# Design of Experiments: A gentle introduction

## — $2^K$ FACTORIAL DESIGN —

Fabio Stella
University of Milano-Bicocca
fabio.stella@unimib.it

# LECTURE LEARNING OBJECTIVES

1) $2^2$ and $2^3$ designs
2) **Residuals** and **model adequacy**
3) **Regression model** and **response surface**
4) **Dispersion effect**
5) **Center points**
6) **First-order** and **second-order models**

# $2^K$ Factorial Design: Basic definitions and principles

In the previous lecture we gave basics of factorial design, while several special cases of the general factorial design are important because they are widely used in research work and also because they form the basis of other designs of considerable practical value.

The most important of these special cases is that of $k$ factors, each at only two levels. These levels may be

- **quantitative**, such as two values of temperature, pressure, or time;
- **qualitative**, such as two machines, two operators, the “high” and “low” levels of a factor, or perhaps the presence and absence of a factor.

A complete replicate of such a design requires $2 \times 2 \times \cdots \times 2 = 2^k$ observations (**$\mathbf{2^k}$ factorial design)**.

The $\mathbf{2^k}$ design is particularly **useful** in the **early stages of experimental work** when **many factors** are likely to be **investigated**.

It provides the **smallest number of runs** with which $\boldsymbol{k}$ **factors** can be **studied** in a complete factorial design. Consequently, these designs are widely used in **factor screening experiments**.

Because there are **only two levels for each factor**, we assume that **the response is approximately linear** over the range of the factor levels chosen. In many factor screening experiments, when we are just **starting to study the process or the system**, this is often a **reasonable assumption**. We will present a **simple method for checking this assumption** and discuss what **action to take if it is violated**.

**Two factors**; (**A**) *reactant concentration*, (**B**) *amount of catalyst*.

**Outcome**; *yield* of a chemical process.

**Goal**; determine if adjustments to either of these two factors would increase the yield.

Factors Levels;

| | | |
|---|---|---|
| (**A**) *reactant concentration,* | Low (15%) | High (25%) |
| (**B**) *amount of catalyst*, | Low (1 pound) | High (2 pounds) |

The experiment is **replicated three times**, so there are **12 runs**.

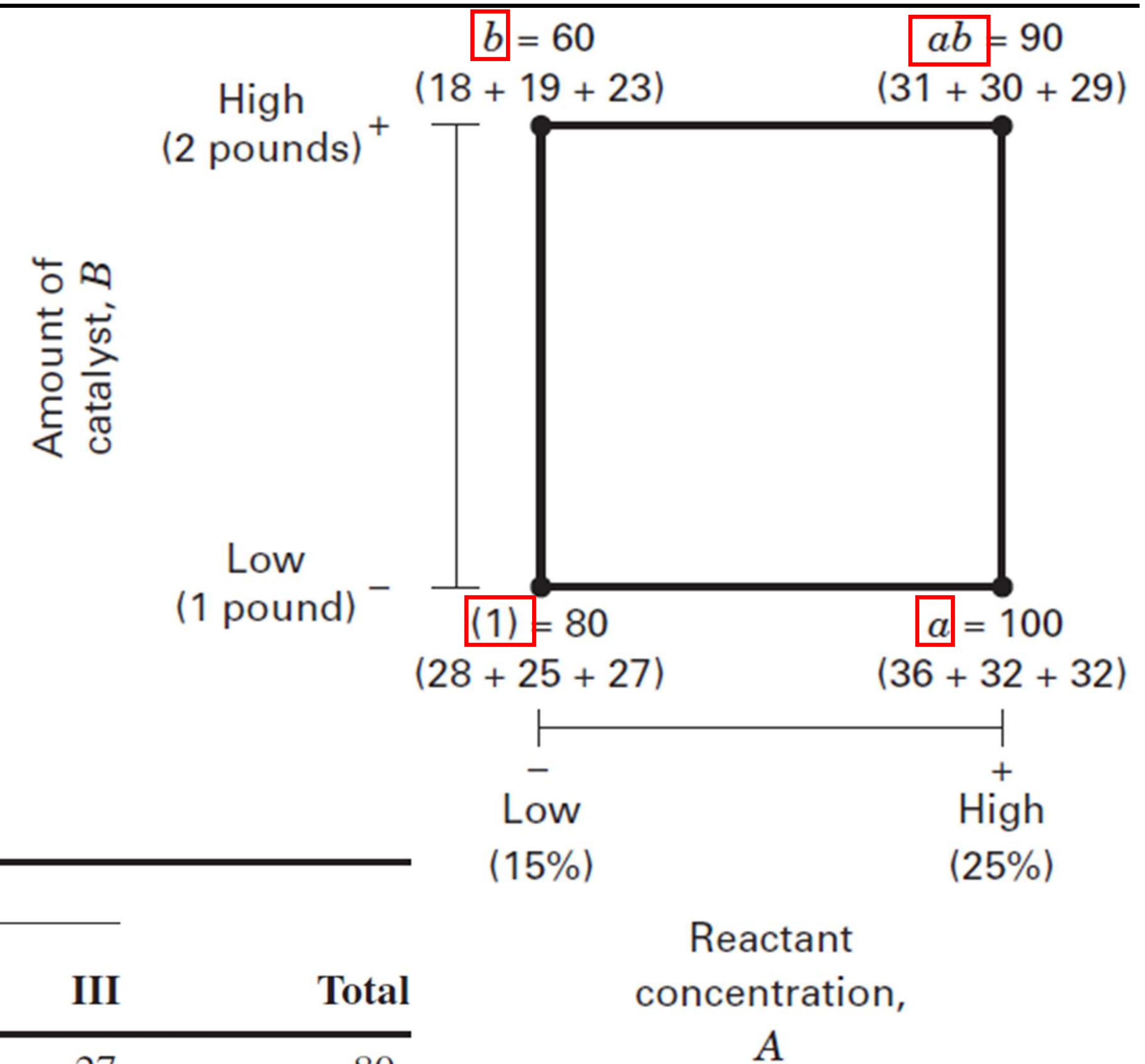
The **order in which the runs are made is random**, so this is a **completely randomized experiment**.

| Factor *A* | Factor *B* | Treatment Combination | Replicate I | Replicate II | Replicate III | Total |
|---|---|---|---|---|---|---|
| − | − | *A* low, *B* low | 28 | 25 | 27 | 80 |
| + | − | *A* high, *B* low | 36 | 32 | 32 | 100 |
| − | + | *A* low, *B* high | 18 | 19 | 23 | 60 |
| + | + | *A* high, *B* high | 31 | 30 | 29 | 90 |

The **four treatment combinations** in this design are shown graphically to the right.

High level of any factor in the treatment combination is denoted by the corresponding lowercase letter and that the low level of a factor in the treatment combination is denoted by the absence of the corresponding letter.

Thus, ***a*** represents the treatment combination of **A** at the **high** level **and B** at the **low** level, ***b*** represents **A** at the **low** level **and B** at the **high** level, and ***ab*** represents **both factors** at the **high** level.




| Factor A | Factor B | Treatment Combination | Replicate I | Replicate II | Replicate III | Total |
|---|---|---|---|---|---|---|
| − | − | $A$ low, $B$ low | 28 | 25 | 27 | 80 |
| + | − | $A$ high, $B$ low | 36 | 32 | 32 | 100 |
| − | + | $A$ low, $B$ high | 18 | 19 | 23 | 60 |
| + | + | $A$ high, $B$ high | 31 | 30 | 29 | 90 |

# $2^K$ Factorial Design: $2^2$ factorial design

In a two-level factorial design, we may define the **average effect of a factor** as the **change in response produced by a change in the level of that factor averaged over the levels of the other factor**.

Main Effect of Factor A

$$A = \frac{1}{2n}\{[ab - b] + [a - (1)]\}$$

$$= \frac{1}{2n}[\boxed{ab + a - b - (1)}]$$

contrast

Main Effect of Factor B

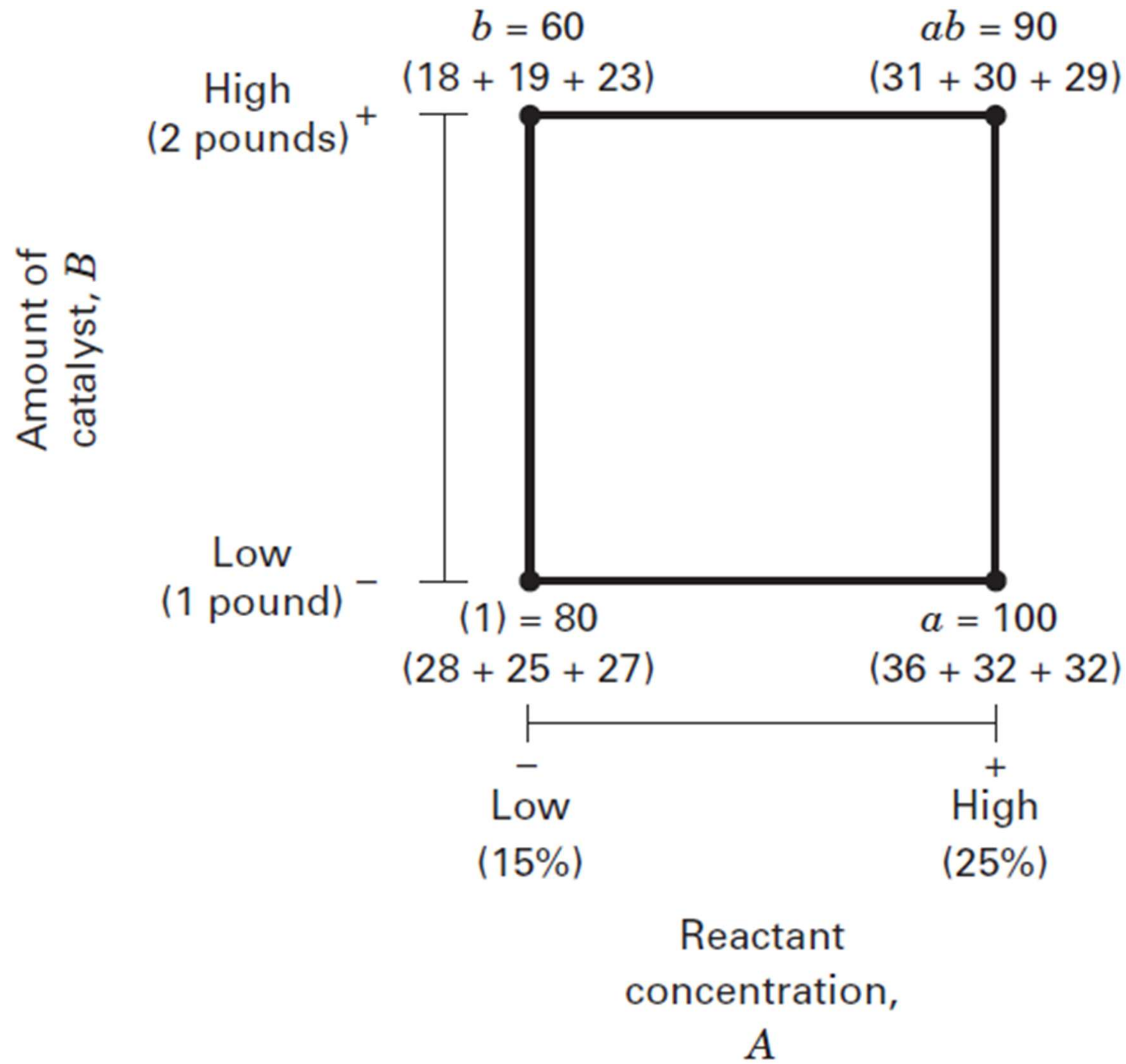
$$B = \frac{1}{2n}\{[ab - a] + [b - (1)]\}$$

$$= \frac{1}{2n}[ab + b - a - (1)]$$

Interaction Effect AB

$$AB = \frac{1}{2n}\{[ab - b] - [a - (1)]\}$$

$$= \frac{1}{2n}[ab + (1) - a - b]$$



In a two-level factorial design, we may define the **average effect of a factor** as the **change in response produced by a change in the level of that factor averaged over the levels of the other factor**.

Main Effect of Factor A

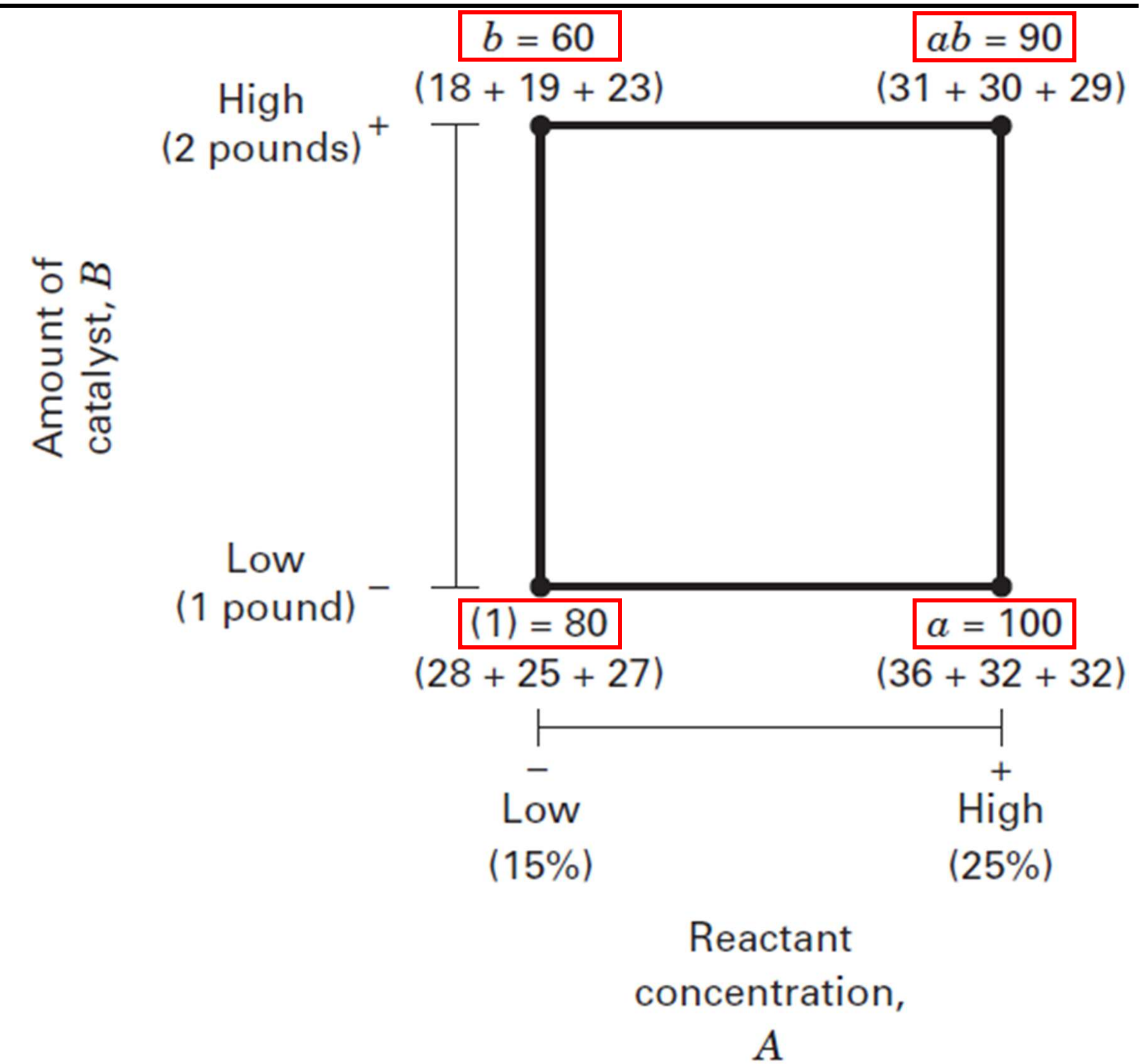
$$A = \frac{1}{2n}\{\boxed{[ab} - \boxed{b]} + \boxed{[a} - \boxed{(1)}]\}$$

