Benefits and Fast Construction of Efficient Two-Level Foldover Designs

Anna Errore
Bradley Jones
William Li
Christopher J. Nachtsheim

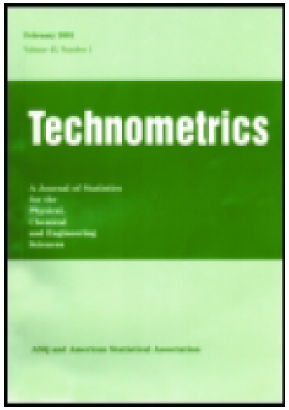
Brad

Anna

William

Chris

Technometrics paper



Taylor & Francis
Taylor & Francis Group

Technometrics

In print January 2017 pp. 48-57

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://www.tandfonline.com/loi/utch20

Benefits and Fast Construction of Efficient Two-Level Foldover Designs

Anna Errore, Bradley Jones, William Li & Christopher J. Nachtsheim

To cite this article: Anna Errore, Bradley Jones, William Li & Christopher J. Nachtsheim (2015): Benefits and Fast Construction of Efficient Two-Level Foldover Designs, Technometrics, DOI: 10.1080/00401706.2015.1124052

To link to this article: http://dx.doi.org/10.1080/00401706.2015.1124052

Outline

- Intro to screening experiments
- Motivation
- Construction algorithm
- EFD evaluation and comparison
- A compound approach
- Discussion
- Conclusions

Screening

Screening experiments

Starting with little prior knowledge and an initial set of potential factors influencing the response

Purpose to identify the smaller set of active factors.

Primary goal : identify active main effects (MEs)

Secondary goal: identify a few active two-factor interactions (2FIs) if possible.

Classical choices for screening

- Resolution III fractional factorial designs (orthogonal ME plans)
- Plackett-Burman designs (nonregular orthogonal ME plans)

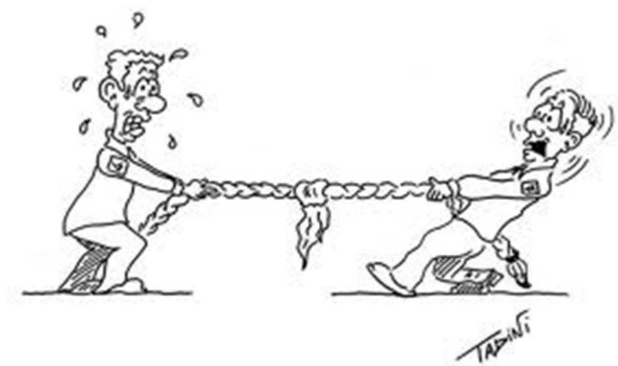
What if there are active 2FIs???
What if we want to estimate them???

The identification of active MEs is the primary goal

Easy to do if there are no active 2FIs.

Desirable design features

- orthogonality of the MEs,
- orthogonality of MEs and 2FIs,
- orthogonality of 2FIs with each other,
- small run size



Motivation

Recent developments in literature:

- Definitive screening designs

  Jones & Nachtsheim (2011, 2013, 2015)

  Foldover plans for three level factors or mixed 2 and 3-level factors.

- Folded-Over Non-orthogonal Designs

  Miller & Sitter (2001, 2005), Lin, Miller & Sitter (2008)

  Advocate the use of non-orthogonal designs for screening

Alan Miller

Devon Lin

Randy Sitter

Research objectives

1. Give more support for the use of non-orthogonal foldover designs,
2. Find a fast algorithm for constructing efficient foldover designs,
3. Expand on the class of small, two-level foldover designs,
4. Develop a compromise algorithm making a trade-off between efficiency of the MEs estimates and correlation of the 2FIs.

Notation

m factors, *n* runs

Linear main effect model (ME)

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i \qquad i = 1, \ldots, n$$

Main effects plus 2FIs model (ME + 2FIs)

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{ij} + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \beta_{jk} x_{ij} x_{ik} + \varepsilon_i \qquad i = 1, \ldots, n.$$

