Chapter 2
Design of Experiments

All life is an experiment.
The more experiments you make the better.

Ralph Waldo Emerson,
Journals

2.1 Introduction to DOE

Within the theory of optimization, an experiment is a series of tests in which the input variables are changed according to a given rule in order to identify the reasons for the changes in the output response. According to Montgomery [4]

> "Experiments are performed in almost any field of enquiry and are used to study the performance of processes and systems. […] The process is a combination of machines, methods, people and other resources that transforms some input into an output that has one or more observable responses. Some of the process variables are controllable, whereas other variables are uncontrollable, although they may be controllable for the purpose of a test. The objectives of the experiment include: determining which variables are most influential on the response, determining where to set the influential controllable variables so that the response is almost always near the desired optimal value, so that the variability in the response is small, so that the effect of uncontrollable variables are minimized."

Thus, the purpose of experiments is essentially optimization and RDA. DOE, or experimental design, is the name given to the techniques used for guiding the choice of the experiments to be performed in an efficient way.

Usually, data subject to experimental error (noise) are involved, and the results can be significantly affected by noise. Thus, it is better to analyze the data with appropriate statistical methods. The basic principles of statistical methods in experimental design are replication, randomization, and blocking. *Replication* is the repetition of the experiment in order to obtain a more precise result (sample *mean value*) and to estimate the experimental error (sample *standard deviation*). *Randomization*

M. Cavazzuti, *Optimization Methods: From Theory to Design*,
DOI: 10.1007/978-3-642-31187-1_2, © Springer-Verlag Berlin Heidelberg 2013

refers to the random order in which the runs of the experiment are to be performed. In this way, the conditions in one run neither depend on the conditions of the previous run nor predict the conditions in the subsequent runs. *Blocking* aims at isolating a known systematic bias effect and prevent it from obscuring the main effects [5]. This is achieved by arranging the experiments in groups that are similar to one another. In this way, the sources of variability are reduced and the precision is improved.

Attention to the statistical issue is generally unnecessary when using numerical simulations in place of experiments, unless it is intended as a way of assessing the influence the noise factors will have in operation, as it is done in MORDO analysis.

Due to the close link between statistics and DOE, it is quite common to find in literature terms like *statistical experimental design*, or statistical DOE. However, since the aim of this chapter is to present some DOE techniques as a mean for collecting data to be used in RSM, we will not enter too deeply in the statistics which lies underneath the topic, since this would require a huge amount of work to be discussed.

Statistical experimental design, together with the basic ideas underlying DOE, was born in the 1920s from the work of Sir Ronald Aylmer Fisher [6]. Fisher was the statistician who created the foundations for modern statistical science. The second era for statistical experimental design began in 1951 with the work of Box and Wilson [7] who applied the idea to industrial experiments and developed the RSM. The work of Genichi Taguchi in the 1980s [8], despite having been very controversial, had a significant impact in making statistical experimental design popular and stressed the importance it can have in terms of quality improvement.

2.2 Terminology in DOE

In order to perform a DOE it is necessary to define the problem and choose the variables, which are called *factors* or parameters by the experimental designer. A design space, or *region of interest*, must be defined, that is, a range of variability must be set for each variable. The number of values the variables can assume in DOE is restricted and generally small. Therefore, we can deal either with qualitative discrete variables, or quantitative discrete variables. Quantitative continuous variables are discretized within their range. At first there is no knowledge on the solution space, and it may happen that the region of interest excludes the optimum design. If this is compatible with design requirements, the region of interest can be adjusted later on, as soon as the wrongness of the choice is perceived. The DOE technique and the number of levels are to be selected according to the number of experiments which can be afforded. By the term *levels* we mean the number of different values a variable can assume according to its discretization. The number of levels usually is the same for all variables, however some DOE techniques allow the differentiation of the number of levels for each variable. In experimental design, the objective function and the set of the experiments to be performed are called *response variable* and *sample space* respectively.

2.3 DOE Techniques

In this section some DOE techniques are presented and discussed. The list of the techniques considered is far from being complete since the aim of the section is just to introduce the reader into the topic showing the main techniques which are used in practice.

2.3.1 Randomized Complete Block Design

Randomized Complete Block Design (RCBD) is a DOE technique based on blocking. In an experiment there are always several factors which can affect the outcome. Some of them cannot be controlled, thus they should be randomized while performing the experiment so that on average, their influence will hopefully be negligible. Some other are controllable. RCBD is useful when we are interested in focusing on one particular factor whose influence on the response variable is supposed to be more relevant. We refer to this parameter with the term *primary factor*, design factor, control factor, or treatment factor. The other factors are called *nuisance factors* or disturbance factors. Since we are interested in focusing our attention on the primary factor, it is of interest to use the blocking technique on the other factors, that is, keeping constant the values of the nuisance factors, a batch of experiments is performed where the primary factor assumes all its possible values. To complete the randomized block design such a batch of experiments is performed for every possible combination of the nuisance factors.

Let us assume that in an experiment there are k controllable factors $X_1, \ldots X_k$ and one of them, X_k, is of primary importance. Let the number of levels of each factor be $L_1, L_2, \ldots, L_k$. If n is the number of replications for each experiment, the overall number of experiments needed to complete a RCBD (*sample size*) is $N = L_1 \cdot L_2 \cdot \ldots \cdot L_k \cdot n$. In the following we will always consider $n = 1$.

Let us assume: $k = 2$, $L_1 = 3$, $L_2 = 4$, X_1 nuisance factor, X_2 primary factor, thus $N = 12$. Let the three levels of X_1 be A, B, and C, and the four levels of X_2 be α, β, γ, and δ. The set of experiments for completing the RCBD DOE is shown in Table 2.1. Other graphical examples are shown in Fig. 2.1.

2.3.2 Latin Square

Using a RCBD, the sample size grows very quickly with the number of factors. Latin square experimental design is based on the same idea as the RCBD but it aims at reducing the number of samples required without confounding too much the importance of the primary factor. The basic idea is not to perform a RCBD but rather a single experiment in each block.

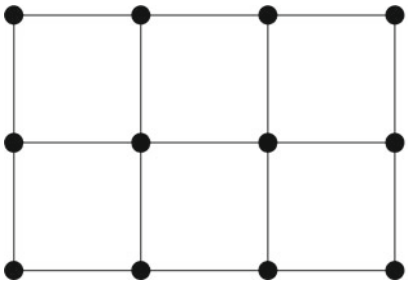
Latin square design requires some conditions to be respected by the problem for being applicable, namely: $k = 3$, X_1 and X_2 nuisance factors, X_3 primary factor, $L_1 = L_2 = L_3 = L$. The sample size of the method is $N = L^2$.

Table 2.1 Example of RCBD experimental design for $k = 2$, $L_1 = 3$, $L_2 = 4$, $N = 12$, nuisance factor X_1, primary factor X_2

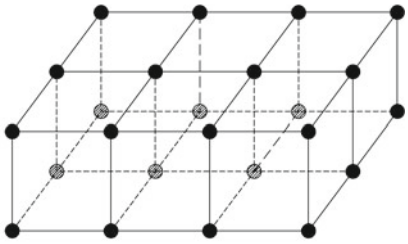
	Experiment Number	Factor Level X_1	Factor Level X_2
Block 1	1	A	α
	2	A	β
	3	A	γ
	4	A	δ
Block 2	5	B	α
	6	B	β
	7	B	γ
	8	B	δ
Block 3	9	C	α
	10	C	β
	11	C	γ
	12	C	δ

$\Longrightarrow$

X_1 \ X_2	α	β	γ	δ
A	Aα	Aβ	Aγ	Aδ
B	Bα	Bβ	Bγ	Bδ
C	Cα	Cβ	Cγ	Cδ



(a) $k = 2, L_1 = 4, L_2 = 3, N = 12$



(b) $k = 3, L_1 = 4, L_2 = 2, L_3 = 3, N = 24$

Fig. 2.1 Examples of RCBD experimental design

For representing the samples in a schematic way, the two nuisance factors are divided into a tabular grid with L rows and L columns. In each cell, a capital latin letter is written so that each row and each column receive the first L letters of the alphabet once. The row number and the column number indicate the level of the nuisance factors, the capital letters the level of the primary factor.

Actually, the idea of Latin square design is applicable for any $k > 3$, however the technique is known with different names, in particular:

- if $k = 3$: *Latin square*,
- if $k = 4$: *Graeco-Latin square*,
- if $k = 5$: *Hyper-Graeco-Latin square*.

Although the technique is still applicable, it is not given a particular name for $k > 5$. In the Graeco-Latin square or the Hyper-Graeco-Latin square designs, the

Table 2.2 Example of Latin square experimental design for $k = 3$, $L = 3$, $N = 9$

A	B	C
C	A	B
B	C	A

$\Longrightarrow$

Experiment Number	Factor Level		
	X_1 (row)	X_2 (column)	X_3 (letter)
1	1	1	A
2	1	2	B
3	1	3	C
4	2	1	C
5	2	2	A
6	2	3	B
7	3	1	B
8	3	2	C
9	3	3	A

additional nuisance factors are added as greek letters and other symbols (small letters, numbers or whatever) to the cells in the table. This is done in respect of the rule that in each row and in each column the levels of the factors must not be repeated, and to the additional rule that each factor must follow a different letters/numbers pattern in the table. The additional rule allows the influence of two variables not to be onfounded completely with each other. To fulfil this rule, it is not possible a Hyper-Graeco-Latin square design with $L = 3$ since there are only two possible letter pattern in a 3×3 table; if $k = 5$, L must be ≥ 4.

The advantage of the Latin square is that the design is able to keep separated several nuisance factors in a relatively cheap way in terms of sample size. On the other hand, since the factors are never changed one at a time from sample to sample, their effect is partially confounded.