$$= \frac{1}{2n}[ab + a - b - (1)]$$

$$A = \frac{1}{2(3)}(90 + 100 - 60 - 80) = \boxed{8.33}$$

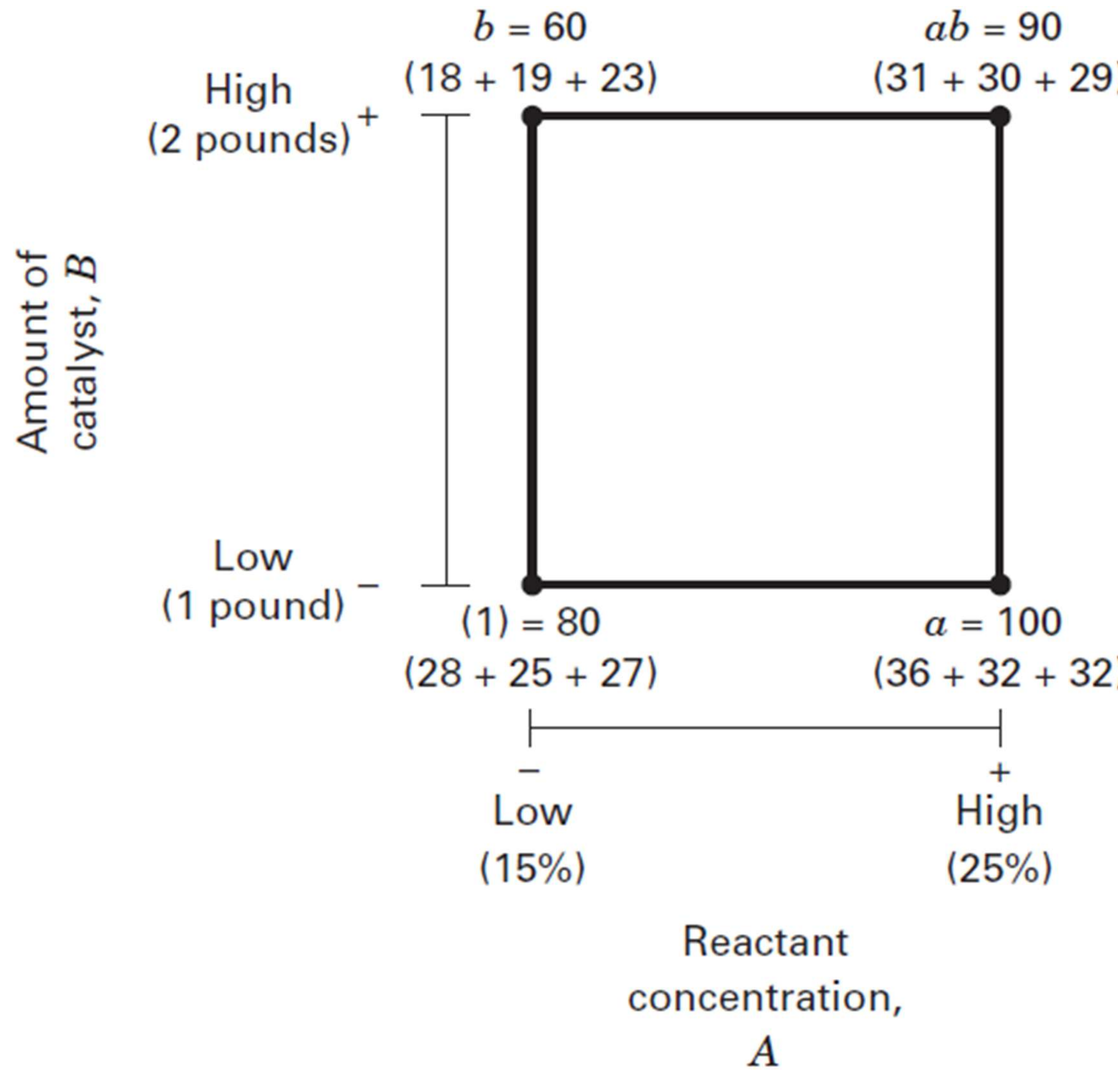
$$B = \frac{1}{2(3)}(90 + 60 - 100 - 80) = \boxed{-5.00}$$

$$AB = \frac{1}{2(3)}(90 + 80 - 100 - 60) = \boxed{1.67}$$




- The **effect of A is positive**; increasing A from the low level (15%) to the high level (25%) will increase the yield.
- The **effect of B is negative**; increasing the amount of B added to the process will decrease the yield.
- The **interaction effect** appears to be **small** relative to the two main effects.

In a two-level factorial design, we may define the **average effect of a factor** as the **change in response produced by a change in the level of that factor averaged over the levels of the other factor**.




Main Effect of Factor A

$$A = \frac{1}{2n}\{[ab - b] + [a - (1)]\}$$
$$= \frac{1}{2n}[ab + a - b - (1)]$$

In experiments involving $2^k$ designs, it is always important to examine the **magnitude and direction of the factor effects** to determine **which variables are likely to be important**.

The **analysis of variance** can generally be used to confirm this interpretation (t-tests could be used too).

Effect magnitude and direction should always be considered along with the ANOVA, because the ANOVA alone does not convey this information.

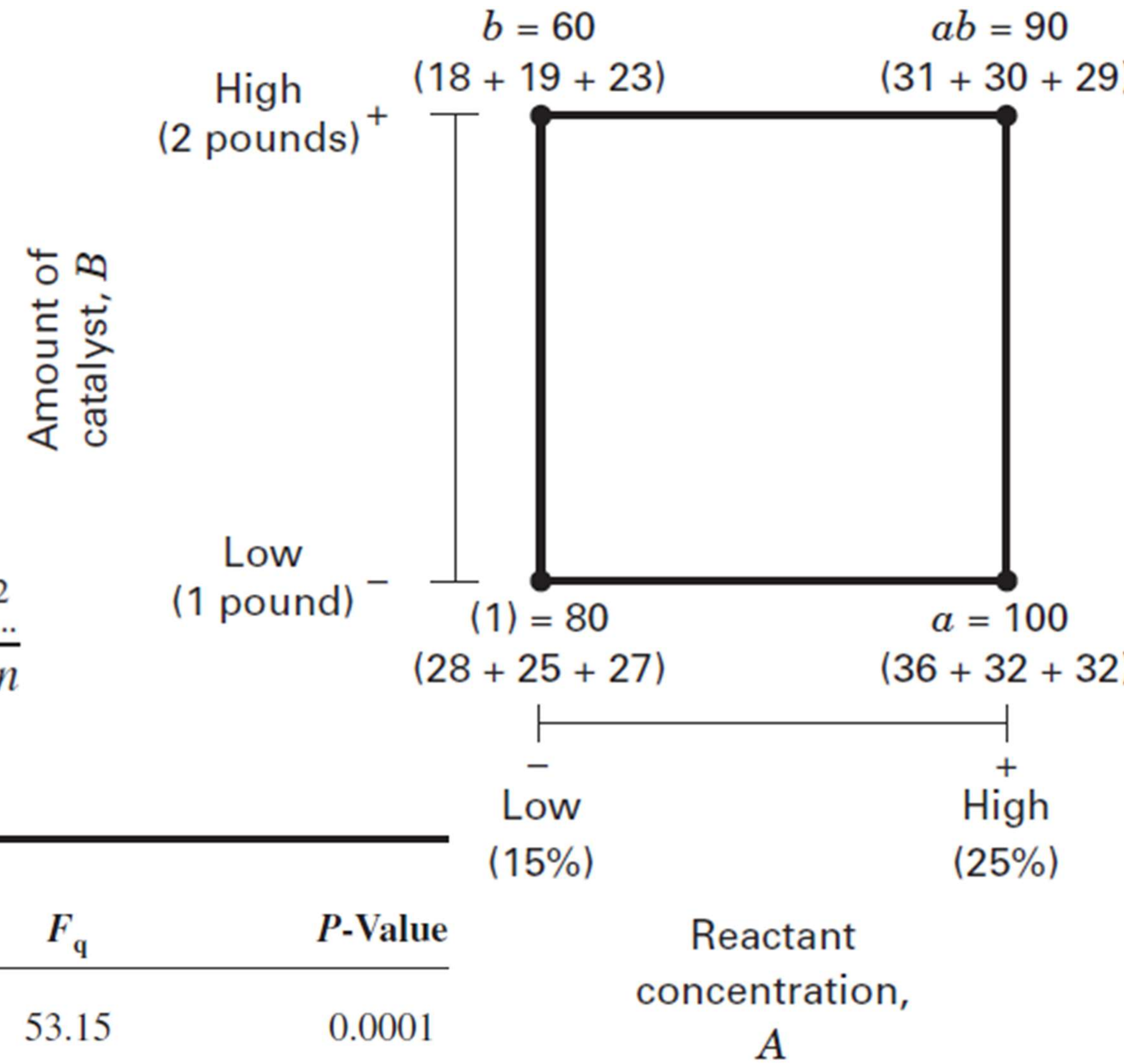
The **sum of squares** for the three **contrasts** are

$$SS_A = \frac{[ab + a - b - (1)]^2}{4n}$$

$$SS_B = \frac{[ab + b - a - (1)]^2}{4n}$$

$$SS_{AB} = \frac{[ab + (1) - a - b]^2}{4n}$$

We can compute sums of squares by only squaring one number.




$$SS_E = SS_T - SS_A - SS_B - SS_{AB} \qquad SS_T = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{4n}$$

**Analysis of Variance for the Experiment in Figure 6.1**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_q$ | *P*-Value |
|---|---|---|---|---|---|
| *A* | 208.33 | 1 | 208.33 | 53.15 | 0.0001 |
| *B* | 75.00 | 1 | 75.00 | 19.13 | 0.0024 |
| *AB* | 8.33 | 1 | 8.33 | 2.13 | 0.1826 |
| Error | 31.34 | 8 | 3.92 | | |
| Total | 323.00 | 11 | | | |

Results of the experiment in terms of a **regression model**

$$y = \beta_0 + \beta_1 \boxed{x_1} + \beta_2 \boxed{x_2} + \epsilon$$

$$\boxed{x_1 = \frac{\text{Conc} - (\text{Conc}_{\text{low}} + \text{Conc}_{\text{high}})/2}{(\text{Conc}_{\text{high}} - \text{Conc}_{\text{low}})/2}} \qquad \boxed{x_2 = \frac{\text{Catalyst} - (\text{Catalyst}_{\text{low}} + \text{Catalyst}_{\text{high}})/2}{(\text{Catalyst}_{\text{high}} - \text{Catalyst}_{\text{low}})/2}}$$

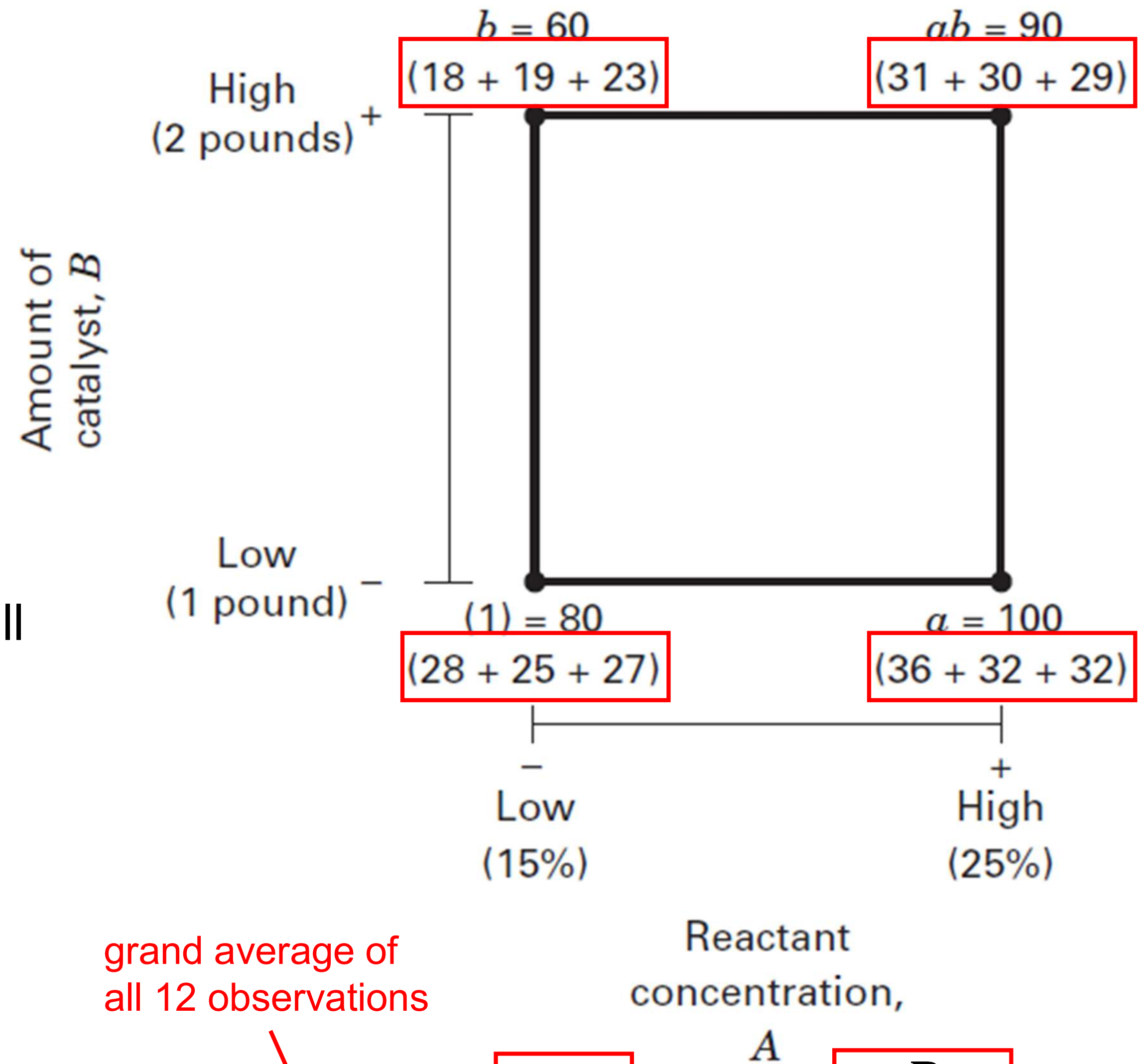
When the natural variables have only two levels, this coding will produce the ±1 notation for the levels of the coded variables.

$$x_1 = \frac{\text{Conc} - (15+25)/2}{(25-15)/2} = \frac{\text{Conc} - 20}{5} \qquad x_2 = \frac{\text{Catalyst} - (1+2)/2}{(2-1)/2} = \frac{\text{Catalyst} - 1.5}{0.5}$$

$$\text{Conc} = 5x_1 + 20 \qquad \text{Catalyst} = 0.5x_2 + 1.5$$

$$x_1 = -1 \Rightarrow \text{Conc} = 15 \qquad x_2 = -1 \Rightarrow \text{Catalyst} = 1$$

$$x_1 = +1 \Rightarrow \text{Conc} = 25 \qquad x_2 = +1 \Rightarrow \text{Catalyst} = 2$$




grand average of
all 12 observations

$$\hat{y} = \boxed{27.5} + \left(\frac{\boxed{\overset{A}{8.33}}}{2}\right)x_1 + \left(\frac{\boxed{\overset{B}{-5.00}}}{2}\right)x_2$$

# $2^K$ Factorial Design: Residuals and model adequacy

The **regression model** can be used to obtain the **predicted or fitted value of $y$** at the four points in the design.