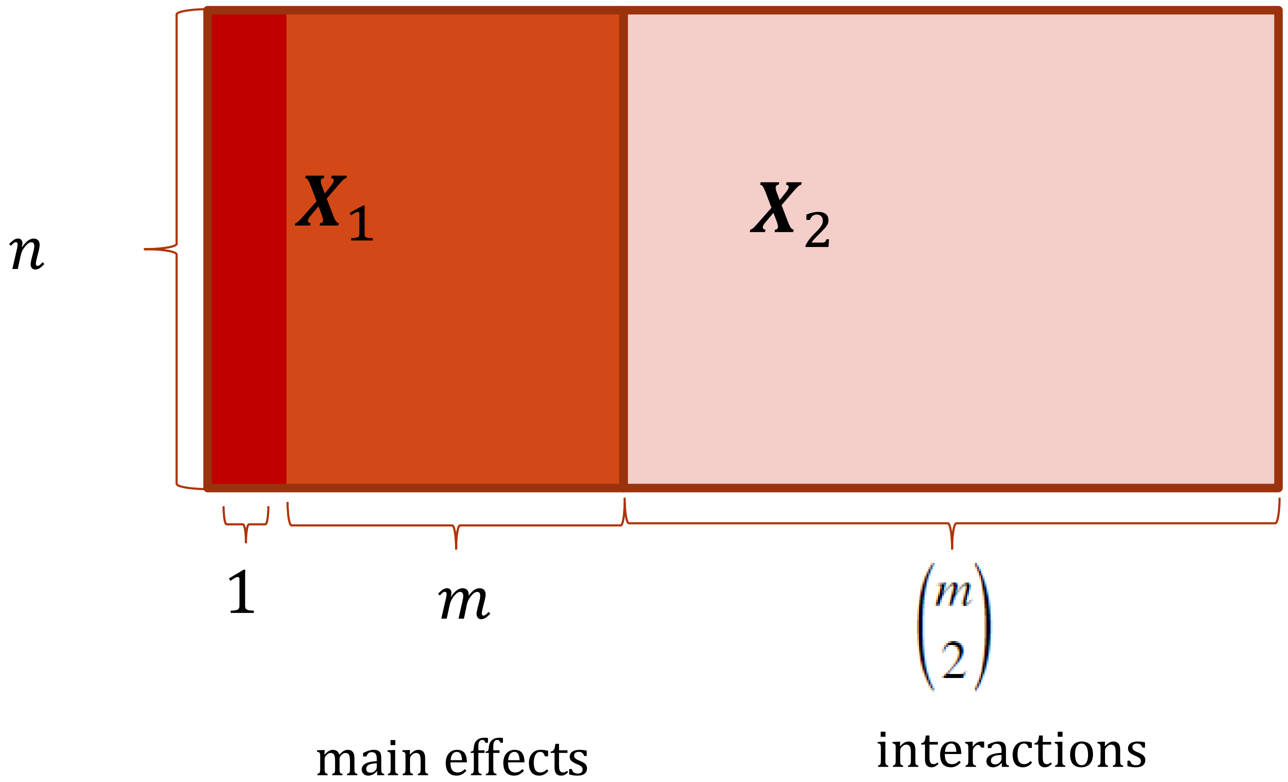
d *nxm* design matrix

$$\mathbf{d} = \begin{bmatrix} x_{11} & \ldots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \ldots & x_{nm} \end{bmatrix},$$

