Dealing with Materials Data: Collection, Analysis and Interpretation

Professor M P Gururajan,
Professor Hina A Gokhale.
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

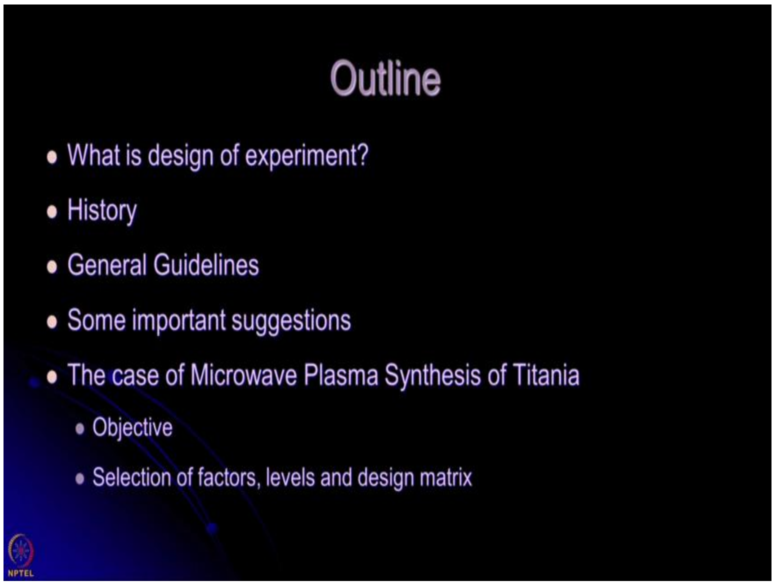
Lecture 87

Design of Experiment – I

Hello and welcome to the course on Dealing with Materials Data. Today we are going to talk about the Design of Experiments. It is a very common technique being used to plan an experiment and work out the analysis of its results to come up with correct set of values or the factor levels at which you can conduct the experiments to give you the best possible results. This design of experiment is a very popular application area of statistics as it is being used properly in number of cases but at times there are certain misconceptions and therefore, the use is not totally complete.

So in the coming few sessions we would like to describe the method of design of experiment through a case study. So first, I will give you an outline.

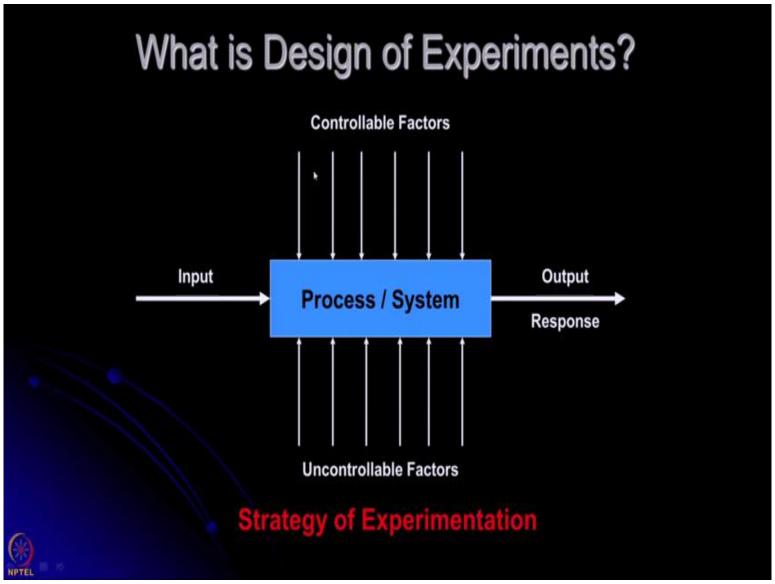
(Refer Slide Time: 1:28)



We would like to, in this session, we would like to discuss what is design of experiment and when you can use it. We will briefly talk about the history of design of experiment, then we will talk about its general guidelines and some important suggestions. We will initiate with the case of

Microwave Plasma Synthesis of Nano Titania, and we will apply the general guidelines and the design of experiment methodology and through this case study, in the subsequent sessions, we will learn how the process of design of experiment gets applied to a real-time situation. Is any experiment amenable to design of experiment? Well, answer is not always, when it is possible?

(Refer Slide Time: 2:38)



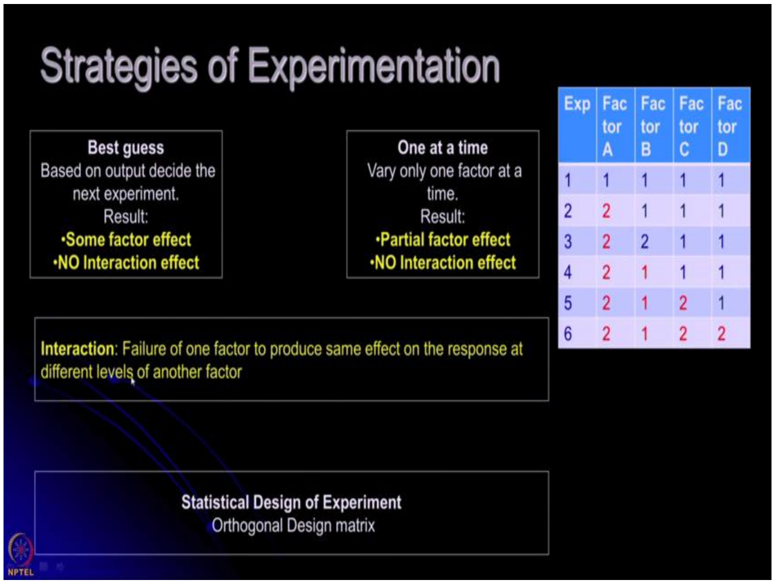
So, the first of all, suppose you have a system, you have a system or a process in which a material input is given and another material output comes out. This process is controlled by many factors. We can divide them into two kinds of factors. One is called controllable factors and the other is called uncontrollable factors. Controllable factors are those, which as the name says you can control its values, its levels. For example, a furnace temperature, or a pressure in a pressurized furnace or a pressure in a forging press or a pass speed in a rolling mill, so these are your controllable factor. Either, the apparatus or the machine itself gives you that possibility or even input becomes a controllable factor, what material you are putting in, that also becomes a controllable factor.

