τὰ ἐνδιαφέροντα τοῦ καθηγητοῦ Λὶν περιστρέφονται γύρω ἀπὸ τὴν μεθοδολογία τῆς στατιστικῆς στὶς ἐφαρμογὲς αὐτῆς στὶς ἐπιχειρήσεις, τὴν βιομηχανία καὶ τὴν διακυβέρνησι (business, industrial and government (BIG) applications). Μέγα μέρος τῆς ἐργασίας του ἔχει ἀφιερωθῇ στοὺς τομεῖς τῆς ἐξορύξεως δεδομένων, τοῦ πειραματικοῦ σχεδιασμοῦ, τῆς μεθοδολογίας ἐπιφανειῶν ἀποκρίσεως, τῆς ποιοτικῆς τεχνολογίας, τοῦ στατιστικοῦ ἐλέγχου διαδικασιῶν καὶ τῆς ἀξιοπιστίας. Οἱ τομεῖς αὐτοὶ ἔχουν ἄμεση σχέσι μὲ τὰ στατιστικὰ ἐργαλεῖα ὅπως ἡ στατιστικὴ τεχνικὴ τῶν προτύπων, ἡ ἐπαγωγὴ κατὰ Bayes, ἡ θεωρία βέλτιστου σχεδιασμοῦ, ἡ βελτιστοποίησι καὶ οἱ χρονοσειρές. Ἔχει δημοσιεύσει ὑπὲρ τὶς 100 ἐργασίες σὲ ποικίλα περιοδικὰ μεταξὺ τῶν ὁποίων τὰ κάτωθι: *Technometrics* and *Statistica Sinica*.

So many (unknown) Parameters at one time!!!

A typical Scientific Study

- Objective of the Study
- Set up Hypothesis
  (or set up underlying model, if possible)
- Prior knowledge
  - Cost (which determines the run size);
  - Number of variables, etc.
  - Not necessary in "functional" form
    (Prior distribution)
- Design of experiment
- Analysis, and etc…Decision making

The Role of Experiment Design When

- When we have extremely well understanding about the problem, or
- When we have only a vague understanding about the problem, or
- Somewhere in between

Important Issues (which will not be addressed here)

- Why experiment
- Why design of experiment? Does it matter?
- What (desirable) properties can be guaranteed?
- How to construct such an (optimal) design?
- Any undesirable properties (side-effect)?

Design Objectives

- Treatment Comparison
- Screening
- Model Building
- Parameter Estimation
- Optimization
- Prediction
- Confirmation
- Discovery (Random Shot)
- etc.

Example: Screening Experiments

- Random Shot
- Group Screening
- Supersaturated Design
- One-At-A-Time Experiment
- Latin Hypercube
- Uniform Design
- etc.

k is large
n is small
p is small ($<<k$)

Which is the "best" (Optimal)?

Some fundamental Issues

- Design comes before data analysis.
  - Design with or without Model
- What can randomization do for you?
  - Design without randomization
- What can orthogonality do you ?
  - Non-orthogonal design

Some fundamental Issues

- Design comes before data analysis.
  - Design with or without Model
- What can randomization do for you?
  - Design without randomization
- What can orthogonality do you ?
  - Non-orthogonal design

Design Criteria

- Combinatorics Design
  - Orthogonality
  - Uniformity
  - Resolution
  - Aberration
  - Projection
  - Clear factors
- Optimal Design
  - Information matrix
  - D-optimality
    (eigenvalue type alphabetic optimalities)
  - Pukelsheim (1993) book
  - Kiefer's original work

Required Information

- Combinatorics Design
  - Number of factors (k)
  - Number of runs (n)
  - Number of Levels
  - Limitation on n
    - $n=4t$ (HM)
    - certain relationship between n and k (e.g., $n=2^{k-p}$)
  - Orthogonality
- Optimal Design
  - Number of factors (k)
  - Number of runs (n)
  - The underlying model
  - Optimality Criterion