Note that typical screening designs are supersaturated in the ME+2FIs model

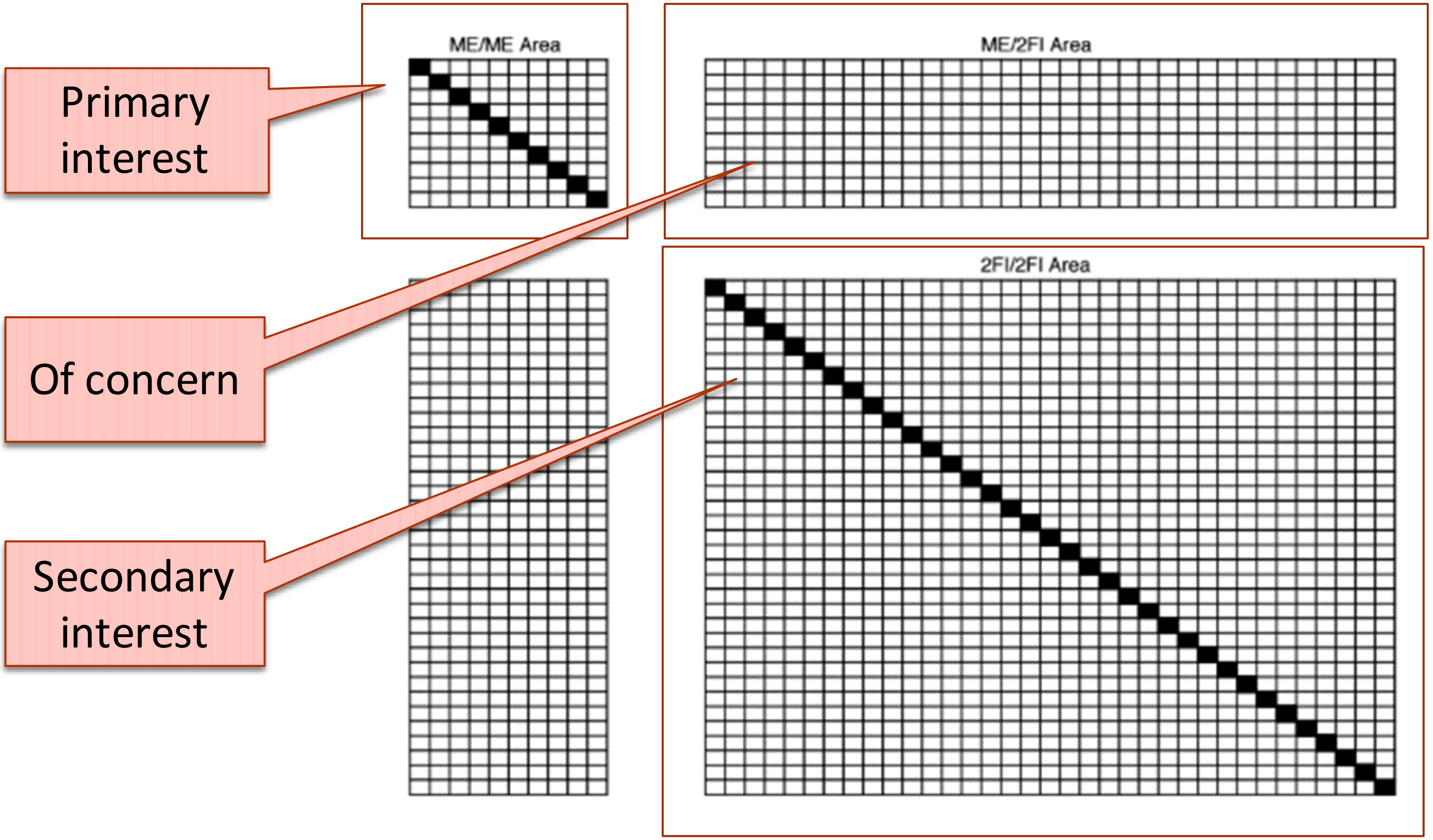
Number of terms by type

m factors *n* runs



Helpful diagnostic plot

correlation cell plot



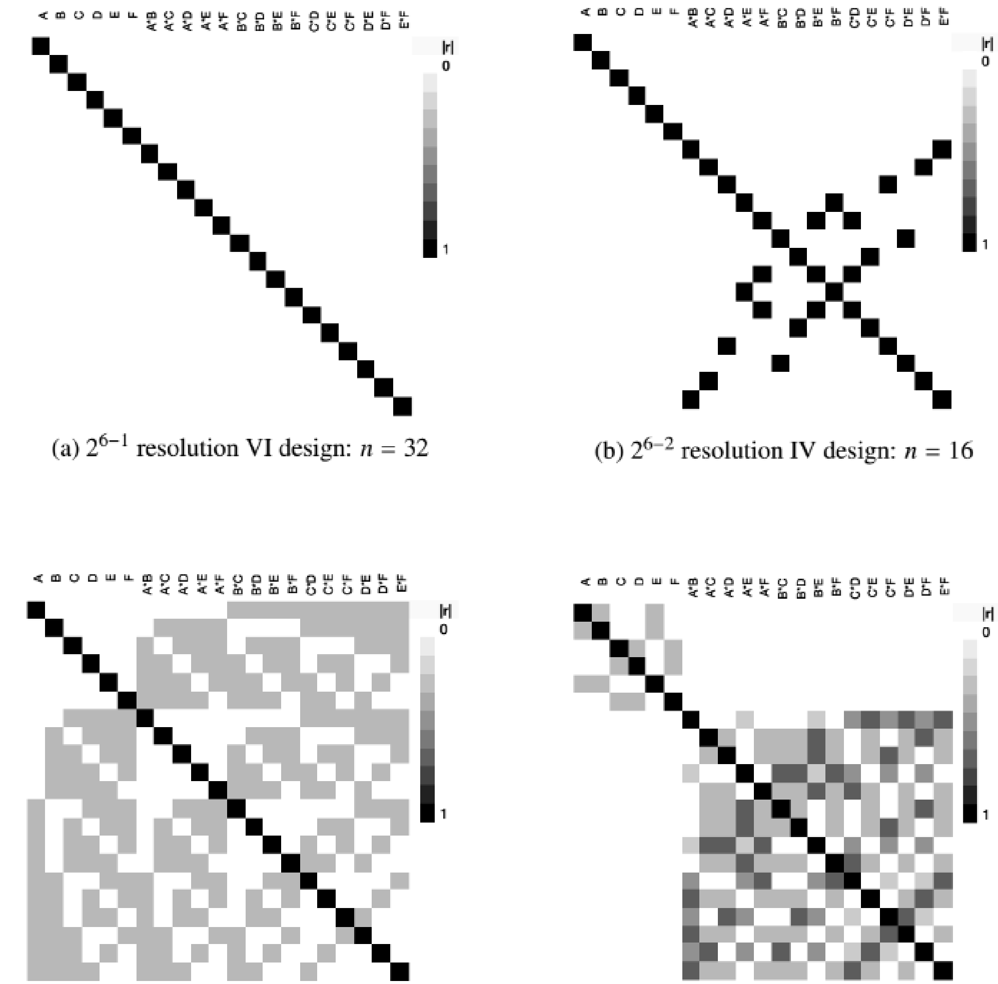
Design comparisons

(a) Resolution VI FF

(b) Resolution IV FF

(c) Placket-Burgman

(d) Efficient Foldover design



(a) 2^{6-1} resolution VI design: $n = 32$

(b) 2^{6-2} resolution IV design: $n = 16$

(c) Plackett-Burman design: $m = 6$ and $n = 12$

(d) EFD: $m = 6$ and $n = 12$

Aliased estimation of MEs

MEs model for OLS estimation: $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}^*$

If there are active 2FIs $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$

Biased estimators: $\mathrm{E}(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2$

Alias matrix: $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$

It would be nice to have
unbiased estimators…

Jones, B., & Nachtsheim, C. J. (2011a). Efficient Designs With Minimal Aliasing. *Technometrics*, *53*(1), 62-71.