While there are many uncontrollable factors, for example, time of the day when you conducted the experiment, or it could be season of the year if the process is affected by the temperature very severely, then season of the year when you are, you know, conducting the experiment is also a factor which you have no control over.

In the factory setup, that is in a regular production setup if you are trying to apply this technique or you try to do this kind of experiment, you will find that the timing of shift, the person operating the system, all these factors also matter. These are called uncontrollable factors. The idea here is what you can control, control in a systematic manner and what is uncontrollable, if possible you account it as a random error.

So this is when such a situation arises when your experiment is of generally this nature, then you can apply what is called design of experiment. And design of experiment that is how it is a kind of a strategy of experimentation. Let us look at different strategies of experimentation.

(Refer Slide Time: 5:25)



The one strategy most commonly used is Best Guess. What does it mean? Based on a specific output you decide what should be your next experiment. So you do not have any plan as such for which experiment will come after the first experiment and which experiment will come after second experiment. You do not even have any idea as to how many number of experiments you are going to conduct, this is called the best guess.

What it gives is that some factor effect, some factor, factor meaning those controllable factor, but it gives no interaction effect. I will define interaction effect soon, so right now we, I would like to emphasize that it does not give any interaction effect. Let us consider the case of one at a time. It is described in this matrix here.

Suppose we have 4 controllable factors A, B, C, D. Then it says that you vary only one factor at a time and conduct the experiment. So, for example, in this matrix, I keep B, C, D fixed at factor at a level 1. So at some value I, keep B, C, D factors fixed and I vary only the factor A at two different values, which I call two different levels.

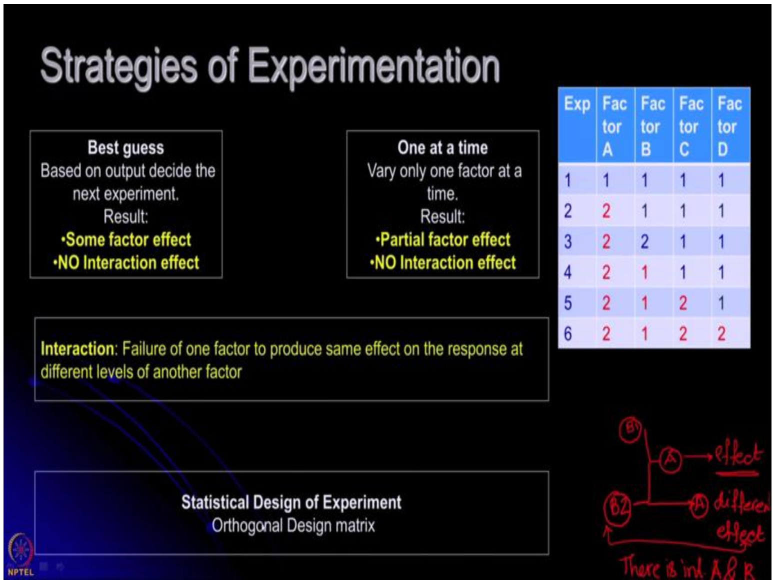
Then, of these two experiment, I find that, the second experiment has met my requirement. It is best among the two, I have some criteria to decide what is best. So the best among the two is this one, so I keep this level 2 fixed for factor A and it is completely written here so I just do not worry about this factor 2 which I have given in red.

Then, I tried and vary the factor B by keeping factor C and D at the same level 1, 1. See I have already conducted an experiment with 2 with level 1 of B, so I conduct another experiment of factor B with level 2, A is fixed at 2 now and this is fixed at 1, 1. And I conduct the two experiment and from these two experiment with respect to factor B, I find that this experiment has performed the best, number 2 has performed the best, so what I do is that I keep that value fixed.

So now, my A factor is fixed at level 2, my B factor is fixed at level 1 and then I conduct the experiment, so I have 2, 1, 1 experiment along with 2, 1 I now I change the factor C, so I conduct the experiment with factor at 2, factor C at 2 level and I find that among the two with the factor 2 of level, sorry, factor C at level 2 gives me the best result, so I choose that, now factor C will remain at level 2 and likewise I conduct another experiment with the factors A, B and C fixed at respectively level 2, 1 and 2.

And I conduct the experiments by varying the levels for factor D and at the end I find that this is the combination which gives me the best results suppose. This is called one factor at a time method. This is called one factor at a time method. It gives again partial factor effect, it is better than this because it gives you some partial factor effect but again, I repeat that it does not give any interaction effect. What is interaction? Well, interaction says that failure of one factor to produce same effect on the response at different levels of another factor.

