CHAPTER 2

THEORY AND LITERATURE SURVEYS

2.1 Theory

According to Lord Kelvin's statement* , it describes the purposes of carrying out experiments, namely to find out what factors have an effect in a given situation and particularly to measure the magnitude of these effects. So, for example, to compare a number of alternative production methods, the factor(s) of interest are varied and the effect on some characteristic of the product is measured.

It is essential that in any investigation the "scientific Method" is employed. Scientific method refers to certain principles of carrying out investigations which have been found to be essential for valid conclusions to be drawn. It involves being objective and unbiased, the onus of proof being on the person putting forward a theory; quantifying (expressing measurements in numbers) wherever possible and constructing and rigorously testing models before using them for investigation.

In the majority of situations, several factors can have an effect on the outcome. The "classical" approach to experimentation as used in physics and chemistry laboratories and in man school and chemistry laboratories is to hold all factors except one constant, vary this factor and measure the response. This is not practical in situation outside laboratories either it is very uneconomic and time consuming or certain factors cannot be controlled. For instance:
In industry, many factors will affect production processes and for practical reasons conditions cannot be controlled so that they will not affect the production processes. Production is essential fro the survival of the organization and so cannot be stopped or interfered with for the sake of an experiment. Also, for investigations into what affects the quality of a process, the process, the process has to be maintained within very close specification limits, so defective items cannot be produced as part of an experiment.

The variability present in these situations is often greater than the response that it is hoped to defect. This is overcome by very careful planning, which bases responses on comparisons and uses statistical method of analysis. These statistical techniques compare the differences in responses with the variability or error in the results to see whether the differences are significantly greater than the error or whether they could have occurred by chance.

Kelvin, Lord W.T., Popular Lectures and Addresses, v1, pp80, Macmillan, London, 1891.

At first sight, this may seem ridiculous – the importance or relevance of factors whose responses are masked by the inherent variability of the situation may be questioned. As an example, consider the comparison of two measuring instruments. The precision of these instruments should be one tenth of the tolerance that they are measuring, A sample of identified components would be masked by the variation in the size of component. This can be overcome, though, by looking whether at the difference between the measurements on each component and statistically testing whether the mean difference is significantly different from zero. The difference between components is thus eliminated from the measurement of the difference experimental method cannot be used in many situations.

Another major criticism of the classical technique is economy. Time is often short, for example in urgent investigations, and resources are always limited. Consequently, experiments which just vary one factor at a time are impractical because they are so wasteful. To overcome this inefficiency, to enable the comparison of several factors, to detect any inter-actions between factors and to get the maximum amount of in formation for the effect put into a experiment, statisticians have developed a series of techniques for both conducting the experiments and analyzing the results. These are known under the collective title of The Statistical design and Analysis of Experiments.

The method were developed in the 1920's and 1930's mainly by Sir R A Fisher, a geneticist and agricultural researcher, and his colleagues at Rothamsted Experimental Station. They were concerned mainly with agriculture and biological experiments so much of the terminology used has agricultural connotations-plots, block, treatments, etc. However, many of the principles and techniques are relevant to other fields of investigation, particularly technological and industrial, although the emphasis is different because certain conditions differ. Agriculturalists have a major constraint in that they often have to wait a whole year to get their results. In industry on the other hand experiments can usually be repeated within a short time. In agricultural experiments the results are often available all at once whereas in industry the results may come in one at a time. In industrial investigations the experimental errors may be smaller in proportion to the effects sought and the cost of the individual experiments may be considerable. Therefore experimental designs which ate appropriate to agriculture may not be very efficient in industry.

Most industrial investigations should be designed and analyzed statistically because the speed, economy and other advantages. Yet, the inefficient and unsatisfactory classical method is used more often than not because the statistical design and analysis of experiments is not widely understood.

The main industrial applications have been in the chemical and process industries where the techniques have been found to yield very big dividends to the companies concerned. In these industries the problems are frequently so complex that to study the underlying causes of all the many effects observed would involve a prohibitive amount of work. In such cases empirical investigations based on statistical principles can be used to find the optimum conditions for operating the process. Having said this the classical approach is sometimes appropriate. For example in laboratory work for the determination of fundamental constants or properties of substances or other circumstances in which the factors concerned are known to be controllable. Though even here statistical methods may be useful in assessing the errors involved. The studies to find out what factors do have some effect.