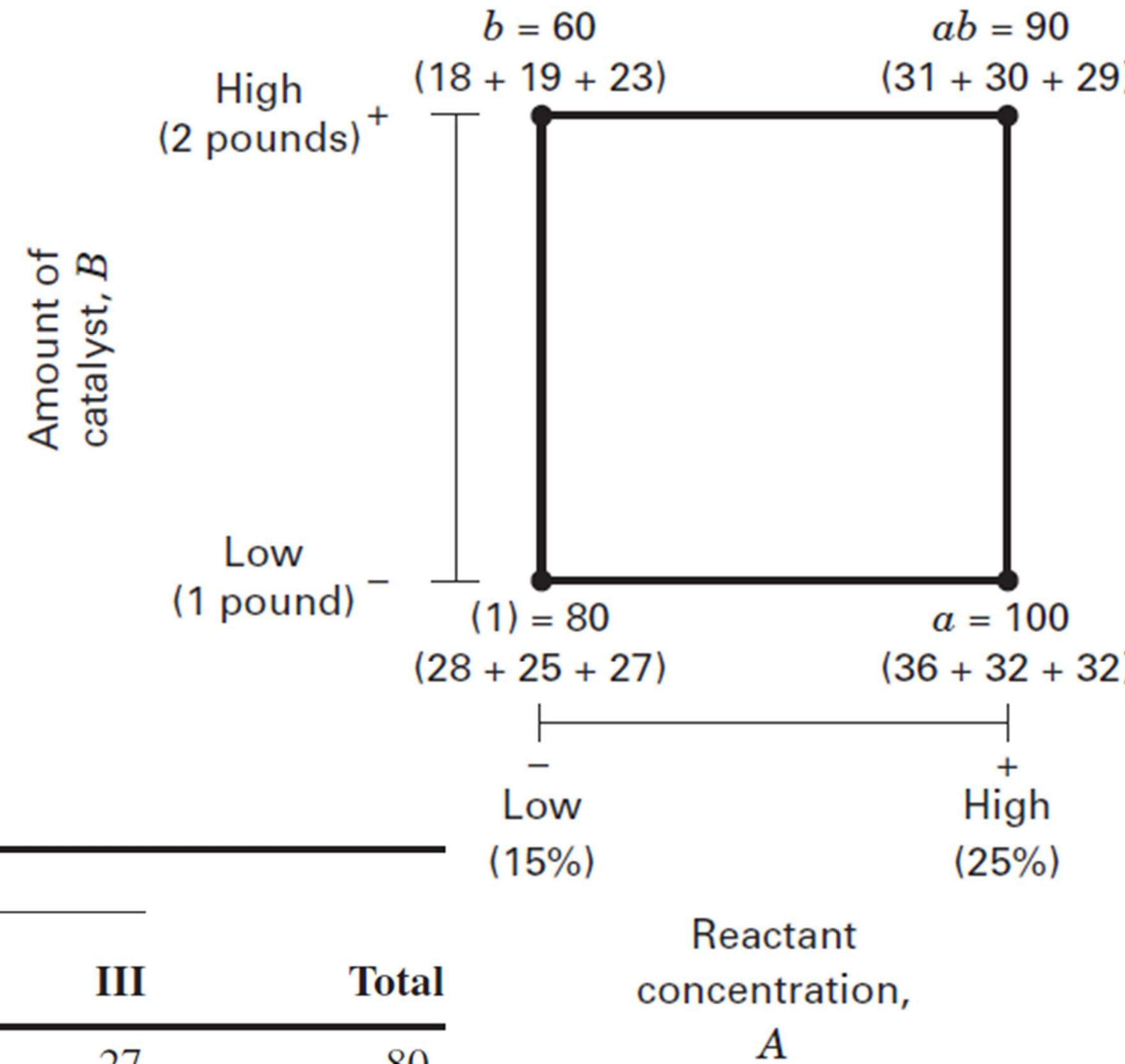
The **residuals** are the **differences between the observed and fitted values of $y$**.

$x_1 = -1$

$x_2 = -1$

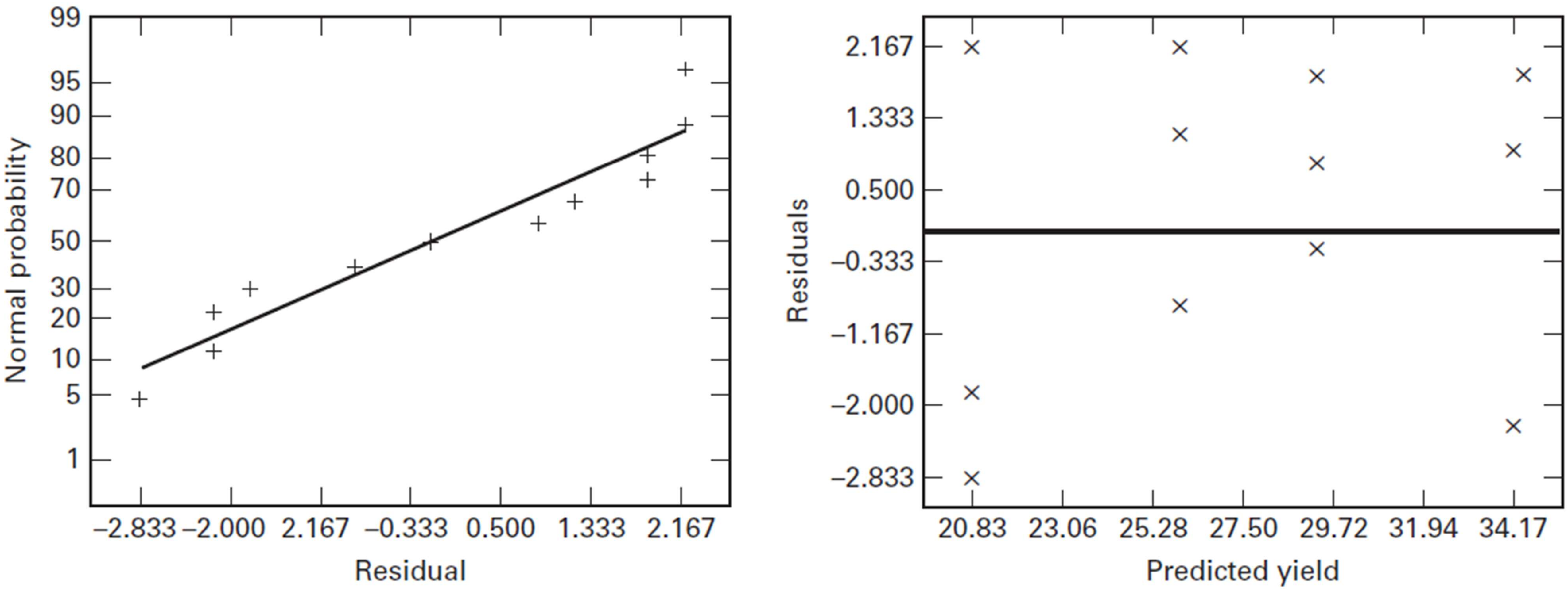
$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)(-1) + \left(\frac{-5.00}{2}\right)(-1) = 25.835$$

$$e_1 = 28 - 25.835 = 2.165$$
$$e_2 = 25 - 25.835 = -0.835$$
$$e_3 = 27 - 25.835 = 1.165$$




| Factor | | | Replicate | | | |
|---|---|---|---|---|---|---|
| $A$ | $B$ | Treatment Combination | I | II | III | Total |
| − | − | $A$ low, $B$ low | 28 | 25 | 27 | 80 |

These plots appear **satisfactory**, so we have **no reason to suspect that there are any problems** with the **validity** of our **conclusions**.




($a$) Normal probability plot

($b$) Residuals versus predicted yield

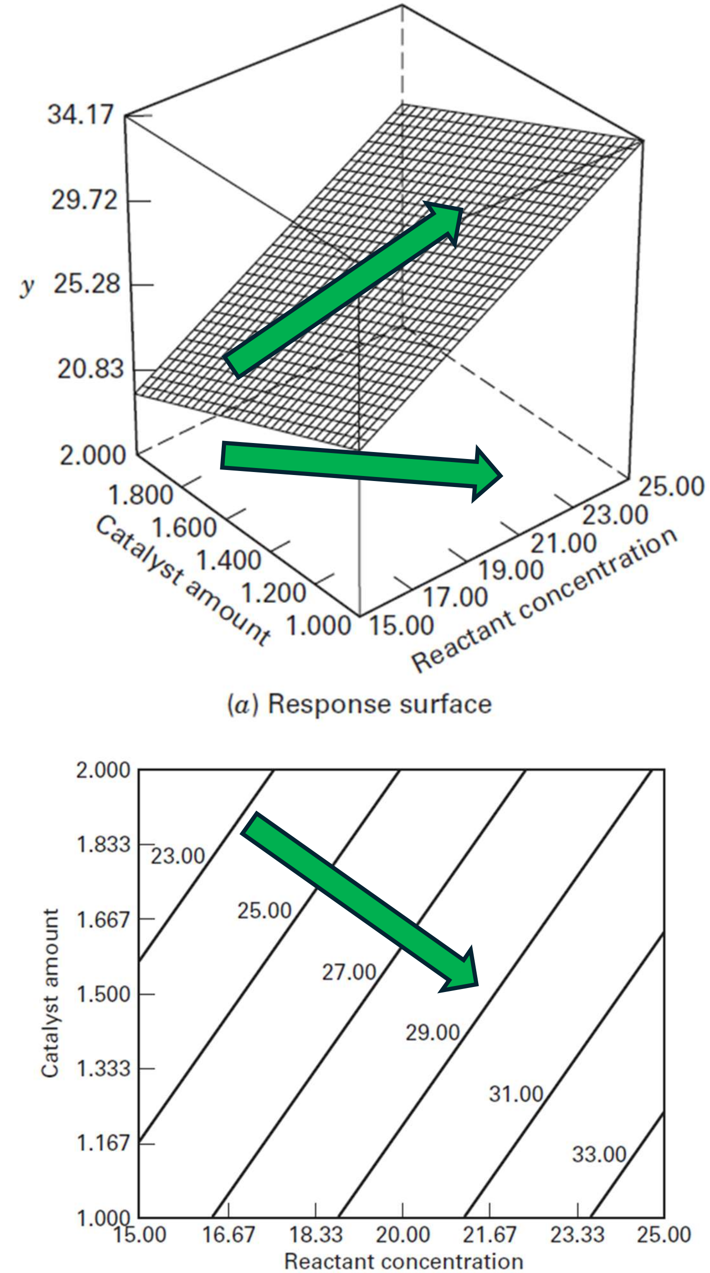
The **regression model** can be used to generate **response surface plots**.

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)x_1 + \left(\frac{-5.00}{2}\right)x_2$$

If it is desirable to construct these plots in terms of the **natural factor levels**, then we simply substitute the relationships between the natural and coded variables that we gave earlier into the regression model, yielding

$$\hat{y} = 27.5 + \left(\frac{8.33}{2}\right)\left(\frac{\text{Conc} - 20}{5}\right) + \left(\frac{-5.00}{2}\right)\left(\frac{\text{Catalyst} - 1.5}{0.5}\right)$$

$$= 18.33 + 0.8333\ \text{Conc} - 5.00\ \text{Catalyst}$$



(*a*) Response surface

(*b*) Contour plot

Suppose that three factors, A, B, and C, each at two levels, are of interest.

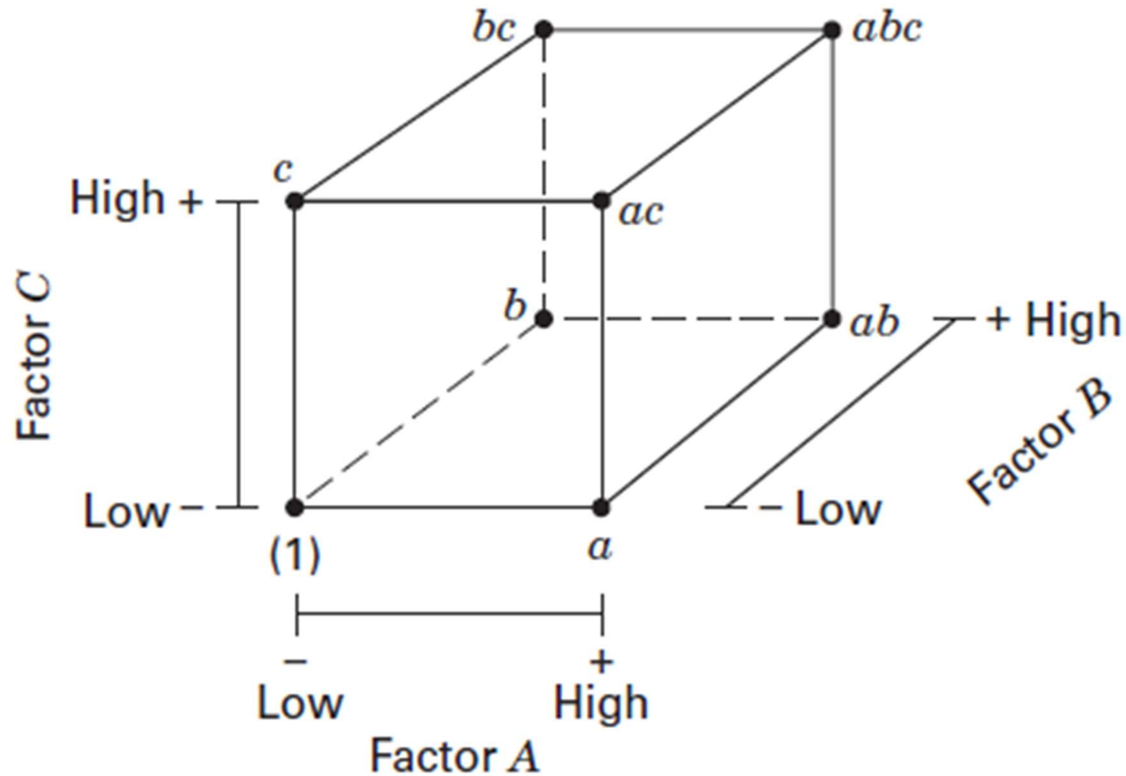
| Run | *A* | *B* | *C* | Labels | *A* | *B* | *C* |
|---|---|---|---|---|---|---|---|
| 1 | − | − | − | (1) | 0 | 0 | 0 |
| 2 | + | − | − | *a* | 1 | 0 | 0 |
| 3 | − | + | − | *b* | 0 | 1 | 0 |
| 4 | + | + | − | *ab* | 1 | 1 | 0 |
| 5 | − | − | + | *c* | 0 | 0 | 1 |
| 6 | + | − | + | *ac* | 1 | 0 | 1 |
| 7 | − | + | + | *bc* | 0 | 1 | 1 |
| 8 | + | + | + | *abc* | 1 | 1 | 1 |