Some Examples in d=1 Dimension

First-Order Model

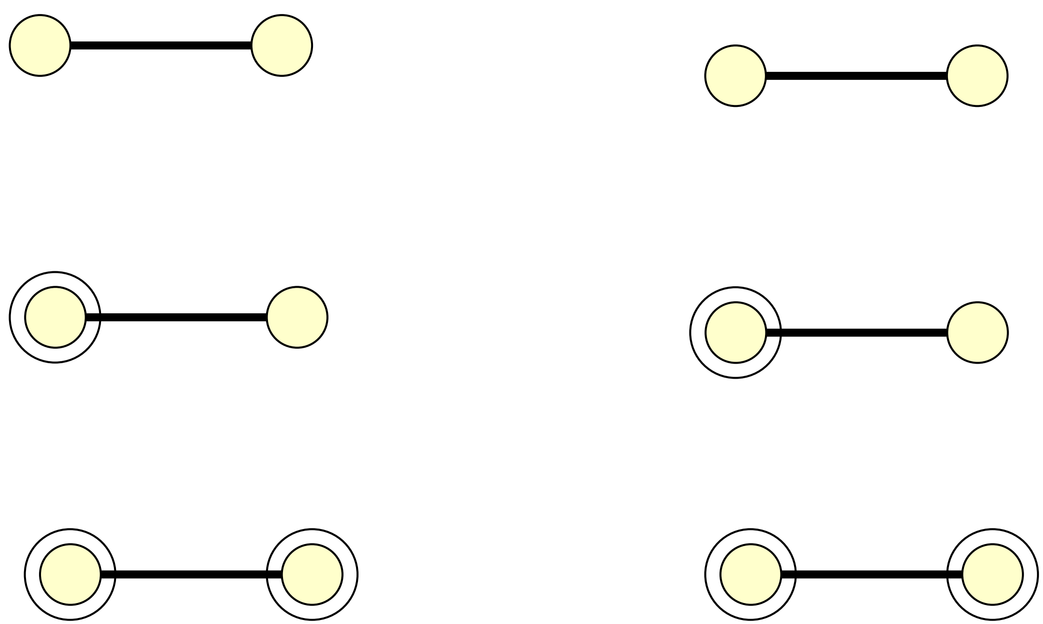
Two-Level Factorial

D-optimal Design

$n=2$

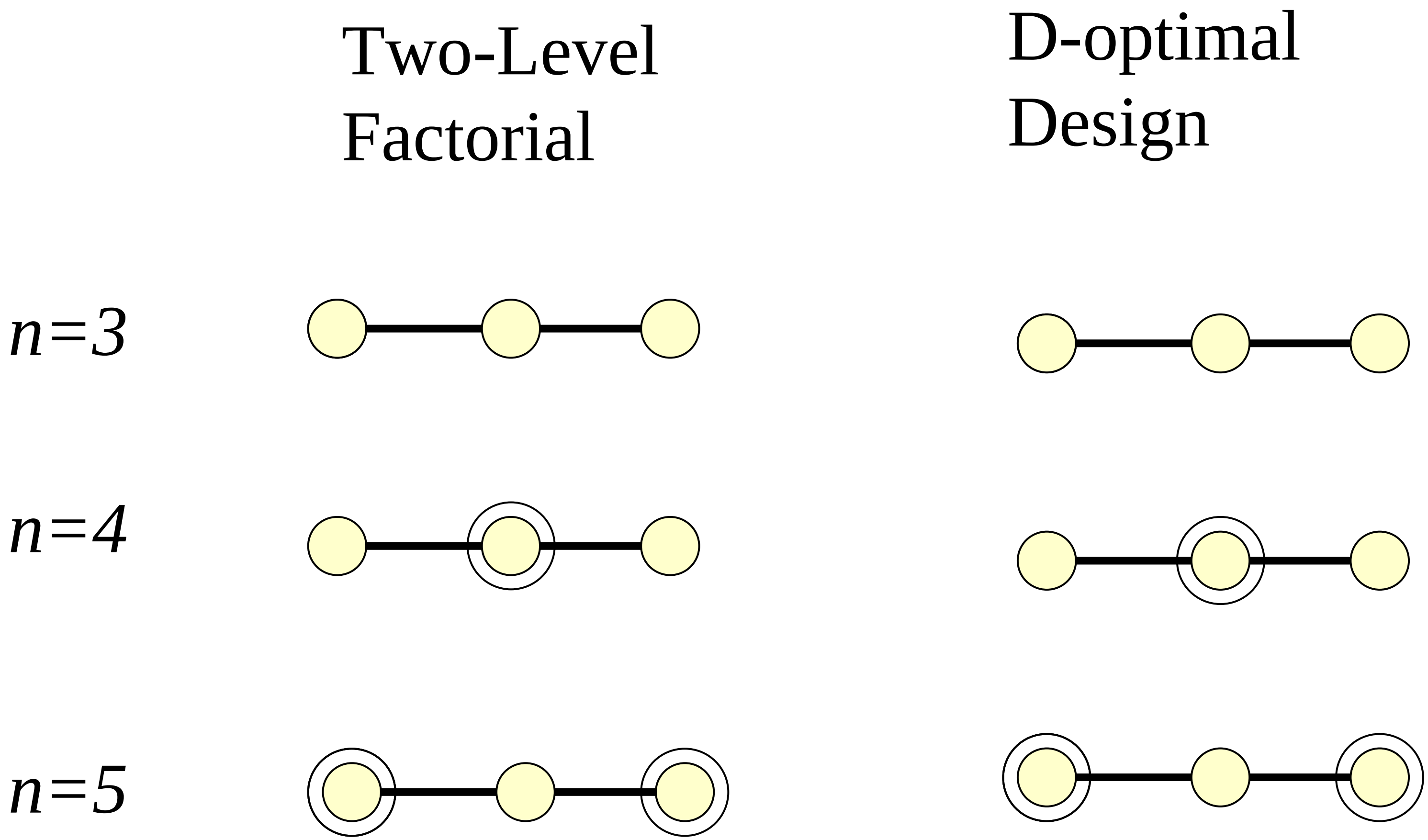
$n=3$

$n=4$



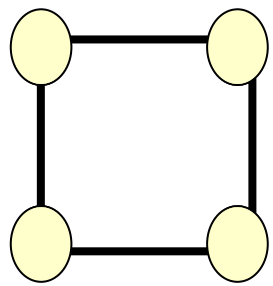
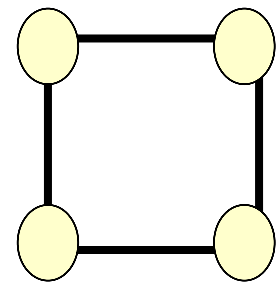
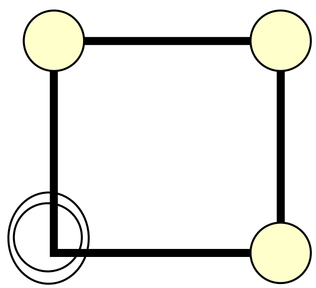
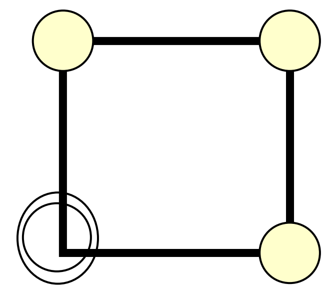
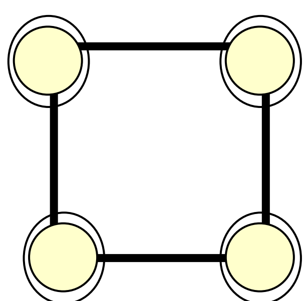
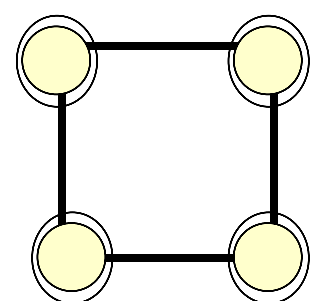
Some Examples in d=1 Dimension