Foldover design structure

1. Half design matrix: $\mathbf{X}_M$ $n \times m$ matrix

2. Model matrix for the main effects model $\mathbf{X}_1 = \begin{bmatrix} 1 & \mathbf{X}_M \\ 1 & -\mathbf{X}_M \end{bmatrix}$

3. Coding $x_{ij} = \pm 1$

Foldover of weighing designs

$2^m//2m$

Minimal-run efficient non-orthogonal designs with no bias between MEs and 2FIs (Margolin, 1969)

Foldover of weighing designs

m=5

+	+	+	+	-
+	+	+	-	+
+	+	-	+	+
+	-	+	+	+
-	+	+	+	+

$2^5//10$

Non-orthogonal

+	+	+	+	-
+	+	+	-	+
+	+	-	+	+
+	-	+	+	+
-	+	+	+	+
-	-	-	-	+
-	-	-	+	-
-	-	+	-	-
-	+	-	-	-
+	-	-	-	-

Foldover of weighing designs

$2^m//2m$

Miller & Sitter (2005) advocate the use of these designs for model-robustness even if non-orthogonal

Non-orthogonal

Foldover of weighing designs

m=5

+	+	+	+	-
+	+	+	-	+
+	+	-	+	+
+	-	+	+	+
-	+	+	+	+

→ $2^5//10$

+	+	+	+	-
+	+	+	-	+
+	+	-	+	+
+	-	+	+	+
-	+	+	+	+
-	-	-	-	+
-	-	-	+	-
-	-	+	-	-
-	+	-	-	-
+	-	-	-	-

Non-orthogonal foldover designs

Lin, Miller, and Sitter (2008)

identify useful non-isomorphic designs for

Number of factors *m* = 4 -12

Number of runs *n* = *2m*

Research questions

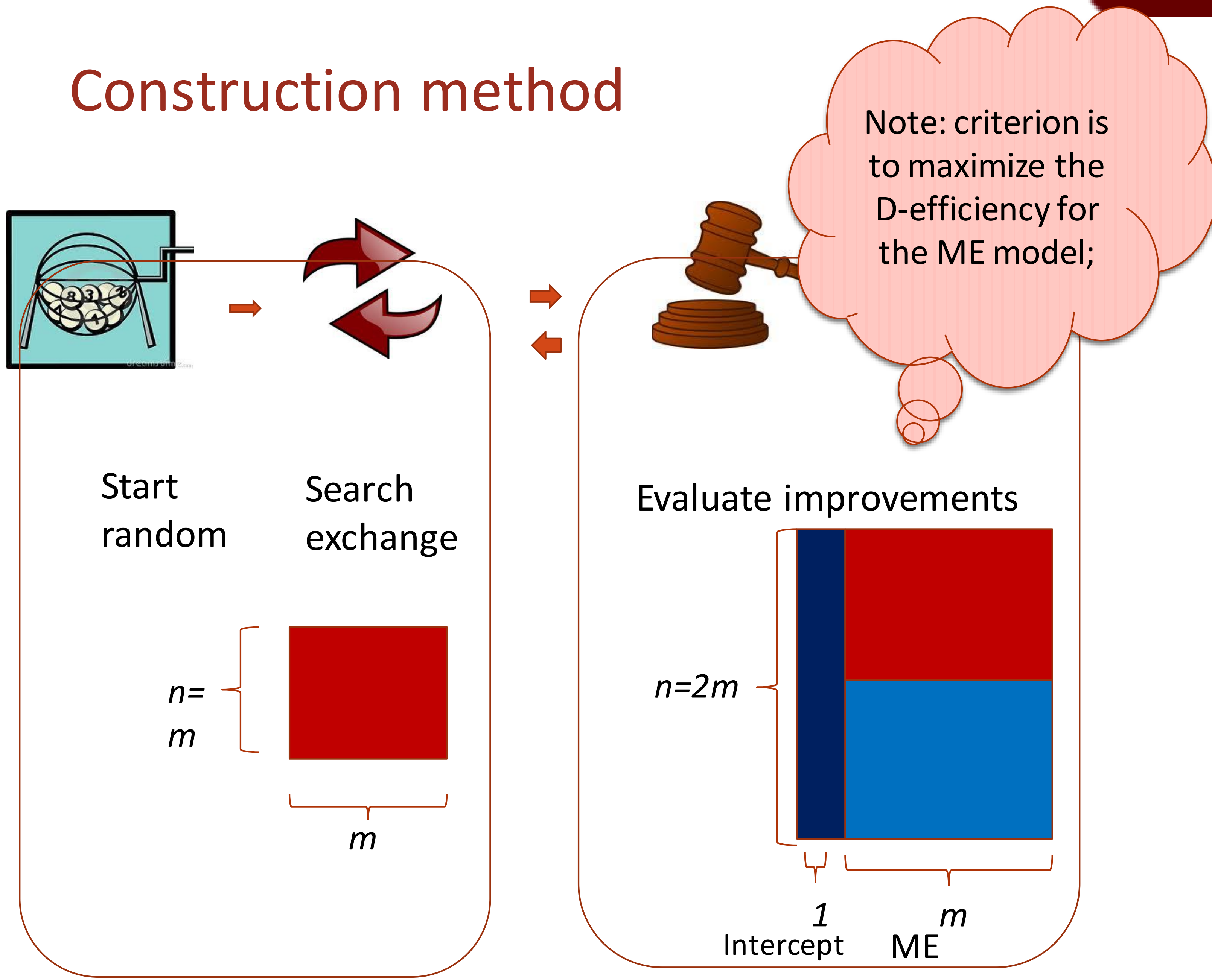
Can we expand upon the $2^m//2m$ class of designs so that we can find a larger class where:

$\mathbf{X}_M$ is not restricted to a known design type (in Margolin it is a weighing design) ?

n is not restricted (in Margolin it is *2m*) ?