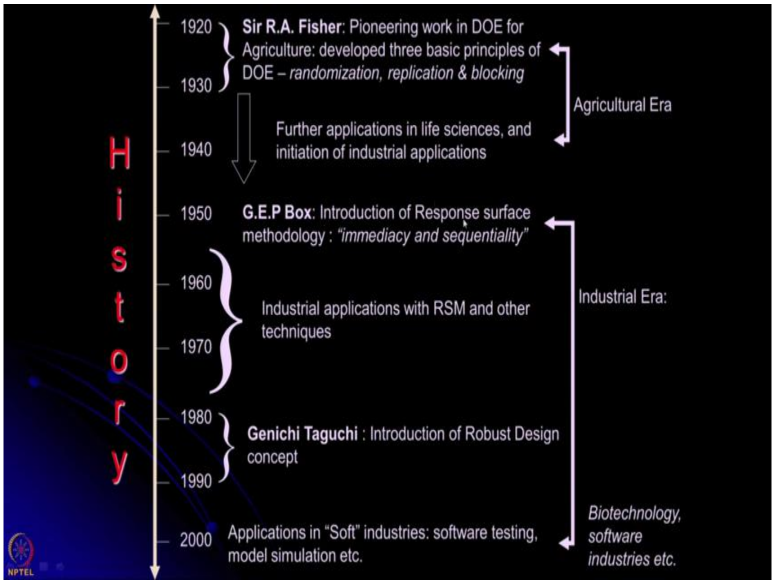
(Refer Slide Time: 9:51)



Let us understand this. Suppose the effect is controlled by factor A. Okay? The effect is control by factor A and then this effect, you want to check, if you vary factor B at level 1, or you keep the factor B at level 2, does it give the same effect or with a factor B, A has a different effect. It means that, if both things are, if B is varied from level 1 to level 2 and factor A gives you only one effect, then we say that there is no interaction effect, but if you vary the factor B at level 1 and 2, and the A starts giving you two different effects, then we say that this situation says that there is interaction between A and B, A and B.

We will explain it once again because interaction effect is very important, so we will time and again if I explain it, so let us move on. When you follow these strategies, you are not able to get any interaction effect and therefore, we are suggesting that statistical design of experiment with orthogonal design matrix is the best possible strategy.

(Refer Slide Time: 12:02)



Historically speaking, it was in 1920s, when Sir R. A. Fisher did the pioneering work in design of experiment for agricultural experimentation and he developed 3 basic principles of design of experiment, they are called randomization, replication and blocking. We will talk about it in future. Then from agriculture it moved to a biological and life-sciences applications and some initiation of industrial application began during 40s, so we can call this an agricultural era of application of design of experiment.

With G.E.P Box, the introduction to Response surface methodology came and it became very popular in the industrial era. The difference why this immediacy and sequentiality came is you can imagine when it was applied by Sir Fisher in agricultural science, you do the, you design the experiment and you have, you sow the seeds and you have to wait until the results came.

So you may have changed the watering plan, you way may have changed the soil, you might have changed the seeds and you might have made some, you know different experimentation using different fertilizers but it would take at least 3 to 4 months time to realize the result and then analyze it.

While applying the same technique in the industries, G.E.P Box realized that there is clear in industry the results come immediately and therefore, you need not plant a very huge experiment but you can plant a smaller experiment and sequentially keep improving yourself on baking different design of experiments step by step and come to a final conclusion.

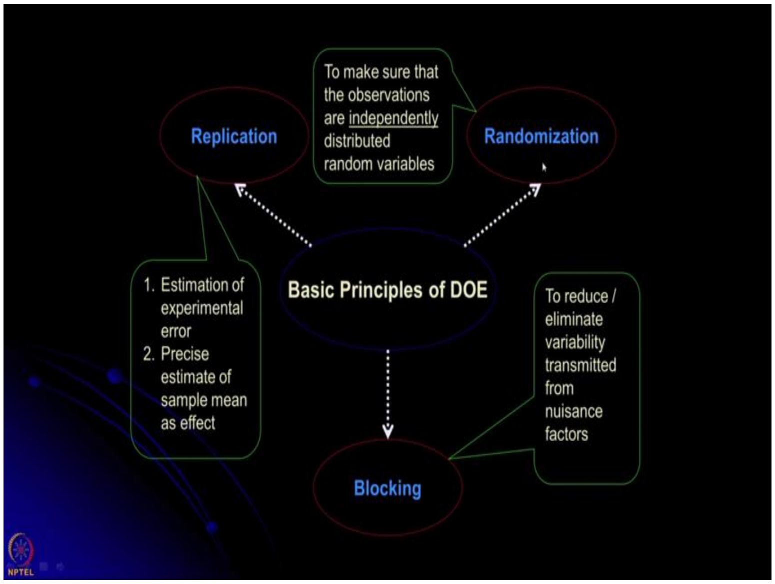
So instead of having a huge design of experiment to come to a one conclusion, he suggested that in the industry, we can go by what is known as response surface methodology, first you conduct a smaller experiment with lesser complicity but more factors and select a few factors which you find is affecting the most. And also you will know in which direction if this factor's value are changed, you will get a better result and then you start working with a fewer and fewer number of factors and come to a better conclusion after 2 or 3 sets of such experiment. So this is called immediacy and sequentiality.

This technique flourished during 60s and 70s for industrial applications and other applications as well. It was in 80s that the Japanese electrical engineer Genichi Taguchi introduced a robust design concept. It would be of interest for you to know, that the first lessons, the few first few lessons of design of experiments Mister Taguchi, Doctor Taguchi gained were very much from India, from Indian Statistical Institute.

And he developed this new technique which is known as a Taguchi Design of Experiments and presently you find that this applications have proliferated in a variety of areas which includes the soft industries such as software testing, model simulation, etc. because the computing power has gone up so high that earlier a smaller the design, more the information. Now we can afford to have a larger design and still have a better information on the, from the experiment analysis.

Now what are the 3 basic principles? Please remember, in throughout this area, the basic principles have not changed. They have remained the same which were laid down by Sir R. A. Fisher, and that is randomization, replication and blocking.