A good experimental design is one which furnishes the required information with the minimum of effort. The requirement of a good experimental design can be summarized:

- The question to be answered must be correctly formulated.

- The experimental method; that is the choice of treatments, experiment units, responses to be measured, etc must take account of the precision required and the various pitfalls and problems which are likely to be encountered.

- Experimental units receiving different treatments should differ in no systematic way from one another- assumptions that certain sources of variability are absent or negligible should, as far as practicable be avoided.

- Random errors of estimation should be suitably small, and this should be achieved with as few experimental units as possible.

- The conclusions should have a wide range of validity and application.

- The experiment should be simple in design and analysis.

- A proper statistical analysis of the results should be possible without making artificial assumptions.

From mathematical theory combined with much practical experience, statisticians have developed a whole series of experimental designs whose properties are known. These have been developed mainly in the areas of agriculture, biology and chemical engineering. The application of these to batch and mass production has been limited and these are plenty of scope for the development of these techniques. So before applying a particular design ensure that the appropriate conditions and assumptions are applicable.

2.2 What is Equipment Design?

A designed experiment is a test or series of test in which purposeful changes are made to the input variables of a process so that we may observe and identify corresponding changes in the output response. In Figure 4, the process can be visualized as some combination of machines, methods and people that transforms an input material into an output product, This output product has one or more observable quality characteristics of responses, Some of the process variable $x_1, x_2, \ldots, x_p$ are controllable, while others $z_1, z_2, \ldots, z_P$ are uncontrollable.

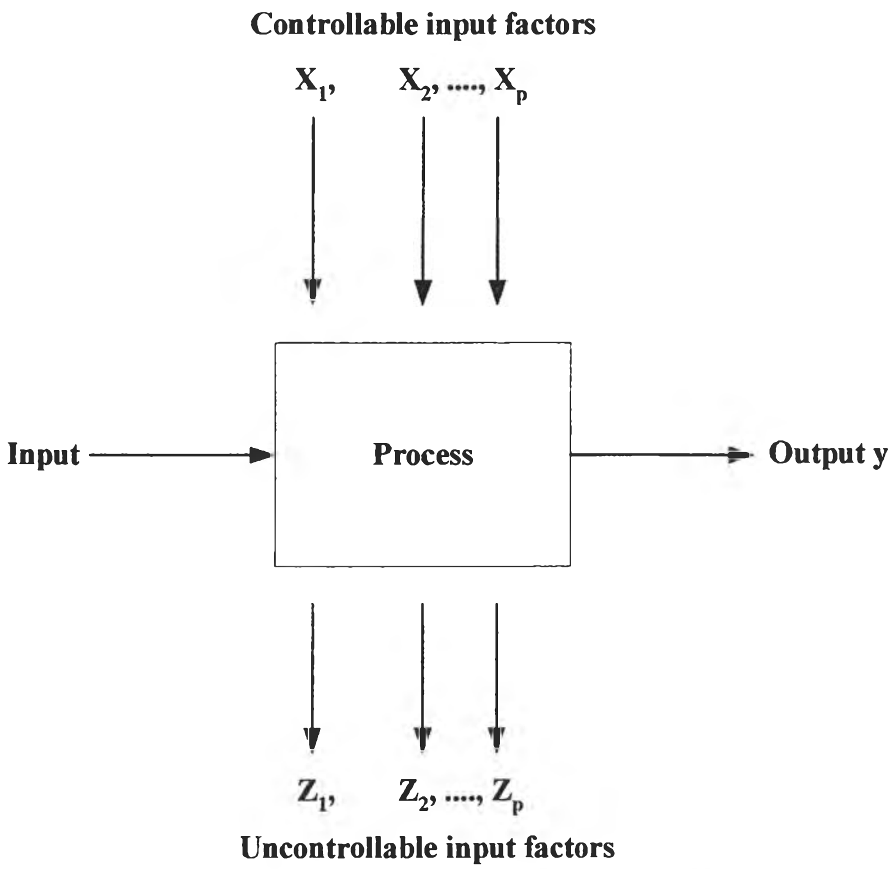

Figure 4 Illustrated General Model of a Process.