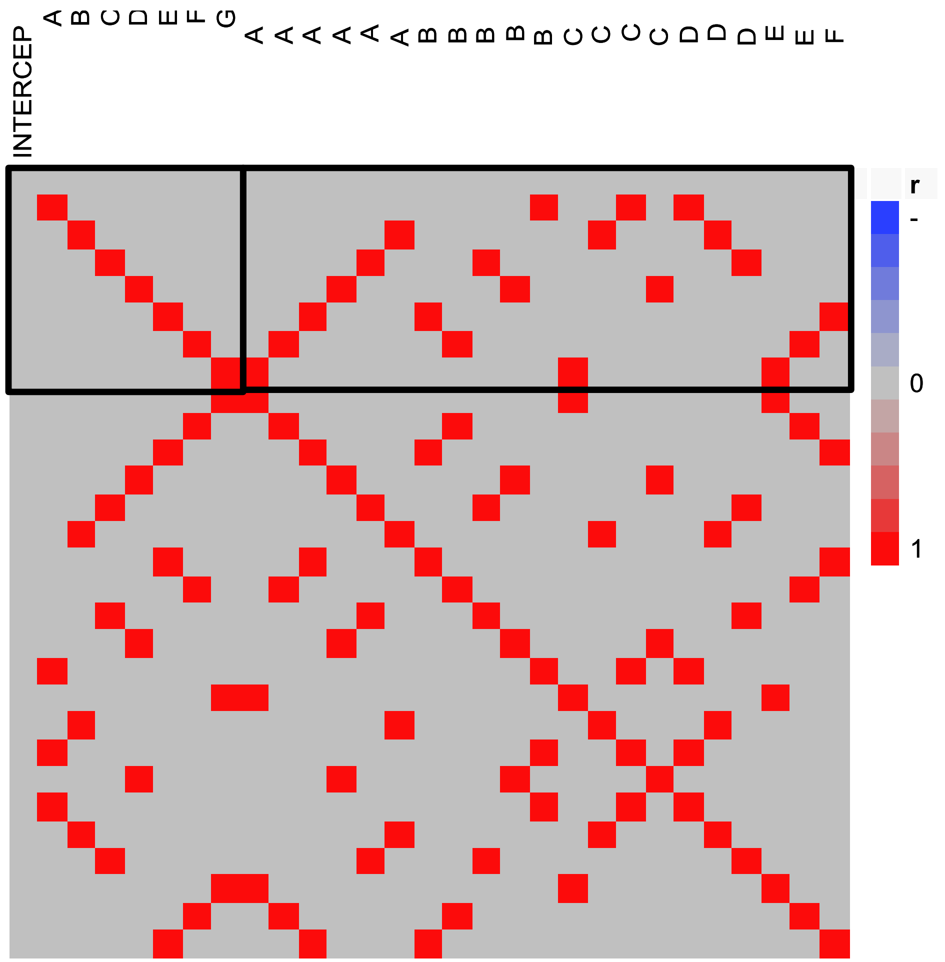
For $2^m//2m$ can we improve the existing results ?