Geometric coding
Orthogonal coding
Effects coding

DESIGN MATRIX

Treatment combinations

Low-High



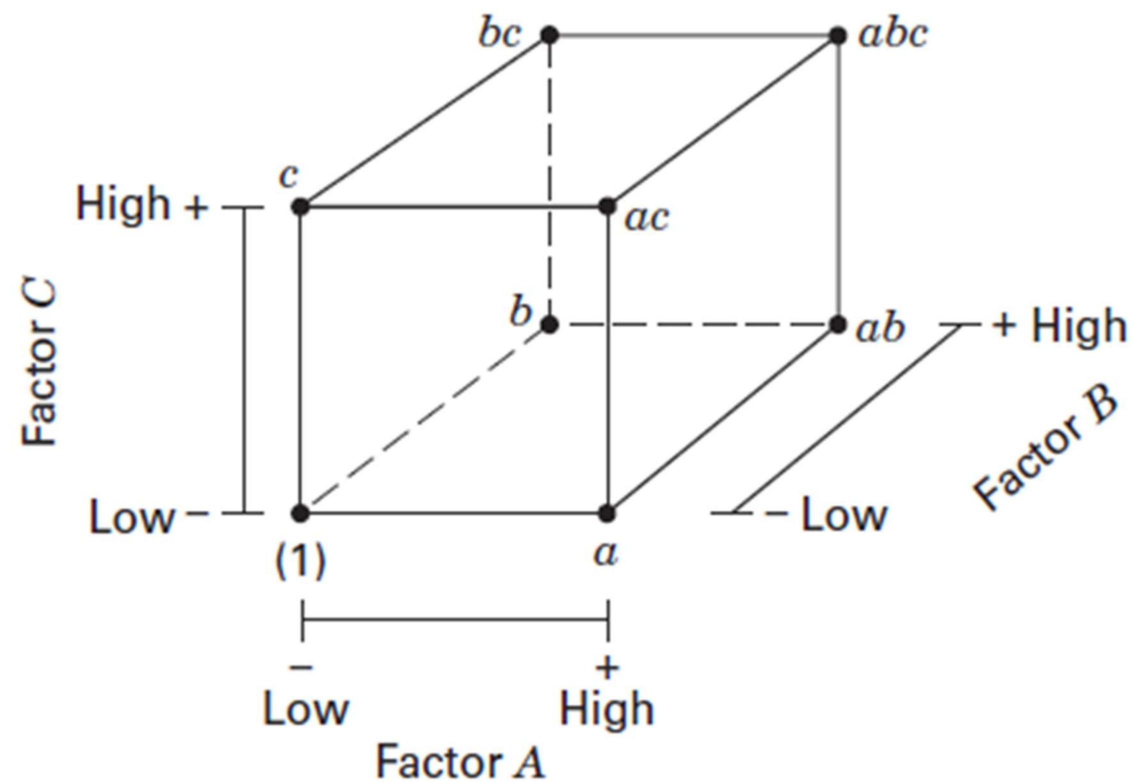
Consider estimating the **main effects**.

First, consider estimating the main effect A.

| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | $a$ |
| 3 | − | + | − | $b$ |
| 4 | + | + | − | $ab$ |
| 5 | − | − | + | $c$ |
| 6 | + | − | + | $ac$ |
| 7 | − | + | + | $bc$ |
| 8 | + | + | + | $abc$ |

DESIGN MATRIX

Main effect A when B=Low and C=Low



# $2^K$ Factorial Design: $2^3$ factorial design

Consider estimating the **main effects**.

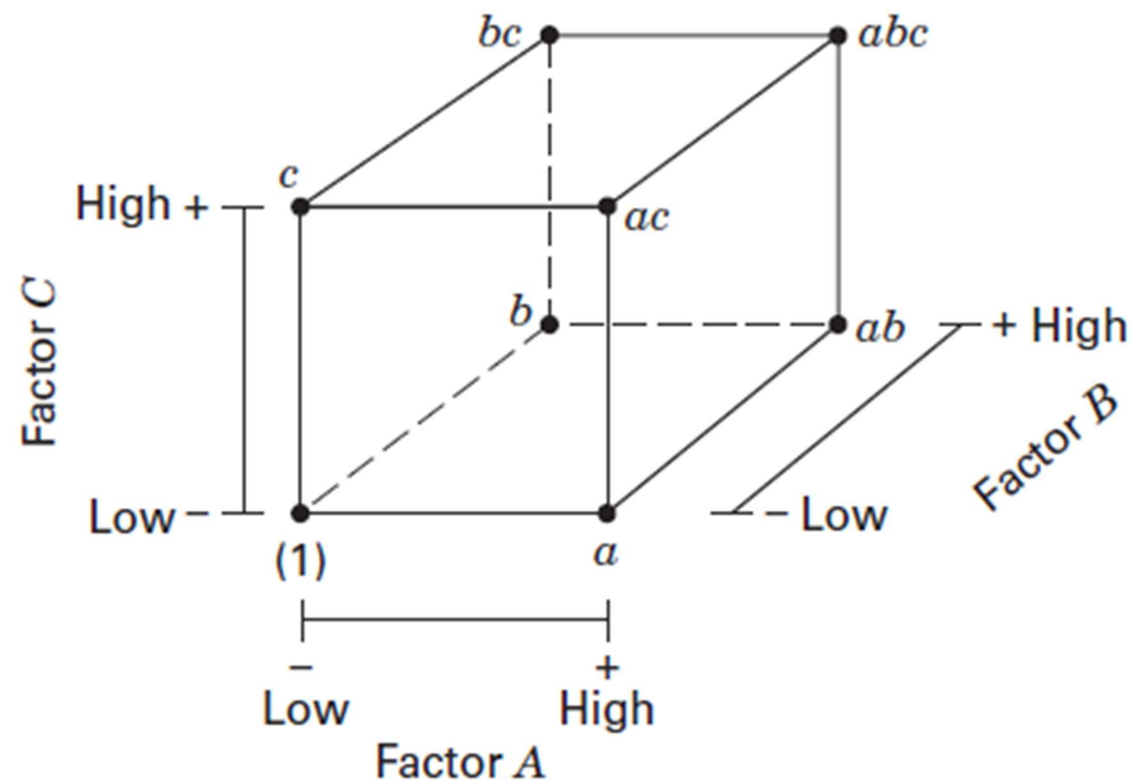
First, consider estimating the main effect A.

| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | $a$ |

DESIGN MATRIX

Main effect A when B=Low and C=Low $\dfrac{a-(1)}{n}$

Main effect A when B=High and C=Low



Consider estimating the **main effects**.

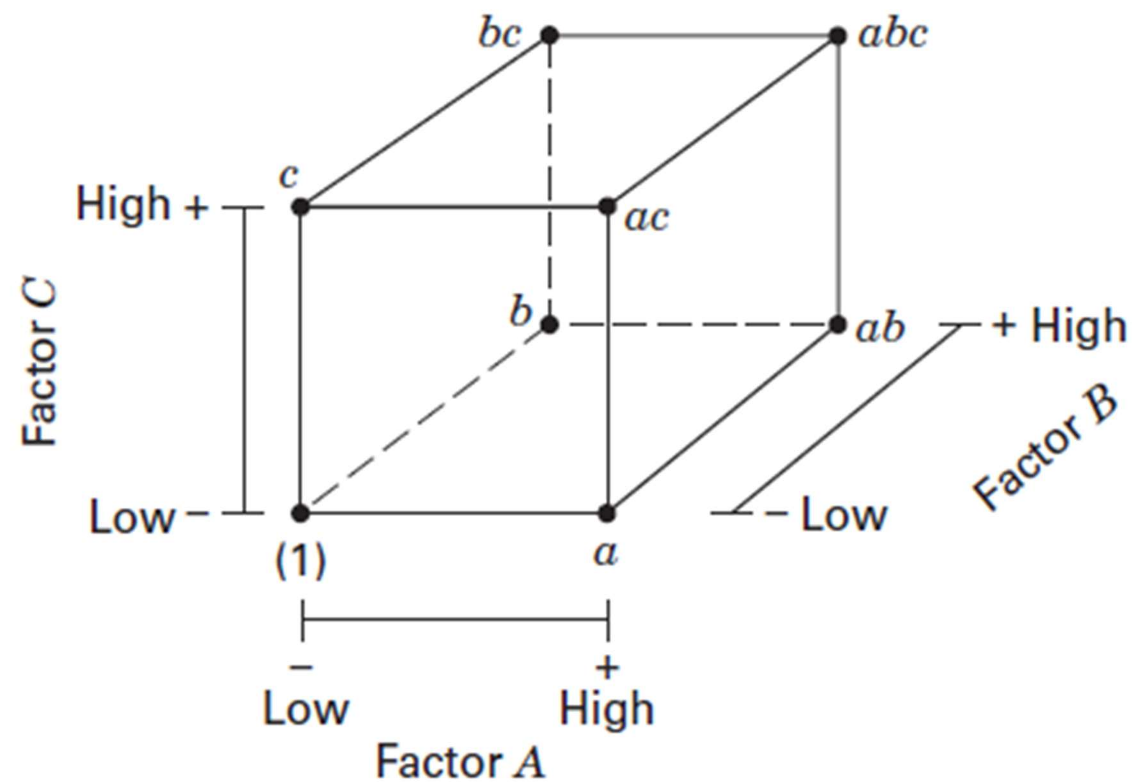
First, consider estimating the main effect A.

| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 3 | − | + | − | $b$ |
| 4 | + | + | − | $ab$ |

DESIGN MATRIX

Main effect A when B=Low and C=Low $\dfrac{a-(1)}{n}$

Main effect A when B=High and C=Low $\dfrac{ab-b}{n}$



# $2^K$ Factorial Design: $2^3$ factorial design

Consider estimating the **main effects**.

First, consider estimating the main effect A.

| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | $a$ |
| 3 | − | + | − | $b$ |
| 4 | + | + | − | $ab$ |
| 5 | − | − | + | $c$ |
| 6 | + | − | + | $ac$ |
| 7 | − | + | + | $bc$ |
| 8 | + | + | + | $abc$ |

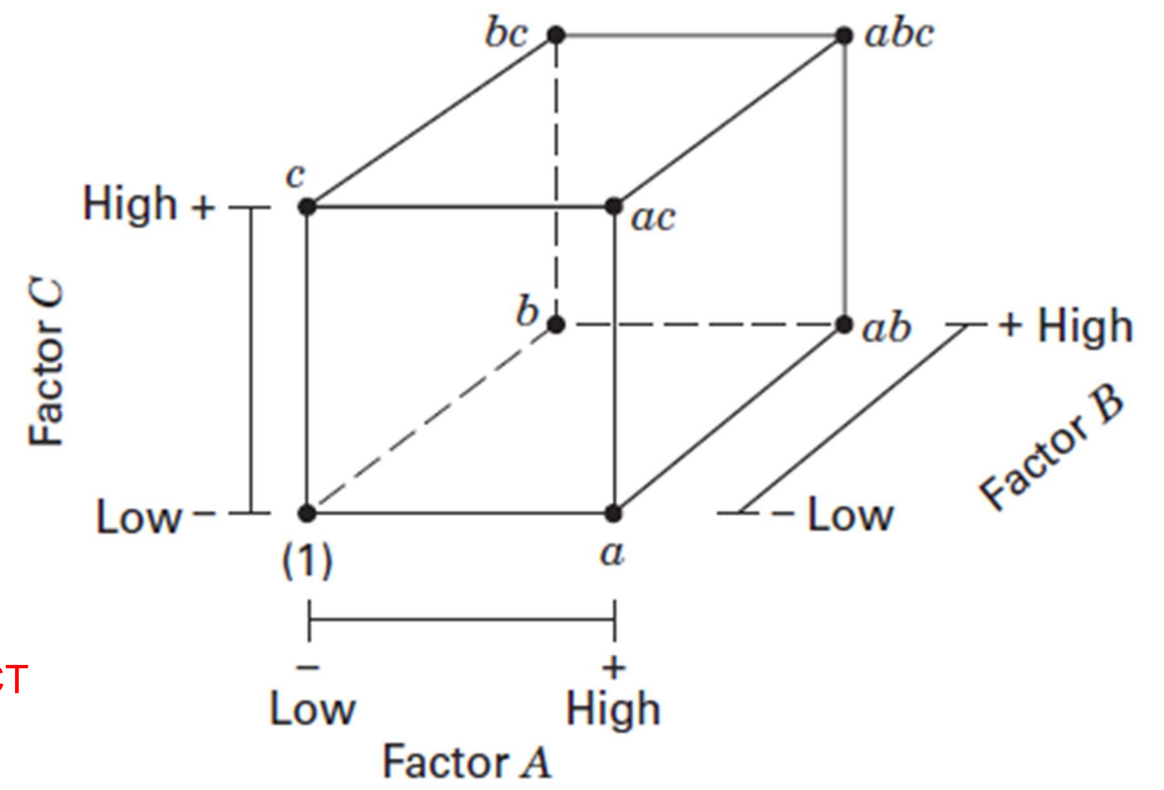
DESIGN MATRIX

Main effect A when B=Low and C=Low $\dfrac{a-(1)}{n}$

Main effect A when B=High and C=Low $\dfrac{ab-b}{n}$

Main effect A when B=Low and C=High $\dfrac{ac-c}{n}$

Main effect A when B=High and C=High $\dfrac{abc-bc}{n}$



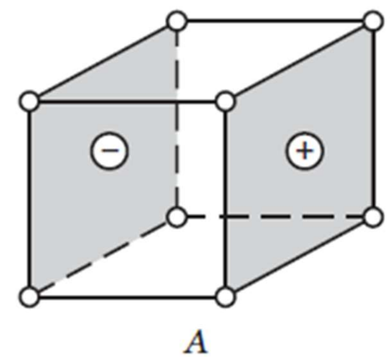

MAIN EFFECT

$$A = \frac{1}{4n}[a - (1) + ab - b + ac - c + abc - bc]$$

# $2^K$ Factorial Design: $2^3$ factorial design

| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | $a$ |
| 3 | − | + | − | $b$ |
| 4 | + | + | − | $ab$ |
| 5 | − | − | + | $c$ |
| 6 | + | − | + | $ac$ |
| 7 | − | + | + | $bc$ |
| 8 | + | + | + | $abc$ |

$$A = \overline{y}_{A+} - \overline{y}_{A-} = \frac{a + ab + ac + abc}{4n} - \frac{(1) + b + c + bc}{4n}$$

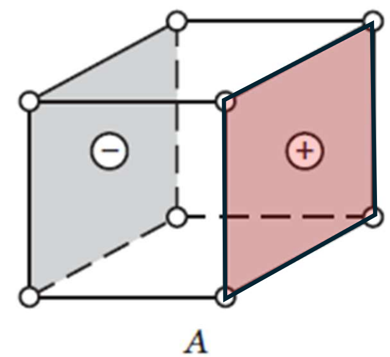



This equation can also be developed as a **contrast** between the **four treatment combinations** in the **right face** of the cube in the figure and the **four in the left face**.

$$A = \frac{1}{4n}[a - (1) + ab - b + ac - c + abc - bc]$$

| Run | *A* | *B* | *C* | Labels |
|---|---|---|---|---|
| 2 | + | − | − | *a* |
| 4 | + | + | − | *ab* |
| 6 | + | − | + | *ac* |
| 8 | + | + | + | *abc* |

$$A = \bar{y}_{A+} - \bar{y}_{A-} = \frac{a + ab + ac + abc}{4n}$$



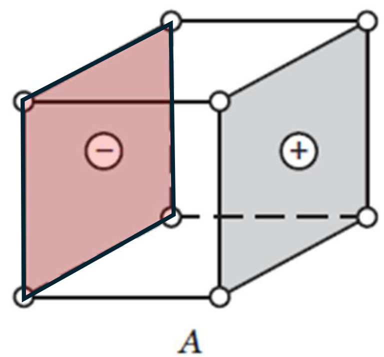

This equation can also be developed as a **contrast** between the **four treatment combinations** in the **right face** of the cube in the figure and the **four in the left face**.