For a better understanding of the way this experimental design works, some examples are given. Let us consider a Latin square design ($k = 3$) with $L = 3$, with X_3 primary factor. Actually, for the way this experimental design is built, the choice of the primary factor does not matter. A possible table pattern and its translation into a list of samples are shown in Table 2.2. The same design is exemplified graphically in Fig. 2.2.

Two more examples are given in Table 2.3, which shows a Graeco-Latin square design with $k = 4$, $L = 5$, $N = 25$, and a Hyper-Graeco-Latin square design with $k = 5$, $L = 4$, $N = 16$. Designs with $k > 5$ are formally possible, although they are usually not discussed in the literature. More design tables are given by Box et al. in [9].

2.3.3 Full Factorial

Full factorial is probably the most common and intuitive strategy of experimental design. In the most simple form, the two-levels full factorial, there are k factors and $L = 2$ levels per factor. The samples are given by every possible combination of the factors values. Therefore, the sample size is $N = 2^k$. Unlike the previous DOE

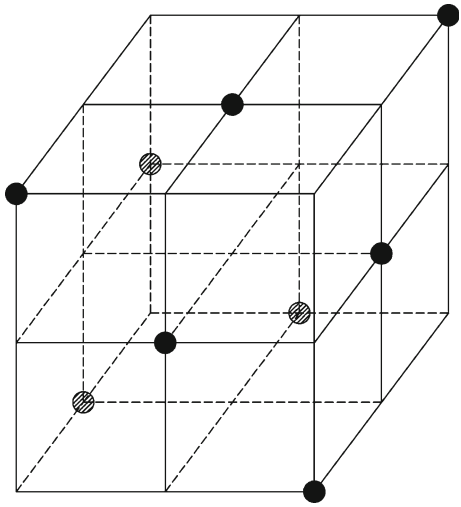


Fig. 2.2 Example of Latin square experimental design for $k = 3$, $L = 3$, $N = 9$

Table 2.3 Example of Graeco-Latin square and Hyper-Graeco-Latin square experimental design

(a) $k = 4$, $L = 5$, $N = 25$

$A\alpha$	$B\beta$	$C\gamma$	$D\delta$	$E\epsilon$
$C\beta$	$D\gamma$	$E\delta$	$A\epsilon$	$B\alpha$
$E\gamma$	$A\delta$	$B\epsilon$	$C\alpha$	$D\beta$
$B\delta$	$C\epsilon$	$D\alpha$	$E\beta$	$A\gamma$
$D\epsilon$	$E\alpha$	$A\beta$	$B\gamma$	$C\delta$

(b) $k = 5$, $L = 4$, $N = 16$

$A\alpha 1$	$B\beta 2$	$C\gamma 3$	$D\delta 4$
$C\delta 2$	$D\gamma 1$	$A\beta 4$	$B\alpha 3$
$D\beta 3$	$C\alpha 4$	$B\delta 1$	$A\gamma 2$
$B\gamma 4$	$A\delta 3$	$D\alpha 2$	$C\beta 1$

methods, this method and the following ones do not distinguish anymore between nuisance and primary factors *a priori*. The two levels are called high ("*h*") and low ("*l*") or, "+1" and "−1". Starting from any sample within the full factorial scheme, the samples in which the factors are changed one at a time are still part of the sample space. This property allows for the effect of each factor over the response variable not to be confounded with the other factors. Sometimes, in literature, it happens to encounter full factorial designs in which also the central point of the design space is added to the samples. The central point is the sample in which all the parameters have a value which is the average between their low and high level and in 2^k full factorial tables can be individuated with "*m*" (mean value) or "0".

Let us consider a full factorial design with three factors and two levels per factor (Table 2.4). The full factorial is an *orthogonal* experimental design method. The term orthogonal derives from the fact that the scalar product of the columns of any two-factors is zero.

We define the *main interaction M* of a variable *X* the difference between the average response variable at the high level samples and the average response at the

Table 2.4 Example of 2^3 full factorial experimental design

Experiment number	Factor level			Response variable	Two- and three-factors interactions			
	X_1	X_2	X_3		$X_1 \cdot X_2$	$X_1 \cdot X_3$	$X_2 \cdot X_3$	$X_1 \cdot X_2 \cdot X_3$
1	-1 (l)	-1 (l)	-1 (l)	$y_{l,l,l}$	$+1$	$+1$	$+1$	-1
2	-1 (l)	-1 (l)	$+1$ (h)	$y_{l,l,h}$	$+1$	-1	-1	$+1$
3	-1 (l)	$+1$ (h)	-1 (l)	$y_{l,h,l}$	-1	$+1$	-1	$+1$
4	-1 (l)	$+1$ (h)	$+1$ (h)	$y_{l,h,h}$	-1	-1	$+1$	-1
5	$+1$ (h)	-1 (l)	-1 (l)	$y_{h,l,l}$	-1	-1	$+1$	$+1$
6	$+1$ (h)	-1 (l)	$+1$ (h)	$y_{h,l,h}$	-1	$+1$	-1	-1
7	$+1$ (h)	$+1$ (h)	-1 (l)	$y_{h,h,l}$	$+1$	-1	-1	-1
8	$+1$ (h)	$+1$ (h)	$+1$ (h)	$y_{h,h,h}$	$+1$	$+1$	$+1$	$+1$

low level samples. In the example in Table 2.4, for X_1 we have

$$M_{X_1} = \frac{y_{h,l,l} + y_{h,l,h} + y_{h,h,l} + y_{h,h,h}}{4} - \frac{y_{l,l,l} + y_{l,l,h} + y_{l,h,l} + y_{l,h,h}}{4}. \quad (2.1)$$

Similar expressions can be derived for M_{X_2} and M_{X_3}. The *interaction effect* of two or more factors is defined similarly as the difference between the average responses at the high level and at the low level in the interaction column. The two-factors interaction effect between X_1 and X_2 following Table 2.4 is

$$M_{X_1,X_2} = \frac{y_{l,l,l} + y_{l,l,h} + y_{h,h,l} + y_{h,h,h}}{4} - \frac{y_{h,l,l} + y_{h,l,h} + y_{l,h,l} + y_{l,h,h}}{4}. \quad (2.2)$$

The main and the interaction effects give a quantitative estimation of the influence the factors, or the interaction of the factors, have upon the response variable. The number of main and interaction effects in a 2^k full factorial design is $2^k - 1$; it is also said that a 2^k full factorial design has $2^k - 1$ *degree of freedom*. The subdivision of the number of main and interaction effects follows the Tartaglia triangle [10], also known as Khayyam triangle, or Pascal triangle: in a 2^k full factorial design there are $\binom{k}{1} = \frac{k!}{1!(k-1)!} = k$ main effects, $\binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{k(k-1)}{2}$ two-factors interactions, $\binom{k}{j} = \frac{k!}{j!(k-j)!}$ j-factors interactions, and so on.

The idea of the 2^k full factorial experimental designs can be easily extended to the general case where there are more than two factors and each of them have a different number of levels. The sample size of the *adjustable* full factorial design with k factors $X_1, \ldots, X_k$, having $L_1, \ldots, L_k$ levels, is $N = \prod_{i=1}^{k} L_i$.

At this point, the careful reader has probably noted that the sample space of the adjustable full factorial design is equivalent to the one of the RCBD. Therefore, we could argue that the RCBD is essentially the more general case of a full factorial design. It is true, however, that in the RCBD the focus is generally on a single variable (the primary factor), and a particular stress is put on blocking and randomization techniques. It is not just a problem of sampling somehow a design space since, in fact, the order of the experiments and the way in which they are performed matter.

In adjustable full factorial designs, it is still possible to compute an estimation of the main and the interaction effects, but the definitions given above must be reformulated in terms of sums of squares. Let us consider the case of $k = 4$, the average of the response variable for all the N samples is

$$\bar{y} = \frac{\sum_{i=1}^{L_1}\sum_{j=1}^{L_2}\sum_{l=1}^{L_3}\sum_{m=1}^{L_4} y_{i,j,l,m}}{N}. \tag{2.3}$$

In order to compute the main effect of X_1, we must evaluate the L_1 averages of the response variables for all the samples where X_1 is fixed to a certain level

$$\bar{y}_{X_1=1} = \frac{\sum_{j=1}^{L_2}\sum_{l=1}^{L_3}\sum_{m=1}^{L_4} y_{1,j,l,m}}{L_2 \cdot L_3 \cdot L_4} \quad \cdots \quad \bar{y}_{X_1=L_1} = \frac{\sum_{j=1}^{L_2}\sum_{l=1}^{L_3}\sum_{m=1}^{L_4} y_{L_1,j,l,m}}{L_2 \cdot L_3 \cdot L_4}. \tag{2.4}$$

The main effect of X_1 is

$$M_{X_1} = \sum_{i=1}^{L_1} \left(\bar{y}_{X_1=i} - \bar{y}\right)^2. \tag{2.5}$$

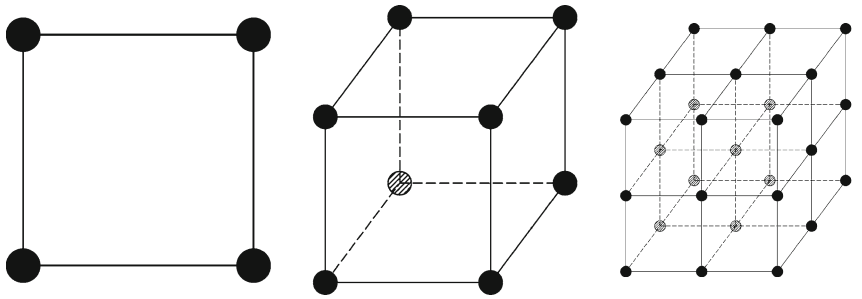
In a similar way, for computing a two-factors interaction effect, namely X_1 and X_2, we need to compute the $L_1 \cdot L_2$ averages

$$\bar{y}_{X_1=1,X_2=1} = \frac{\sum_{l=1}^{L_3}\sum_{m=1}^{L_4} y_{1,1,l,m}}{L_3 \cdot L_4} \quad \cdots \quad \bar{y}_{X_1=L_1,X_2=L_2} = \frac{\sum_{l=1}^{L_3}\sum_{m=1}^{L_4} y_{L_1,L_2,l,m}}{L_3 \cdot L_4}. \tag{2.6}$$