The objectives of designed experiment may include

1. Determining which variables are most influential on the response, y.
2. Determining where to set the influential x's so that y is near the normal requirement.
3. Determining where to set the influential x's so that variability in y is small.
4. Determining where to set the influential x's that that the effects of the uncontrollable variables z are minimized.

Experimental design is a critically important engineering tool for improving a manufacturing process. It also has extensive application in the development of new processes. Application of these techniques early in process development can result in

1. Improve yield.
2. Reduced variability and closer conformance to nominal.
3. Reduced overall costs.

2.3 Basic Principles

The three basic principles of experimental design are replications, randomization, and blocking.

Replication means a repetition of the experiment. Replication has two important benefits. First, it allows the experimenter to obtain an estimate of the experimental error. This estimate of error becomes a basic unit of measurement for determining whether observed differences in the data are really statistically different. Second, if the sample mean is used to estimate the effect of a factor in the experiment, the replication permits the experimenter to obtain a more precise estimate of this effect.

Randomization means that the order in which the individual runs or trials of the experiment are to be performed are randomly determined. Statistical methods require that the observations or errors be independently distributed random variables. Randomization usually makes this assumption valid. By properly randomizing the experiment, the effects of nuisance variable is balanced out.

Blocking is a technique used to increase the precision of an experiment. This technique is used in order to control or remove variability arising from nuisance variables. A block is a portion of the experimental material that should be more homogeneous than the entire set of material. Blocking involves making comparisons among the conditions of interest in the experiment within each block.

2.4 Type of Designed Experiment

By the number of factor, the designed experiment can be classified as single factor experiment, factorial experiment, and 2^k factorial experiment,

Single-factor experiment is the designed experiment for testing effect of a factor, which has more than two levels, on responses.

Factorial experiment is used to study the effects of two or more factors. The effects of factors include a main effect and an interaction effect on the interesting responses. This experiment is suitable for more than two levels of each factor.

Only two levels of each factor, for two or more factors, especially in several factors (K factorial), 2^K factorial experiment is widely used to study the joint effect of the factors on a response. The 2^k design is particularly useful in the early stages of experimental work to screen factor that does not affect an response variable out, so called the factor screening experiment.

To illustrate the concept of interaction. Suppose that both of our design factors are quantitative (such as temperature, pressure, time, etc.). Then a regression model representation of the two-factor factorial experiment could be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

In Figure 5, the main effect of factor A and B have on an response when factor levels (A^-, A^+, B^+, and B^-) change, but there is no interaction effect.

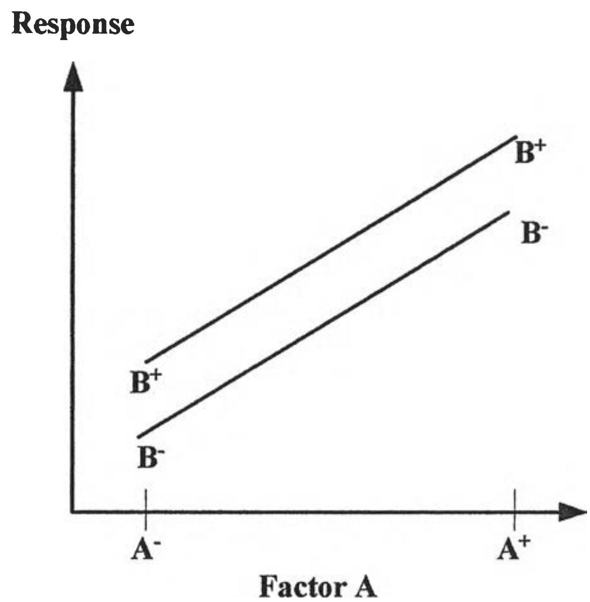

Figure 5 illustrated A Factorial Experiment without Interaction.
Source : Douglas C. Montgomery, Design and Analysis of Experiments, Introduction to Factorial Design, page 171

In Figure 6, the main effect of factor A and B have on an response when factor levels (A^-, A^+, B^+, and B^-) change . Also there is the interaction effect between factors A and B, the two lines are not parallel.

