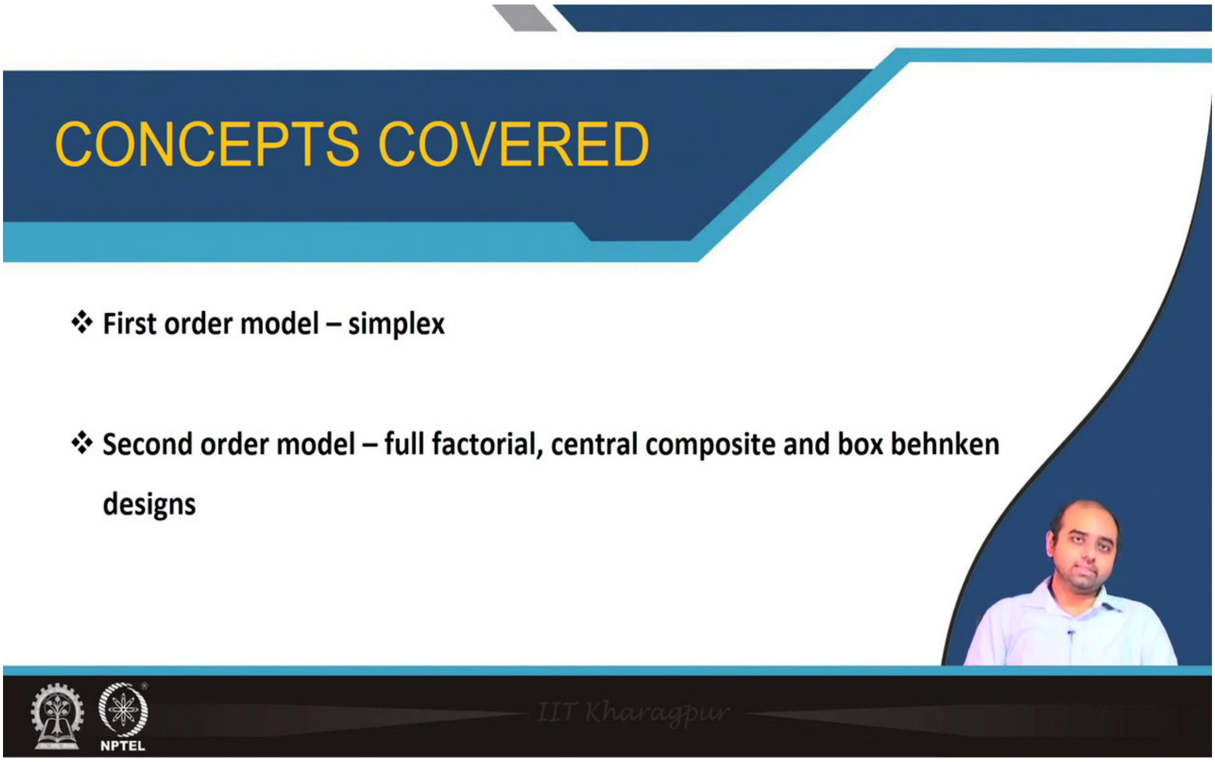
**Mathematical Modelling and Simulation of Chemical Engineering Process**
**Professor Doctor Sourav Mondal**
**Department of Chemical Engineering**
**Indian Institute of Technology, Kharagpur**
**Lecture 58**
**Design of experiments**

Hello everyone, design of experiments is a very important aspect for response surface methodology, because we want to evaluate or we want to understand that what are the possible number of experimental data that is needed for a particular response model to be used or to be developed. So, generally experiments where resources are very limited or it is too time consuming, this becomes more significant and you do not have the possibilities to generate as many data points as you want, this becomes quite necessary.

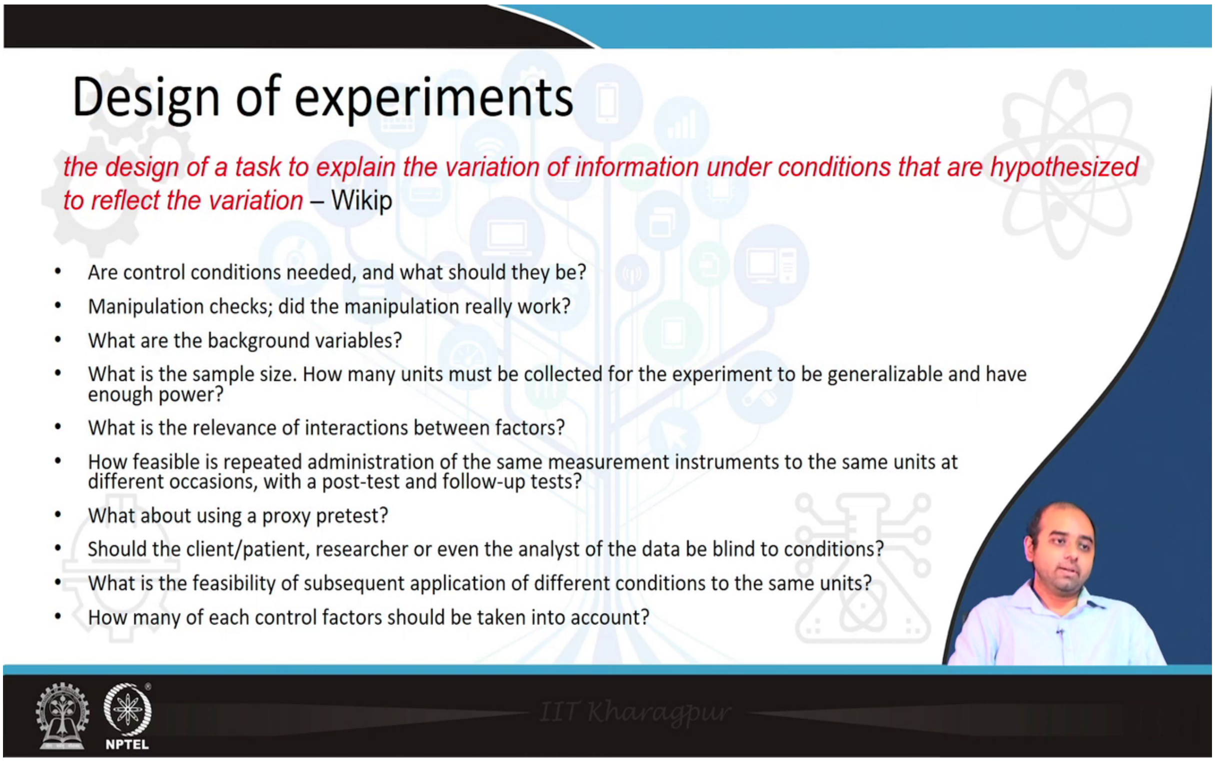
So, in most chemical or biological or biochemical processes, these becomes I mean, it is not only time consuming, but it also requires a lot of resource, a lot of manpower. So, then proper choice of your process variables or the values of the variables is very very vital, that is where the knowledge of design of experiments becomes very crucial.