$$A = \frac{1}{4n}[a - (1) + ab - b + ac - c + abc - bc]$$

| Run | A | B | C | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 3 | − | + | − | $b$ |
| 5 | − | − | + | $c$ |
| 7 | − | + | + | $bc$ |

$$A = \bar{y}_{A+} - \bar{y}_{A-}$$

$$-\frac{(1) + b + c + bc}{4n}$$



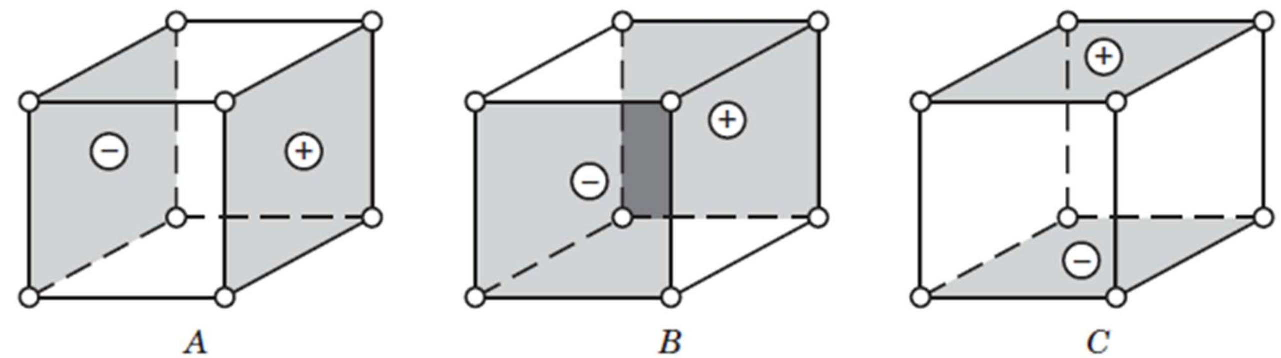

This equation can also be developed as a **contrast** between the **four treatment combinations** in the **right face** of the cube in the figure and the **four in the left face**.

$$A = \frac{1}{4n}[a - (1) + ab - b + ac - c + abc - bc]$$

# $2^K$ Factorial Design: $2^3$ factorial design

| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | $a$ |
| 3 | − | + | − | $b$ |
| 4 | + | + | − | $ab$ |
| 5 | − | − | + | $c$ |
| 6 | + | − | + | $ac$ |
| 7 | − | + | + | $bc$ |
| 8 | + | + | + | $abc$ |

$A = \overline{y}_{A+} - \overline{y}_{A-}$ $B = \overline{y}_{B+} - \overline{y}_{B-}$ $C = \overline{y}_{C+} = \overline{y}_{C-}$

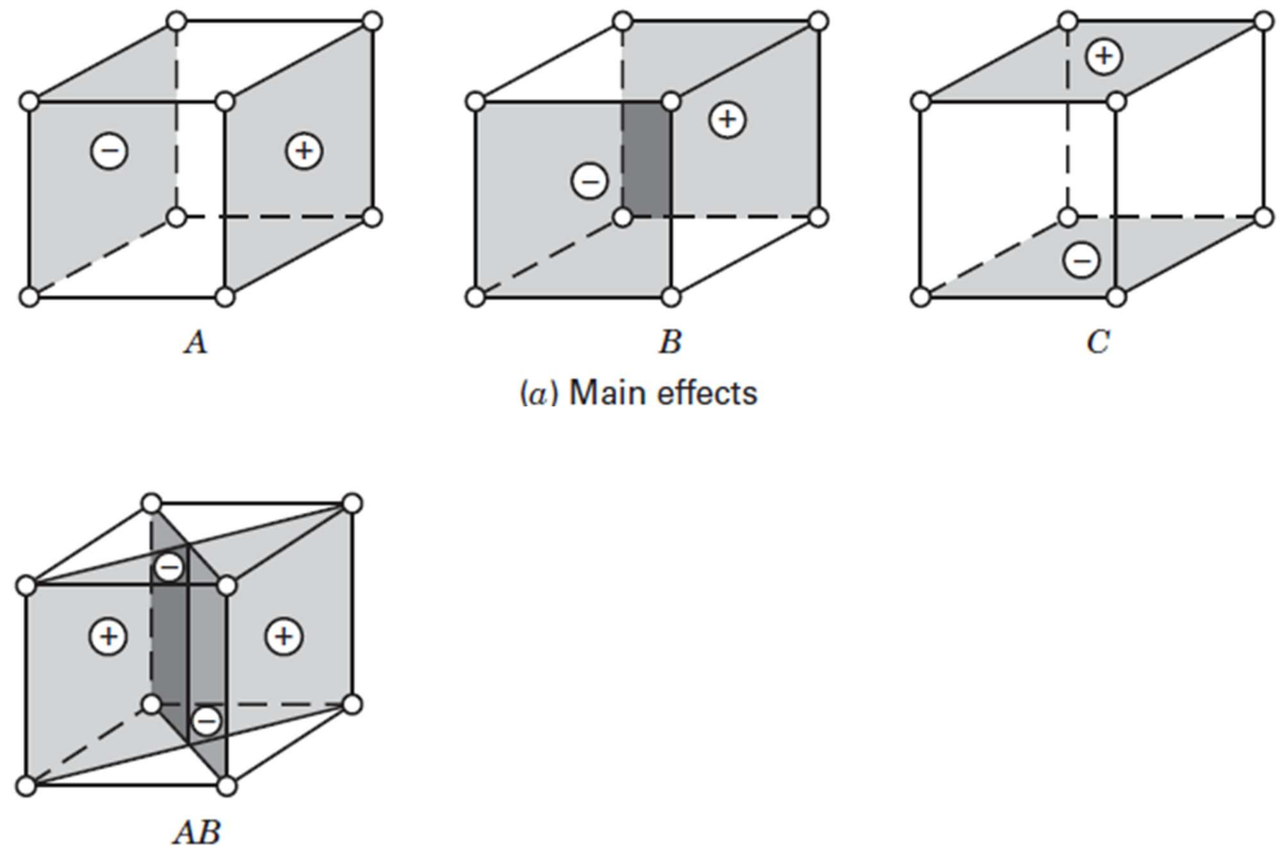


($a$) Main effects

The **two-factor interaction effects** may be computed easily.

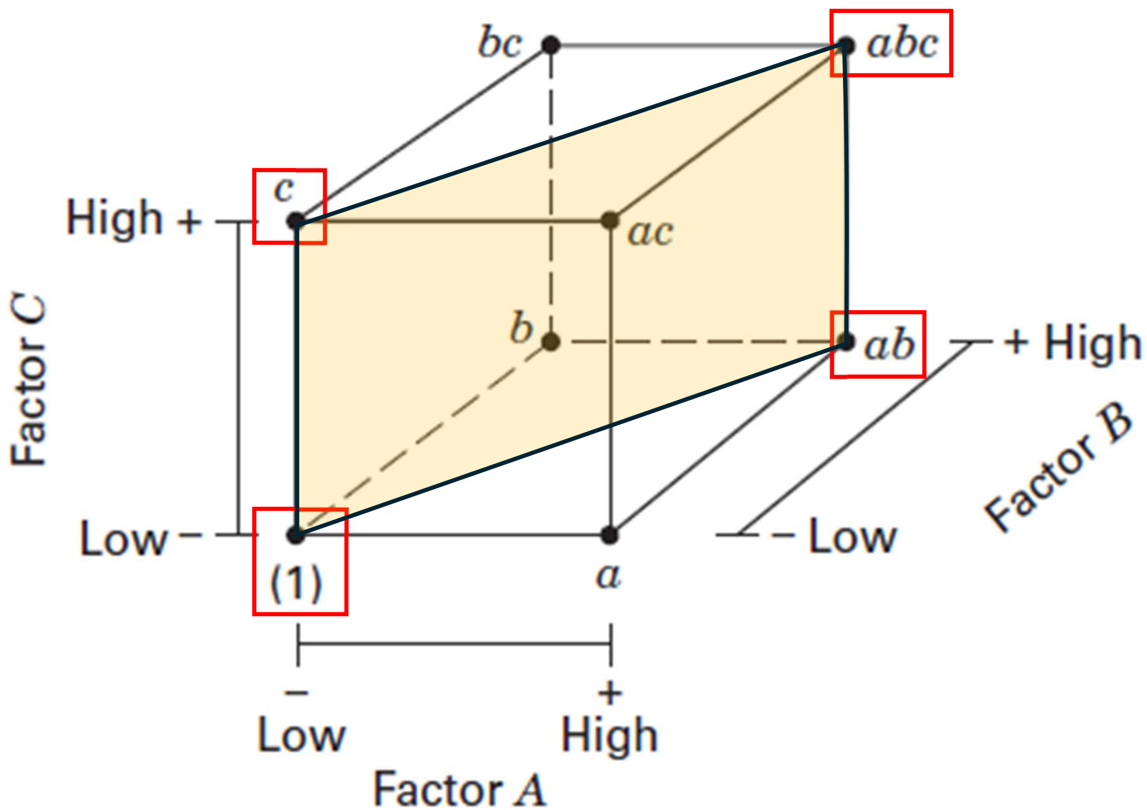
$$AB = \frac{abc + ab + c + (1)}{4n} - \frac{bc + b + ac + a}{4n}$$

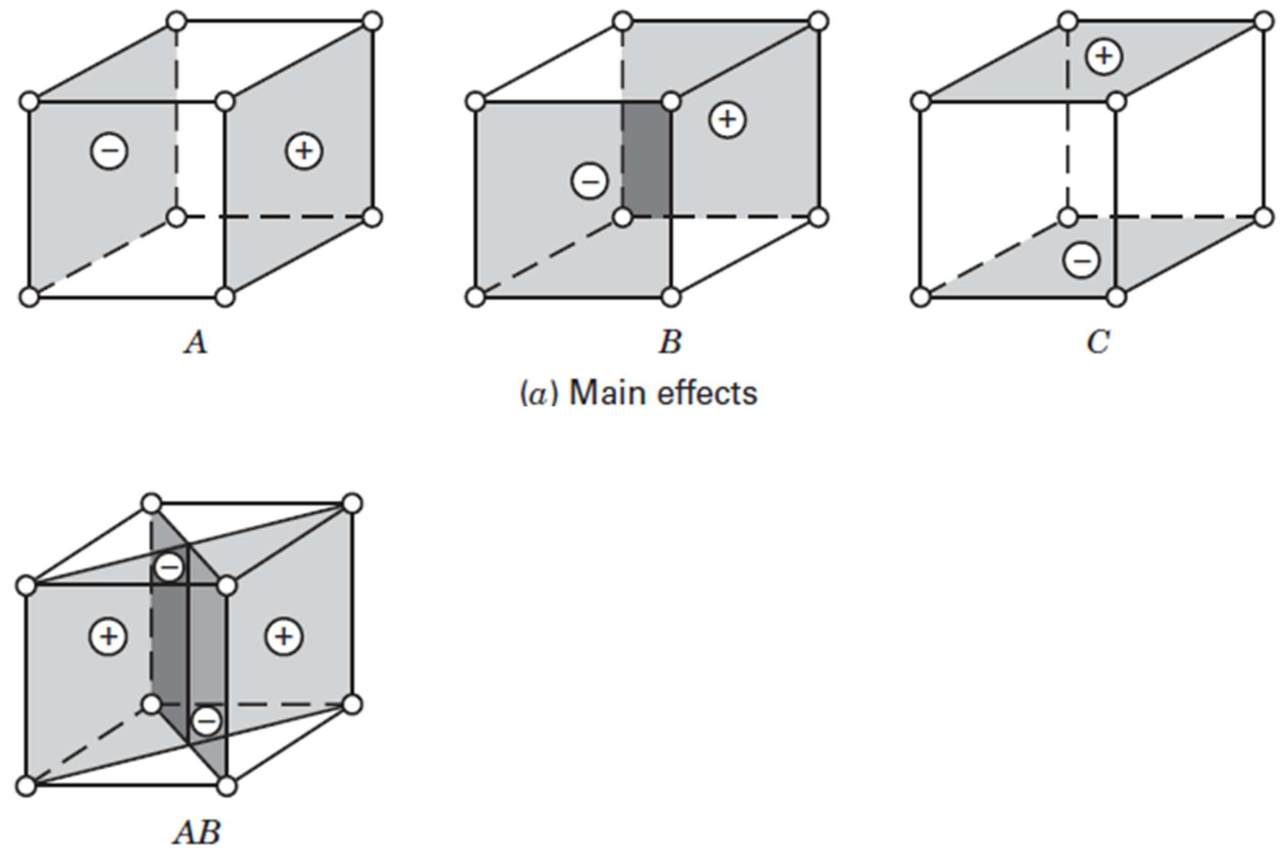
| Run | $A$ | $B$ | $C$ | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | $a$ |
| 3 | − | + | − | $b$ |
| 4 | + | + | − | $ab$ |
| 5 | − | − | + | $c$ |
| 6 | + | − | + | $ac$ |
| 7 | − | + | + | $bc$ |
| 8 | + | + | + | $abc$ |



The **two-factor interaction effects** may be computed easily.