Second-Order Model

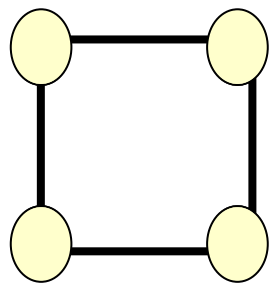
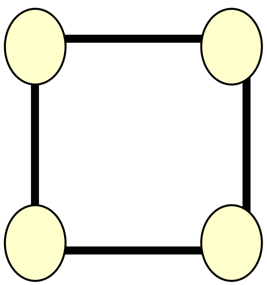
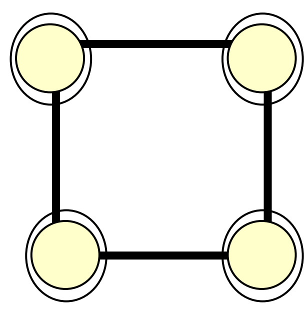
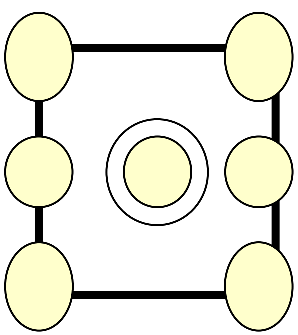


Some Examples in d=2 Dimension

First-Order Model

	Two-Level Factorial	D-optimal Design
n=4		
n=5		
n=8		

Some Examples in d=2 Dimension
Second-Order Model

	Two-Level Factorial	D-optimal Design
n=4		
n=8		

D-optimal Designs in higher degree (d-1 Dimension)

Pukelsheim (1993)