Construction method



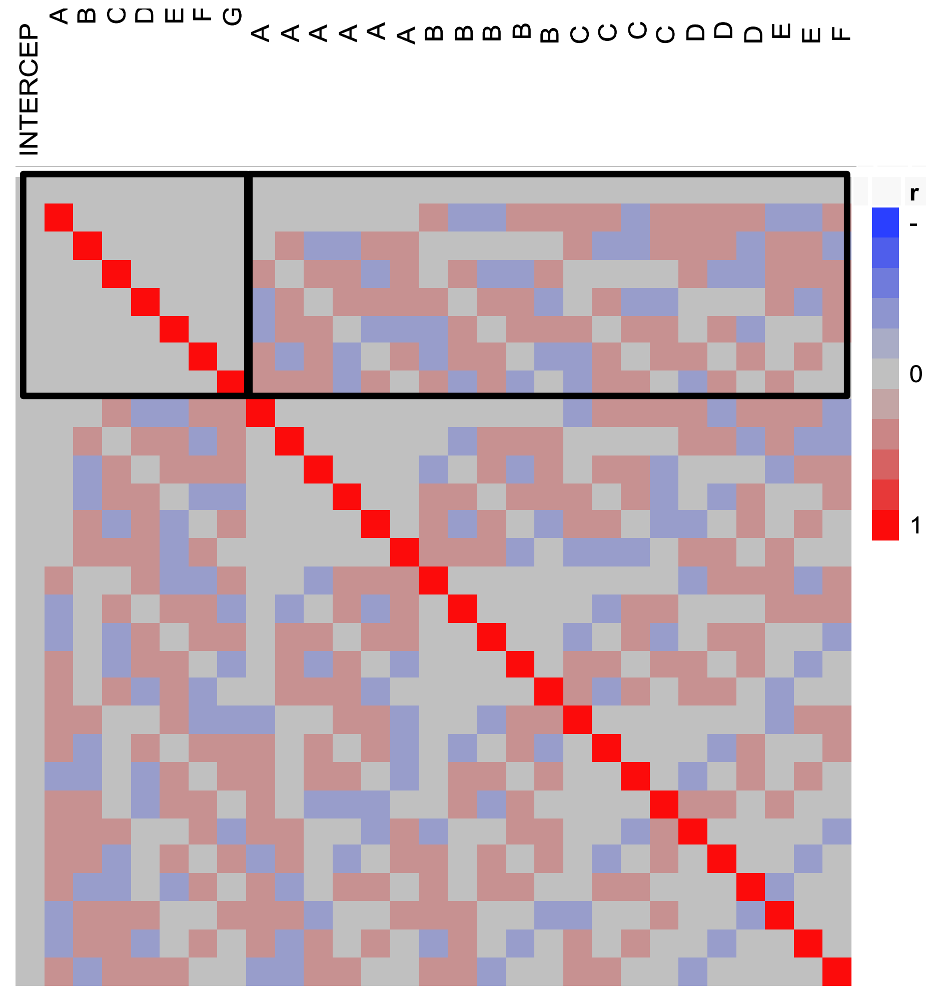
Scenario – 7 Factor Screening Example 1

8 run Fractional Factorial



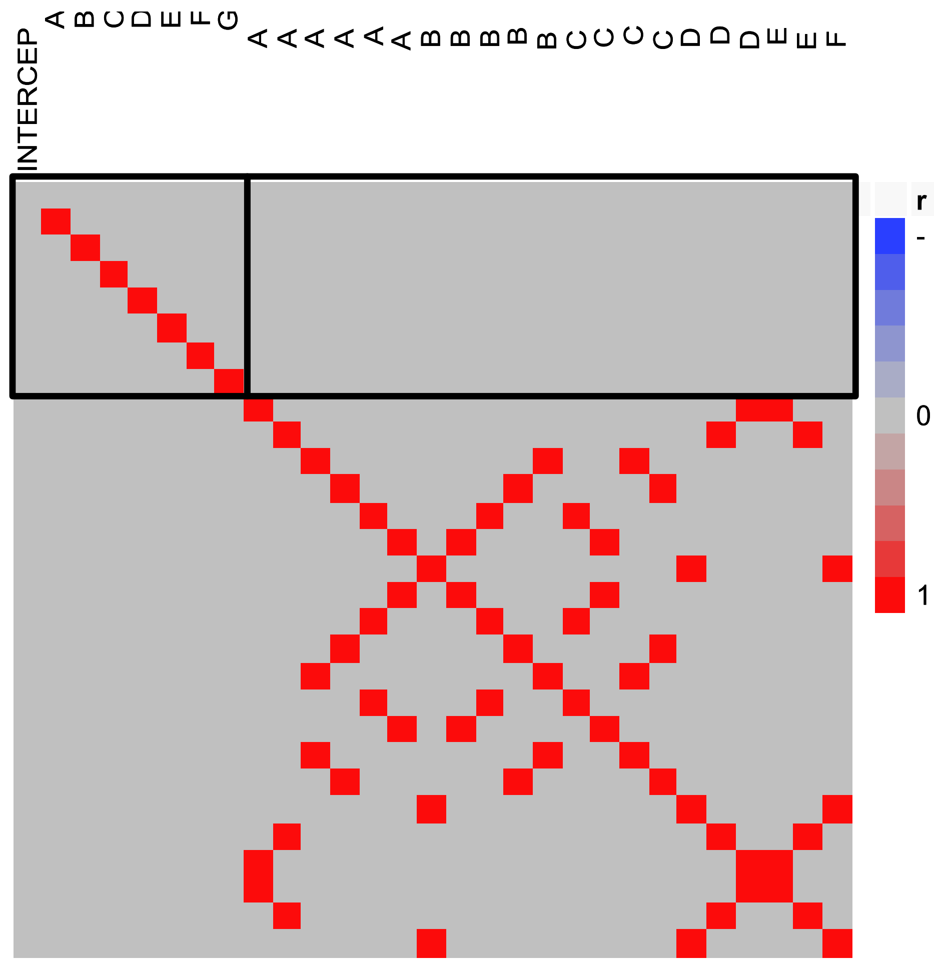
Scenario – 7 Factor Screening Example 2

12 run Plackett Burman



Scenario – 7 Factor Screening Example 3

16 run Fractional Factorial



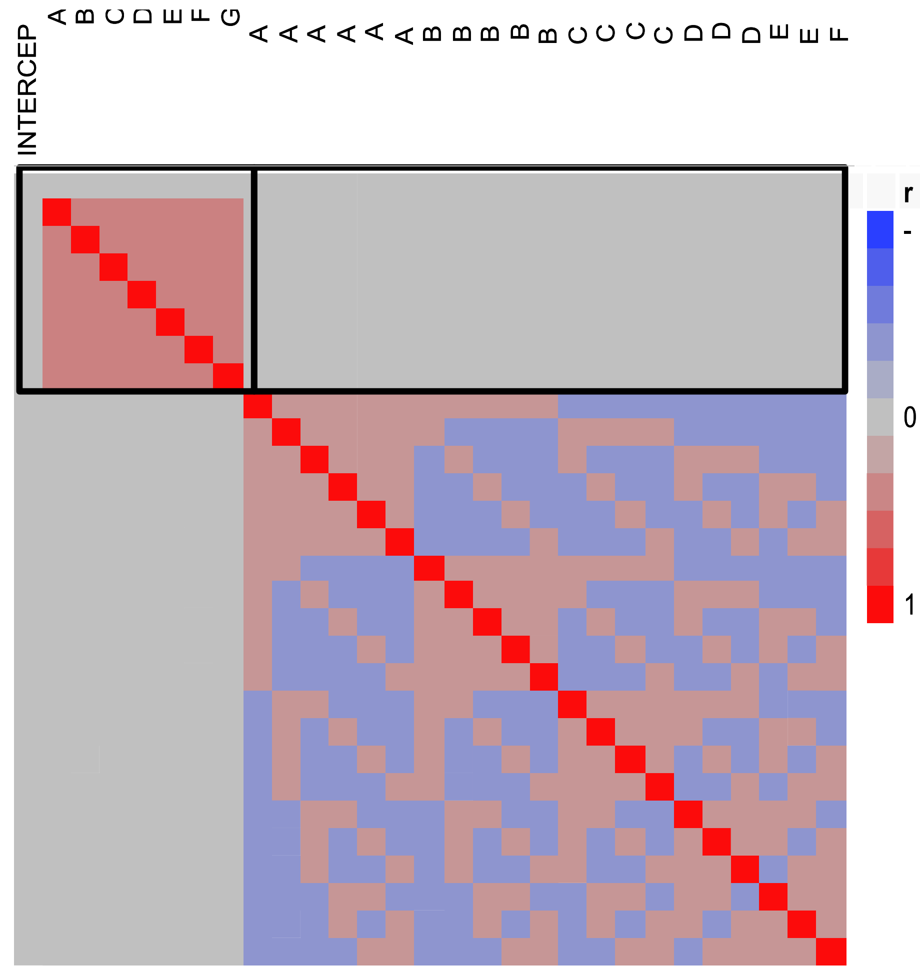
Scenario – 7 Factor Screening Example 4

$2^7//14$

Margolin (1969)

Foldover of

weighing design