(Refer Slide Time: 01:37)




So, generally the two types of models that we will talk about, one is the first order model and another one is the second order model. So, first order model we have the two to the power k models as well as the simplex one which we will talk about and in the second order, we will have the ones like full factorial, fractional factorial, central composite and box behnken designs.

(Refer Slide Time: 02:01)




So, what exactly we look for when we try to think about the design of experiments. So, these I have listed down several factors that needs to be considered when we try to get our experimental data or while we try to design our experiments. The first thing is that, are there any controlled experiments or controlled conditions which are needed or not? And if yes, then that is something that we need to find out that what that could be. So, for example, in any first order, second order model, you always have a constant term.

So, the constant term is sort of the bias to the problem or a bias to the system or something that, a response you get even without any input. If the input conditions are still 0, you still manage to get that bias or that constant. So, that means, a controlled experiment can be useful to find out that even without an input variable, are we getting any output for the process? And if yes, then how do you design that experiment? Or how to set up that experiment.

Second is that manipulation, the checks, like whether there are any manipulations that you feel can be done to the process so that the accuracy of the response is improved or the sensitivity of the response is improved. Then effect of background variables. So, we may be tracking two variables. Like for the last class, if you realize in that table where we listed down the time and the temperature of a reaction process at the bottom 4 or 5 values, they are all same, but still the response is different.

So, that means, there is an issue with some background effect, that is something that needs to be defined. And then there are a lot of these factors that you can see, it is already listed down in the screen that what are the interaction between the factors? Like, how important are they? If you have some background information on that, that would be helpful. Then the question comes that how much reproducible are these results. So, that is very important to identify the effect of the background variable or it is the error in the reproducibility of an experimental data.

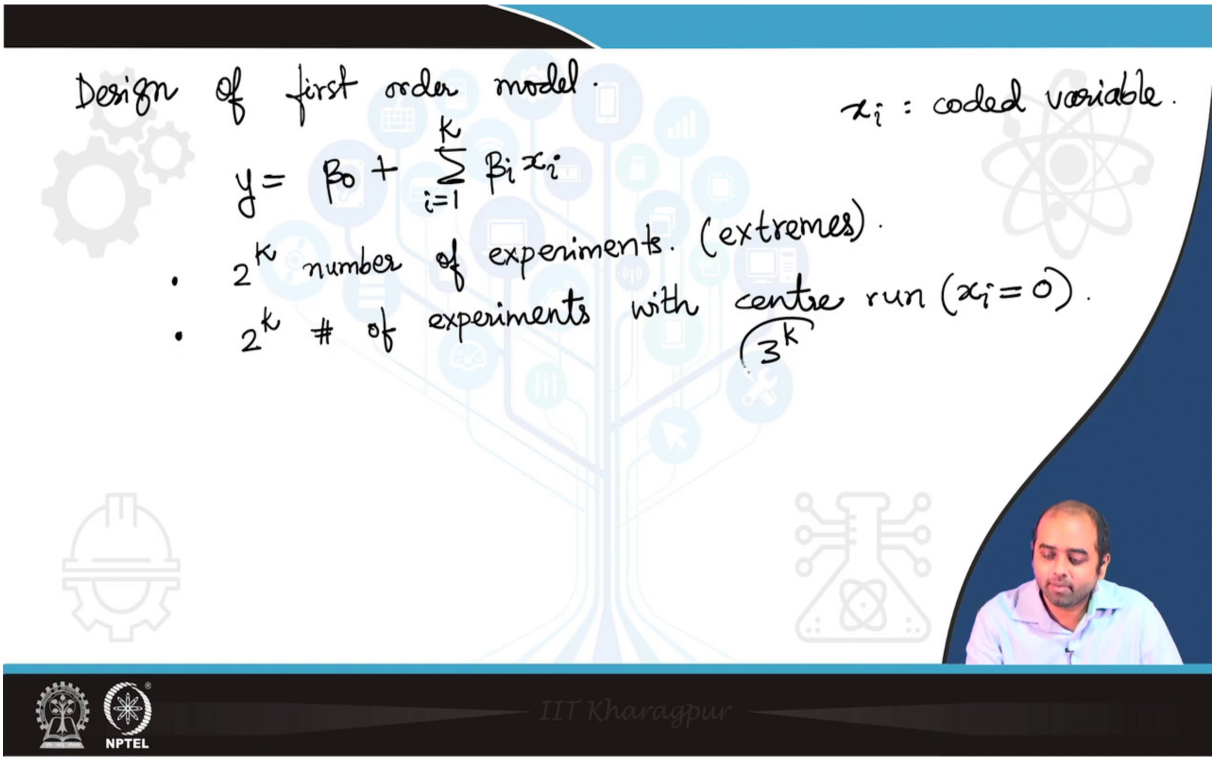
Similarly, the question comes that what is the experimental error associated with a particular data? Then some other more questions, which is very important and which is something that also brings variability to the final result is that are these experiments or these observations influenced by environmental factors. For example, in clinical data test, it all, the result depends a lot on the patient or the client and their condition.