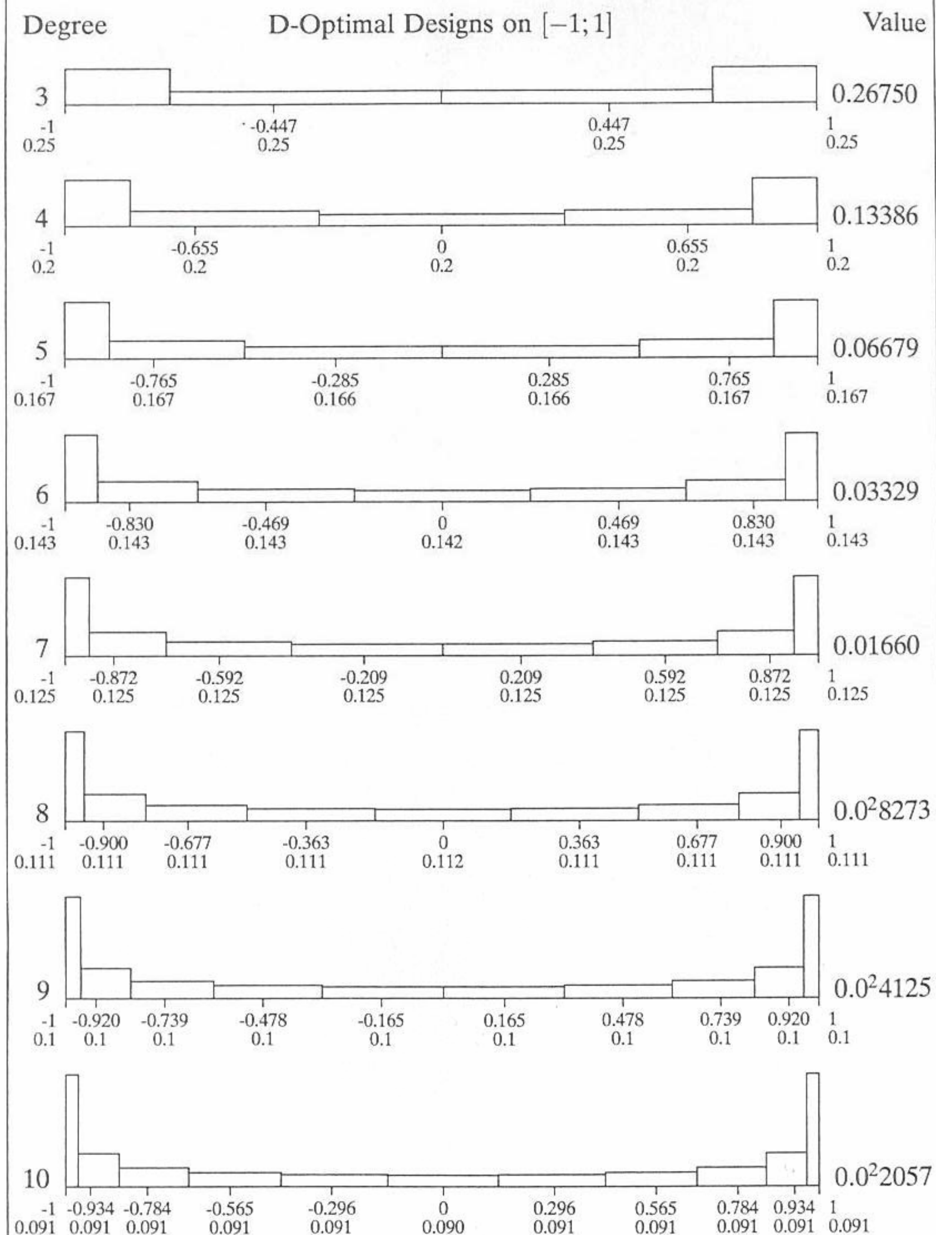
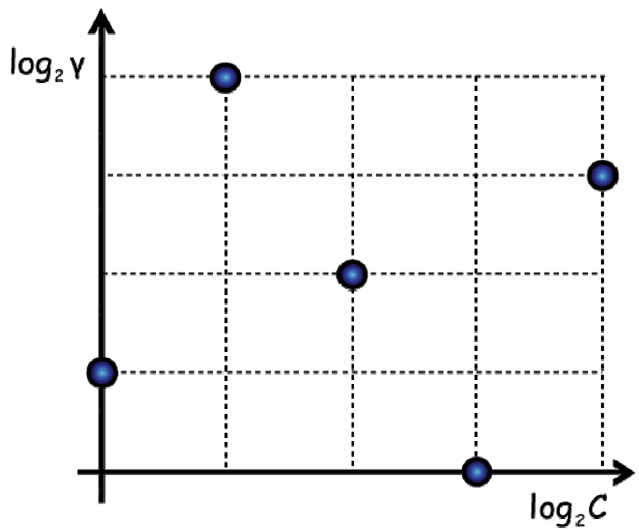


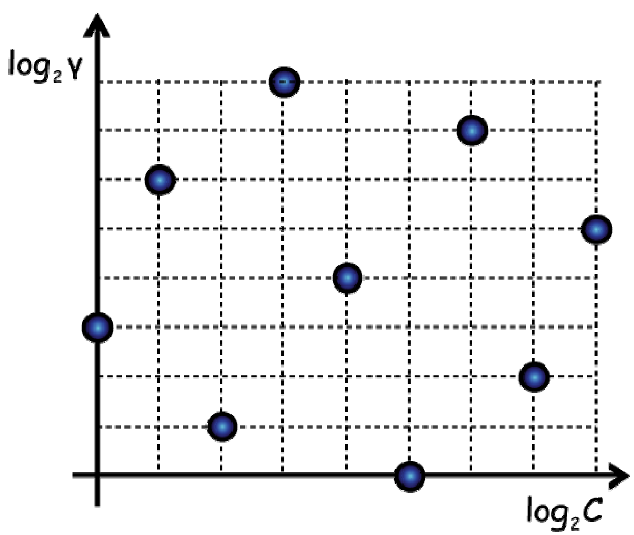
EXHIBIT 9.2 Polynomial fits over $[-1;1]$: ϕ_0-optimal designs τ_0^d for θ in T. Left: degree d of the fitted polynomial. Middle: support points and weights of τ_0^d, and a histogram representation. Right: optimal value $v_d(\phi_0)$ of the determinant criterion.