(Refer Slide Time: 16:48)



So these basic principles are the first of all is randomization. See, the idea is that, what you are conducting in a laboratory or on a small scale in an industry, the results and the analysis of it you would like to apply in the real life situation. And as we discussed in real life situation, there are many factors which are uncontrollable, they are kind of a random factors. Some factors you can account for in what we will now discuss as blocking but there will be certain human error, there will be certain machine error built into your experimental result.

These are called random errors and if you want to control the random error, it is suggested that instead of conducting the experiment in a one particular sequential order, you try to conduct the experiment in a randomized order. You do it in a random way, so that the random effect also gets mixed up, it does not remain a systematic effect that from experiment 1 to experiment 16, systematically the machine, for example, if you use the machine often and often, the machine's certain working parameters get derailed, they slowly get away from their from their average way of working.

Therefore, in order to avoid this kind of a systematic error, if you randomize the experiment, then you take into account this change in a random fashion which may happen in real life and therefore, this is called to make sure that the observations are independently distributed random variables to make sure that what you are getting does not have any systematic error but it has a random error in it, you go through what is called randomization.

Then comes replication. This is not something very new, we are quite habituated that you do a simple physics experiment. You take 3 observations and you take an average of it. Okay? This is called observation's replication, repetition, you repeat the observations and then you take an average. The idea is again the same that there will be some random error in it and you want to average it out.

Well, when you when you do the experiment, the whole experiment itself need to be repeated. The whole experiment, I am repeating, whole experiment need to be repeated. Each experiment need to be repeated so that you conducted the experiment or I conducted the experiment in a certain circumstances, the effect of those circumstances which is also random can be neutral out or it can be averaged out if you do the replication.

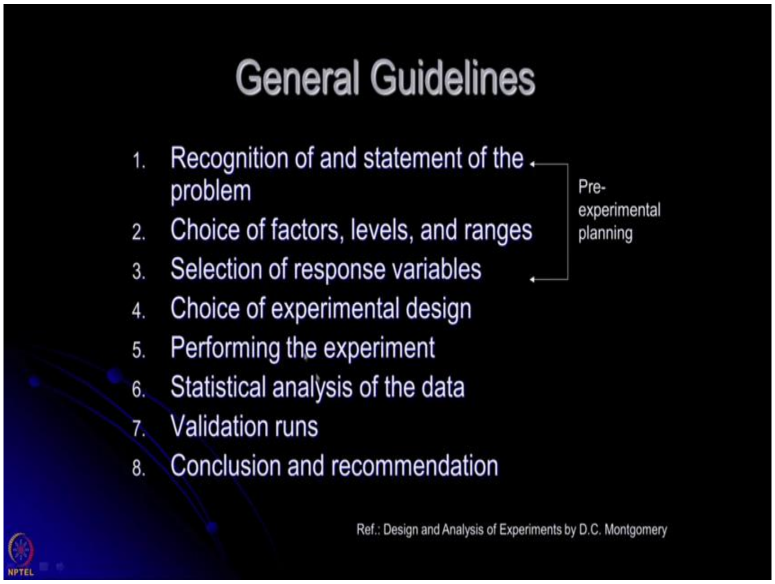
So replication is says that, to estimation of experimental error and precise estimate of a sample mean as an effect. So, what you are estimating, that itself has an error in it, your experimentation itself has an error in it, that whole experiment if you repeat, as many times as you can afford, it is called replication. It is very common to mix up the repetition with the replication. Observations are repeated in order to get away from the random errors in the observation, which is you average out the random error in the observation, by replication we mean that the whole experiments are repeated so that the experimental error gets averaged out.

Then comes blocking. You remember we talked about certain factors which are uncontrollable. Some we have taken care in randomization and replication, but, for example, if you are conducting an experiment in a laboratory and there are two operators for the machine, okay? Each one have their own style of working and they cause their kind of an error but you cannot really control it.

Particularly, when you conduct the experiment in the industrial environment, the shift matters, what time of the day the person started working on your experiment matters. If it is his first experiment of the day, it is one kind. If it is just before he wants to go home and his shift is getting over, his metal set setup will be likely to be different and so his experiment, way of you know handling the equipment would be different. So there are many such situations which are uncontrollable.

So for that it says that you block them, it means that you systematically assign certain experiments to different shifts or so that there are different persons in it and you reduce or at least if not eliminate, you reduce the variability caused by uncontrollable factors which are known as nuisance factors.

(Refer Slide Time: 22:34)



So, here is a general guideline, I have picked it up all this I have picked it up from Montgomery's book. You are all welcome to go through it, it gives extremely good treatment of design of experiments. So, coming to general guideline, it says that first recognize the and make the statement of the problem.

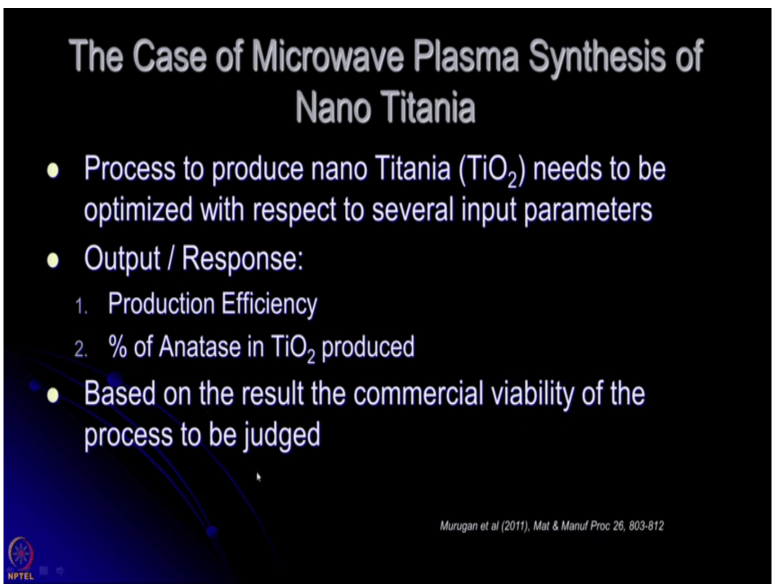
So you must understand what is the problem that we are trying to solve. You should recognize it and you should make a clear statement of it. You should know what are your factors, what levels you would like to vary them, at which range you can vary them and then you know what is your response variable.