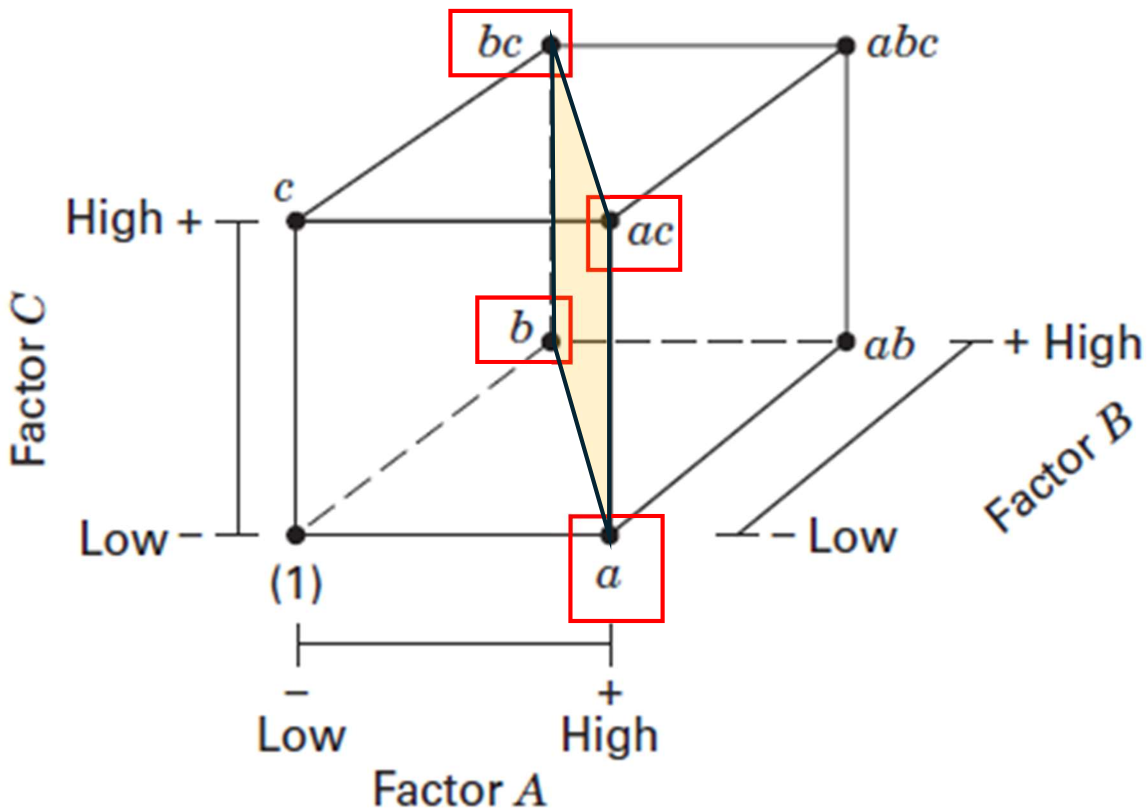
$$AB = \frac{\boxed{abc} + \boxed{ab} + \boxed{c} + \boxed{(1)}}{4n}$$

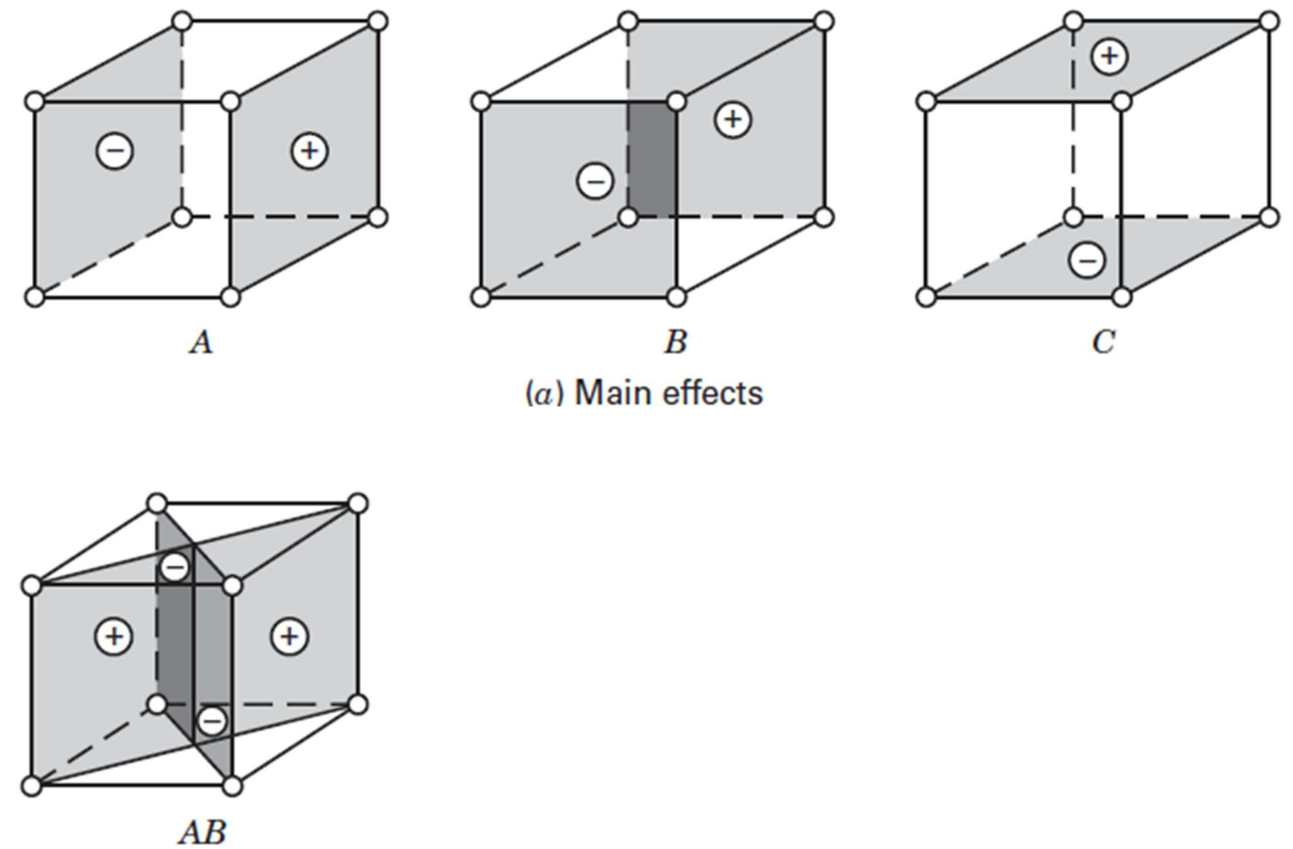


The **two-factor interaction effects** may be computed easily.

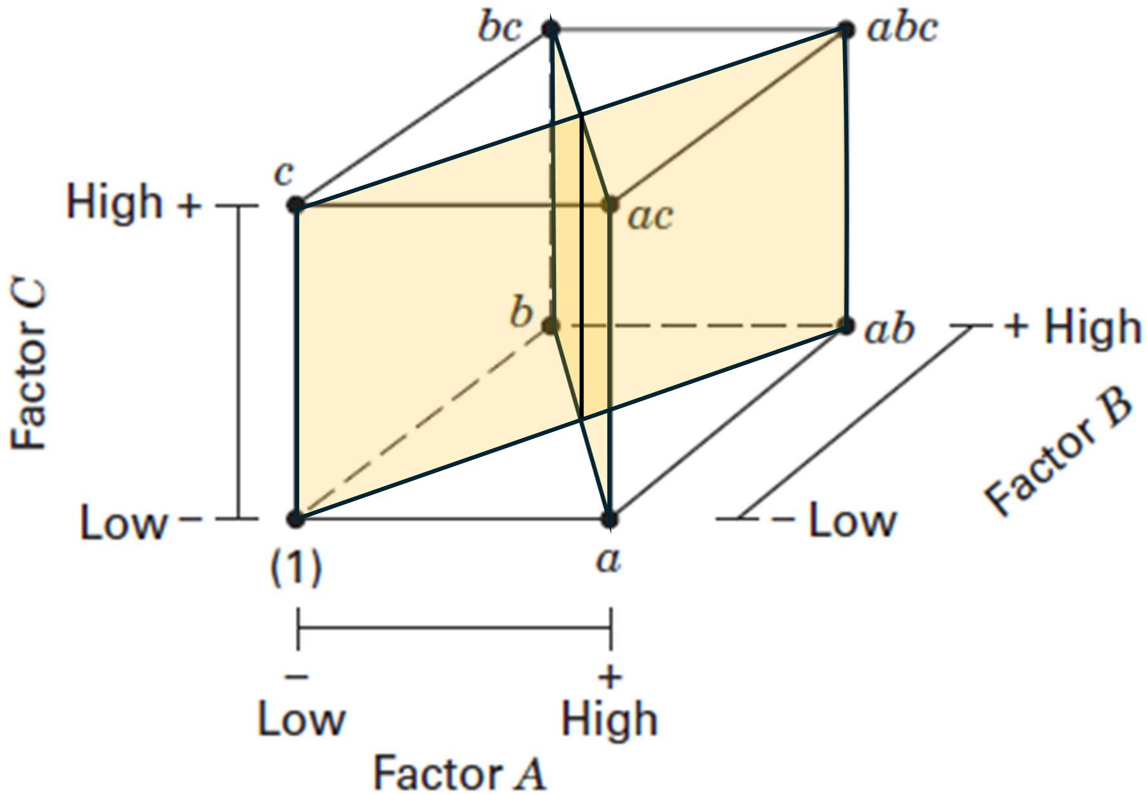
$$\frac{\boxed{bc}+\boxed{b}+\boxed{ac}+\boxed{a}}{4n}$$

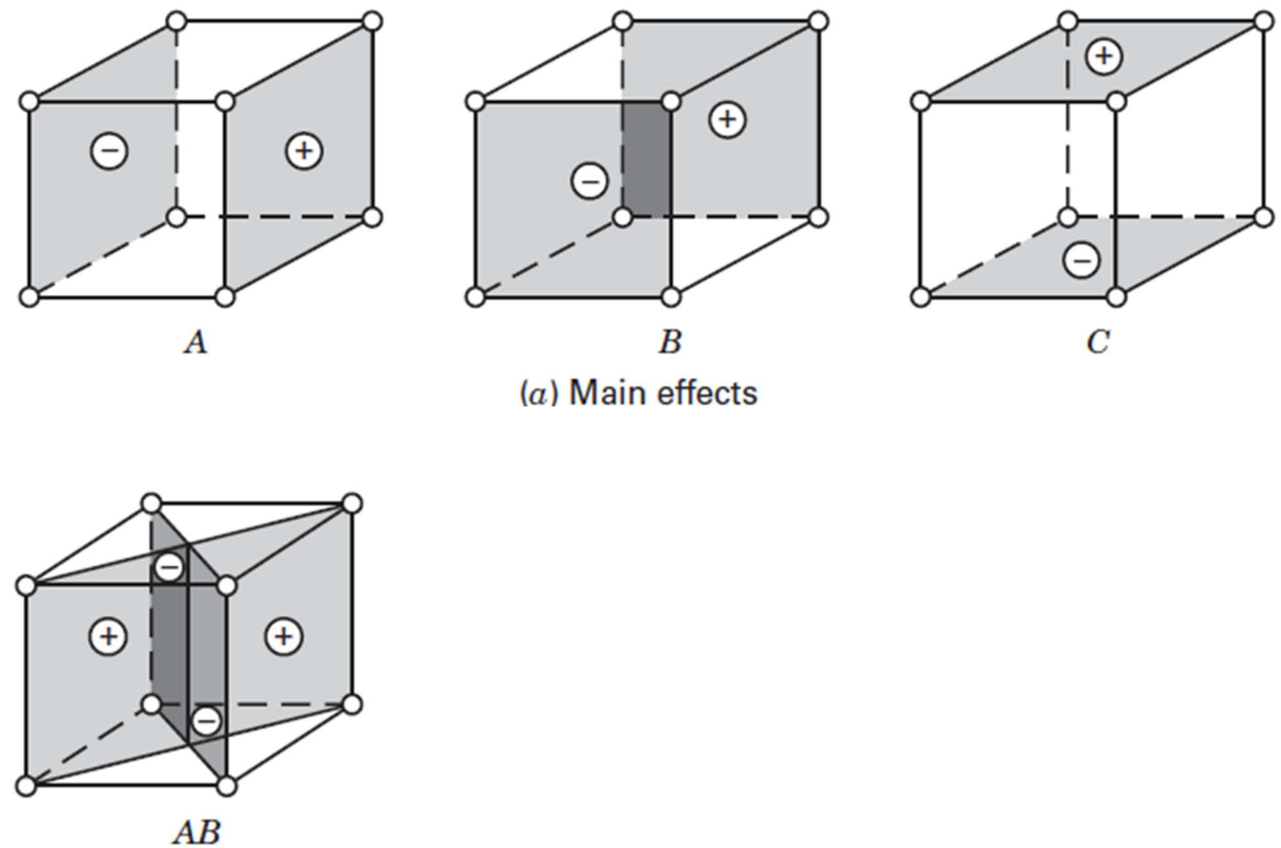


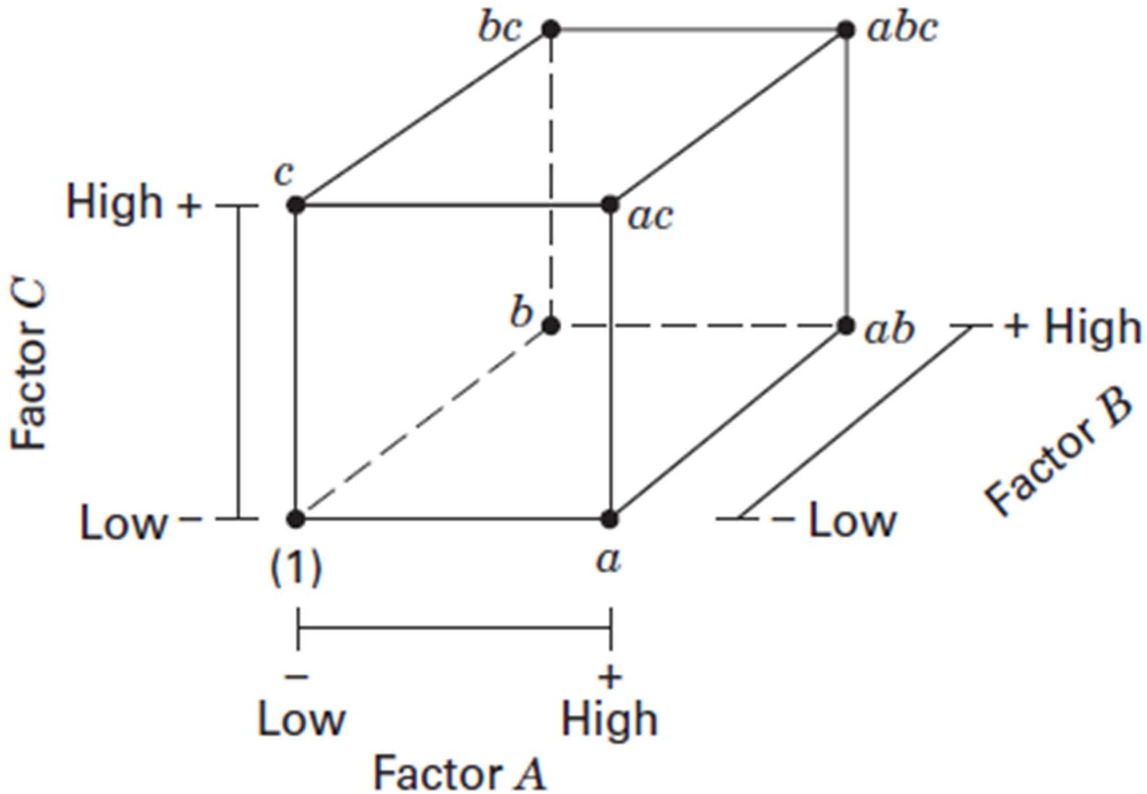
The **two-factor interaction effects** may be computed easily.