Did you know:
In many “practical” cases,
they are identical!!!

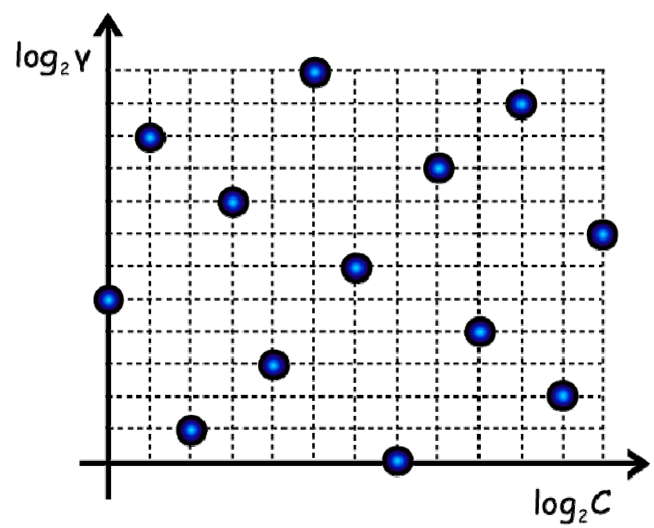
Uniform Designs



The 5 runs UD sampling pattern



The 9 runs UD sampling pattern



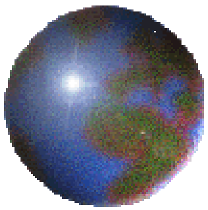
The 13 runs UD sampling pattern

Design of Experiment

- Model (M)
- Number of Factors (*k*)
- Number of runs (*n*)

Models

Statistics vs. Engineering Models



$$y = f(x, \theta) + \varepsilon$$

Statistical Model

$$y = \beta_0 + \Sigma\beta_i x_i + \Sigma\beta_{ij} x_i x_j + \varepsilon$$

A Typical Engineering Model (page 1 of 3, in Liao and Wang, 1995)

$$
\begin{aligned}
&\rho_b A_b \frac{\partial^2 w}{\partial t^2} + E_b I_b \frac{\partial^4 w}{\partial x^4} \\
&+\{(\rho_s A_s + \rho_c A_c)\frac{\partial^2 w}{\partial t^2} + \rho_s A_s\left(\frac{t_b+t_s}{2}\right)\left(\frac{\partial^3 u_b}{\partial x \partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^4 w}{\partial x^2 \partial t^2} + \frac{t_s}{2}\frac{\partial^3 \beta}{\partial x \partial t^2}\right) \\
&+\rho_c A_c a\left(\frac{\partial^3 u_b}{\partial x \partial t^2} - a\frac{\partial^4 w}{\partial x^2 \partial t^2} + t_s \frac{\partial^3 \beta}{\partial x \partial t^2}\right) + C_{11}^0 I_c \frac{\partial^4 w}{\partial x^4} - E_c A_c a\left(\frac{\partial^3 u_b}{\partial x^3} - a\frac{\partial^4 w}{\partial x^4} + t_s\frac{\partial^3 \beta}{\partial x^3}\right)\}[H(x-x_1) - H(x-x_2)] \\
&+\{\rho_s A_s\left(\frac{t_b+t_s}{2}\right)\left(\frac{\partial^2 u_b}{\partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^3 w}{\partial x \partial t^2} + \frac{t_s}{2}\frac{\partial^2 \beta}{\partial t^2}\right) + \rho_c A_c a\left(\frac{\partial^2 u_b}{\partial t^2} - a\frac{\partial^3 w}{\partial x \partial t^2} + t_s\frac{\partial^2 \beta}{\partial t^2}\right) \\
&+2C_{11}^0 I_c \frac{\partial^3 w}{\partial x^3} - 2E_c A_c a\left(\frac{\partial^2 u_b}{\partial x^2} - a\frac{\partial^3 w}{\partial x^3} + t_s\frac{\partial^2 \beta}{\partial x^2}\right)\}[\delta(x-x_1) - \delta(x-x_2)] \\
&+\{C_{11}^0 I_c \frac{\partial^2 w}{\partial x^2} - E_c A_c a\left(\frac{\partial u_b}{\partial x} - a\frac{\partial^2 w}{\partial x^2} + t_s\frac{\partial \beta}{\partial x}\right) + b d_{31} E_c a V(t)\}[\delta'(x-x_1) - \delta'(x-x_2)] = f(x,t)
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&\rho_b A_b \frac{\partial^2 u_b}{\partial t^2} - E_b A_b \frac{\partial^2 u_b}{\partial x^2} \\
&+\{\rho_s A_s\left(\frac{\partial^2 u_b}{\partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^3 w}{\partial x \partial t^2} + \frac{t_s}{2}\frac{\partial^2 \beta}{\partial t^2}\right) + \rho_c A_c\left(\frac{\partial^2 u_b}{\partial t^2} - a\frac{\partial^3 w}{\partial x \partial t^2} + t_s\frac{\partial^2 \beta}{\partial t^2}\right) \\
&-E_c A_c\left(\frac{\partial^2 u_b}{\partial x^2} - a\frac{\partial^3 w}{\partial x^3} + t_s\frac{\partial^2 \beta}{\partial x^2}\right)\}[H(x-x_1) - H(x-x_2)] \\
&+\{-E_c A_c\left(\frac{\partial u_b}{\partial x} - a\frac{\partial^2 w}{\partial x^2} + t_s\frac{\partial \beta}{\partial x}\right) + b d_{31} E_c V(t)\}[\delta(x-x_1) - \delta(x-x_2)] = 0
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
&[\rho_s A_s \frac{t_s}{2}\left(\frac{\partial^2 u_b}{\partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^3 w}{\partial x \partial t^2} + \frac{t_s}{2}\frac{\partial^2 \beta}{\partial t^2}\right) + \rho_c A_c t_s\left(\frac{\partial^2 u_b}{\partial t^2} - a\frac{\partial^3 w}{\partial x \partial t^2} + t_s\frac{\partial^2 \beta}{\partial t^2}\right) \\
&+A_s(G \cdot \beta) - E_c A_c t_s\left(\frac{\partial^2 u_b}{\partial x^2} - a\frac{\partial^3 w}{\partial x^3} + t_s\frac{\partial^2 \beta}{\partial x^2}\right)][H(x-x_1) - H(x-x_2)]
\end{aligned}
\tag{3}
$$