So, that also needs to be factored in. So, these are the different issues that comes up often with experiments and we have to take into account of all of these very carefully so as to understand that the final model that we create, what are the influence of these issues behind the experiment and whether they are accurately accounted for, and is there something that can make the output to be more sensitive towards the process variable is something that we want to achieve.

So, we want to really see the effect of the process variables to the final response. So, these sorts of other factors can always play a big role and that is something we need to be very careful about, as well as the error associated with the particular data in the experiment is also something which can influence the effect of the background. So, whether it is an experimental error or whether it is the effect of the background is something that needs to be also explicitly and exclusively identified and understood.

So, all of these comes under the purview of Design of Experiments. So, that is something we can, so, the idea is that with design of experiments, we can get the data set that helps us to explain or relate the final response or the experimental observation with respect to our process variable or change in the process variable.

(Refer Slide Time: 06:47)



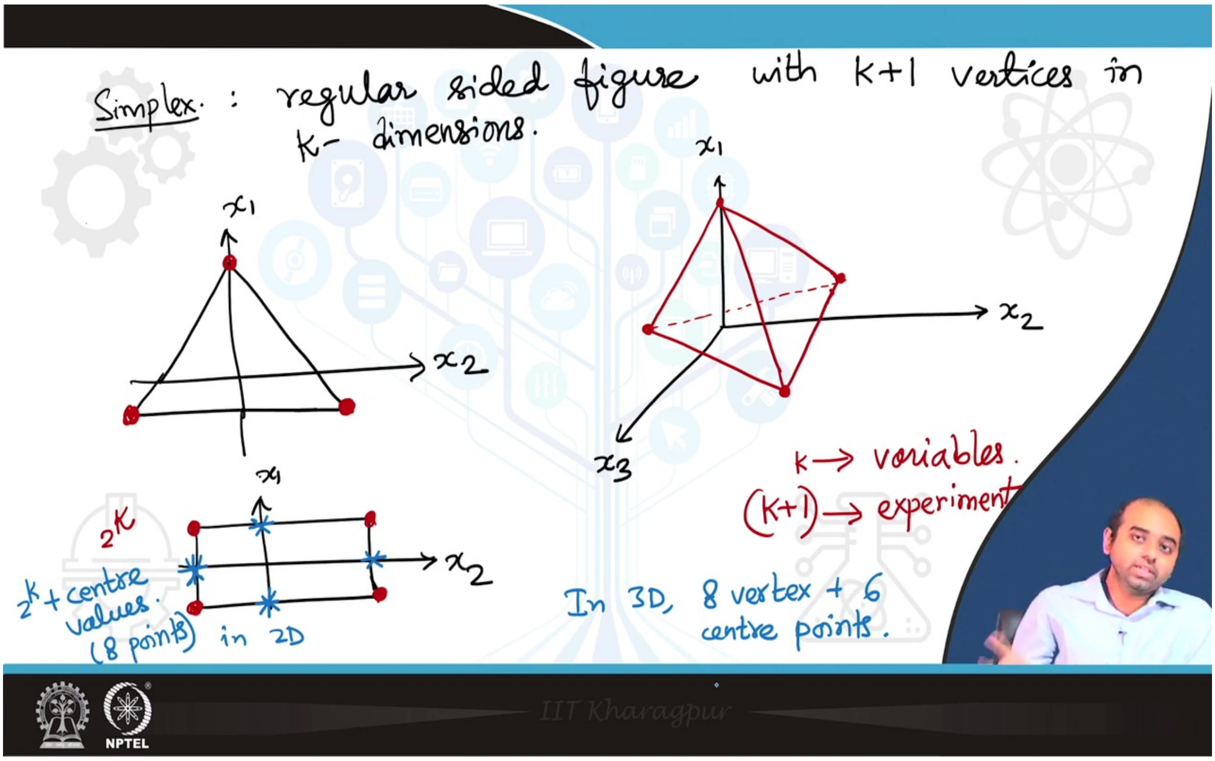

Now, what about design of first order models? So, first always one should try of a faster model and see whether it can explain all the possible scenarios or like the experimental data. If it fails, then we move to the second order models. So, as you know the first order model is y is equal to beta 0 0 plus beta I xi and this is like i is equal to 1 to k number of experiments, if you have, this is the first order model. There are no interaction terms, no second order terms.

So, the easiest way to have the minimum number of experiments is to have 2 to the power k number of experiments-one high, one low. One at the highest and the lowest value. So, for each process variable all the extremes are covered. The next option is to have 2 to the power k number of experiments with center run. So, this means that apart from the extreme, we also will be needing value at the center.

So, here x whatever I have written is in terms of coded variable. So, x is equal to 0 represents the mean value of that process parameter. So, that is something so, additional one run maybe added and this is something that can help us in estimating the situation when one of the process parameters or the effect of one of the process parameter is turned off. So, setting the center run, so, if xi is 0, right, but let us say x1 is 0, but x2 is not, so, it is not only the issue of the combined extra extremes, it is also the issue of one of the effects of one of the parameters is not present is something which can also be tracked down.

So, it helps us to determine the estimate, the variance of the system and it does not influence the regression coefficient. So, that is something which is very important and the estimation of the beta 0 also becomes more accurate. So, having a center run is sometimes very useful. Apart from the extremes, it is something we get in 2 the power k experiment. So, this will of course, lead to 3 to the power k number of experiments with the addition of one center. And if you think that so much of experiments are not possible or you cannot afford, then you have the simplex design. So, what is the simplex design?

(Refer Slide Time: 10:07)




So, instead of, let us talk about the simplex design in a little bit more detail. So, in the simplex design you consider a regular sided figure essentially with k plus 1 vertices in k dimensions. So, let us say x1 vary, forget x1, let us try to draw this and so, this is my x1 and this is my x2. So, these are the three points in this region or combination of three points which will give us at least a close bounded region of the parameter space.