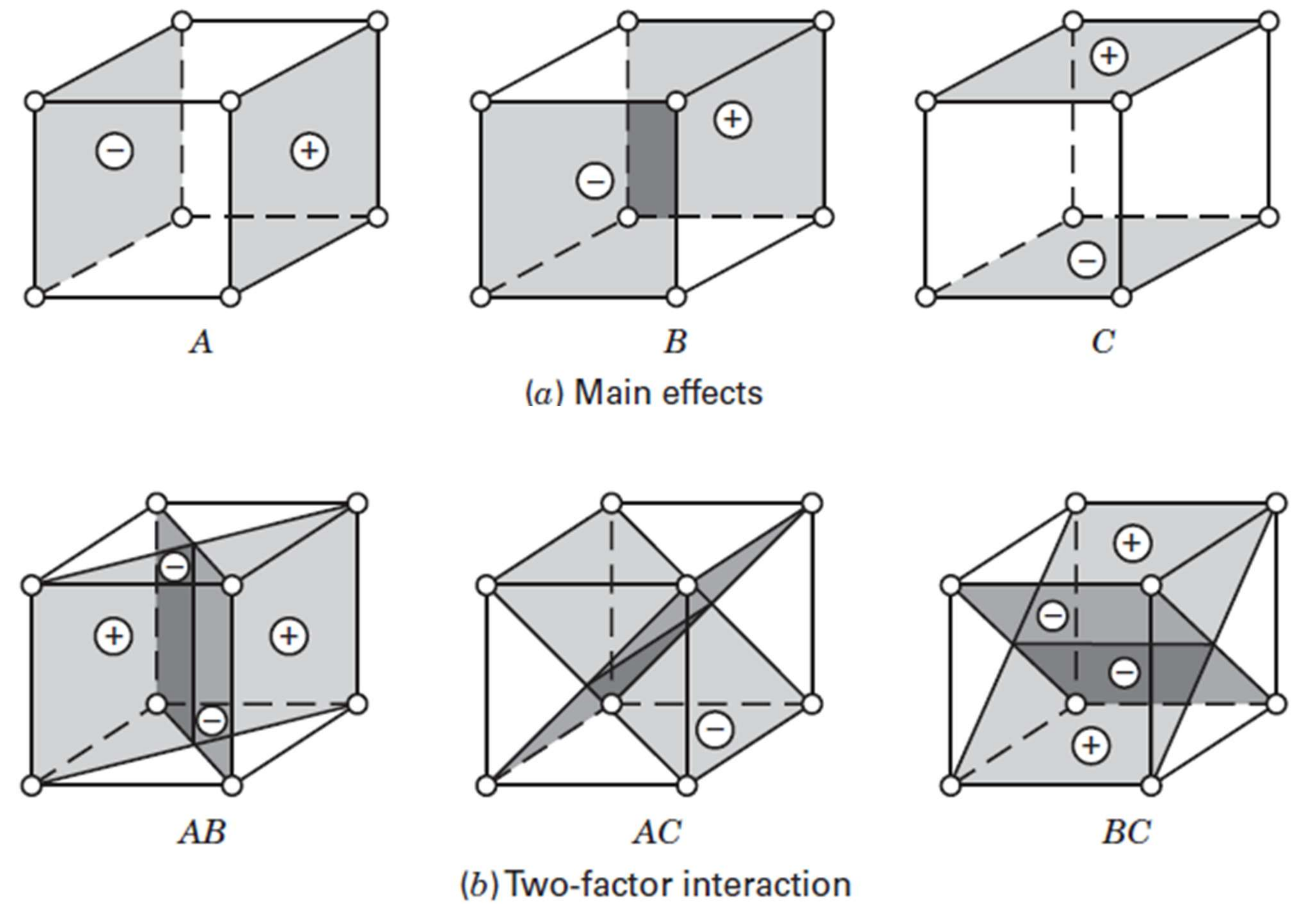
$$AB = \frac{abc + ab + c + (1)}{4n} - \frac{bc + b + ac + a}{4n}$$



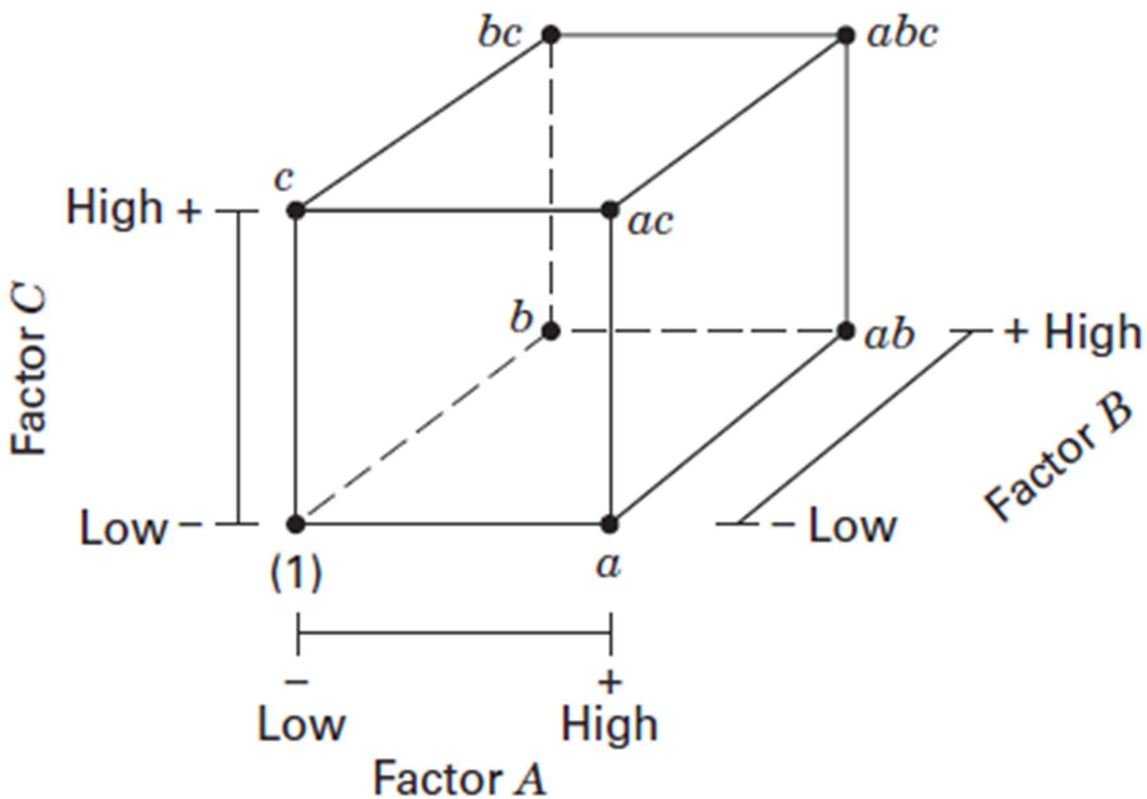
The **two-factor interaction effects** may be computed easily.



The **three-factor interaction effects may be computed not that easily** but it worth of your attention to check whether the above reasoning was clear enough.

$$ABC = \frac{1}{4n}\{[abc - bc] - [ac - c] - [ab - b] + [a - (1)]\}$$
$$= \frac{1}{4n}[abc - bc - ac + c - ab + b + a - (1)]$$




| Run | A | B | C | Labels |
|---|---|---|---|---|
| 1 | − | − | − | (1) |
| 2 | + | − | − | a |
| 3 | − | + | − | b |
| 4 | + | + | − | ab |
| 5 | − | − | + | c |
| 6 | + | − | + | ac |
| 7 | − | + | + | bc |
| 8 | + | + | + | abc |

# $2^K$ Factorial Design: $2^3$ factorial design

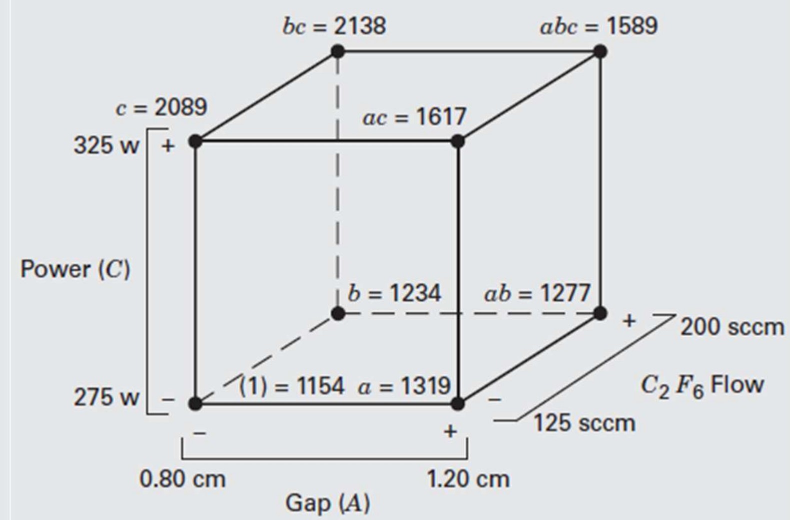
A $2^3$ factorial design was used to develop a nitride etch process on a single-wafer plasma etching tool. The design factors are the gap between the electrodes, the gas flow ($C_2F_6$ is used as the reactant gas), and the RF power applied to the cathode (see Figure 3.1 for a schematic of the plasma etch tool). Each factor is run at two levels, and the design is replicated twice. The response variable is the etch rate for silicon nitride (Å/m). The etch rate data are shown in Table 6.4, and the design is shown geometrically in Figure 6.6.

■ **TABLE 6.4**
**The Plasma Etch Experiment, Example 6.1**

| | Coded Factors | | | Etch Rate | | | Factor Levels | | |
|---|---|---|---|---|---|---|---|---|---|
| **Run** | ***A*** | ***B*** | ***C*** | **Replicate 1** | **Replicate 2** | **Total** | | **Low (−1)** | **High (+1)** |
| 1 | −1 | −1 | −1 | 550 | 604 | (1) = 1154 | *A* (Gap, cm) | 0.80 | 1.20 |
| 2 | 1 | −1 | −1 | 669 | 650 | *a* = 1319 | *B* ($C_2F_6$ flow, SCCM) | 125 | 200 |
| 3 | −1 | 1 | −1 | 633 | 601 | *b* = 1234 | *C* (Power, W) | 275 | 325 |
| 4 | 1 | 1 | −1 | 642 | 635 | *ab* = 1277 | | | |
| 5 | −1 | −1 | 1 | 1037 | 1052 | *c* = 2089 | | | |
| 6 | 1 | −1 | 1 | 749 | 868 | *ac* = 1617 | | | |
| 7 | −1 | 1 | 1 | 1075 | 1063 | *bc* = 2138 | | | |
| 8 | 1 | 1 | 1 | 729 | 860 | *abc* = 1589 | | | |

# $2^K$ Factorial Design: $2^3$ factorial design

| Run | Coded Factors A | B | C | Etch Rate Replicate 1 | Replicate 2 | Total | Factor Levels | Low (−1) | High (+1) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | −1 | −1 | −1 | 550 | 604 | (1) = 1154 | $A$ (Gap, cm) | 0.80 | 1.20 |
| 2 | 1 | −1 | −1 | 669 | 650 | $a$ = 1319 | $B$ ($C_2F_6$ flow, SCCM) | 125 | 200 |
| 3 | −1 | 1 | −1 | 633 | 601 | $b$ = 1234 | $C$ (Power, W) | 275 | 325 |
| 4 | 1 | 1 | −1 | 642 | 635 | $ab$ = 1277 | | | |
| 5 | −1 | −1 | 1 | 1037 | 1052 | $c$ = 2089 | | | |
| 6 | 1 | −1 | 1 | 749 | 868 | $ac$ = 1617 | | | |
| 7 | −1 | 1 | 1 | 1075 | 1063 | $bc$ = 2138 | | | |
| 8 | 1 | 1 | 1 | 729 | 860 | $abc$ = 1589 | | | |




**Effect Estimate Summary for Example 6.1**

| Factor | Effect Estimate | Sum of Squares | Percent Contribution |
|---|---|---|---|
| $A$ | −101.625 | 41,310.5625 | 7.7736 |
| $B$ | 7.375 | 217.5625 | 0.0409 |
| $C$ | 306.125 | 374,850.0625 | 70.5373 |
| $AB$ | −24.875 | 2475.0625 | 0.4657 |
| $AC$ | −153.625 | 94,402.5625 | 17.7642 |
| $BC$ | −2.125 | 18.0625 | 0.0034 |
| $ABC$ | 5.625 | 126.5625 | 0.0238 |

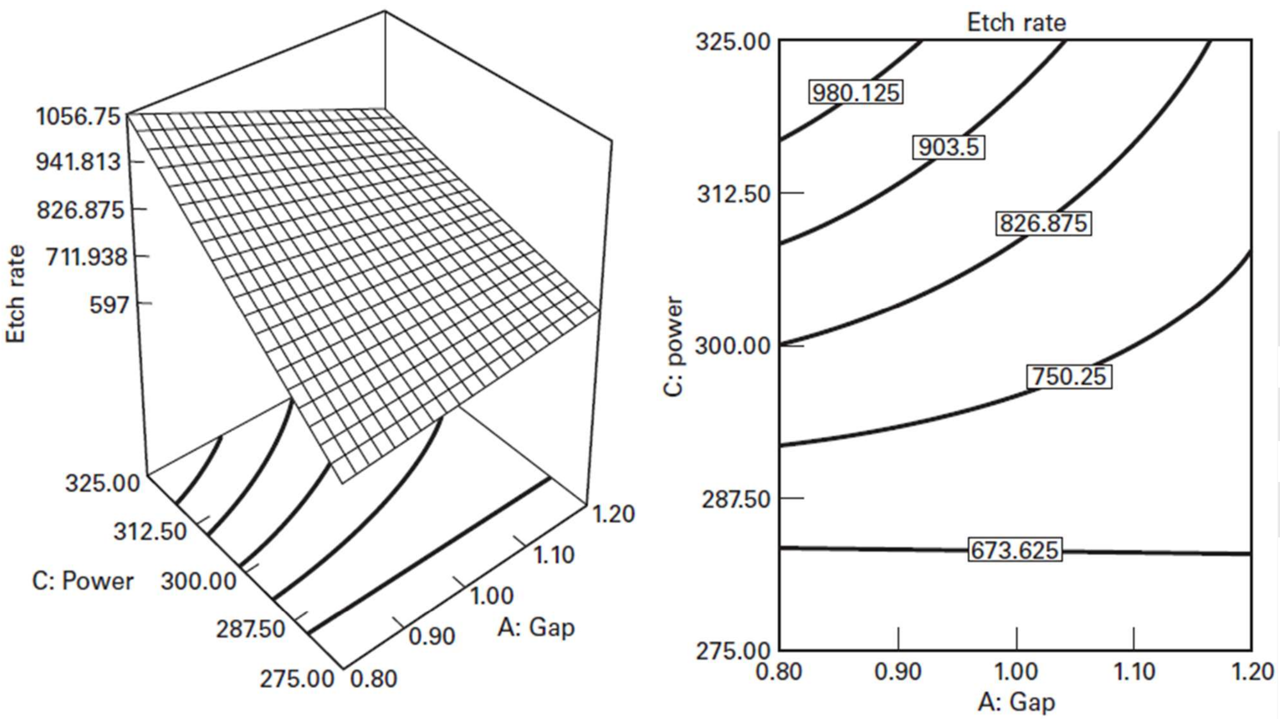
**Analysis of Variance for the Plasma Etching Experiment**

| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ | $P$-Value |
|---|---|---|---|---|---|
| Gap ($A$) | 41,310.5625 | 1 | 41,310.5625 | 18.34 | 0.0027 |
| Gas flow ($B$) | 217.5625 | 1 | 217.5625 | 0.10 | 0.7639 |
| Power ($C$) | 374,850.0625 | 1 | 374,850.0625 | 166.41 | 0.0001 |
| $AB$ | 2475.0625 | 1 | 2475.0625 | 1.10 | 0.3252 |
| $AC$ | 94,402.5625 | 1 | 94,402.5625 | 41.91 | 0.0002 |
| $BC$ | 18.0625 | 1 | 18.0625 | 0.01 | 0.9308 |
| $ABC$ | 126.5625 | 1 | 126.5625 | 0.06 | 0.8186 |
| Error | 18,020.5000 | 8 | 2252.5625 | | |
| Total | 531,420.9375 | 15 | | | |

# $2^K$ Factorial Design: The Regression Model and Response Surface

The **regression model for predicting etch rate** is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \hat{\beta}_{13} x_1 x_3 = 776.0625 + \left(\frac{-101.625}{2}\right) x_1 + \left(\frac{306.125}{2}\right) x_3 + \left(\frac{-153.625}{2}\right) x_1 x_3$$



($a$) The response surface
($b$) The contour plot

| Analysis of Variance for the Plasma Etching Experiment | | | | | |
|---|---|---|---|---|---|
| Source of Variation | Sum of Squares | Degrees of Freedom | Mean Square | $F_0$ | *P*-Value |
| Gap (*A*) | 41,310.5625 | 1 | 41,310.5625 | 18.34 | 0.0027 |
| Power (*C*) | 374,850.0625 | 1 | 374,850.0625 | 166.41 | 0.0001 |
| *AC* | 94,402.5625 | 1 | 94,402.5625 | 41.91 | 0.0002 |