Number of Factors (k)

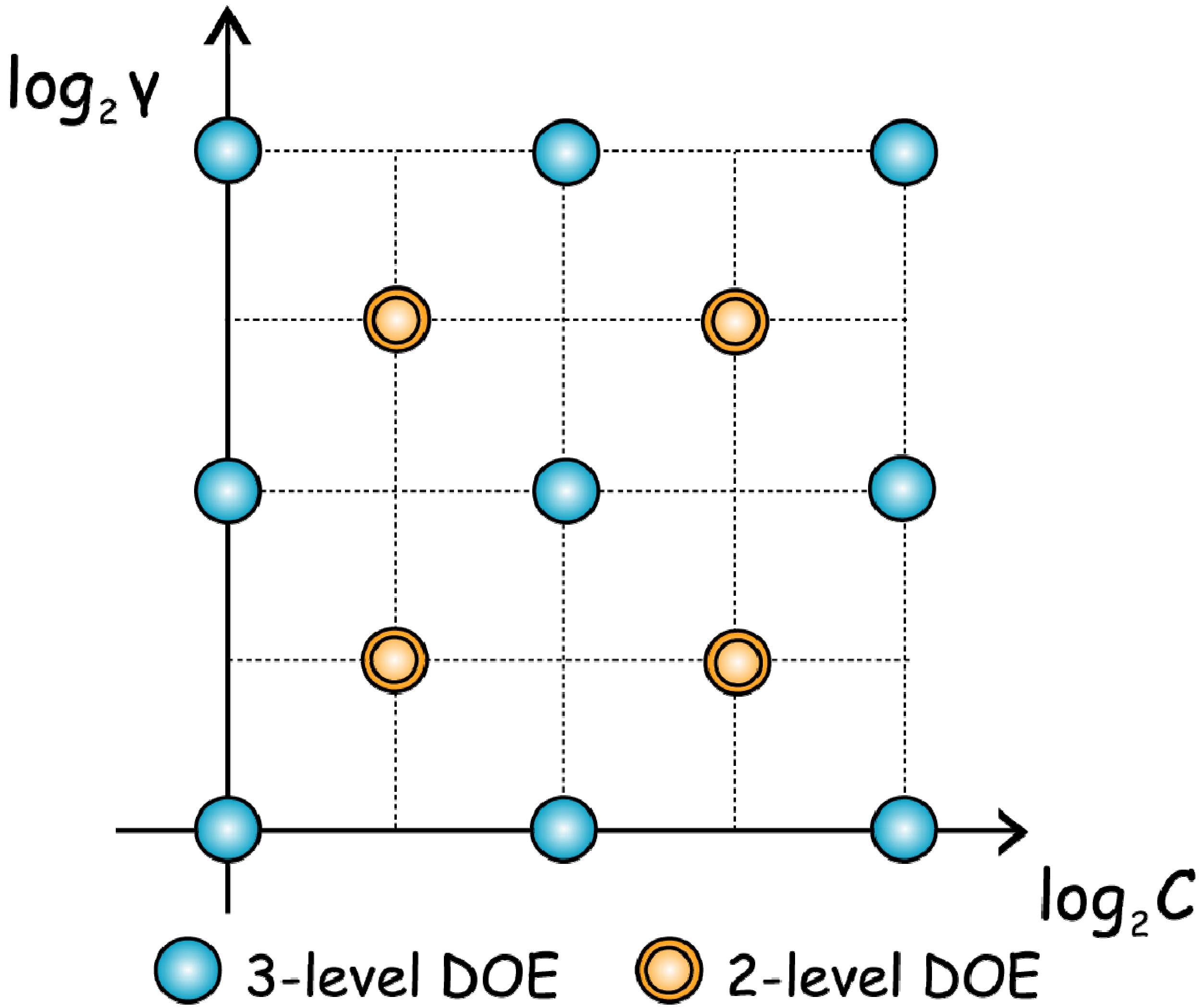
- Agricultural Experiment vs
  Industrial Experiment
- Screening Design vs
  High-Level (Full) Factorial Design
- Grouping Design
- Textbook k=3 to 5
  Practical k=20 to 200 (to some 6000+)

