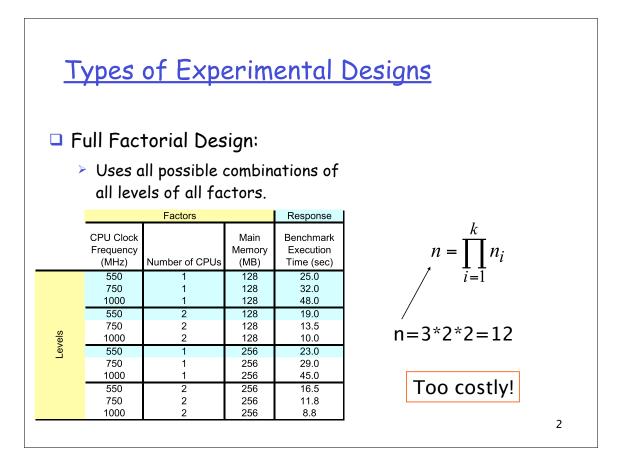
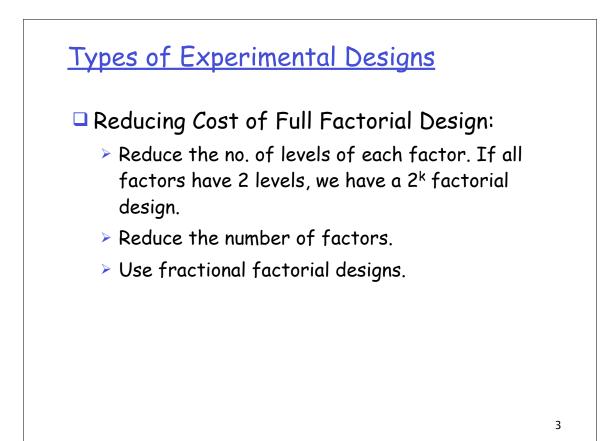
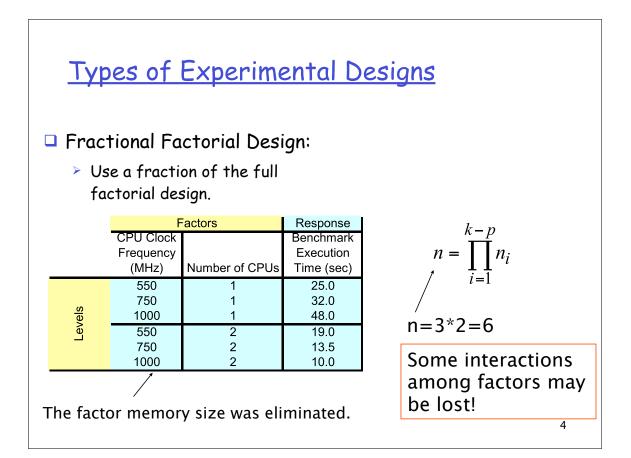
# 2<sup>k</sup>, 2<sup>k</sup>r and 2<sup>k-p</sup> Factorial Designs







# 2<sup>k</sup> Factorial Designs

## 2<sup>k</sup> Factorial Designs

2<sup>k</sup> designs are used to determine the effects of k factors, each of which have two alternatives or levels

- > Easier to analyze than full factorial designs
- Help sort out factors in the order of their impact, especially when there are a large number of factors

# 2<sup>2</sup> Factorial Designs

Special case of 2<sup>k</sup> design with k = 2
 Example: impact of cache size and memory size on the performance of a computer

Cache	Memory Size						
size (KB)	4 MB	16 MB					
1	15	45					
2	25	75					

Performance in MIPS

**Regression Model** Let y denote the performance of the computer. We can model y using a non-linear regression model as follows:  $y = q_0 + q_a x_a + q_b x_b + q_{ab} x_a x_b$ where x<sub>a</sub> and x<sub>b</sub> are variables that represent the factors memory size and cache size respectively, and the q's are called effects

Regression Model (con'td)