These three are pre-experimental planning. Before you plan anything, you must have this basic idea. See, number of times, this seems very simple and very straightforward, but in the haste of doing experiment, number of times if these things are not clear, the design itself is not having any clarity and the analysis becomes a problem. And this is what has been observed several time, so it is very important to go through these first three stages before you design the experiment. Afterwards, you design, you make a choice of experimental design.

Experimental design is that, how many levels you want to vary your factors. Number 2, in that case how many minimum number of experiments you must conduct. Number 3, how many replications you are going to make. Number 4, how you are going to randomize it. All these aspects you have to cover in choice of experimental design.

Then comes performing the experiments. Then statistical analysis of the data, you will come up at the end, you with what is called predictive interval, so you will say that if I do everything in this manner, my result should fall in this interval. So, you are actually giving a interval estimation of your result. Then you must run the validation runs, we must make sure that what I predicted by doing the first 6 steps, what I did is indeed true. So I must have some validation run and at the end, I can make conclusion and recommendation. So now we are ready to start with the case of Microwave Plasma Synthesis of Nano Titania.

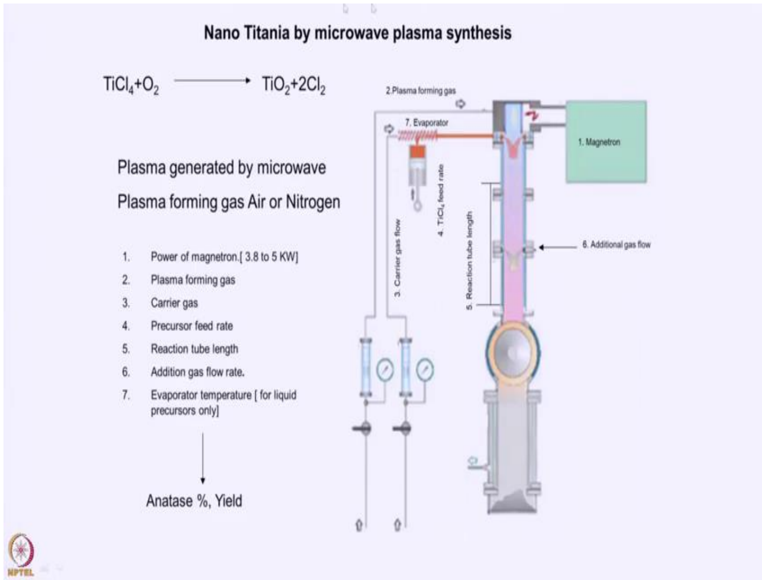
(Refer Slide Time: 25:29)



This experiment, design of experiment was conducted at ARCI in Hyderabad and it has been published in this paper in the Materials and Manufacturing Processes. Here, the idea is to process, we have to optimize the process to produce Nano Titania with several inputs parameters and we have 2 responses, one is a production efficiency and the second one is percentage of Anatase in

Nano Titania powder produced. Based on the result we have to give a commercial viability of the process to the experimenter.

(Refer Slide Time: 26:16)



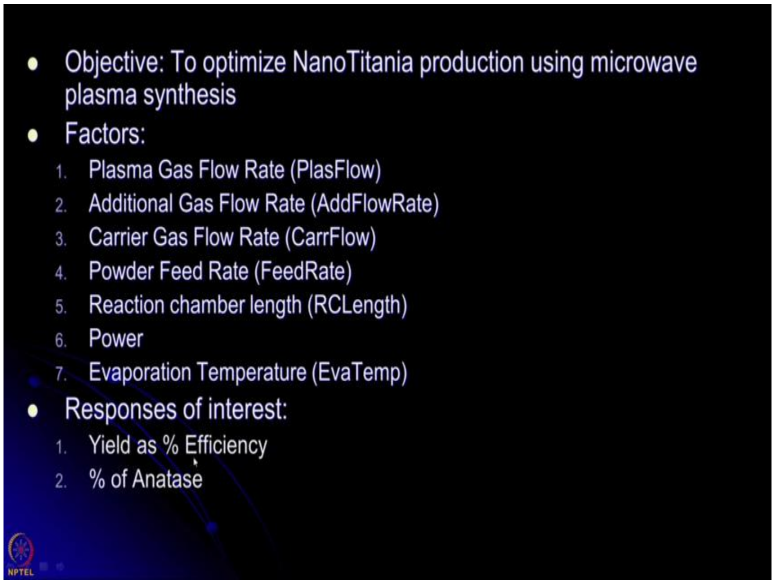
This is the schematic of the Microwave Plasma Synthesis apparatus. There are 7 parameters or 7 factors that that can be controlled.

First one is called power of magnetron, so this is a magnetron power. Then there is a plasma forming gas which goes from here and enters into the chamber. Then there is a carrier gas which carries the, the feed, the feed powder or the whether if it is in a powdered form or TiCl4, which is in the either powdered form or liquid form. This is a carrier gas, if it is liquid, it has there is an evaporator in it.

So there is a carrier gas, there is a precursor feed rate. So this is a precursor, TiCl4 is the precursor, its feed rate is another factor which can be controlled. Reaction tube length, this is a reaction tube and this is whole thing is a reaction chamber. So there is a reaction tube length which you can change and depending on that the powder quality may differ.