Number of Experimental Runs

What is the optimal number of runs (n)?

- As many as possible (*n*>30)
  - Typical Scientists (College of Science)
- $how me the Money
  - Project Manager (Business School)
- 1
  - Engineers—for confirmation only
- 0
  - Mathematical Statistician—We've proved it!

Mixed Level Designs



Summary: Design of Experiment

- ***Model is known***
  - Optimal design
  - Optimality Criteria
    (e.g., alphabetical optimalities)
- ***Model is unknown***

  (or is not completely known)
  - Bayesian Design
  - Robustness
  - Robustness Criterion
  - Representative Points

Optimal vs Robust

Some Minor Points

- Some solutions for Optimal Design may be involved with $i=\sqrt{-1}$
  which may not be easy to implement.

- *Rounding* may be sensitive for the optimality evaluation, although not much impact in practice
  —x=1.1234 vs x=1.1100.

DOE for Future Advanced Technology

(Partial List)

- Design for Quality (in FDA)
- Computer Experiment
  - Expensive simulation
  - Design with very large n
- MicroArray Experiment
  - Comments from Stan Young
- Split-Plot type Design
  - Hard-To-Change Variable etc.
- Nano-Technology Problem

How would optimal design fit here?

Where have all the Data gone?

- No need for data (Theoretical Development)
- Survey Sampling and Design of Experiment (Physical data collection)
- Computer Simulation (Experiment)
  - Statistical Simulation

    (Random Number Generation)
  - Engineering Simulation
- Data from Internet
  - On-line auction
  - Search Engine

Computer Experiments

- "Inexpensive" Computer Simulations
  - Large combinatorics designs are not too difficult to obtain
  - Large "optimal design may have some difficulties
- "Expensive" Computer Simulations
  - Smaller run size is preferable
  - Model is typically (very) complicated → known, but kind of unknown
- In the middle ground
  - Model is known → optimal design
  - Model is unknown → combinatorics design

A Typical Engineering Model (page 1 of 3, in Liao and Wang, 1995)