The X_1, X_2 interaction effect is

$$M_{X_1,X_2} = \sum_{i=1}^{L_1}\sum_{j=1}^{L_2} \left(\bar{y}_{X_1=i,X_2=j} - \bar{y}\right)^2 - M_{X_1} - M_{X_2}. \tag{2.7}$$

The advantage of full factorial designs is that they make a very efficient use of the data and do not confound the effects of the parameters, so that it is possible to evaluate the main and the interaction effects clearly. On the other hand, the sample size grows exponentially with the number of parameters and the number of levels. The family of the L^k designs, that is, the full factorial designs where the number of levels is the same for each factor, is particularly suitable for interpolation by polynomial *response surfaces*, since a 2^k design can be interpolated with a complete bilinear form, a 3^k design with a complete biquadratic form, a 4^k with a complete



(a) 2^2 **(b)** 2^3 **(c)** 3^3

Fig. 2.3 Example of L^k full factorial experimental designs

bicubic, and so on. However, bilinear and biquadratic interpolations are generally poor for a good response surface to be generated. We refer to the terms bilinear, biquadratic, and bicubic broadly speaking, since the number of factors is k, not two, and we should better speak of k-linear, k-quadratic, and k-cubic interpolations.

Figure 2.3 shows graphical representations for the 2^2, the 2^3 and the 3^3 full factorial designs.

2.3.4 Fractional Factorial

As the number of parameters increases, a full factorial design may become very onerous to be completed. The idea of the *fractional factorial* design is to run only a subset of the full factorial experiments. Doing so, it is still possible to provide quite good information on the main effects and some information about interaction effects. The sample size of the fractional factorial can be *one-half*, or *one-quarter*, and so on, of the full factorial one. The fractional factorial samples must be properly chosen, in particular they have to be *balanced* and orthogonal. By balanced we mean that the sample space is made in such a manner so that each factor has the same number of samples for each of its levels.

Let us consider a one-half fractional factorial of a 2^k full factorial design. The one-half is referred to as 2^{k-1} fractional factorial. Let us assume $k = 3$. In order to build the list of the samples, we start with a regular full factorial 2^{k-1} (Table 2.5), the levels for the additional parameter are chosen as an interaction of some of the other parameters. In our case, we could add the product $X_1 \cdot X_2$ or $-X_1 \cdot X_2$.

The fractional factorial design in Table 2.5 is said to have *generator* or *word* $+ABC$ because the element-by-element multiplication of the first (A), the second (B), and the third (C) column is equal to the identity column I. The main and the interaction effects are computed as in the previous paragraph. However, the price to pay, in such an experimental design, is that it is not possible to distinguish between the main effect of X_3 (C) and the $X_1 \cdot X_2$ (AB) interaction effect. In technical terms we say that X_3 has been *confounded*, or *aliased* with $X_1 \cdot X_2$. However, this is not the

Table 2.5 Example of 2^{3-1} fractional factorial experimental design

Experiment number	Factor level X_1 (A)	X_2 (B)	$X_3 = X_1 \cdot X_2$ (C)	$I = X_1 \cdot X_2 \cdot X_3$
1	-1	-1	$+1$	$+1$
2	-1	$+1$	-1	$+1$
3	$+1$	-1	-1	$+1$
4	$+1$	$+1$	$+1$	$+1$

only confounded term: multiplying the columns suitably, we realize that, if $C = AB$, we have $AC = A \cdot AB = B$ and $BC = B \cdot AB = A$, that is, every main effect is confounded with a two-factors interaction effect.

The 2^{3-1} design with generator $I = +ABC$ (or $I = -ABC$) is a *resolution III design*. For denoting the design resolution a roman numeral subscript is used (2^{3-1}_{III}). A design is said to be of resolution R if no q-factors effect is aliased with another effect with less than $R - q$ factors. This means that:

- in a resolution III design the main effects are aliased with at least 2-factors effects,
- in a resolution IV design the main effects are aliased with at least 3-factors effects and the 2-factors effects are aliased with each other,
- in a resolution V design the main effects are aliased with at least 4-factors effects and the 2-factors effects are aliased with at least 3-factors effects.

In general, the definition of a 2^{k-p} design requires p "words" to be given. Considering all the possible aliases these become $2^p - 1$ words. The resolution is equal to the smallest number of letters in any of the $2^p - 1$ defining words. The $2^p - 1$ words are found multiplying the p original words with each other in every possible combination. The resolution tells how badly the design is confounded. The higher is the resolution of the method, the better the results are expected to be. It must be considered that resolution depends on the choice of the defining words, therefore the words must be chosen accurately in order to reach the highest possible resolution.

Table 2.6 shows an example of a 2^{6-2} design with the evaluation of its resolution and the list of the main effect and the two-factors interaction aliases.

The same idea for building fractional factorial designs can be generalized to a L^{k-p} design, or to factorial designs with a different number of levels for each factor. We start writing down the set of samples for a L^{k-p} full factorial design, then the levels for the remaining p columns are obtained from particular combinations of the other $k - p$ columns. In the same way shown above, it is possible to compute the aliases and the resolution of the design. Although the concept is the same, things are a bit more complicated since the formulas giving the last p columns are not defined on a sort of binary numeral system anymore, but need to be defined according to different systems with different number of levels.

Figure 2.4 show a few graphical examples of fractional factorial designs. A wide list of tables for the most common designs can be found in literature [4, 5] .

Table 2.6 Example of 2^{6-2} fractional factorial experimental design and evaluation of the design resolution

Design 2^{6-2}	Main effect aliases	Two-factors interaction aliases
	$A = BCE = ABCDF = DEF$	$AB = CE = ACDF = BDEF$
Defining Words	$B = ACE = CDF = ABDEF$	$AC = BE = ABDF = CDEF$
$I = ABCE$	$C = ABE = BDF = ACDEF$	$AD = EF = BCDE = ABCF$
$I = BCDF$	$D = ABCDE = BCF = AEF$	$AE = BC = DF = ABCDEF$
$I = ADEF$	$E = ABC = BCDEF = ADF$	$AF = DE = BDEF = ABCD$
Resolution	$F = ABCEF = BCD = ADE$	$BD = CF = ACDE = ABEF$
IV		$BF = CD = ACEF = ABDE$

Experiment number	Factor level $X_1(A)$	$X_2(B)$	$X_3(C)$	$X_4(D)$	$X_5(E)$	$X_6(F)$
1	−1	−1	−1	−1	−1	−1
2	−1	−1	−1	+1	−1	+1
3	−1	−1	+1	−1	+1	+1
4	−1	−1	+1	+1	+1	−1
5	−1	+1	−1	−1	+1	+1
6	−1	+1	−1	+1	+1	−1
7	−1	+1	+1	−1	−1	−1
8	−1	+1	+1	+1	−1	+1
9	+1	−1	−1	−1	+1	−1
10	+1	−1	−1	+1	+1	+1
11	+1	−1	+1	−1	−1	+1
12	+1	−1	+1	+1	−1	−1
13	+1	+1	−1	−1	−1	+1
14	+1	+1	−1	+1	−1	−1
15	+1	+1	+1	−1	+1	−1
16	+1	+1	+1	+1	+1	+1

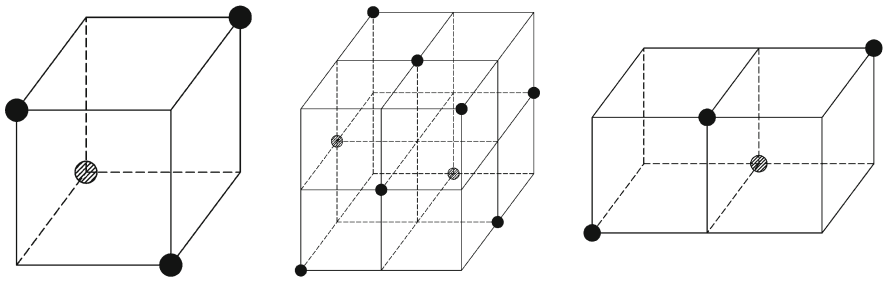
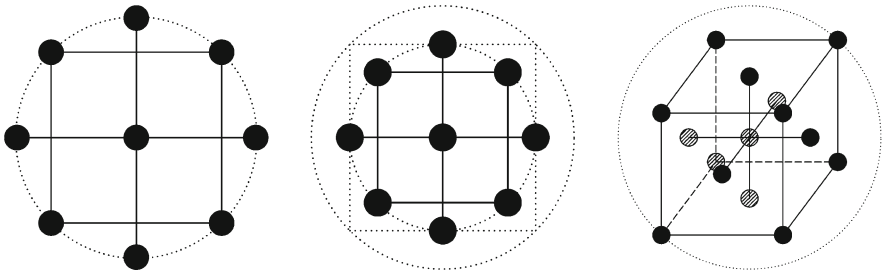


Fig. 2.4 Example of fractional factorial experimental designs

It must be noted that Latin square designs are equivalent to specific fractional factorial designs. For instance, a Latin square with L levels per factor is the same as a L^{3-1} fractional factorial design.