There is evaporating temperature in case for liquid precursor only, so there is an evaporating temperature here. And eventually what is collected here is Nano Titania powder, and we would like to look into the percentage of Anatase in the powder and the yield in terms of percentage efficiency of the process.

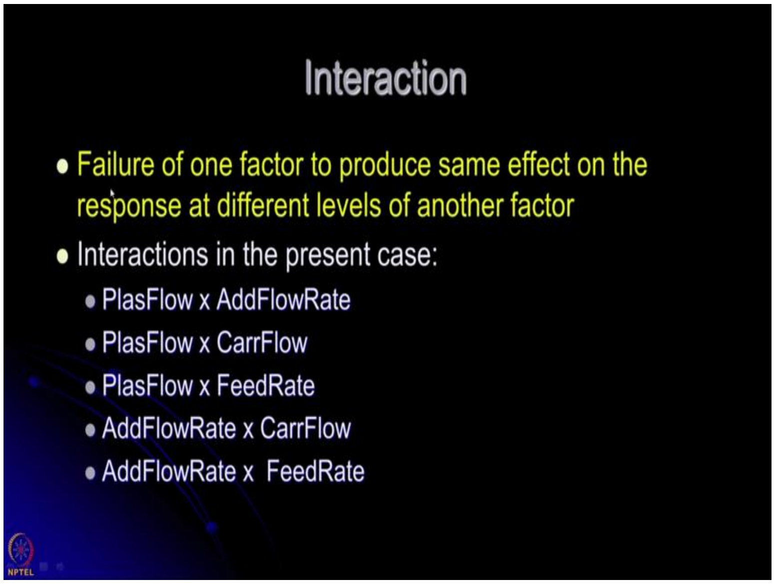
(Refer Slide Time: 28:05)



So the objective is to optimize Nano Titania production using Microwave Plasma Synthesis. These are the 7 factors which is Plasma Gas Flow Rate, I have given you here everyone as a short form, so there is a PA, actually it is a PFR, Plasma Gas Flow Rate, Additional Gas Flow Rate, now Additional Gas Flow Rate is comes straight to the reaction chamber and it increases the time to for the powder to go through the reaction, go through the reaction chamber.

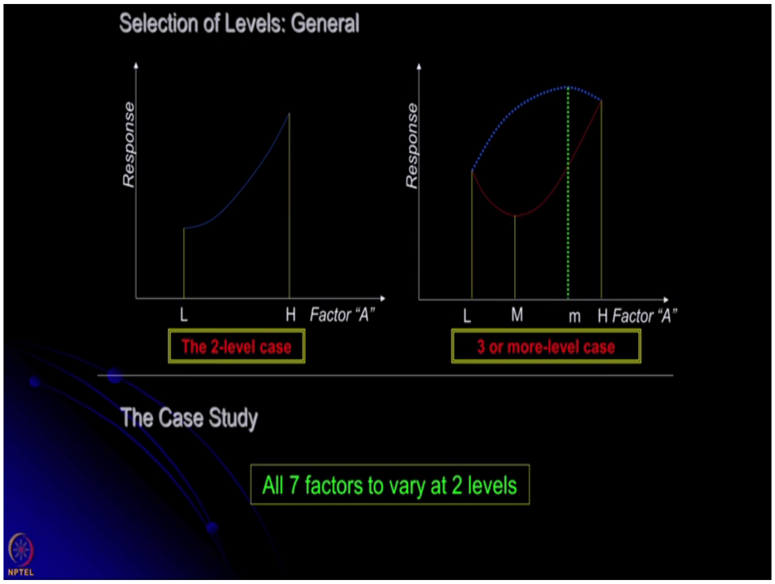
Then there is a Carrier Gas Flow Rate, there is a Powder Feed Rate, Reaction Chamber Length, the Power of magnetron and Evaporating Temperature. And there are two responses, one is a Yield as a percentage of Efficiency and the second one is percentage of Anatase in the powder.

(Refer Slide Time: 29:08)



Interaction, I repeat, failure of one factor to produce the same effect on the response at the different levels of another factor is called an interaction. So here, we found that the plasma flow rate with additional gas flow rate interaction, interaction of plasma flow rate with carrier gas flow rate, plasma flow rate with feed rate, additional gas flow rate with carrier gas flow rate and additional gas flow rate with the feed rate. These were the 5 interactions found important in this case. Next, so these are the interactions which are show in the schematic.

(Refer Slide Time: 30:02)



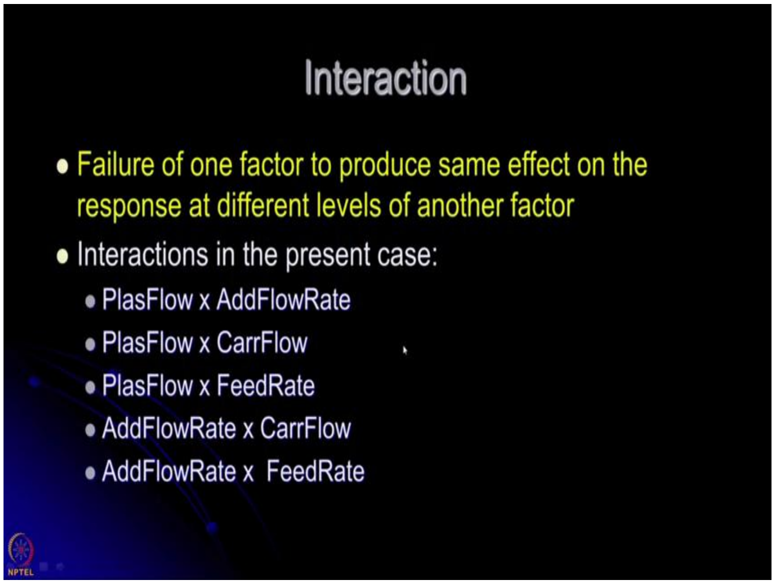
We have to select the levels. Now I am, this is a very general philosophy how you can select the level with a statistical point of view. Suppose, design of experiment point of view, I correct myself. Suppose, you are varying a factor A from level low to level high. And this is your response axis, and you find that the response graph is something like this, which is a steadily increasing function or it may be a steadily decreasing function, okay?