ME D-efficiency = 79.12 %

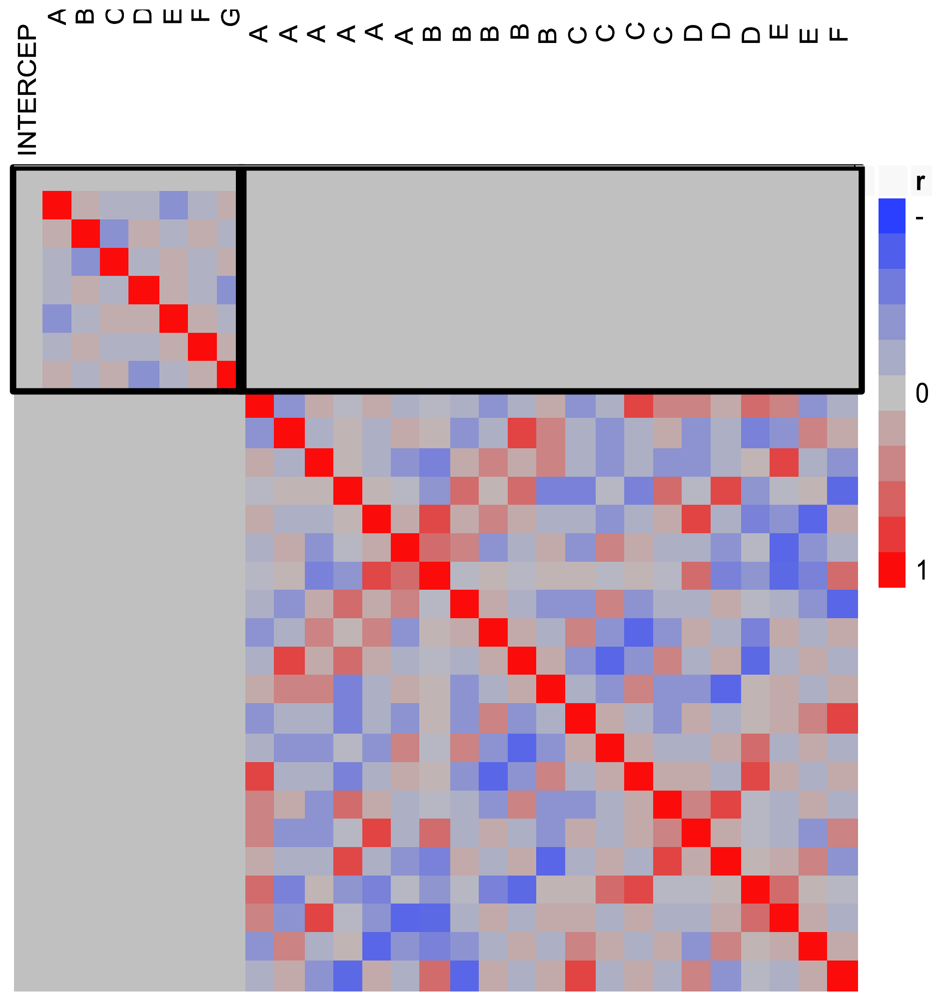
ME $|\rho|$ = 3/7

Scenario – 7 Factor Screening Example 5

14 run EFD

D-efficiency = 91.65%

ME average $|\rho|$ = 0.1837



Goal 2: Eliminate fully aliased 2FIs

Goal 1: find active MEs, unbiased by any active 2FIs

Goal 2: estimate a few of non-negligible 2FIs without ambiguity

Necessary to eliminate identical 2FI columns

Maximizing ME D-efficiency does not guarantee this.