$$
\begin{aligned}
&\rho_b A_b \frac{\partial^2 w}{\partial t^2} + E_b I_b \frac{\partial^4 w}{\partial x^4} \\
&+\{(\rho_s A_s + \rho_c A_c)\frac{\partial^2 w}{\partial t^2} + \rho_s A_s\left(\frac{t_b+t_s}{2}\right)\left(\frac{\partial^3 u_b}{\partial x \partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^4 w}{\partial x^2 \partial t^2} + \frac{t_s}{2}\frac{\partial^3 \beta}{\partial x \partial t^2}\right) \\
&+\rho_c A_c a\left(\frac{\partial^3 u_b}{\partial x \partial t^2} - a\frac{\partial^4 w}{\partial x^2 \partial t^2} + t_s\frac{\partial^3 \beta}{\partial x \partial t^2}\right) + C_{11}^0 I_c \frac{\partial^4 w}{\partial x^4} - E_c A_c a\left(\frac{\partial^3 u_b}{\partial x^3} - a\frac{\partial^4 w}{\partial x^4} + t_s\frac{\partial^3 \beta}{\partial x^3}\right)\}[H(x-x_1) - H(x-x_2)] \\
&+\{\rho_s A_s\left(\frac{t_b+t_s}{2}\right)\left(\frac{\partial^2 u_b}{\partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^3 w}{\partial x \partial t^2} + \frac{t_s}{2}\frac{\partial^2 \beta}{\partial t^2}\right) + \rho_c A_c a\left(\frac{\partial^2 u_b}{\partial t^2} - a\frac{\partial^3 w}{\partial x \partial t^2} + t_s\frac{\partial^2 \beta}{\partial t^2}\right) \\
&+2C_{11}^0 I_c \frac{\partial^3 w}{\partial x^3} - 2E_c A_c a\left(\frac{\partial^2 u_b}{\partial x^2} - a\frac{\partial^3 w}{\partial x^3} + t_s\frac{\partial^2 \beta}{\partial x^2}\right)\}[\delta(x-x_1) - \delta(x-x_2)] \\
&+\{C_{11}^0 I_c \frac{\partial^2 w}{\partial x^2} - E_c A_c a\left(\frac{\partial u_b}{\partial x} - a\frac{\partial^2 w}{\partial x^2} + t_s\frac{\partial \beta}{\partial x}\right) + b d_{31} E_c a V(t)\}[\delta'(x-x_1) - \delta'(x-x_2)] = f(x,t)
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
&\rho_b A_b \frac{\partial^2 u_b}{\partial t^2} - E_b A_b \frac{\partial^2 u_b}{\partial x^2} \\
&+\{\rho_s A_s\left(\frac{\partial^2 u_b}{\partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^3 w}{\partial x \partial t^2} + \frac{t_s}{2}\frac{\partial^2 \beta}{\partial t^2}\right) + \rho_c A_c\left(\frac{\partial^2 u_b}{\partial t^2} - a\frac{\partial^3 w}{\partial x \partial t^2} + t_s\frac{\partial^2 \beta}{\partial t^2}\right) \\
&-E_c A_c\left(\frac{\partial^2 u_b}{\partial x^2} - a\frac{\partial^3 w}{\partial x^3} + t_s\frac{\partial^2 \beta}{\partial x^2}\right)\}[H(x-x_1) - H(x-x_2)] \\
&+\{-E_c A_c\left(\frac{\partial u_b}{\partial x} - a\frac{\partial^2 w}{\partial x^2} + t_s\frac{\partial \beta}{\partial x}\right) + b d_{31} E_c V(t)\}[\delta(x-x_1) - \delta(x-x_2)] = 0
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
&[\rho_s A_s \frac{t_s}{2}\left(\frac{\partial^2 u_b}{\partial t^2} - \frac{t_b+t_s}{2}\frac{\partial^3 w}{\partial x \partial t^2} + \frac{t_s}{2}\frac{\partial^2 \beta}{\partial t^2}\right) + \rho_c A_c t_s\left(\frac{\partial^2 u_b}{\partial t^2} - a\frac{\partial^3 w}{\partial x \partial t^2} + t_s\frac{\partial^2 \beta}{\partial t^2}\right) \\
&+A_s(G * \beta) - E_c A_c t_s\left(\frac{\partial^2 u_b}{\partial x^2} - a\frac{\partial^3 w}{\partial x^3} + t_s\frac{\partial^2 \beta}{\partial x^2}\right)][H(x-x_1) - H(x-x_2)]
\end{aligned}
\tag{3}
$$

Design of Experiment in Marketing and Social Studies:

where the “model” is not in mathematical functional form!!!
(Have a fun with HR people lately?)

Genetic Studies

- Model is completed unknown
- Classical (agricultural) designs are popular here—these are typically combinatorics designs.

Micro-Array Design (SS Young)

- How do you pick the sequences that you are going to use.
  - They have to be unique vs other sub-sequences and they have to work at common temperature/chemical fluid concentrations.
- Placement of sub sequences on the chip and the number of rep spots.
- A no-no is possible reuse of chips. How to wash would be a good DOE problem.
  - Technically I think the chips can be reused, but contract requires only one use.
- There is very large variation in response among the sub-sequences within a gene. Does this relate to the sequence in some way?
  - If so, it would be of interest to try to figure out the factors influencing the among sub-sequence variation. So the question would be DOE on the selection of the subsequence and then DOE on the assay conditions.

Type of Designs

- Combinatorics Design
  - (Fractional) Factorial
  - Orthogonal Arrays
  - Latin Hypercube
  - Uniform Design
- Optimal Design (Bayesian Design)

- Middle Ground?
  - Chen, Lin and Tsai (2005)

Common Ground:

Conditional Optimal Design
by Chen, Lin & Tsai (2005)

- Use combinatroric design at the first stage (typically for screening and Steepest ascent)—model is not clear.
- When curvature is found and the second-order model is called for—model is known, use optimal design for additional experiments.
- Life is beautiful!!!

Practical & Theoretical

- Practical: We know how to make it work, but we don't know why.
- Theoretical: We know exactly why, but we don't know how to make it work.

- Now with Practical and Theoretical combined…

  It doesn't work and we don't know why!

Some Optimal Designs

(and some not-so-optimal design)

Conclusion

- Every design has its own properties and values.
- There is more than one way for success.
- Those who can tell good advice from bad advice probably don’t need any advice!

  For others, cook book is helpful.

STILL QUESTION?

Send $500 to

- Dennis Lin

  University Distinguished Professor

  483 Business Building

  Department of Supply Chain & Information Systems

  Penn State University

- +1 814 865-0377 (phone)
- +1 814 863-7076 (fax)
- DKL5@psu.edu

(Customer Satisfaction or your money back!)