Let  $x_a = \begin{cases} -1 & \text{if memory size} = 4 \text{ MB} \\ 1 & \text{if memory size} = 16 \text{ MB} \end{cases}$  $x_b = \begin{cases} -1 & \text{if cache size} = 1 \text{ KB} \\ 1 & \text{if cache size} = 2 \text{ KB} \end{cases}$ Substituting the four observations yields

> $15 = q_0 - q_a - q_b + q_{ab}$   $45 = q_0 + q_a - q_b - q_{ab}$   $25 = q_0 - q_a + q_b - q_{ab}$  $75 = q_0 + q_a + q_b + q_{ab}$

Regression Model (cont'd)

There is a unique solution for the four effects:  $q_0 = 40$ ,  $q_a = 20$ ,  $q_b = 10$ ,  $q_{ab} = 5$ . So, we have:

 $y = 40 + 20x_a + 10x_b + 5x_a x_b$ 

Thus the mean performance is 40 MIPS, the effect of memory size is 20 MIPS, the effect of cache size 10 MIPS, and the interaction between memory and cache size accounts for 5 MIPS

# <u>Computing effects</u>

In general, the model for a 2<sup>2</sup> design can be solved to obtain:

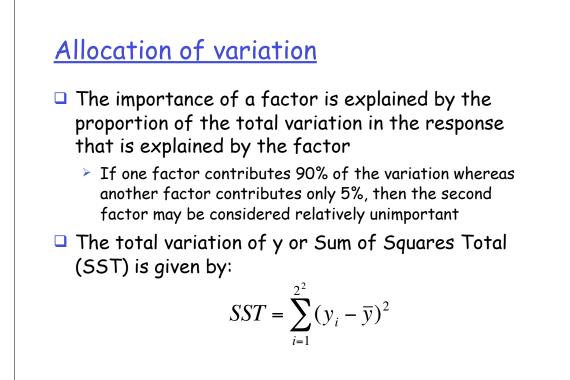
$$q_{0} = \frac{1}{4}(y_{1}+y_{2}+y_{3}+y_{4})$$

$$q_{A} = \frac{1}{4}(-y_{1}+y_{2}-y_{3}+y_{4})$$

$$q_{B} = \frac{1}{4}(-y_{1}-y_{2}+y_{3}+y_{4})$$

$$q_{AB} = \frac{1}{4}(y_{1}-y_{2}-y_{3}+y_{4})$$

I	А	В	AB	У
1	-1	-1	1	15
1	1	-1	-1	45
1	-1	1	-1	25
1	1	1	1	75
160	80	40	20	Total
40	20	10	5	Total/4





### <u>Allocation of variation (cont'd)</u>

For a 2<sup>2</sup> design, the variation can be divided into three parts:

$$SST = 2^2 q_A^2 + 2^2 q_B^2 + 2^2 q_{AB}^2$$

where the three terms on the RHS represent the portion of the variation that is explained by the effects of A, B, and the interaction AB respectively (see derivation in textbook). Thus