Compromise designs

Compound optimization approach

$$C_w = w * C_1^s + (1 - w) * C_2^s$$

Maximize D-eff in the MEs model

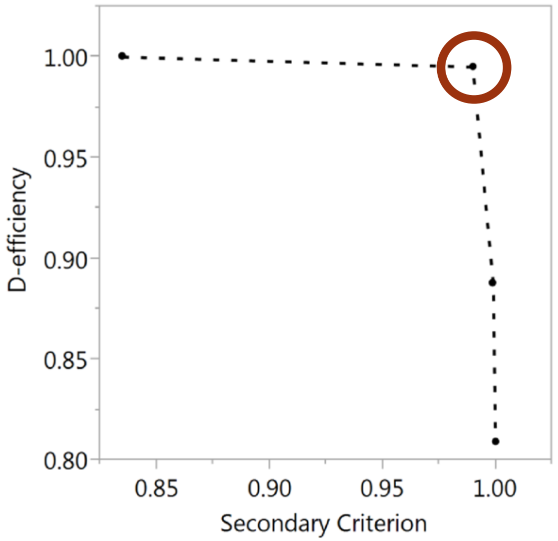
$$C_1 = \frac{|\mathbf{X}_1'\mathbf{X}_1|^{\frac{1}{m+1}}}{n}$$

Maximize D-efficiency

$$C_2 = \left(\sum_{k=1}^{g} |c_k|^r \right)^{\frac{-1}{r}}$$

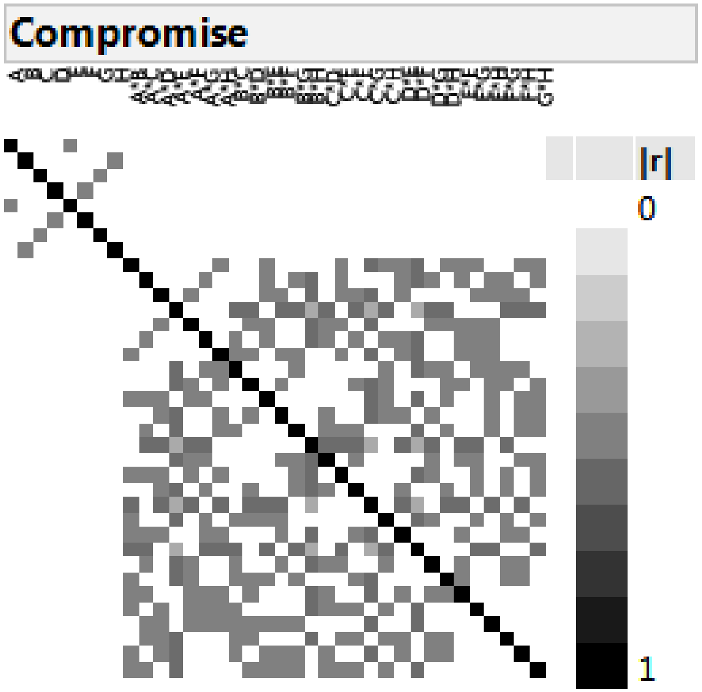
Penalize full confounding between 2FIs

Huge penalty for large |ρ|



Pareto frontier of criterion values for non-dominated designs for m = 9 and n = 22

8 Factor 16 Run Example



Maximum|r| = 0.5774

For 3 circled dark cells |r| = 1

JMP Demonstration – Analysis and Empirical Power

Fit EFD - true model terms A, B, C, D, AB, AC, BC

Stage 1 - Main Effect Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
A	2.0163	0.1689	11.936	0.0003*
B	-1.669	0.1689	-9.879	0.0006*
C	-1.121	0.1689	-6.638	0.0027*
D	2.4563	0.1689	14.54	0.0001*

Statistic	Value
RMSE	0.6757
DF	4

☑ Quadratic Terms Obey Strong Heredity

☑ Interactions Obey Strong Heredity

Stage 2 - Even Order Effect Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	50.05	0.1354	369.75	<.0001*
A*B	-2.913	0.1354	-21.52	0.0002*
A*C	-2.003	0.1354	-14.79	0.0007*
B*C	3.8575	0.1354	28.498	<.0001*
B*D	0.315	0.1354	2.3271	0.1024

Statistic	Value
RMSE	0.5414
DF	3

Combined Model Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	50.05	0.1554	322.01	<.0001*
A	2.0163	0.1554	12.972	<.0001*
B	-1.669	0.1554	-10.74	<.0001*
C	-1.121	0.1554	-7.214	0.0002*
D	2.4563	0.1554	15.803	<.0001*
A*B	-2.913	0.1554	-18.74	<.0001*
A*C	-2.003	0.1554	-12.88	<.0001*
B*C	3.8575	0.1554	24.818	<.0001*
B*D	0.315	0.1554	2.0266	0.0823

Statistic	Value
RMSE	0.6217
DF	7

Trade off

Lost orthogonality for MEs

In exchange for orthogonal MEs and 2FIs

Cost:

- small loss in power to identify the active MEs when no 2FIs are active
- wider CI for parameter estimates

Assuming an orthogonal design for the same number of runs exists.

m	n	Upper Bound On Fractional Increase in Maximum Width of the Confidence Intervals for Main Effects
5	10	0.05
	12	0.09
	14	0.08
6	12	0.10
	14	0.15
7	14	0.14
9	18	0.21
	20	0.05
	22	0.07
10	20	0.05
	22	0.07
11	22	0.10
13	26	0.02
	28	0.04
	30	0.06

Conclusions…

1. DSDs are excellent choice when all factors are continuous
2. For two level factors standard choices are small orthogonal plans in which MEs are fully or partially aliased by 2FIs
3. Resolution IV designs require larger sample size and 2FIs are fully confounded with each other…
4. Higher resolution designs are generally too costly for screening
5. EFDs are preferable if most or all factors are categorical