We can choose this x1 at least in one of the points, one of the values would be extreme. So, this is something known as a simplex design where you can have as the smallest bounded figure can be attained by three sides, that is a triangle we all know. So, in two-dimensional space a closed object can be produced with only three sides, I mean with more number of sides it is always possible but the minimum number of three sides is near and that is the basis of the simplex.

Similarly, in the case of three-dimensional space let me try to draw it. So, let us say this is x1, this is x2 and this is x3. So, if you are having a three-dimensional space, then it is a sort of the you can think of this as the pyramid. I hope you are getting this pyramid. So, a pyramid is actually a closed bounded figure with four surfaces. So, this is what is essentially so, it essentially relates down to k, so, if you are having k variables in your system, with the simplex design you can have k plus 1 number of experiments to have, close boundary of the parameter space.

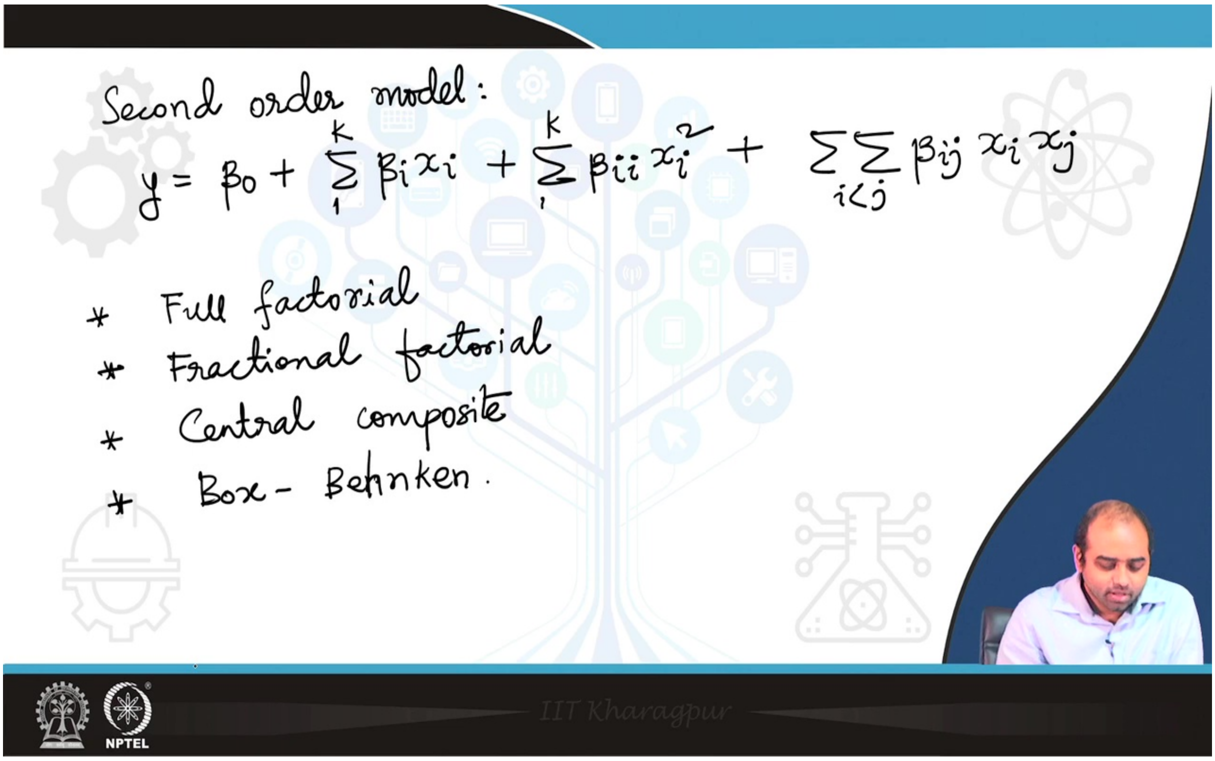
So ultimately, we want to find out the parameter space with which we can have the model of the first order equation which we are trying to estimate that. So, minimum k plus 1 number of data points would be needed that would follow from the simplex design. So, in this context, you can also think that the two-factor process, so, if you have the 2 to the power k model that could be equivalently thought of as an this kind of rectangle. So, this has 2 to the power k number of points.

Similarly, in the case of three factors, it is a cube. So, this is a simplex and this is 2 to the power k model and you have, if you want to bring in additional center points, center runs, so these are the center runs. So, this one is 2 to the power k plus center values. This will be approximate design. Similarly, in the case of 3d, it would be like the center of the faces of each surfaces of the cube that needs to be also factored in.

I hope all of you realize what do we mean by these centers spaces again. So, as I can see that if you have 2 the power k plus center values it is not coming out to be a 3 to the power k. So, I should not write this, this is a mistake. 2 to the power k with center runs in the case of three variables. Two variables it comes out to be this 12 and in the case of eight. So, in the case of three variables, it turns out to be you have eight edge points and six center points, total 14 number of points you will be getting with 2 to the power k plus center values.