If that is the case, then there is no point taking more than two levels of experiment because it is a steadily increasing function so optimum value is going to occur either at one end or the other end and therefore, it calls for a 2-level case that is, your factor need to be vary, factor A need to be vary at 2 levels.

Suppose it happens that it reaches a minimum or a maximum while varying from levels, low level to a high level for another factor A, in that case it is necessary to find an optimum that we must include a medium, a third level in the middle. And therefore, we may have to consider 3 or more level case depending if it is very wavy, then you would like to have more levels but generally if it is deeps and it takes up or it takes up and then comes down you are going to consider it as a case for 3 or more levels.

Now all this definitely tells you that this cannot be said by the experimenter or by the scientist who would like to understand this Microwave Plasma Synthesis system for Nano producing Nano Titania that this cannot happen unless he has done some experiments before.

(Refer Slide Time: 32:50)

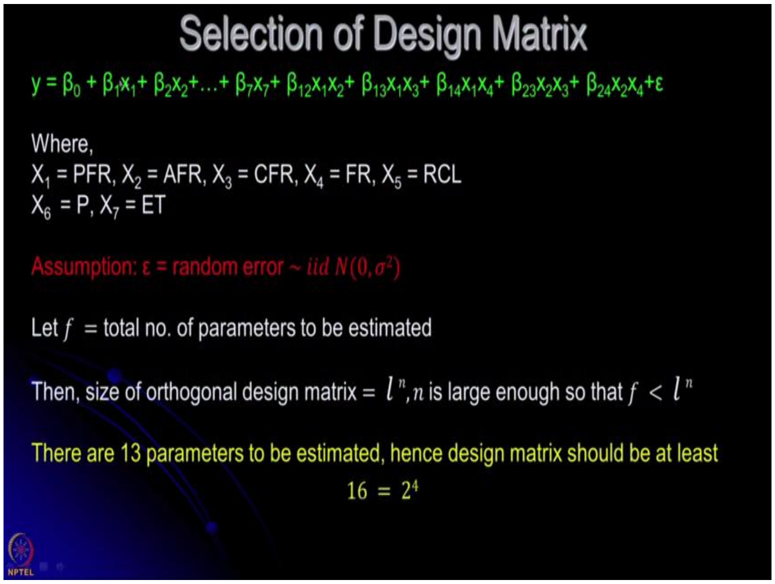


And let me emphasize that designed experiment is never your first experiment. Designed experiment is your experiment which comes after several of those preliminary experimentation to understand the system itself, to understand the process, then only you would know what are the interactions which are going to, which are playing the role and what are the, what are the levels that you can work with. So please remember designed experiment is never the first experiment that you ever conduct. It has to be preceded by several experimentations which will make you understand this system.

Design of experiment is meant to optimize the system, your optimize the results, your optimize the responses from the process that you want to go through, it is, it cannot be the first experiment and as we saw that at times if we are conducting the response surface kind of analysis then it is actually a sequential set of experimentation, so any designed experiment may not be the last one either.

But this is worth remembering, that design of experiment is never the first experiment. One has to have a several experiments before to understand the system.

(Refer Slide Time: 34:05)



Now having done that, we have a selection of design matrix. So this is the model that we are going to fit. Please remember our case of multiple regression model.

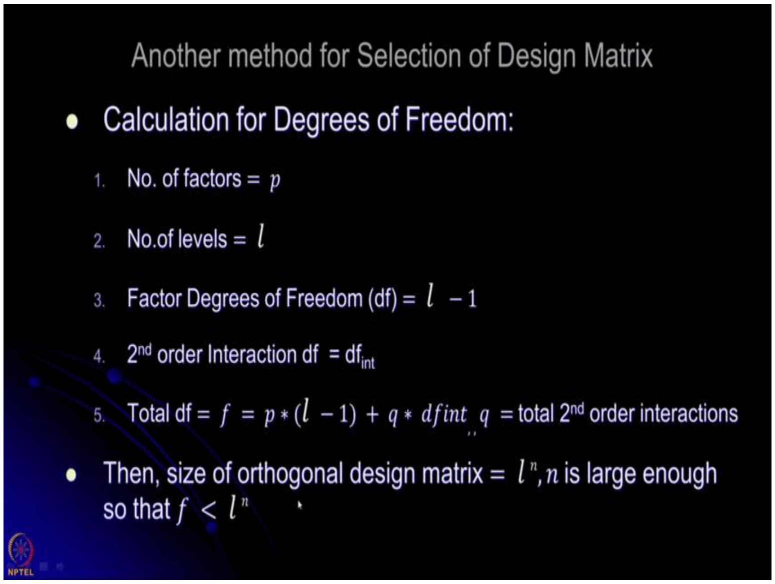
y is your response and the model I am assuming is beta 0 plus beta 1 x1 beta 2 x2 plus beta 7 x7 up to this point I am talking about the 7 main factors, then I am talking about beta 1, beta 12 which is an interaction between x1 and x2 so I write it as beta 1 x1 times x2, beta 13 x1 times x3, etc. And for example, x1 x4, it means that I am talking about interaction between powder feed, it is a plasma flow rate, I am sorry, it is a plasma flow rate with the feed rate of the powder.