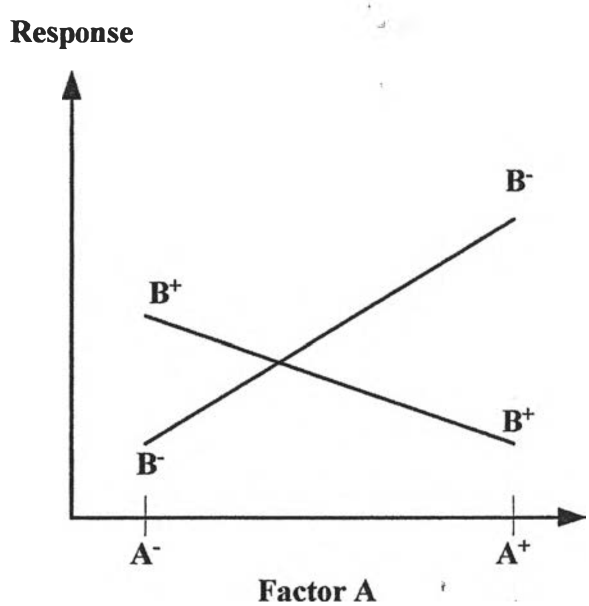

Figure 6 illustrated A Factorial Experiment without Interaction.
Source : Douglas C. Montgomery, Design and Analysis of Experiments, Introduction to Factorial Design, page 171

2.5 Analysis of Variance

The method of Analysis of Variance (ANOVA) is applied to the designed experiment to draw conclusion about the effect of factors on an response.

For example, experiment with two factors, the hypotheses about the model of observations, which will be tested by ANOVA, are as follows.

The observations resulting from the experiment are showed in Figure ? may be described by the model

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \text{ by } i = 1,2, \ldots\ldots, a$$
$$j = 1,2, \ldots\ldots, b$$
$$k = 1,2, \ldots\ldots, n$$

where μ is the overall mean effect, τ_i is the effect of the ith level of the row factor A, β_j is the effect of the j th level of column factor B, $(\tau\,\beta)_{ij}$ is the effect of the interaction between τ_i and β_j, and ε_{ijk} is a random error component. Both factors are assumed to be fixed, and treatment effects are defined as deviations from the overall mean, so $\sum_{i=1}^{a} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_i = 0$. Similarly, the interaction effects are fixed and are defined such that $\sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$. Since, there are n replicates of the experiment, there are abn total observations.

Factor B

Factor A	**1**	**2**	**.....**	**b**
1	$y_{111}, y_{112}, \ldots, y_{11n}$	$y_{121}, y_{122}, \ldots, y_{12n}$		$y_{1b1}, y_{1b2}, \ldots, y_{1bn}$
2	$Y_{211}, y_{212}, \ldots, y_{21n}$	$Y_{221}, y_{222}, \ldots, y_{22n}$		$y_{2b1}, y_{12b2}, \ldots, y_{2bn}$
. **.**				
a	$y_{a11}, y_{a12}, \ldots, y_{a1n}$	$Y_{a21}, y_{a22}, \ldots, y_{a2n}$		$y_{ab1}, y_{ab2}, \ldots, y_{abn}$

Figure 7 illustrated General Arrangement for a two-factor Factorial Design.
Source : Douglas C. Montgomery, Design and Analysis of Experiments, Introduction to Factorial Design, page 176

In the factorial experiment, both row and column factors or treatment, A and B, are of equal interest. So the hypotheses about the equality of row treatment effects will be tested.

H_o : $\tau_1 = \tau_2 = \ldots = \tau_a = 0$
H_1 : at lease one $\tau_i \neq 0$

And the hypotheses about the equality of column treatment effects will be tested

H_o : $\beta_1 = \beta_2 = \ldots = \beta_b = 0$
H_1 : at lease one $\beta_j \neq 0$

Finally, the hypotheses about interaction effect will also be tested.

H_o : $(\tau\beta)_{ij} = 0$ for all I, j
H_1 : at lease one : $(\tau\beta)_{ij} \neq 0$

These hypotheses are tested using ANOVA of the fixed effects model by computing sum of squares, mean squares, and ratio of mean squares (F_o) as follows.