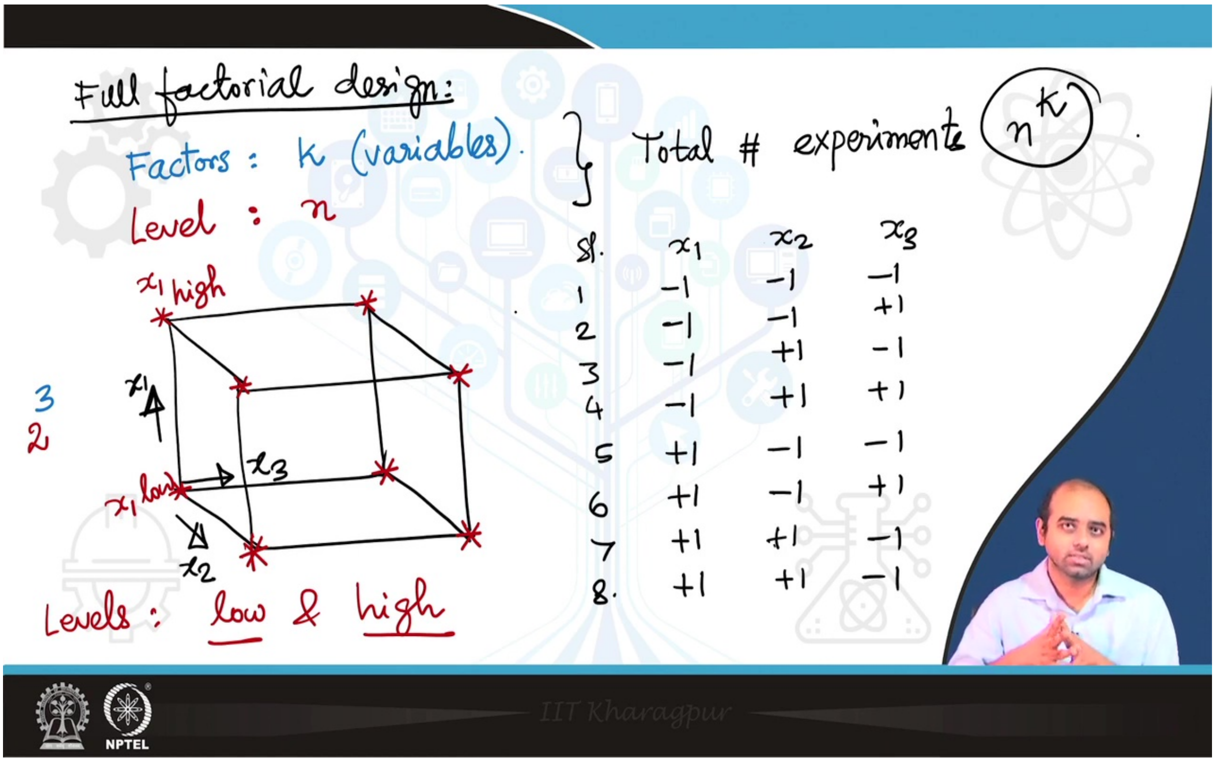
So, in this case, you will be getting eight points in 2d. So, in 3d you will be getting eight vertex plus six center points. So, total 14 number of data points that would be needed. Whereas, in the case of simplex design you will be getting only four data points as opposed to 14 data points that is needed with 2 to the power k plus center runs.

(Refer Slide Time: 16:24)




So, now next let us move to the second order model. The second order models, the general equation for two variables is 0 plus sigma beta i xi to k to k. So, what are the different types of second order model that we have that is generally popular? One is known as full factorial design, next is known as fractional factorial, another one is central composite, then you have the Box Behnken models. There are some more also but at least these four we will study today itself.

(Refer Slide Time: 18:11)




So, what is the full factorial design? Let us say you have the factors, k number of factors or k number of variables in your process and level, so, each you can consider the levels to be as the lowest value or the highest value, low, mid and high values, something like that as the levels. So, if this is the case, I mean if you have n number of levels in your system, then the full factorial design tells you that total number of experiments is n to the power k.

So, what is the common way to visualize this? So, let us say if you have three parameters in your system or three factors in your system and based on their extreme values, this side is x1, this side is x2 and this side is x3. So, these are sort of the lowest and the highest values for each of the parameter variables. This is like x1 high this, is like x1 low and this is low for everyone. So, all of these points, the corner points of this parameter box or the phase box represents the number of experiments that you need to do.

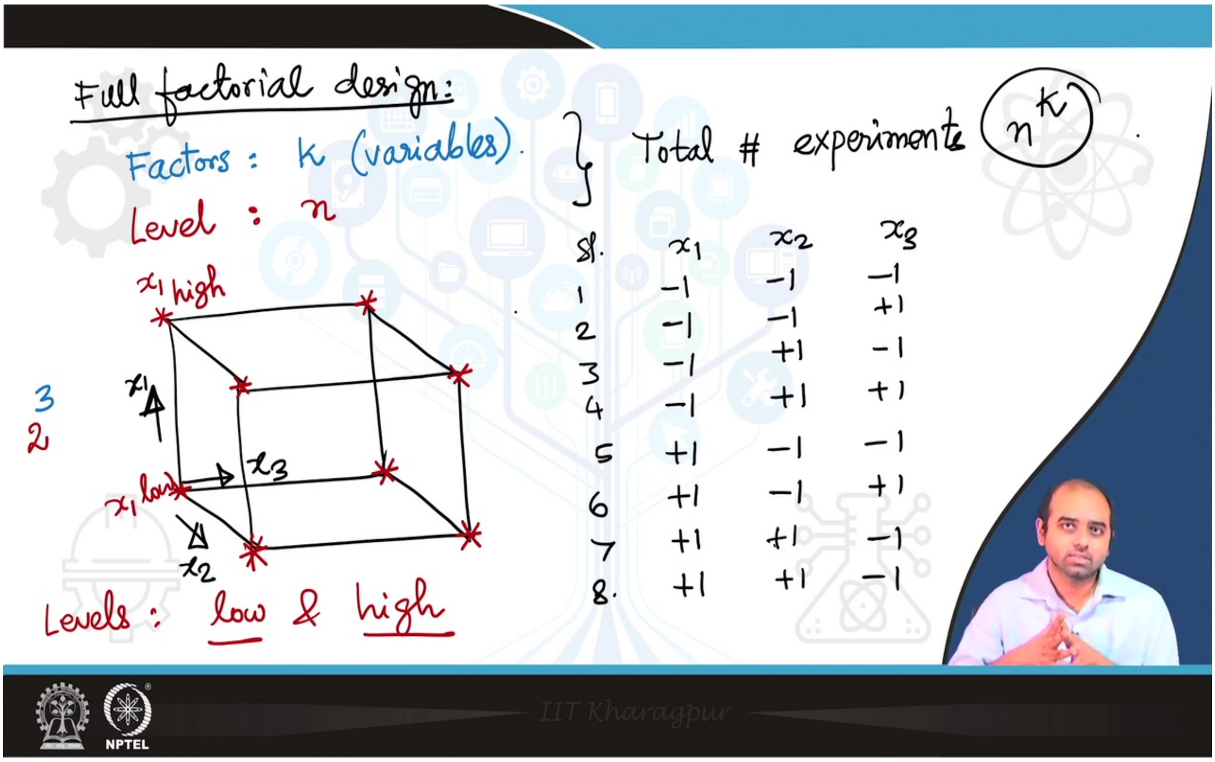
So, in this case it is there are three factors right and two extremes or two levels. So, in this case, we have considered that the levels as low and high and this is normally the case in sometimes you have mid values. So, in this case you are having two levels and three factors. So, 2 to the power 3 number of experimental data points which is needed if you follow the full factorial design process.