(a) 2^2CCC **(b)** 2^2CCI **(c)** 2^3CCF

Fig. 2.5 Example of central composite experimental designs

2.3.5 Central Composite

A *central composite* design is a 2^k full factorial to which the central point and the *star points* are added. The star points are the sample points in which all the parameters but one are set at the mean level "*m*". The value of the remaining parameter is given in terms of distance from the central point. If the distance between the central point and each full factorial sample is normalized to 1, the distance of the star points from the central point can be chosen in different ways:

- if it is set to 1, all the samples are placed on a hypersphere centered in the central point (*central composite circumscribed*, or CCC). The method requires five levels for each factor, namely *ll*, *l*, *m*, *h*, *hh*,
- if it is set to $\frac{\sqrt{k}}{k}$, the value of the parameter remains on the same levels of the 2^k full factorial (*central composite faced*, or CCF). The method requires three levels for each factor, namely *l*, *m*, *h*,
- if a sampling like the central composite circumscribed is desired, but the limits specified for the levels cannot be violated, the CCC design can be scaled down so that all the samples have distance from the central point equal to $\frac{\sqrt{k}}{k}$ (*central composite inscribed*, or CCI). The method requires five levels for each factor, namely *l*, *lm*, *m*, *mh*, *h*,
- if the distance is set to any other value, whether it is $<\frac{\sqrt{k}}{k}$ (star points inside the design space), <1 (star points inside the hypersphere), or >1 (star points outside the hypersphere), we talk of *central composite scaled*, or CCS. The method requires five levels for each factor.

For k parameters, $2k$ star points and one central point are added to the 2^k full factorial, bringing the sample size for the central composite design to 2^k+2k+1. The fact of having more samples than those strictly necessary for a bilinear interpolation (which are 2^k), allows the curvature of the design space to be estimated.

Figure 2.5 shows a few graphical examples of central composite experimental designs.

Table 2.7 Box-Behnken tables for $k = 3$, $k = 4$, $k = 5$, and $k = 6$

(a) 3 parameters: 3 out of 3 blocks with 2^2 full factorial, plus the central point, 13 samples overall, 10 coefficients needed for fitting a 2^{nd} order interpolating polynomial

Block \ Parameter			
	±	±	0
	±	0	±
	0	±	±
	0	0	0

(b) 4 parameters: 6 out of 6 blocks with 2^2 full factorial, plus the central point, 25 samples overall, 15 coefficients needed for fitting a 2^{nd} order interpolating polynomial

±	±	0	0
±	0	±	0
±	0	0	±
0	±	±	0
0	±	0	±
0	0	±	±
0	0	0	0

(c) 5 parameters: 10 out of 10 blocks with 2^2 full factorial, plus the central point, 41 samples overall, 21 coefficients needed for fitting a 2^{nd} order interpolating polynomial

±	±	0	0	0
±	0	±	0	0
±	0	0	±	0
±	0	0	0	±
0	±	±	0	0
0	±	0	±	0
0	±	0	0	±
0	0	±	±	0
0	0	±	0	±
0	0	0	±	±
0	0	0	0	0

(d) 6 parameters: 6 out of 20 blocks with 2^3 full factorial, plus the central point, 49 samples overall, 28 coefficients needed for fitting a 2^{nd} order interpolating polynomial

±	±	0	±	0	0
0	±	±	0	±	0
0	0	±	±	0	±
±	0	0	±	±	0
0	±	0	0	±	±
±	0	±	0	0	±
0	0	0	0	0	0

2.3.6 Box-Behnken

Box-Behnken [11] are incomplete three-levels factorial designs. They are built combining two-levels factorial designs with incomplete block designs in a particular manner. Box-Behnken designs were introduced in order to limit the sample size as the number of parameters grows. The sample size is kept to a value which is sufficient for the estimation of the coefficients in a second degree least squares approximating polynomial. In Box-Behnken designs, a block of samples corresponding to a two-levels factorial design is repeated over different sets of parameters. The parameters which are not included in the factorial design remain at their mean level throughout the block. The type (full or fractional), the size of the factorial, and the number of blocks which are evaluated, depend on the number of parameters and it is chosen so that the design meets, exactly or approximately, the criterion of *rotatability*. An experimental design is said to be rotatable if the variance of the predicted response at any point is a function of the distance from the central point alone.

Since there is not a general rule for defining the samples of the Box-Behnken designs, tables are given by the authors for the range from three to seven, from nine to twelve and for sixteen parameters. For better understandability of this experimental design technique, Table 2.7 shows a few examples. In the table, each line stands for a factorial design block, the symbol "±" individuates the parameters on which the

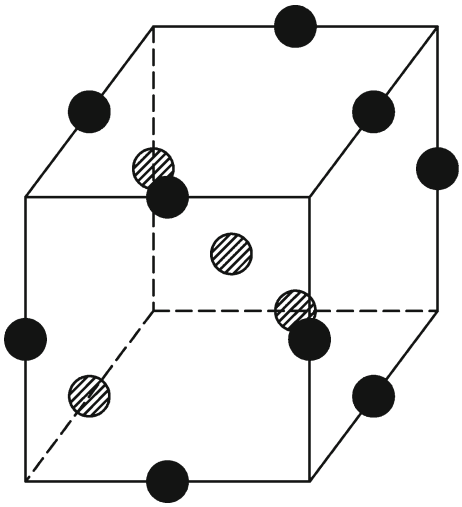


Fig. 2.6 Example of Box-Behnken experimental design for $k = 3$

factorial design is made, "0" stands for the variables which are blocked at the mean level.

Let us consider the Box-Behnken design with three parameters (Table 2.7a), in this case a 2^2 full factorial is repeated three times:

i. on the first and the second parameters keeping the third parameter at the mean level (samples: llm, lhm, hlm, hhm),
ii. on the first and the third parameters keeping the second parameter at the mean level (samples: lml, lmh, hml, hmh),
iii. on the second and the third parameters keeping the first parameter at the mean level (samples: mll, mlh, mhl, mhh),

then the central point (mmm) is added. Graphically, the samples are at the midpoints of the edges of the design space and in the centre (Fig. 2.6). An hypothetical graphical interpretation for the $k = 4$ case is that the samples are placed at each midpoint of the twenty-four two-dimensional faces of the four-dimensional design space and in the centre.

As for the CCC and the CCI, all the samples have the same distance from the central point. The vertices of the design space lie relatively far from the samples and on the outside of their convex hull, for this reason a response surface based on a Box-Behnken experimental design may be inaccurate near the vertices of the design space. The same happens for CCI designs.

2.3.7 Plackett-Burman

Plackett-Burman are very economical, two-levels, resolution III designs [12]. The sample size must be a multiple of four up to thirty-six, and a design with N samples can be used to study up to $k = N - 1$ parameters. Of course, as the method requires

a very small number of experiments, the main effects are heavily confounded with two-factors interactions and Plackett-Burman designs are useful just for screening the design space to detect large main effects. As in the case of Box-Behnken, Plackett-Burman designs do not have a clear defining relation and tables for a different number of factors are given by the authors. For N which is a power of two, the designs are equivalent to 2^{k-p}_{III} fractional factorial designs, where $2^{k-p} = N$. In Plackett-Burman designs, a main effect column X_i is either orthogonal to any $X_i \cdot X_j$ two-factors interaction or identical to plus or minus $X_i \cdot X_j$.

The cases $N = 4$, $N = 8$, $N = 16$, $N = 32$ are equivalent to 2^{3-1}, 2^{7-4}, 2^{15-11}, 2^{31-26} fractional factorial designs. For the cases $N = 12$, $N = 20$, $N = 24$, $N = 36$ a row of 11, 19, 23, and 35 plus (high level) and minus signs (low level) is given (Table 2.8). The Plackett-Burman designs are obtained writing the appropriate row as the first row of the design table. The second row is generated by shifting the elements of the first row one place right, and so on for the other rows. In the end, a row of minus signs is added. Table 2.8 shows the Plackett-Burman patterns for $N = 12$, $N = 20$, $N = 24$, $N = 36$, and the sample space for the case $N = 12$. The designs for the $N = 28$ case are built in a different way: three patterns of 9×9 plus and minus signs are given, and these patterns are assembled in a 27×27 table, then a row of minus signs is added in the end as usual. In Plackett-Burman designs, if the parameters are less than $N - 1$, the first k columns are taken and the $N - 1 - k$ last columns of the design table are discarded.

2.3.8 Taguchi

The *Taguchi* method was developed by Genichi Taguchi [8] in Japan to improve the implementation of off-line total quality control. The method is related to finding the best values of the controllable factors to make the problem less sensitive to the variations in uncontrollable factors. This kind of problem was called by Taguchi *robust parameter design problem*.

Taguchi method is based on mixed levels, highly fractional factorial designs, and other orthogonal designs. It distinguishes between *control variables*, which are the factors that can be controlled, and *noise variables*, which are the factors that cannot be controlled except during experiments in the lab. Two different orthogonal designs are chosen for the two sets of parameters. We call *inner array* the design chosen for the controllable variables, and *outer array* the design chosen for the noise variables. The combination of the inner and the outer arrays give the *crossed array* which is the list of all the samples scheduled by the Taguchi method. By combination we mean that for each sample in the inner array the full set of experiments of the outer array is performed. An important point about the crossed array Taguchi design is that, in this way, it provides information about the interaction between the controllable variables and the noise variables. These interactions are crucial for a robust solution.