The total sum of squares, mean squares is completed as usual by

$$SS_T = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y^2_{ijk} - \frac{y^2_{...}}{abn},$$

$$y_{...} = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}$$

The sum of square for the main effects are

$$SS_A = \frac{1}{bn} \sum_{i=1}^{a} y^2_{i..} - \frac{y^2_{...}}{abn}$$

$$SS_B = \frac{1}{an} \sum_{j=1}^{b} y^2_{.j.} - \frac{y^2_{...}}{abn}$$

The sum of squares for the interaction effect is

$$SS_{AB} = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} y^2_{ij.} - \frac{y^2_{...}}{abn} - SS_A - SS_B$$

The sum of squares of error is

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
A treatments	SS_A	**a-1**	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_E}$
B treatments	SS_B	**b-1**	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_E}$
Interaction	SS_{AB}	**(a-1)(b-1)**	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_E}$
Error	SS_E	**ab(n-1)**	$MSE = \frac{SS_E}{ab(n-1)}$	
Total	SS_T	**abn-1**		

Table 3 The Analysis of Variance Table for the Two Factor.
Source : Douglas C. Montgomery, Design and Analysis of Experiments, Introduction to Factorial Design, page 180

And $F_{\alpha, v1, v2}$ can be obtained from the table of percentage points of the F distribution, α is the signification level, and v_1 and v_2 are the degrees of freedom.

In table3, we would reject H_0 if F_0 of A treatments is more than $F_{\alpha, a-1, ab(n-1)}$, we conclude that factor A significantly affects an response. In the same way, we would reject H_0 if F_0 of B treatments is more than $F_{\alpha, b-1, ab(n-1)}$, we conclude that factor B significantly affects an response.
And we would reject H0 of F0 of interaction is more than Fα, (a-1)(b-1), ab(n-1), we conclude that there is an interaction effect between the two factor on an response.

2.6 Model Adequacy Checking
As the analysis of variance assumes that the model errors are normally and independently distributed with the same variance in each factor level, abbreviated NID(0, σ^2), these assumptions can be checked by examining the residuals. A residual is defined as the difference between the actual observation and the value that would be obtained from a least-squares fit of the underlying analysis of variance model to the sample data. For example, the residuals for the two-factorial model are

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} \quad \text{or}$$

$$e_{ijk} = y_{ijk} - \bar{y}_{ijk}$$

The normality assumption can be checked by construction a normal probability plot of the residuals, plotting residuals ranked in ascending order (k) versus their cumulative probability points Pk = (k-0.5)/n, n is number of all observations in the experiment.

To check the assumption of equal variance at each factor level, plot the residuals against the factor level and the fitted values, and then compare the spread in the residuals.

2.7 Duncan's Multiple Range Test

A procedure is widely used for comparing individual means of either factor, either the row averages or the column averages, when using the fixed effects models.

For example, two-factor factorial experiment, R_P in the equation below are used to compare with difference between two means.

$$R_P = r_\alpha(p,f)\, S_{\bar{y}_{ij.}} \text{ for } p=2,3 \ldots\ldots, a \text{ or } b$$

From Duncan's table of significant ranges, obtain the value $r_\alpha(p,f)$, for p=2, 3, ……, a or b, where α is the significance level and f is the number of degrees of freedom for error, $S_{\bar{y}_{ij.}} = \sqrt{\frac{MS_E}{n}}$, and n replicates.

2.8 Choice of number of replicates

Operating characteristic curve can be used to find the number of replicates for the designed experiment, for the two-factor factorial experiment, using the following formula.

$$\Phi^2 = \frac{naD^2}{2b\sigma^2}$$

where n id the number of replicates, a levels of factor A, b levels of factor B,D is the difference in mean, σ is standard deviation, $\upsilon_1 = b-1$, and $\upsilon_2 = ab(n-1)$. Using Φ resulting from trials of n, α, υ_1 and υ_2 in the operating characteristic curve leads to β risk that could be acceptable to select the number of replicates.

2.9 Guideline for Designing Experiments

Montgomery (1991) gives an outline of the recommended procedure as follows.

2.9.1 Recognition of and Statement of the problem

In practice, it is often difficult to realize that a problem requiring formal designed experiments exists, so it may not be easy to develop a clear and generally accepted statement of the problem. However, it is absolutely essential to fully develop all ideas about the problem and about the specific objectives of the experiment.