So, if you have the experimental data set coded variables x1 x2 x3, then all possible comparisons of the low and high, so, which means we will be having minus 1 minus 1 minus 1, plus 1 plus 1 plus 1 plus 1 plus 1 minus 1 plus 1. So, all these combinations, so, if you count them you will see total eight possible values or combinations that we can get here. So, this list downs all the possible combinations of these extremes of this x1 and x2.

So, what is the advantage? The advantage is that you get all the possible compounded effects of the variables. So, this is a very good way to find out that the compounded effects are actually captured nicely in the final response and also, we see that the main and the interactive effects are also understood quite properly because this is a combination of all the low and the possible high values.

So, that is why the main value since… and they are also operated at the extreme point, so, the main effect would also be there and the component effect will also be there. So, both are quite helpful. One downside is that if the number of process variables increase, say if you have a number of factors in the process increase like 4 or 5, then this becomes too large as the possible number of experiments.

(Refer Slide Time: 23:14)



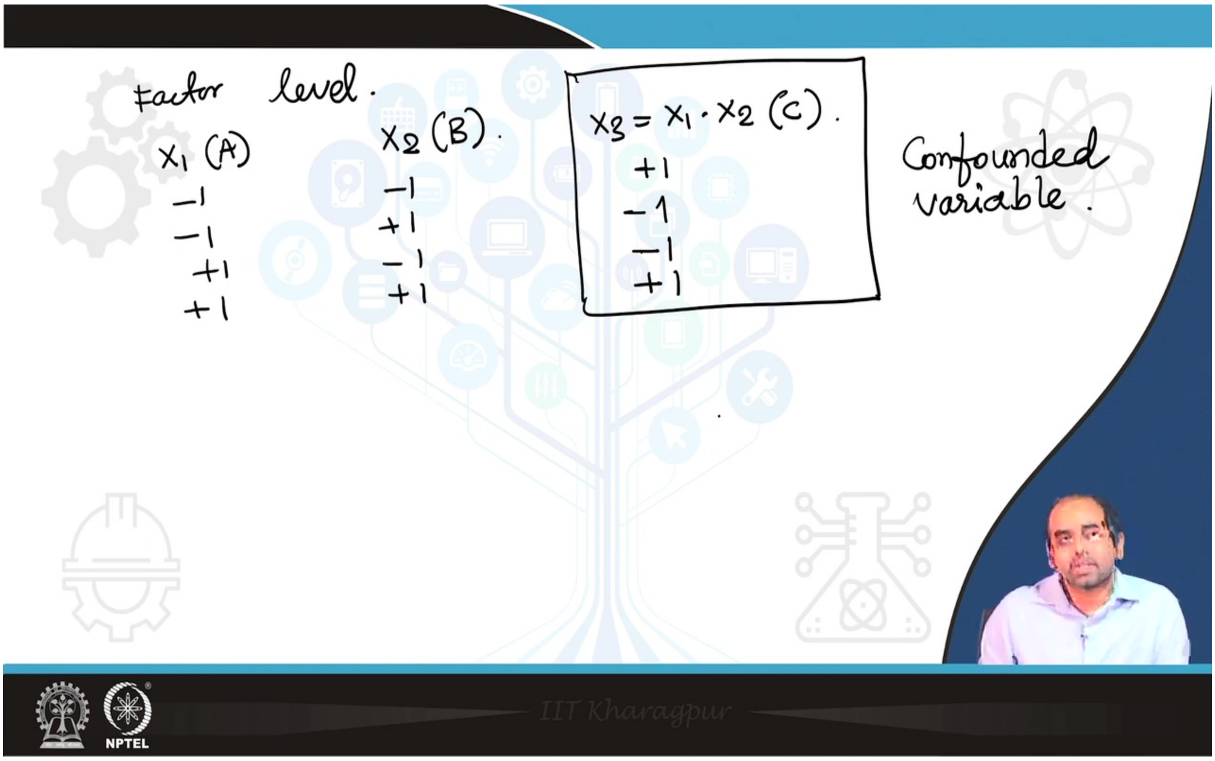

So, that is why you have the other design known as the fractional factorial. So, in this case of the fractional factorial, it is actually a subset of the full factorial if I may put it that way and what is done is that for one half fractional factorial design, you reduce the total number of extreme ends from full factorial, it is 2 to the power k, it becomes half of 2 to the power k. So, it is 2 to the power k minus 1 and how it is that possible?

You actually confound the effect of two variables. So, if there are two variables, so, how it is possible? Every main effect is confounded with two interactions, two factor interaction. So, which means that if you are having three parameters in your system, let us say A, B and C, so, in the fractional factorial, we assume that this third variable is confounded. So, C is actually related to A and B in some way.