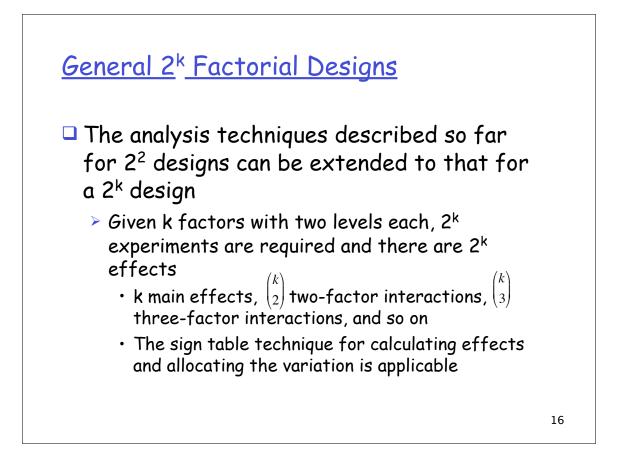
SST = SSA + SSB + SSAB

## Example

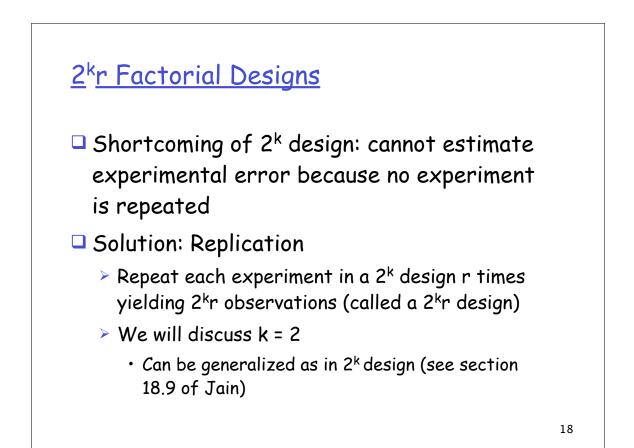
For the memory-cache example,

$$\overline{y} = \frac{1}{4}(15 + 55 + 25 + 75) = 40$$
  
Total variation 
$$= \sum_{i=1}^{4} (y_i - \overline{y})^2 = (25^2 + 15^2 + 15^2 + 35^2)$$
$$= 2100 = 4 \times 20^2 + 4 \times 10^2 + 4 \times 5^2$$

Thus, 76% (1600) of the total variation can be attributed to memory size, 19% (400) can be attributed to cache, and only 5% (100) can be attributed to the interaction between memory and cache.







## 2<sup>2</sup>r Factorial Design

Model for 2<sup>2</sup> design is extended to add an error term

 $y = q_0 + q_A x_A + q_B x_B + q_{AB} x_A x_B + e$ where the q's are effects as before and e is the experimental error

- Holding the factor level constant and repeating the experiment yields samples of the response y<sub>i</sub>
- Statistical analysis of the y<sub>i</sub>'s yields the fraction of variation due to experimental error, and confidence intervals for y

19

## **Computation of Effects**

The effects can be calculated using a sign table as before except that in the y column we put the sample mean of r measurements at the given factor level

I	Α	В	AB	У	Mean $\overline{y}$
1	-1	-1	1	(15,18,12)	15
1	1	-1	-1	(45,48,51)	48
1	-1	1	-1	(25,28,19)	24
1	1	1	1	(75,75,81)	77
164	86	38	20		Total
41	21.5	9.5	5		Total/4

# **Estimation of experimental errors** Once the effects have been computed, the model can be used to estimate the response for any given factor values as: $\hat{y}_i = q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$ The difference between the estimate and the measured value $y_{ij}$ in the jth replication of the ith experiment represents the experimental error $e_{ij} = y_{ij} - \hat{y}_i = y_{ij} - q_0 + q_A x_{Ai} + q_B x_{Bi} + q_{AB} x_{Ai} x_{Bi}$

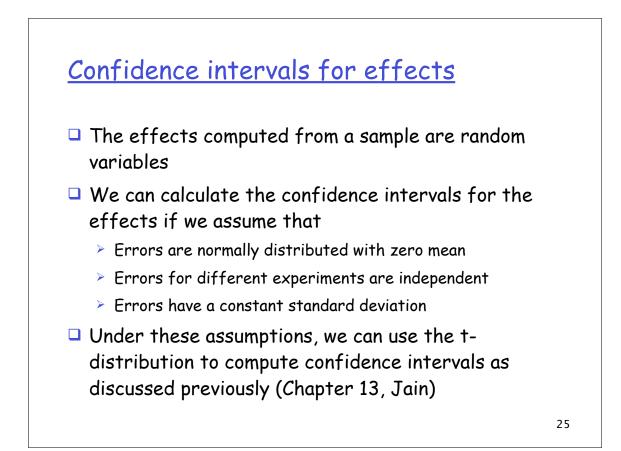
Ι	A	В	AB	(y <sub>i1</sub> ,y <sub>i2</sub> ,y <sub>i3</sub> )	$\hat{y}_i$	e <sub>i1</sub>	e <sub>i2</sub>	e <sub>i3</sub>
1	-1	-1	1	(15,18,12)	15	0	3	-3
1	1	-1	-1	(45,48,51)	48	-3	0	3
1	-1	1	-1	(25,28,19)	24	1	4	-5
1	1	1	1	(75,75,81)	77	-2	-2	4
41	21.5	9.5	5					