# $2^K$ Factorial Design: The Regression Model and Response Surface

**ANOVA for Selected Factorial Model**
**Analysis of variance table [Partial sum of squares]**

| Source | Sum of Squares | DF | Mean Square | F Value | Prob > F |
|---|---|---|---|---|---|
| Model | 5.106E + 005 | 3 | 1.702E + 005 | 97.91 | < 0.0001 |
| *A* | *41310.56* | *1* | *41310.56* | *23.77* | *0.0004* |
| *C* | *3.749E + 005* | *1* | *3.749E + 005* | *215.66* | < 0.0001 |
| *AC* | *94402.56* | *1* | *94402.56* | *54.31* | < 0.0001 |
| Residual | 20857.75 | 12 | 1738.15 | | |
| *Lack of Fit* | *2837.25* | *4* | *709.31* | *0.31* | *0.8604* |
| *Pure Error* | *18020.50* | *8* | *2252.56* | | |
| Cor Total | 5.314E + 005 | 15 | | | |

| | |
|---|---|
| Std. Dev. | 41.69 |
| Mean | 776.06 |
| C.V. | 5.37 |
| PRESS | 37080.44 |

$$R^2 = \frac{SS_{\text{Model}}}{SS_{\text{Total}}} \qquad R^2_{\text{Adj}} = 1 - \frac{SS_E/df_E}{SS_{\text{Total}}/df_{\text{Total}}}$$

$$R^2_{\text{Pred}} = 1 - \frac{\text{PRESS}}{SS_{\text{Total}}}$$

| | |
|---|---|
| R-Squared | 0.9608 |
| Adj R-Squared | 0.9509 |
| Pred R-Squared | 0.9302 |
| Adeq Precision | 22.055 |

| Factor | Coefficient Estimated | DF | Standard Error | 95% CI Low | 95% CI High | VIF |
|---|---|---|---|---|---|---|
| Intercept | 776.06 | 1 | 10.42 | 753.35 | 798.77 | |
| *A*-Gap | −50.81 | 1 | 10.42 | −73.52 | 28.10 | 1.00 |
| *C*-Power | 153.06 | 1 | 10.42 | 130.35 | 175.77 | 1.00 |
| *AC* | −76.81 | 1 | 10.42 | −99.52 | −54.10 | 1.00 |

**Final Equation in Terms of Coded Factors:**

$$\text{Etch rate} = \boxed{\begin{array}{r} +776.06 \\ -50.81 \\ +153.06 \\ -76.81 \end{array}} \quad \boxed{\begin{array}{r} \\ * A \\ * C \\ * A * C \end{array}}$$

**Final Equation in Terms of Actual Factors:**

$$\text{Etch rate} = \boxed{\begin{array}{r} -5415.37500 \\ +4354.68750 \\ +21.48500 \\ -15.36250 \end{array}} \quad \boxed{\begin{array}{r} \\ * \text{Gap} \\ * \text{Power} \\ * \text{Gap} * \text{Power} \end{array}}$$

| Factor | Coefficient Estimated | DF | Standard Error | 95% CI Low | 95% CI High | VIF |
|---|---|---|---|---|---|---|
| Intercept | 776.06 | 1 | 10.42 | 753.35 | 798.77 | |
| *A*-Gap | −50.81 | 1 | 10.42 | −73.52 | 28.10 | 1.00 |
| *C*-Power | 153.06 | 1 | 10.42 | 130.35 | 175.77 | 1.00 |
| *AC* | −76.81 | 1 | 10.42 | −99.52 | −54.10 | 1.00 |

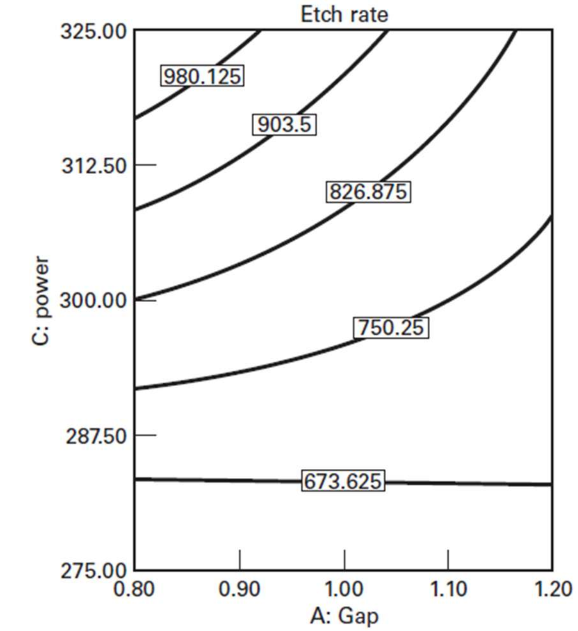
# $2^K$ Factorial Design: The Regression Model and Response Surface

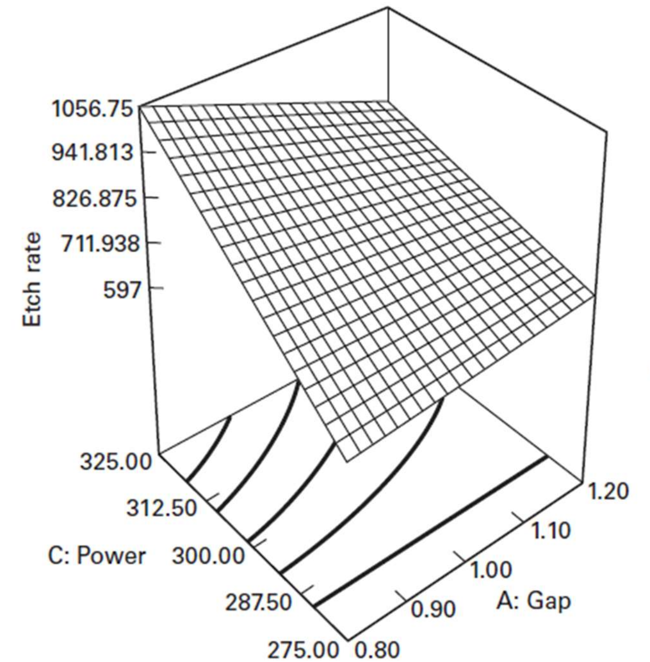
| | Coded Factors | | | Etch Rate | | | Factor Levels | | |
|---|---|---|---|---|---|---|---|---|---|
| Run | $A$ | $B$ | $C$ | Replicate 1 | Replicate 2 | Total | | Low (−1) | High (+1) |
| 1 | −1 | −1 | −1 | 550 | 604 | (1) = 1154 | $A$ (Gap, cm) | 0.80 | 1.20 |
| 2 | 1 | −1 | −1 | 669 | 650 | $a$ = 1319 | $B$ ($C_2F_6$ flow, SCCM) | 125 | 200 |
| 3 | −1 | 1 | −1 | 633 | 601 | $b$ = 1234 | $C$ (Power, W) | 275 | 325 |
| 4 | 1 | 1 | −1 | 642 | 635 | $ab$ = 1277 | | | |
| 5 | −1 | −1 | 1 | 1037 | 1052 | $c$ = 2089 | | | |
| 6 | 1 | −1 | 1 | 749 | 868 | $ac$ = 1617 | | | |
| 7 | −1 | 1 | 1 | 1075 | 1063 | $bc$ = 2138 | | | |
| 8 | 1 | 1 | 1 | 729 | 860 | $abc$ = 1589 | | | |

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_3 x_3 + \hat{\beta}_{13} x_1 x_3 = 776.0625 + \left(\frac{-101.625}{2}\right) x_1 + \left(\frac{306.125}{2}\right) x_3 + \left(\frac{-153.625}{2}\right) x_1 x_3$$

**Diagnostics Case Statistics**

| Standard Order | Actual Value | Predicted Value | Residual | Leverage | Student Residual | Cook's Distance | Outlier $t$ | Run Order |
|---|---|---|---|---|---|---|---|---|
| 1 | 550.00 | 597.00 | −47.00 | 0.250 | −1.302 | 0.141 | −1.345 | 9 |
| 2 | 604.00 | 597.00 | 7.00 | 0.250 | 0.194 | 0.003 | 0.186 | 6 |
| 3 | 669.00 | 649.00 | 20.00 | 0.250 | 0.554 | 0.026 | 0.537 | 14 |
| 4 | 650.00 | 649.00 | 1.00 | 0.250 | 0.028 | 0.000 | 0.027 | 1 |
| 5 | 633.00 | 597.00 | 36.00 | 0.250 | 0.997 | 0.083 | 0.997 | 3 |
| 6 | 601.00 | 597.00 | 4.00 | 0.250 | 0.111 | 0.001 | 0.106 | 12 |
| 7 | 642.00 | 649.00 | −7.00 | 0.250 | −0.194 | 0.003 | −0.186 | 13 |
| 8 | 635.00 | 649.00 | −14.00 | 0.250 | −0.388 | 0.013 | −0.374 | 8 |
| 9 | 1037.00 | 1056.75 | −19.75 | 0.250 | −0.547 | 0.025 | −0.530 | 5 |
| 10 | 1052.00 | 1056.75 | −4.75 | 0.250 | −0.132 | 0.001 | −0.126 | 16 |
| 11 | 749.00 | 801.50 | −52.50 | 0.250 | −1.454 | 0.176 | −1.534 | 2 |
| 12 | 868.00 | 801.50 | 66.50 | 0.250 | 1.842 | 0.283 | 2.082 | 15 |
| 13 | 1075.00 | 1056.75 | 18.25 | 0.250 | 0.505 | 0.021 | 0.489 | 4 |
| 14 | 1063.00 | 1056.75 | 6.25 | 0.250 | 0.173 | 0.002 | 0.166 | 7 |
| 15 | 729.00 | 801.50 | −72.50 | 0.250 | −2.008 | 0.336 | −2.359 | 10 |
| 16 | 860.00 | 801.50 | 58.50 | 0.250 | 1.620 | 0.219 | 1.755 | 11 |



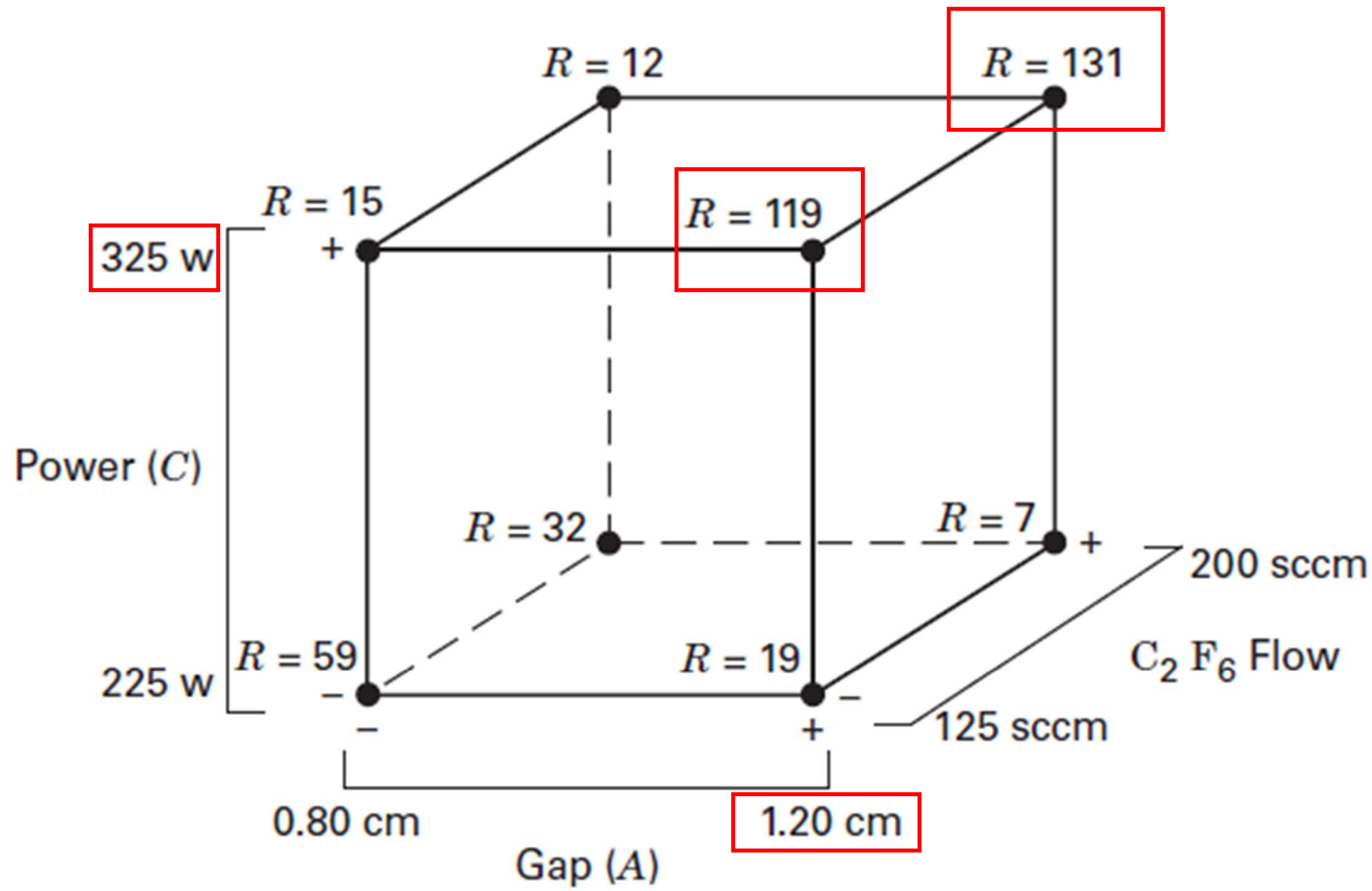



The process engineer working on the plasma etching tool was also interested in **dispersion effects**; that is, **do any of the factors affect variability in etch rate from run to run**?

One way to answer the question is to look at the **range of etch rates for each of the eight runs in the $2^3$ design**.

These ranges are plotted on the cube to the right (***R***).

Notice that the ranges in **etch rates are much larger when both Gap and Power are at their high levels**, indicating that this **combination of factor levels may lead to more variability in etch rate than other recipes**.



A potential concern in the use of two-level factorial designs is the **assumption of linearity in the factor effects**.

Of course, perfect linearity is unnecessary, and the $2^k$ system will work quite well even when the **linearity assumption holds only very approximately**.

In fact, we have noted that if interaction terms are added to a main effect or first-order model, resulting in

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j}\sum \beta_{ij} x_i x_j + \epsilon$$

then we have a model capable of **representing some curvature in the response function**.
This curvature, of course, results from the twisting of the plane induced by the interaction terms $\beta_{ij}x_i x_j$.

In some situations, the **curvature in the response function will not be adequately modeled by the above equation**.

In such cases, a logical model to consider is the **second-order response surface model**.

second-order or **quadratic effects**

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j}\sum \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} \boxed{x_j^2} + \varepsilon$$

In running a two-level factorial experiment, we usually anticipate fitting the **first-order model** in

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum\sum_{i<j} \beta_{ij} x_i x_j + \epsilon$$

but we should be alert to the possibility that the **second-order model** in

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum\sum_{i<j} \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \varepsilon$$

There is a method of **replicating certain points in a $2^k$ factorial that will provide protection against curvature from second-order effects** as well as allow an **independent estimate of error** to be obtained.

The method consists of adding **center points** to the $2^k$ design.