So, therefore, this effect of AC right if we try to calculate let us say AC, so, that is what? A into AB which is essentially B. And similarly, BC. So, that is B into AB which can be eventually related as A or B or whatever. So, this can be related as AB. Similarly, BC can also be related as AB. So, the effect of the third variable does not exist explicitly and that is something which we confound it. So, this is an assumption we make of course and in so, many cases this becomes quite logical also that the third effect is not so important in this problem.

So, that is why the fractional factorial can be helpful when you have more than at least three or four factors in your system. So, one effect cannot be studied explicitly. It needs to be aliased with the interaction term. So, based on that whether you want one factor or whether you want two factor in your problem, that is something it needs to be decided by you.

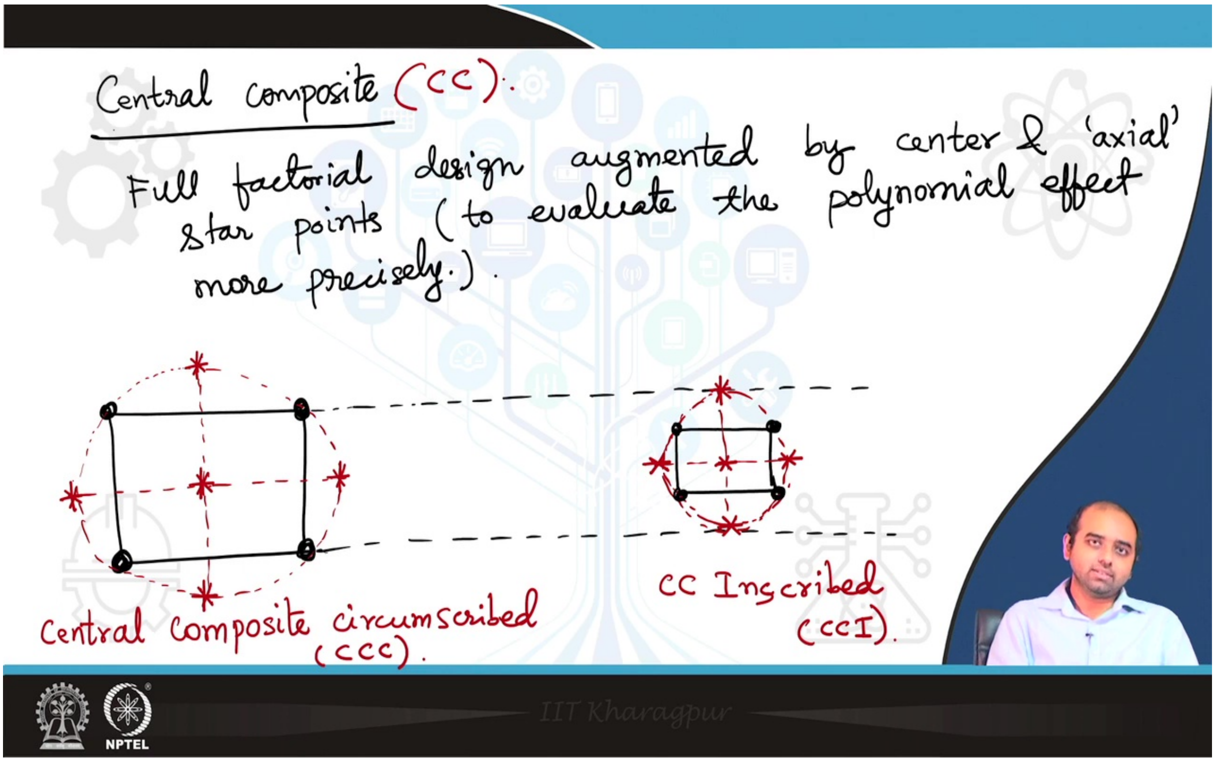
(Refer Slide Time: 27:29)




So, an example of the third I mean this fractional factorial is that let us say the factor level that you have in your problem let us say I have x1 which is like A and x2 which is like B, then I have only two parameters to be explicitly determined. So, I will be having only four combinations. So, this is like the confounded variable, so, this becomes plus 1, then I have minus 1 minus 1 plus 1. So, this is the effect of the confounded variable in this case in the full, this fractional factorial design.

So, you do not explicitly set or change that value of that variable, but based on the values of A and B you can essentially estimate out what would be the value of the C for which you want to study, if you are following the fractional factorial design.

(Refer Slide Time: 28:54)




Now, let us move to the central composite design. So, the central composite design tells you that apart from the full factorial design you will also be having some star points in the system. So, it is like full factorial design. Let us say you can afford more experiments, full factorial design and it is further augmented by center and axial star points to evaluate the polynomial effect more precisely.

So, what is the scenario here? So, let us say we have the full factorial design, for two factors I draw, so, this is full factorial design with two variables, x1 and x2, is not it? This is a full factorial design. Now, what are the star points? One-star point is the center. So, this is one star point, I mean the additional points in the central composite and then you have star points on the axial positions in each dimensions.

So, if I try to draw a sort of a circle where the circle radius is the half diagonal of this square, so, then these axial points are also located at the same distance from the center as the corner points are the full factorial points, is not it? So, the circle is drawn considering the radius as the half diagonal. That is the reason why I am saying this distance of these axial points are same compared to the corner points.

So, this is what the central composite design known as. Here you have for 2 to the power k number of full factorial points plus a central point plus these axial star points. Now, since these axial star points are outside the bounded regime of the full factorial or the corner points, this is known as circumscribed. So, this is known as central composite circumscribed. So, central composite circumscribed design, CCC. But, if it is possible that these axial points cannot be outside the maximum value of any particular process variable, then you scale down the distance of these axial point or essentially the entire box, you scale down that distance.