Let us consider a problem with five parameters ($k = 5$), three of which are controllable ($k_{in} = 3$) and two uncontrollable ($k_{out} = 2$), and let us consider two-levels

Table 2.8 Plackett-Burman patterns for $N = 12$, $N = 20$, $N = 24$, $N = 36$, and example of Plackett-Burman experimental design for $k = 11$

k	N	Plackett-Burman pattern
11	12	$+\,+\,-\,+\,+\,+\,-\,-\,-\,+\,-$
19	20	$+\,+\,-\,-\,+\,+\,+\,+\,-\,+\,-\,+\,-\,-\,-\,-\,-\,+\,+\,-$
23	24	$+\,+\,+\,+\,+\,-\,+\,-\,+\,+\,-\,-\,+\,+\,-\,-\,+\,-\,+\,-\,-\,-\,-$
35	36	$-\,+\,-\,+\,+\,+\,-\,-\,-\,+\,+\,+\,+\,+\,-\,+\,+\,+\,-\,-\,+\,-\,-\,-\,-\,-\,+\,-\,+\,-\,+$ $+\,-\,-\,+\,-$

Experiment number	Parameter X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	+1	+1	−1	+1	+1	+1	−1	−1	−1	+1	−1
2	−1	+1	+1	−1	+1	+1	+1	−1	−1	−1	+1
3	+1	−1	+1	+1	−1	+1	+1	+1	−1	−1	−1
4	−1	+1	−1	+1	+1	−1	+1	+1	+1	−1	−1
5	−1	−1	+1	−1	+1	+1	−1	+1	+1	+1	−1
6	−1	−1	−1	+1	−1	+1	+1	−1	+1	+1	+1
7	+1	−1	−1	−1	+1	−1	+1	+1	−1	+1	+1
8	+1	+1	−1	−1	−1	+1	−1	+1	+1	−1	+1
9	+1	+1	+1	−1	−1	−1	+1	−1	+1	+1	−1
10	−1	+1	+1	+1	−1	−1	−1	+1	−1	+1	+1
11	+1	−1	+1	+1	+1	−1	−1	−1	+1	−1	+1
12	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1

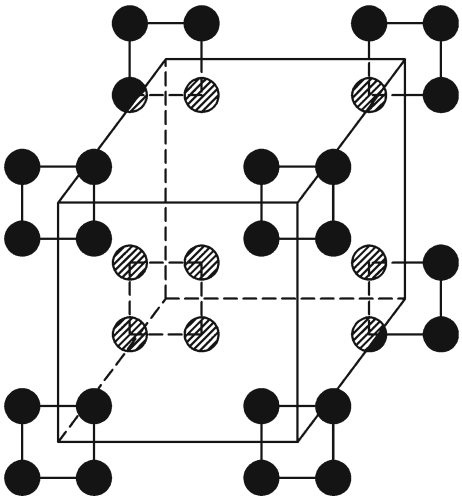


Fig. 2.7 Example of Taguchi DOE for $k_{in} = 3$, $k_{out} = 2$, 2^3 full factorial inner array, 2^2 full factorial outer array

full factorial experimental designs for the inner and the outer arrays. We assume full factorial designs for simplicity, even though they are never taken into consideration by the Taguchi method. Therefore, we must perform a full 2^2 factorial design (outer array) for each sample of the 2^3 inner array. We can graphically represent the situation as in Fig. 2.7.

Table 2.9 Example of Taguchi DOE for $k_{in} = 3$, $k_{out} = 2$, 2^3 full factorial inner array, 2^2 full factorial outer array

	Inner aray			Outer array						Output	
Exp. num	Parameter			Exp.num		1	2	3	4	Mean	Std. deviation
	$X_{in,1}$	$X_{in,2}$	$X_{in,3}$	Par.	$X_{out,1}$	-1	-1	$+1$	$+1$		
					$X_{out,2}$	-1	$+1$	-1	$+1$		
1	-1	-1	-1			$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{1,4}$	$\mathrm{E}[y_1]$	$\mathrm{E}[(y_1 - \mathrm{E}[y_1])^2]$
2	-1	-1	$+1$			$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{2,4}$	$\mathrm{E}[y_2]$	$\mathrm{E}[(y_2 - \mathrm{E}[y_2])^2]$
3	-1	$+1$	-1			$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	$y_{3,4}$	$\mathrm{E}[y_3]$	$\mathrm{E}[(y_3 - \mathrm{E}[y_3])^2]$
4	-1	$+1$	$+1$			$y_{4,1}$	$y_{4,2}$	$y_{4,3}$	$y_{4,4}$	$\mathrm{E}[y_4]$	$\mathrm{E}[(y_4 - \mathrm{E}[y_4])^2]$
5	$+1$	-1	-1			$y_{5,1}$	$y_{5,2}$	$y_{5,3}$	$y_{5,4}$	$\mathrm{E}[y_5]$	$\mathrm{E}[(y_5 - \mathrm{E}[y_5])^2]$
6	$+1$	-1	$+1$			$y_{6,1}$	$y_{6,2}$	$y_{6,3}$	$y_{6,4}$	$\mathrm{E}[y_6]$	$\mathrm{E}[(y_6 - \mathrm{E}[y_6])^2]$
7	$+1$	$+1$	-1			$y_{7,1}$	$y_{7,2}$	$y_{7,3}$	$y_{7,4}$	$\mathrm{E}[y_7]$	$\mathrm{E}[(y_7 - \mathrm{E}[y_7])^2]$
8	$+1$	$+1$	$+1$			$y_{8,1}$	$y_{8,2}$	$y_{8,3}$	$y_{8,4}$	$\mathrm{E}[y_8]$	$\mathrm{E}[(y_8 - \mathrm{E}[y_8])^2]$

Using $L^{k_{in}}$ and $L^{k_{out}}$ full factorial designs the Taguchi method is equivalent to a generic $L^{k_{in}+k_{out}}$ full factorial, and using fractional factorial designs or other orthogonal designs, the outcome in terms of number and distribution of the samples would not be too different from some fractional factorial over the whole number of parameters $k_{in}+k_{out}$. However, the stress is on the distinction between controllable variables and noise variables. Looking at the design as a way of performing a set of samples (outer array) for each sample in the inner array allows us to estimate the mean value and the standard deviation, or other statistical values for each design point as noise enters the system. The aim then is to improve the average performance of the problem while keeping the standard deviation low. This idea is shown in Table 2.9 for the example given above and summarized in Fig. 2.7. Actually, Taguchi did not consider the mean response variable and its standard deviation as performance measures. He introduced more than sixty different performance measures to be maximized, which he called *signal-to-noise ratios* (SN). Depending on the nature of the investigated problem, an appropriate ratio can be chosen. These performance measures, however, have not met much success in that their responses are not always meaningful for the problem. The most well-known signal-to-noise ratios are [13]:

- *smaller-the-better*: to be used when the response variable is to be minimized.

$$\mathrm{SN}_{stb} = -10 \log_{10} \mathrm{E}\left[y_i^2\right] \tag{2.8}$$

- *larger-the-better*: to be used then the response variable is to be maximized.

$$\mathrm{SN}_{ltb} = -10 \log_{10} \mathrm{E}\left[\frac{1}{y_i^2}\right] \tag{2.9}$$

Table 2.10 Taguchi designs synoptic table

Number of variables	Number of levels			
	2	3	4	5
2, 3	L4	L9	LP16	L25
4	L8	L9	LP16	L25
5	L8	L18	LP16	L25
6	L8	L18	LP32	L25
7	L8	L18	LP32	L50
8	L12	L18	LP32	L50
9, 10	L12	L27	LP32	L50
11	L12	L27	N./A.	L50
12	L16	L27	N./A.	L50
13	L16	L27	N./A.	N./A.
14, 15	L16	L36	N./A.	N./A.
from 16 to 23	L32	L36	N./A.	N./A.
from 24 to 31	L32	N./A.	N./A.	N./A.

- *nominal-the-best*: to be used when a target value is sought for the response variable.

$$\mathrm{SN}_{ntb} = -10 \log_{10} \frac{\mathrm{E}^2 [y_i]}{\mathrm{E}\left[(y_i - \mathrm{E}[y_i])^2\right]} \tag{2.10}$$

E stands for the *expected value*. According to the Taguchi method, the inner and the outer arrays are to be chosen from a list of published orthogonal arrays. The Taguchi orthogonal arrays, are individuated in the literature with the letter L, or LP for the four-levels ones, followed by their sample size. Suggestions on which array to use depending on the number of parameters and on the numbers of levels are provided in [14] and are summarized in Table 2.10. L8 and L9 Taguchi arrays are reported as an example in Table 2.11. Whenever the number of variables is lower than the number of columns in the table the last columns are discarded.

2.3.9 Random

The DOE techniques discussed so far are experimental design methods which originated in the field of statistics. Another family of methods is given by the *space filling* DOE techniques. These rely on different methods for filling uniformly the design space. For this reason, they are not based on the concept of levels, do not require discretized parameters, and the sample size is chosen by the experimenter independently from the number of parameters of the problem. Space filling techniques are generally a good choice for creating response surfaces. This is due to the fact that, for a given N, empty areas, which are far from any sample and in which the interpolation may be inaccurate, are unlikely to occur. However, as space filling techniques

Table 2.11 Example of Taguchi arrays

L8 (2 levels) Experiment number	Variables X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

L9 (3 levels) Experiment number	Variables X_1	X_2	X_3	X_4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

LP16 (4 levels) Experiment number	Variables X_1	X_2	X_3	X_4	X_5
1	1	1	1	1	1
2	1	2	2	2	2
3	1	3	3	3	3
4	1	4	4	4	4
5	2	1	2	3	4
6	2	2	1	4	3
7	2	3	4	1	2
8	2	4	3	2	1
9	3	1	3	4	2
10	3	2	4	3	1
11	3	3	1	2	4
12	3	4	2	1	3
13	4	1	4	2	3
14	4	2	3	1	4
15	4	3	2	4	1
16	4	4	1	3	2

are not level-based it is not possible to evaluate the parameters main effects and the interaction effects as easily as in the case of factorial experimental designs.