A clear statement of the problem and the objectives of the experiment often contribute substantially to better process understanding and eventual solution of the problem.

2.9.2 Choice of Factor and levels

The experimenter must choose the factor to be varied in the experiment, the range over which these factors will be varied, and the specific levels at which runs will be made. Process knowledge including practical experience and theoretical understanding is required to do this. This step determines type of experiment whether single-factor experiment or factorial experiment or 2^k factorial experiment.

2.9.3 Selection of the Response Variable

Inn selecting the response variable, the experimenter should be certain that the variable really provides useful information about the process under study. Most often the average or standard deviation (or both) of the measured characteristic will be the response variable.

2.9.4 Choice of Experience Design

Choice if design involves consideration of number of replicates, selection of a suitable run order for the experimental trials, and whether or nor blocking or other randomization restrictions are involved.

2.9.5 Performing the experiment

When running the experiment, it is vital to carefully monitor the process to ensure that everything is being done according to plan. Errors Error in experimental procedure at this stage will usually destroy experimental validity. Up-front planning is crucial to success. It is easy to underestimate the logistical and planning aspects of running a designed experiment in a complex manufacturing environment.

2.9.6 Data Analysis

Statistical methods should be used to analyze the data so that results and conclusions are objective rather than judgment. If the experiment has been designed correctly and if it has been performed according to the design, then the type of statistical methods required is not elaborate. Many excellent software package are available to assist in the data analysis, and simple graphical methods play an important role in data interpretation. Residual analysis and model validity checking are also important.

2.9.7 Conclusions and Recommendations

Once the data have been analyzed, the experiment must draw practical conclusions about the results and recommend a course of action, Graphical methods are often useful in this stage, particularly in presenting the results to other. Follow-up runs and confirmation testing should also be performed to validate the conclusions from the experiment.

2.10 Literature Surveys

Literature Survey

Factorial designs are most efficient for the this type of experiment. The several factors is to conduct a factorial experiment. This is an experimental strategy in which factors are varied together, instead of one at a time. The factorial experimental design concept is extremely important for studying CO2 cleaner implementation factors.
Douglas C. Montgomery, 2001, Design and Analysis of Experiments, Fifth edition, JOHN WILLY & SONS, INC.

Advancing Applications in Contamination Control, address Cleaning with CO2 and Dry Ice Particles, cleaning is one of the most important steps in the manufacturing of semiconductors. As the devices grow smaller, their sensitivity to dust, bacteria, and certain gasses Simply blowing air or nitrogen across a surface will remove the larger particles (5 microns), but these days we are concerned about particles as small as 0.1 micron.
Site: http:// www.a2c2.com

CleanTech (2001), address Carbon dioxide (CO2) can be used in three distinct states in precision cleaning applications: in liquid form, where CO2 acts to perform surface cleaning and degreasing; as a gas, which is ejected as "snow" from specialised nozzles; and in a "super critical" form for chemical extraction cleaning.
Site: http:// www.precisioncleaningweb.com/tech_carbod.cfm

Using Solid-state CO2 in Critical cleaning (2001), addresses Carbon dioxide (CO2) snow cleaning is, for some manufactures, a prominent entry on their list of viable alternatives to current ODC-based operations. Over the past several years, there has been considerable investigation into the effectiveness of the technology. Users in several high-tech markets are taking advantage of the unique surface cleaning capabilities of CO2 snow to improve to existing cleaning standards. In fact, progressive design improvement and process optimizations have earned this relatively new cleaning method a position on the production line of some major microdevice manufactures world-wide. This article explores the science of snow cleaning and offers some case reports based on a patented CO2 snow cleaning technology, which uses a thermally ionized gas (TIG) snow-or TIG-snow.
Site: http:// www.precisioncleaningweb.com/article_index.cfm?article=322

IDEMA (2002), address knowledge precision cleaning for Data Storage Components. As head/disk flying heights continue to shrink, even trace amounts of contaminants become unacceptable. As a result, cleaning of components is becoming increasingly critical. This course provides a thorough examination of the technologies, the equipment, the techniques, the options, and the trade-offs for precision cleaning in disk drive industry.
http:// www.IDEMA.org.com