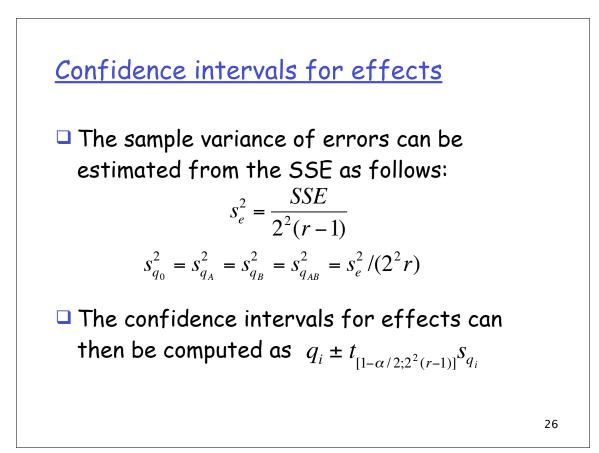
## <u>Allocation of variation</u>

Let  $\overline{y}_{...}$  represent the mean of responses from all replications of all experiments. Then  $SST = \sum_{i,j} (y_{ij} - \overline{y}_{..}) = 2^2 r q_A^2 + 2^2 r q_B^2 + 2^2 r q_{AB}^2 + \sum_{i,j} e_{ij}^2$ SST = SSA + SSB + SSAB + SSEWe can also show that SST = SSY - SSO, giving us an easier way to compute SSE: SSE = SSY - (SSO + SSA + SSB + SSAB)where  $SSY = \sum_{i,j} y_{ij}^2$  and  $SSO = 2^2 r q_0^2$ 

### <u>Example</u>

For our memory-cache example  $SSY = 15^2 + 18^2 + 12^2 + 45^2 + ... + 75^2 + 75^2 + 81^2 = 27,204$   $SSO = 12 \times 41^2 = 20,172$  SSA = 5547, SSB = 1083, SSAB = 300  $SSE = 27,204 - 2^2 \times 3(41^2 + 21.5^2 + 9.5^2 + 5^2) = 102$  SST = SSY - SSO = 27,204 - 20,172 = 7032Thus, factor A explains 5547/7032 or 78.88% of the variation, factor B explains 15.4%, interaction AB explains 4.27% of the variation. The remaining 1.45% is attributed to experimental errors



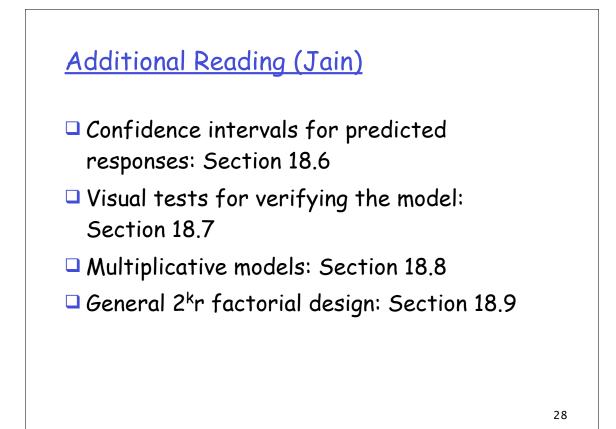


## Example

For the memory-cache example,

$$s_e = \sqrt{\frac{SSE}{2^2(r-1)}} = \sqrt{\frac{102}{8}} = \sqrt{12.75} = 3.57$$
$$s_{q_i} = s_e / \sqrt{(2^2 r)} = 3.57 / \sqrt{12} = 1.03$$