The most obvious space filling technique is the *random* one, by which the design space is filled with uniformly distibuted, randomly created samples. Nevertheless, the random DOE is not particularly efficient, in that the randomness of the method does not guarantee that some samples will not be clustered near to each other, so that they will fail in the aim of uniformly filling the design space.

2.3.10 Halton, Faure, and Sobol Sequences

Several efficient space filling techniques are based on *pseudo-random numbers generators*. The quality of random numbers is checked by special tests. Pseudo-random numbers generators are mathematical series generating sets of numbers which are able to pass the randomness tests. A pseudo-random number generator is essentially a function $\Phi : [0, 1) \longrightarrow [0, 1)$ which is applied iteratively in order to find a serie of γ_k values

$$\gamma_k = \Phi\left(\gamma_{k-1}\right), \qquad \text{for} \qquad k = 1, 2, \ldots \tag{2.11}$$

starting from a given γ_0. The difficulty is to choose Φ in order to have a uniform distribution of γ_k. Some of the most popular space filling techniques make use of the quasi-random low-discrepancy mono-dimensional *Van der Corput sequence* [15, 16].

In the Van der Corput sequence, a base $b \geq 2$ is given and successive integer numbers n are expressed in their b-adic expansion form

$$n = \sum_{j=1}^{T} a_j b^{j-1} \tag{2.12}$$

where a_j are the coefficients of the expansion. The function

$$\begin{aligned} &\varphi_b : \mathbb{N}_0 \longrightarrow [0, 1) \\ &\varphi_b\left(n\right) = \sum_{j=1}^{T} \frac{a_j}{b^j} \end{aligned} \tag{2.13}$$

gives the numbers of the sequence.

Let us consider $b = 2$ and $n = 4$: 4 has binary expansion 100, the coefficients of the expansion are $a_1 = 0, a_2 = 0, a_3 = 1$. The fourth number of the sequence is $\varphi_2\left(4\right) = \frac{0}{2} + \frac{0}{4} + \frac{1}{8} = \frac{1}{8}$. The numbers of the base-two Van der Corput sequence are: $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \frac{7}{8}, \ldots$

The basic idea of the multi-dimensional space filling techniques based on the Van der Corput sequence is to subdivide the design space into sub-volumes and put a sample in each of them before moving on to a finer grid.

Halton sequence [17] uses base-two Van der Corput sequence for the first dimension, base-three sequence in the second dimension, base-five in the third dimension, and so on, using the prime numbers for base. The main challenge is to avoid multi-dimensional clustering. In fact, the Halton sequence shows strong correlations between the dimensions in high-dimensional spaces. Other sequences try to avoid this problem.

Faure [18, 19] and *Sobol sequences* [20] use only one base for all dimensions and a different permutation of the vector elements for each dimension.

The base of a Faure sequence is the smallest prime number ≥ 2 that is larger or equal to the number of dimensions of the problem. For reordering the sequence, a recursive equation is applied to the a_j coefficients. Passing from dimension $d-1$ to dimension d the reordering equation is

$$a_i^{(d)}(n) = \sum_{j=i}^{T} \frac{(j-1)!}{(i-1)!\,(j-i)!} a_j^{(d-1)} \bmod b. \tag{2.14}$$

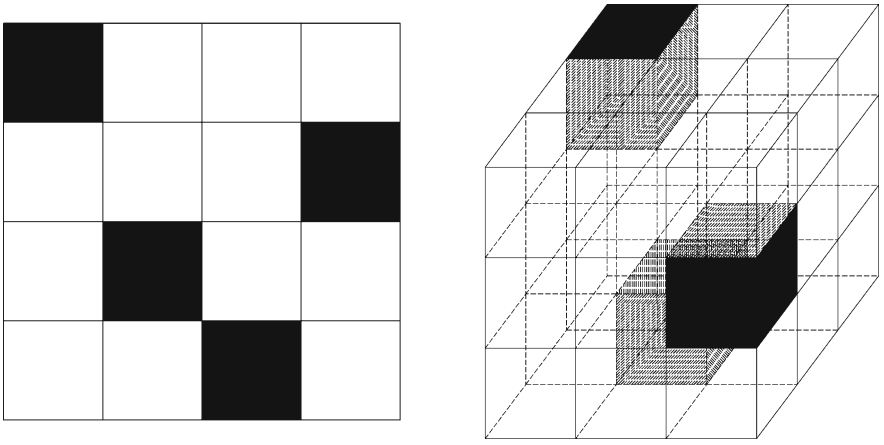
Sobol sequence uses base two for all dimensions and the reordering task is much more complex than the one adopted by Faure sequence, and is not reported here. Sobol sequence is the more resistant to the high-dimensional degradation.

2.3.11 Latin Hypercube

In *latin hypercube* DOE the design space is subdivided into an orthogonal grid with N elements of the same length per parameter. Within the multi-dimensional grid, N sub-volumes are invididuated so that along each row and column of the grid only one sub-volume is chosen. In Fig. 2.8, by painting the chosen sub volumes black gives, in two dimensions, the typical crosswords-like graphical representation of latin hypercube designs. Inside each sub-volume a sample is randomly chosen.

It is important to choose the sub-volumes in order to have no spurious correlations between the dimensions or, which is almost equivalent, in order to spread the samples all over the design space. For instance, a set of samples along the design space diagonal would satisfy the requirements of a latin hypercube DOE, although it would show a strong correlation between the dimensions and would leave most of the design space unexplored. There are techniques which are used to reduce the correlations in latin hypercube designs.

Let us assume the case of k parameters and N samples. In order to compute a set of Latin hypercube samples [21] two matrices $\mathbf{Q}_{N\times k}$ and $\mathbf{R}_{N\times k}$ are built. The columns of $\mathbf{Q}$ are random permutations of the integer values from 1 to N. The elements of $\mathbf{R}$ are random values uniformly distributed in [0, 1]. Assuming each parameter has range [0, 1], the *sampling map* $\mathbf{S}$ is given by



(a) $k = 2\,,\ N = 4$ **(b)** $k = 3\,,\ N = 3$

Fig. 2.8 Example of latin hypercube designs

$$\mathbf{S} = \frac{1}{N}\left(\mathbf{Q} - \mathbf{R}\right). \tag{2.15}$$

In case the elements are to be spread on $\mathbb{R}^k$ according to a certain distribution function, each element of $\mathbf{S}$ is mapped over a matrix $\mathbf{X}$ through the cumulative distribution function D. Different distributions can be chosen for each parameter (D_j, $j = 1, \ldots, k$)

$$x_{i,j} = D_j^{-1}\left(s_{i,j}\right). \tag{2.16}$$

In case of *normal Gaussian distribution*, the cumulative function is

$$D\left(x\right) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x - \mu}{\sigma\sqrt{2}}\right)\right) \tag{2.17}$$

with μ mean value and σ standard deviation. $\mathbf{X}$ is the matrix whose rows are the samples of the latin hypercube DOE. In case of uniformly distributed parameters on the interval $[0, 1]$, $\mathbf{X} = \mathbf{S}$ is taken. The *correlation reduction* operation is essentially an operation on $\mathbf{Q}$. We map the elements of $\mathbf{Q}$ divided by $N + 1$ over a matrix $\mathbf{Y}$ through the normal Gaussian cumulative distribution function D_{norm}

$$y_{i,j} = D_{norm}^{-1}\left(\frac{q_{i,j}}{N+1}\right). \tag{2.18}$$

Then the *covariance matrix* of $\mathbf{Y}$ is computed and Choleski decomposed

$$\mathbf{C} = \text{cov}\mathbf{Y} = \mathbf{L}\mathbf{L}^T. \tag{2.19}$$

Table 2.12 Example of latin hypercube samples computation for $k = 2$, $N = 5$

$$\mathbf{Q} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \\ 4 & 5 \\ 5 & 4 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 0.60 & 0.83 \\ 0.42 & 0.11 \\ 0.69 & 0.51 \\ 0.32 & 0.58 \\ 0.83 & 0.32 \end{bmatrix} \Rightarrow \mathbf{Y} = \begin{bmatrix} -0.97 & -0.43 \\ -0.43 & -0.97 \\ 0.00 & 0.00 \\ 0.43 & 0.97 \\ 0.97 & 0.43 \end{bmatrix} \Rightarrow$$

$$\mathbf{C} = \begin{bmatrix} 0.45 & 0.33 \\ 0.33 & 0.45 \end{bmatrix} \Rightarrow \mathbf{L} = \begin{bmatrix} 0.67 & 0.00 \\ 0.50 & 0.45 \end{bmatrix} \Rightarrow \mathbf{Y}^* = \begin{bmatrix} -1.44 & 0.64 \\ -0.64 & -1.44 \\ 0.00 & 0.00 \\ 0.64 & 1.44 \\ 1.44 & -0.64 \end{bmatrix} \Rightarrow$$

$$\mathbf{Q}^* = \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 3 & 3 \\ 4 & 1 \\ 5 & 4 \end{bmatrix} \Rightarrow \mathbf{S} = \begin{bmatrix} 0.08 & 0.23 \\ 0.32 & 0.98 \\ 0.46 & 0.50 \\ 0.74 & 0.08 \\ 0.83 & 0.74 \end{bmatrix} \Rightarrow \mathbf{X} = \begin{bmatrix} -1.41 & 0.73 \\ -0.48 & 2.01 \\ -0.10 & -0.01 \\ 0.63 & -1.38 \\ 0.97 & 0.63 \end{bmatrix}$$

The covariance matrix is the $k \times k$ matrix whose elements are

$$c_{i,j} = \frac{1}{N} \sum_{l=1}^{N} \left(y_{l,i} - \mu_i \right) \left(y_{l,j} - \mu_j \right) \tag{2.20}$$

where μ_i is the average of the values in the ith column of $\mathbf{Y}$. The *Choleski decomposition* requires $\mathbf{C}$ to be positive definite. For the way the matrix is built this is guaranteed if $N > k$. A new matrix $\mathbf{Y}^*$ is computed so that

$$\mathbf{Y}^* = \mathbf{Y} \left(\mathbf{L}^{-1} \right)^T \tag{2.21}$$

and the ranks of the elements of the columns of $\mathbf{Y}^*$ become the elements in the columns of the matrix $\mathbf{Q}^*$ which is used in place of $\mathbf{Q}$ in order to compute the samples.