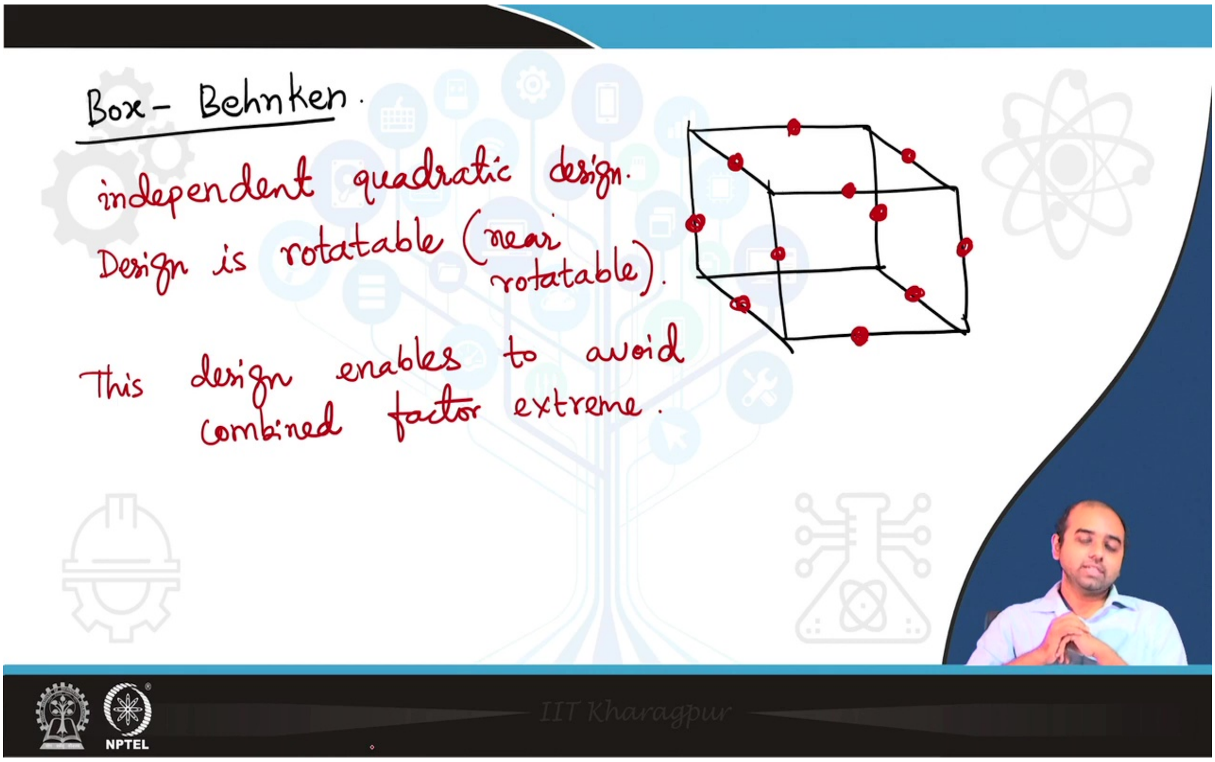
So, how? So, now, if it is I mean if we see that it is not possible to exceed the limits of the extreme value of the process variables, in that case what we do is that the start points cannot be outside of that rectangle box. So, now the box is actually scaled down to something like this and now this circle that whatever we have is inscribed. So, the star points now do not cross the limits of the extreme values.

Of course, this means that the corner points of the full factorial… So, if you see that in this case I have scaled down my box such that the circle the actual circle where we are locating the axial points, please excuse my drawing, these axial points do not exceed the physical space limit of the actual process variable. So, this means that essentially the corner points of the full factorial design is also reduced and the axial points is now located on the parameter extremes and not outside of it.

So, this is something known as the central composite inscribed, CCI design. So, these two are generally the very popular modes of the design, there is also something known as a central composite face design which is applicable for 3 and more dimensions where you take points on the center of the faces and not on these axial limits. So, that is something known as the central composite face.

So, please remember that in this case we are having 2 to the power k number of full factorial points plus these additional star points. So, in the case of two variables or two factors we have additional five-star points. Similarly, in the case of three dimensions it would be more and three factors and four factors, it would be generally. So central composite is not very popular for more than three factors of course.

(Refer Slide Time: 36:48)




Last one is the Box Behnken design. So, in this case, this is applicable for three factors and onwards. We try to draw a cube and the factors are located or the points are located at the center point of the edges. So, here these are not the corner points, but the center points of the edges. So, what is the advantage of this compared to the full factorial design? Because here also you get eight number of points. So, this is generally known as the independent quadratic model design. So, it does not represent the full factorial or the fractional factorial because it is not at the corner points but at the midpoint edges.

So, here it actually calculates or takes into account of the interactive effect very nicely. And one of the best thing about this is that these designs is rotatable or near rotatable which means if you rotate this box, gift positions or the locations does not change. So, this is very vital and this is not the case with any of the designs. So, even if you rotate this box in any way, even if you change the XYZ directions also, the combination of the midpoint combination still remain the same.

So, with these midpoints, the box can be rotated. So, this is known as a rotated box design. It is a multi level, multi factor experiment of course, that you have realized, where you see that confound, I mean interactive effects of two variable extremes is very difficult to adjust in the experiment, then this is something which is very useful because here none of the factors is at the extreme level.

And this linear and quadratic both these effects are quite nicely handled here. And of course, as I said one of the biggest advantage is that you can avoid the combined factor extremes. This design enables to avoid combined factor extremes. So, it may happen that this experiment becomes too difficult to handle when all the parameters are at extreme levels and this is where this turns out to be very useful.

So, these are the different types of design experiments, experimental designs, which can be made for the second order design and first order also we have talked about. So, with this, I hope all of you got a fair exposure and understanding of how the response surface methodology actually works, what are the different kinds of design or the experimental design that can be made or prepared.

And I want you to also realize that this is a Blackbox model, but a very careful or a very methodical approach needs to be done so that you can have the appropriate choice of your experimental data, particularly when it is resource limited, time consuming, expensive, laborious. I hope all of you found this quite useful. And in the next class, in the next couple of classes, we will talk about a more advanced method of modeling this sort of stochastic process using neural networks. Thank you everyone, and see you in the next class.