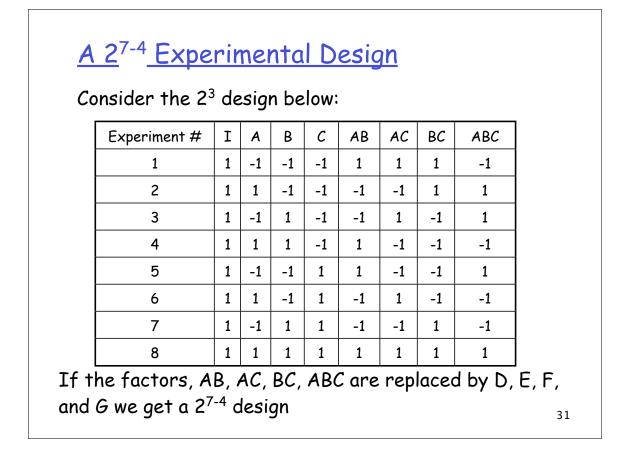
□ The t-value for 8 degrees of freedom and 90% confidence is 1.86. Thus, the confidence intervals for the effects are  $q_i \pm (1.86)(1.03) = q_i \pm 1.92$  that is (39.08,42.91), (19.58,23.41), (7.58,11.41), (3.08,6.91) for  $q_0$ ,  $q_A$ ,  $q_B$ , and  $q_{AB}$  respectively







# Fractional Factorial Designs If we have 7 factors, a 2<sup>7</sup> factorial design will require 128 experiments How much information can we obtain from fewer experiments, e.g. 2<sup>7-4</sup> = 8 experiments? A 2<sup>k-p</sup> design allows the analysis of k two-level factors with fewer experiments



# <u>A 2<sup>7-4</sup> design</u>

If the interactions AB, AC, AD,..., ABCD are negligible we can use the table

below										
DEIOW	Experiment #	I	A	В	С	D	E	F	G	у
	1	1	-1	-1	-1	1	1	1	-1	20
	2	1	1	-1	-1	-1	-1	1	1	35
	3	1	-1	1	-1	-1	1	-1	1	7
	4	1	1	1	-1	1	-1	-1	-1	42
	5	1	-1	-1	1	1	-1	-1	1	36
	6	1	1	-1	1	-1	1	-1	-1	50
	7	1	-1	1	1	-1	-1	1	-1	45
	8	1	1	1	1	1	1	1	1	82
	Total	317	101	35	109	43	1	47	3	
	Total/8	39.62 (	12.62	4.37	13.62	5.37	0.12	5.9	0.37	
	Percent variation		37.26	4.74	43.4	6.75	0	8.1	0.03	32
		1								<u> </u>

## Preparing the sign table for a 2<sup>k-p</sup> design

1. Choose k-p factors and prepare a complete sign table for a full factorial design with k-p factors.

There are  $2^{k-p}$  rows and columns in the table.

The first column is marked I and consists of all 1's.

The next k-p columns correspond to the k-p selected factors. The remaining columns correspond to the products of these factors.

 Of the 2<sup>k-p</sup>-k+p-1 remaining columns, select p columns corresponding to the p factors that were not chosen in step 1.

Note: there are several possibilities; the columns corresponding to negligible interactions should be chosen.

33

Experiment #	I	A	В	С	AB	AC	ВС	ABC
1	1	-1	-1	-1	1	1	1	-1
2	1	1	-1	-1	-1	-1	1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	1	-1	1	-1	-1	-1
5	1	-1	-1	1	1	-1	-1	1
6	1	1	-1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1	1

If the ABC interaction is negligible, we should replace ABC with D. If AB is negligible, we can replace AB with D.

## <u>A 2<sup>4-1</sup> design</u>