A Matlab/Octave script implementing the method is reported in Appendix A.1 and a numerical example in Table 2.12. Figure 2.9 shows the effect of the correlation reduction procedure for a case with two parameters and ten samples. The correlation reduction was obtained using the above-mentioned script. Figure 2.10 shows a comparison between random, Sobol, and latin hypercube space filling DOE techniques on a case with two parameters and a thousand samples. It is clear that the random method is not able to completely avoid samples clustering. Using latin hypercubes the samples are more uniformly spread in the design space. The Sobol sequence gives the most uniformly distributed samples.

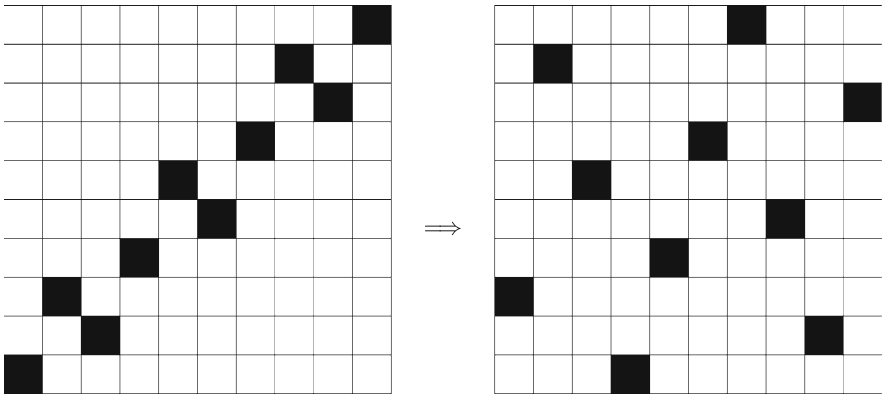
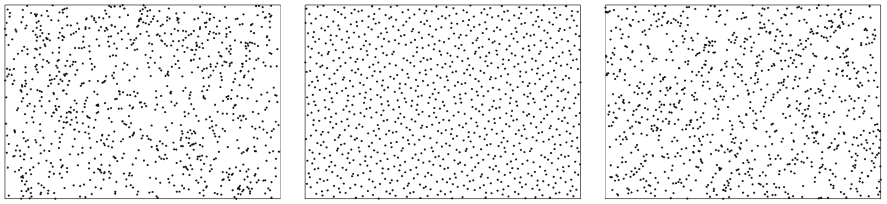


Fig. 2.9 Example of correlation reduction in a latin hypercube DOE with $k = 2$, $N = 10$



(a) Random **(b)** Sobol **(c)** Latin hypercube

Fig. 2.10 A comparison between different space filling DOE techniques for $k = 2$, $N = 1{,}000$

2.3.12 Optimal Design

Optimal design [22, 23] is a good DOE method whenever the classical orthogonal methods may fail due to the presence of constraints on the design space. It is a response-surface-oriented method whose output depends on the RSM technique which is intended to be used later. A set of candidate samples is needed at the beginning. This is usually given by an adjustable full factorial experimental design with many levels for each parameter. Optimal design tests different sets of samples looking for the one minimizing a certain function. It is an iterative method which involves an onerous computation and could require a lot of time to be completed. For instance, consider that for k parameters, with L levels each, the number of possible combinations of N samples in the set are $\frac{L^{kN}}{N!}$: for the very simple case of $k = 3$, $L = 4$, $N = 10$ this would mean $3.2 \cdot 10^{11}$ sets to be tested. For this reason, optimization algorithms are usually applied to the search procedure. The procedure is stopped after a certain number of iterations, and the best solution found is taken as the optimal. The output of the method is a set of samples spread through the whole design space. As the number of samples grows, optimal designs often include repeated samples.

Example 2.1 Let us consider a piston pin as described in Example 1.1 at p. 4. The following tables show the samples list and the results of the simulations according to different DOE techniques.

2^3 Full factorial					
Experiment	Parameters [mm]			Results	
number	L	D_{in}	D_{out}	M [g]	σ_{max} [MPa]
1	80	13	17	59.19	189.04
2	80	13	19	94.70	114.11
3	80	16	17	16.28	577.68
4	80	16	19	51.79	179.24
5	100	13	17	73.98	236.30
6	100	13	19	118.4	142.64
7	100	16	17	20.35	722.10
8	100	16	19	64.74	224.05

2^{3-1}_{III}, $I = ABC$ Fractional factorial					
Experiment	Parameters [mm]			Results	
number	L	D_{in}	D_{out}	M [g]	σ_{max} [MPa]
1	80	13	19	94.70	114.11
2	80	16	17	16.28	577.68
3	100	13	17	73.98	236.30
4	100	16	19	64.74	224.05
Central composite circumscribed					
Experiment	Parameters [mm]			Results	
number	L	D_{in}	D_{out}	M [g]	σ_{max} [MPa]
1–8	as the 2^3 full factorial				
9	90	14.5	18	63.12	203.65
10	90	14.5	16.27	30.22	432.45
11	90	14.5	19.73	99.34	126.39
12	90	17.10	18	17.53	635.56
13	90	11.90	18	101.2	145.73
14	72.68	14.5	18	50.97	164.46
15	107.3	14.5	18	75.26	242.84
Box-Behnken					
Experiment	Parameters [mm]			Results	
number	L	D_{in}	D_{out}	M [g]	σ_{max} [MPa]
1	80	13	18	76.45	143.96
2	80	16	18	33.54	278.92
3	100	13	18	95.56	179.95
4	100	16	18	41.92	346.09
5	80	14.50	17	38.84	264.26
6	80	14.50	19	74.35	134.84
7	100	14.50	17	48.55	330.33
8	100	14.50	19	92.94	168.55
9	90	13	17	66.59	212.67
10	90	13	19	106.5	128.37
11	90	16	17	18.31	649.89
12	90	16	19	58.26	201.64
13	90	14.50	18	63.12	203.65

Latin hypercube Experiment number	Parameters [mm] L	D_{in}	D_{out}	Results M [g]	σ_{max} [MPa]
1	81.59	14.04	18.76	77.88	137.56
2	83.25	14.33	18.54	71.03	155.18
3	84.24	15.39	17.05	27.97	386.23
4	86.93	13.76	17.54	63.41	198.10
5	88.88	14.59	17.84	57.76	216.38
6	91.58	13.48	17.21	64.63	220.09
7	92.89	15.86	17.61	33.54	379.86
8	95.35	15.61	18.85	65.64	205.31
9	97.07	13.29	18.20	92.53	171.88
10	98.81	14.81	18.15	67.06	226.79

Different optimal design methods involve different optimality criteria. The most popular is the *I-optimal* which aims at the minimization of the normalized average, or *integrated prediction variance*. In I-optimal designs of multivariate functions, the *variance* of the predicted response variable

$$\text{var}\left(y\left(\mathbf{x}\right)\right) \approx \nabla y\left(\mathbf{x}_0\right)^T \cdot \text{var}\left(\mathbf{x}\right) \cdot \nabla y\left(\mathbf{x}_0\right) \tag{2.22}$$

is integrated over the design space. Equation 2.22 comes from the delta method for deriving an approximate probability distribution for a function of a statistical estimator. $\mathbf{x} = [x_1, \ldots, x_k]$ is a point in the design space in the neighbourhood of $\mathbf{x}_0 = \left[x_{0,1}, \ldots, x_{0,k}\right]$, and var $(\mathbf{x})$ is the covariance matrix

$$\begin{pmatrix} \text{var}\left(x_1\right) & \text{cov}\left(x_1, x_2\right) & \ldots & \text{cov}\left(x_1, x_k\right) \\ \text{cov}\left(x_2, x_1\right) & \text{var}\left(x_2\right) & \ldots & \text{cov}\left(x_2, x_k\right) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}\left(x_k, x_1\right) & \text{cov}\left(x_k, x_2\right) & \ldots & \text{var}\left(x_k\right) \end{pmatrix} \tag{2.23}$$

where $x_i, i = 1, \ldots, k$ are the parameters. The variance of the ith parameter and the covariance of the ith and the jth parameters are defined as

$$\text{var}\left(x_i\right) = \text{E}\left[\left(x_i - \mu_i\right)^2\right] = \frac{\sum_{l=1}^{N}\left(x_{l,i} - \mu_i\right)^2}{N} \tag{2.24}$$

$$\text{cov}\left(x_i, x_j\right) = \text{E}\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\right] = \frac{\sum_{l=1}^{N}\left(x_{l,i} - \mu_i\right)\left(x_{l,j} - \mu_j\right)}{N} \tag{2.25}$$

where E is the expected value of the quantity in brackets and $\mu_i = E[x_i] = \frac{\sum_{i=1}^{N} x_i}{N}$ is the mean value, or the expected value, of x_i.