So it is the interaction between these two. So likewise, and then at the end we put an epsilon. And I assume that epsilon is a random error, it is iid, that is if I conduct n experiments here then I will have n1, n2, n3, y1, y2, y3, yn results and therefore, I will have the random error epsilon 1, epsilon 2, epsilon 3, epsilon n, all of them are independently, identically distributed with mean 0 and a variance sigma square, where I do not know what is sigma square, sigma square is unknown.

I call f as a total number of parameters to be estimated, then the size of orthogonal matrix, this is my model, how many experiments do I need to conduct? So, I calculate it in this way, that if f is a total number of parameters that I have to estimate, then the site, size of orthogonal matrix should be say l to the power n, where l is the level, number of levels that you have chosen and n is the size of the experiment. Then you have to choose, your l in this case is 2 because we have taken 2 there and you have to choose n so that your number of parameters that you are going to estimate is smaller than l to the power n.

So here we have 13 parameters, so f is equal to 13. Your levels are 2, so 13 is less than 16 which is 2 to the power 4, and therefore, you need to conduct 16 experiments. We need to conduct 16 experiment. So you have to choose n large enough so that 2 to the, l to the power n becomes smaller than f, then l to the power n are the number of experiments you have to conduct and therefore, in this case we have to conduct 16 experiments.

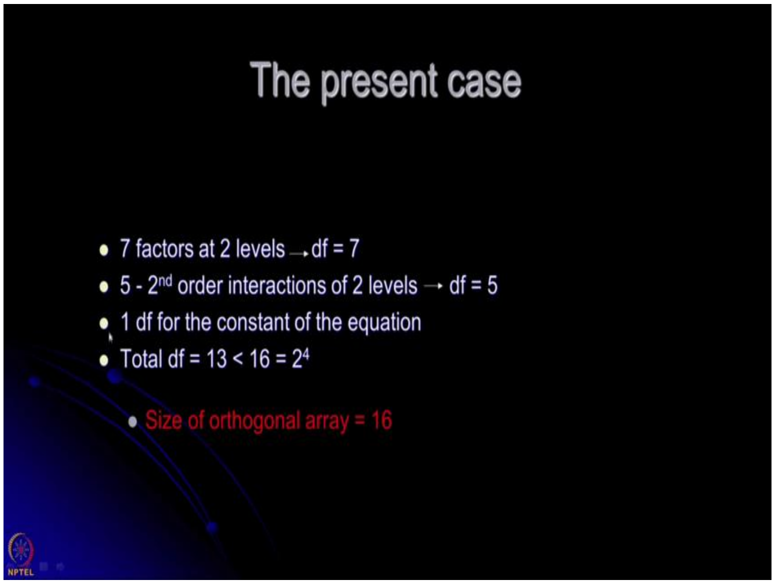
(Refer Slide Time: 37:25)



There are other methods of calculating the size of design matrix. Suppose, number of factors are p, number of levels are l, then factor degrees of freedom is L minus 1. Second order interaction, we call it degrees of freedom of interaction. Then the degree of freedom f is p multiplied by, p is the number of factor multiplied by L minus 1 plus q, which is total number of interactions, multiplied by degrees of freedom of interaction.

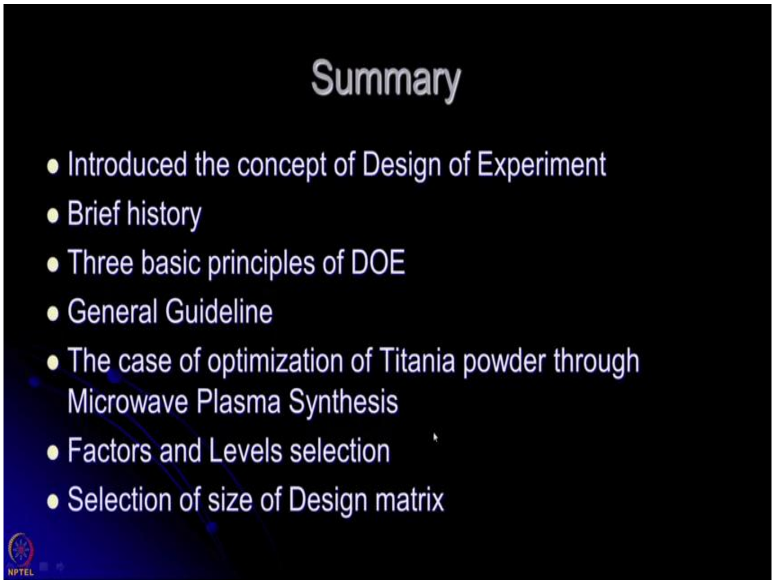
And then, the size of orthogonal matrix is L to the power n where n is large enough so that f is smaller than l by n.

(Refer Slide Time: 38:16)



So accordingly, we have 7 factors at 2 level so degrees of freedom is 7. We have 5 second level interaction, so degrees of freedom is 5. 1 degree of freedom for the constant of the equation, you remember we have a constant here, beta 0, so for that we have to add 1 into it, so total degrees of freedom is 13 and therefore you have to choose n large enough so that l to the power n is just larger than 13 and therefore the size of your orthogonal array is 16.

(Refer Slide Time: 38:56)



So let us summarize. We introduced the concept of Design of Experiment. We gave a brief history of it. The three basic principles of design of experiment randomization, replication and blocking, we explained. We explained the general guideline and we introduced the case of optimization of Titania powder through Microwave Plasma Synthesis. Factor and level selection we discussed and selection of design of size of design matrix we discussed. Thank you.