Let us assume that we wish to construct a design for fitting a full quadratic polynomial response surface on a k-dimensional design space

$$y(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{i,i} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{i,j} x_i x_j + \epsilon \tag{2.26}$$

where $y(\mathbf{x})$ is the response variable, $x_1, \ldots, x_k$ are the parameters, ϵ are the errors of the quadratic model which are independent, with zero mean value, and σ^2 variance. β are the $p = \frac{(k+1)(k+2)}{2}$ unknown coefficients. Assuming that the design consists of $N \geq p$ samples

$$\mathbf{x}_j = \left[x_{j,1}, \ldots, x_{j,k}\right], \quad j = 1, \ldots N \tag{2.27}$$

let $\mathbf{X}_{N \times p}$ be the *expanded design matrix* containing one row

$$\mathbf{f}\left(\mathbf{x}_j\right) = \left[1, x_{j,1}, \ldots, x_{j,k}, x_{j,1}^2, \ldots, x_{j,k}^2, x_{j,1}x_{j,2}, \ldots, x_{j,k-1}x_{j,k}\right] \tag{2.28}$$

for each design point. The *moment matrix* is defined as

$$\mathbf{M}_\mathbf{X} = \frac{1}{N}\mathbf{X}^T\mathbf{X}. \tag{2.29}$$

The prediction variance at an arbitrary point x and the integrated prediction variance, which is the objective to be minimized in a I-optimal design, are

$$\text{var}\, y(\mathbf{x}) = \frac{\sigma^2}{N}\mathbf{f}(\mathbf{x})\,{\mathbf{M}_\mathbf{X}}^{-1}\mathbf{f}(\mathbf{x})^T \tag{2.30}$$

$$I = \frac{n}{\sigma^2}\int_R \text{var}y(\mathbf{x})\,\mathrm{d}r(\mathbf{x}) = \text{trace}\left\{\mathbf{M}{\mathbf{M}_\mathbf{X}}^{-1}\right\} \tag{2.31}$$

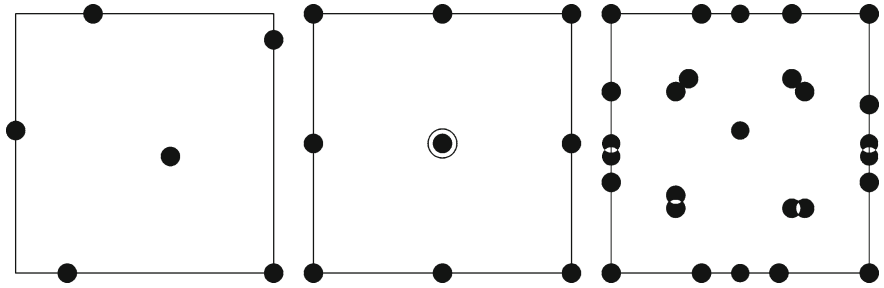
where R is the design space and

$$\mathbf{M} = \int_R \mathbf{f}(\mathbf{x})^T\,\mathbf{f}(\mathbf{x})\,\mathrm{d}r(\mathbf{x}). \tag{2.32}$$

Optimal designs and their objectives are summarized in Table 2.13 for the case of a polynomial response surface. A Maxima script for computing the matrix $\mathbf{M}$ and a Matlab/Octave script implementing the above equations for finding the I-optimal set of samples are presented in Appendix A.2 for either full quadratic or cubic polynomial response with two parameters. Figure 2.11 shows three I-optimal designs obtained using the script for the cases $k = 2$, $L = 21$ with $N = 6$, and with $N = 10$ for a full

Table 2.13 Optimal designs synoptic table

Optimal design	Objective
A-optimal	minimize trace $\left\{\mathbf{M}_X{}^{-1}\right\}$
D-optimal	minimize $\{\det \mathbf{M}_\mathbf{X}\}^{-\frac{1}{p}}$
E-optimal	minimize max eigenvalue $\left\{\mathbf{M}_X{}^{-1}\right\}$
G-optimal	minimize max var $\{\mathbf{f}(\mathbf{x})\}, \mathbf{x} \in R$
I-optimal	minimize trace $\left\{\mathbf{M}\mathbf{M}_\mathbf{X}{}^{-1}\right\}$



(a) $N = 6$, $I = 4.60$, quadratic polynomial **(b)** $N = 10$, $I = 3.66$, quadratic polynomial, the sample point in the centre is repeated **(c)** $N = 20$, $I = 6.49$, cubic polinomial

Fig. 2.11 Example of I-optimal designs for $k = 2$, $L = 21$, polynomial response surface

quadratic polynomial response surface, and with $N = 20$ for a full cubic polynomial response surface.

2.4 Conclusions

Several DOE techniques are available to the experimental designer. However, as it always happens in optimization, there is no best choice. Much depends on the problem to be investigated and on the aim of the experimentation. Items to be considered are:

- the *number of experiments N* which can be afforded.
  In determining the number of experiments, an important issue is the time required for a single experiment. There is a lot of difference between whether the response variable is extracted from a quick simulation in which a number is computed or taken from a spreadsheet or it involves the setting up of a complex laboratory experiment. In the former case it could take a fraction of a second to obtain a response, in the latter one each experiment could take days.
- the *number of parameters k* of the experiment.
  For many DOE techniques, the number of experiments required grows exponentially with the number of parameters (Fig. 2.12). Not necessarily to use a cheap

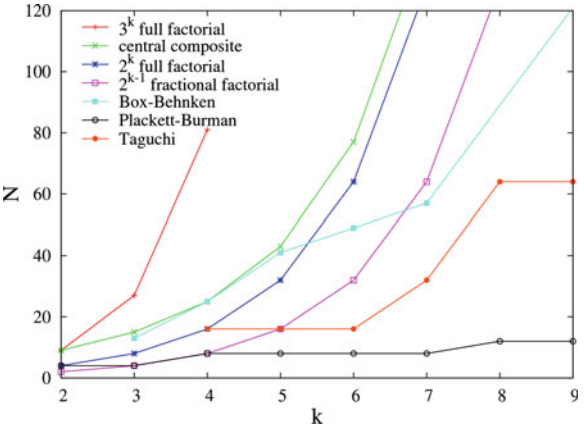


Fig. 2.12 Number of experiments required by the DOE techniques

Table 2.14 DOE methods synoptic table

Method	Number of experiments	Suitability
RCBD	$N(L_i) = \prod_{i=1}^{k} L_i$	Focusing on a primary factor using blocking techniques
Latin squares	$N(L) = L^2$	Focusing on a primary factor cheaply
Full factorial	$N(L, k) = L^k$	Computing the main and the interaction effects, building response surfaces
Fractional factorial	$N(L, k, p) = L^{k-p}$	Estimating the main and the interaction effects
Central composite	$N(k) = 2^k + 2k + 1$	Building response surfaces
Box-Behnken	$N(k)$ from tables	Building quadratic response surfaces
Plackett-Burman	$N(k) = k + 4 - \mod\left(\frac{k}{4}\right)$	Estimating the main effects
Taguchi	$N(k_{in}, k_{out}, L) = N_{in} N_{out}$, $N_{in}(k_{in}, L)$, $N_{out}(k_{out}, L)$ from tables	Addressing the influence of noise variables
Random	chosen by the experimenter	Building response surfaces
Halton, Faure, Sobol	chosen by the experimenter	Building response surfaces
Latin hypercube	chosen by the experimenter	Building response surfaces
Optimal design	chosen by the experimenter	Building response surfaces

technique is the best choice, because a cheap technique means imprecise results and insufficient design space exploration. Unless the number of experiments which can be afforded is high, it is important to limit the number of parameters as much as possible in order to reduce the size of the problem and the effort required to solve it. Of course the choice of the parameters to be discarded can be a particularly delicate issue. This could done by applying a cheap technique (like Plackett-Burman) as a preliminary study for estimating the main effects.

- the *number of levels* L for each parameter.

The number of experiments also grows very quickly with the number of levels admitted for each factor. However, a small number of levels does not allow a good interpolation to be performed on the design space. For this reason, the number of levels must be chosen carefully: it must be limited when possible, and it has to be kept higher if an irregular behaviour of the response variable is expected. If the DOE is carried out for RSM purpose, it must be kept in mind that a two-levels method allows approximately a linear or bilinear response surface to be built, a three-levels method allows a quadratic or biquadratic response surface, and so on. This is just a rough hint on how to choose the number of levels depending on the expected regularity of the response variable.

- the *aim* of the DOE.
  The choice of a suitable DOE technique depends also on the aim of the experimentation. If a rough estimate of the main effects is sufficient, a Plackett-Burman method would be preferable. If a more precise computation of the main and some interaction effects must be accounted for, a fractional or a full factorial method is better. If the aim is to focus on a primary factor a latin square or a randomized complete block design would be suitable. If noise variables could influence significantly the problem a Taguchi method is suggested, even if a relatively cheap method also brings drawbacks. For RSM purposes, a Box-Behnken, a full factorial, a central composite, or a space filling technique has to be chosen. Table 2.14 summarizes the various methods, their cost in term of number of experiments, and their aims. The suitability column is not to be intended in a restrictive way. It is just an hint on how to use DOE techniques since, as reminded above, much depends on the complexity of the problem, the availability of resources and the experimenter sensitivity. To the author's experience, for a given number of experiments and for RSM purpose, space filling Sobol and Latin hypercube DOE always over-perform the other techniques. It is also to be reminded that when dealing with response surfaces it is not just a matter of choosing the appropriate DOE technique, also the RSM technique which is coupled to the DOE data can influence significantly the overall result. This issue takes us to the next chapter.

http://www.springer.com/978-3-642-31186-4

Optimization Methods
From Theory to Design Scientific and Technological Aspects in Mechanics
Cavazzuti, M.
2013, XVIII, 262 p., Hardcover
ISBN: 978-3-642-31186-4