Chair for Network Architectures and Services—Prof. Carle
Department of Computer Science
TU München

Experiment planning: Factorial design, factor analysis

Most slides/figures taken from:
Oliver Rose
Averill Law, David Kelton

DOE (Design of Experiments)

- Introduction and motivation
- Comparing two alternative systems
- Linear and nonlinear regression
- Analysis of Variance (ANOVA)
  - One-way ANOVA
  - Two-way ANOVA
- Factorial designs
  - 2^k factorial designs
  - Fractional factorial designs
- Important background information (within above topics):
  Hypothesis testing

Literature

- A. Law, D. Kelton: „Simulation Modeling & Analysis“, McGraw-Hill, 1991.
- G. Box, W. Hunter, J. Hunter: „Statistics for Experimenters“,
  Wiley, 1978.
- D. Montgomery: „Design and Analysis of Experiments“,
  Wiley, 1997.
- D. Goldberg: „Genetic Algorithms in Search, Optimization, and Machine Learning“, Addison-Wesley, 1989.

Motivation

- Statistics chapter:
  - Basic statistical concepts
  - Hypothesis testing
  - Analysis of a single simulation run
- But: Simulation not only used for single runs
  - We want to compare alternative designs!
- Approach for comparison
  - Explorative approach (“Fiddle around with parameters” / “Hit or Miss” strategy) = inefficient or even dangerous
  - → Methodic design of Experiments (DOE)

Why compare system alternatives?

- Goals:
  - Better understanding of system
  - Better control of system
  - Better performance of system

- Methods:
  - Try out in different simulated environments
    - Try out different workloads with different characteristics
    - Try out different network topologies
  - Try out with different system parameters

Terminology

- **factor**: input variable (e.g., TCP window size), condition, structural assumption (e.g., TCP congestion control algorithm)
- **level:** one factor value that is used in our experiments
- **response:** system parameter of interest that depends on given set of factors (e.g., achieved TCP throughput)
- **run**: evaluation of response for a given set of factor values
  - i.e., the analysed simulation result
  - There will (should!) be multiple runs

Remember:

- In simulation experiments, responses vary for runs of the same factor values due to random effects
- Therefore: several runs have to be performed!

Comparing two alternative systems

- Comparison of two systems:
  Is there a difference in value for a given response variable?
  - e.g., difference in achieved network throughput

- Test criterion:
  1. Calculate difference between the two response variables
  2. This difference is statistically significant if its confidence interval (CI) does not contain 0
     - e.g.: CI ($\text{throughput}_{\text{TCP Reno}} - \text{throughput}_{\text{TCP Vegas}}$) $\not\ni 0$
       → We can assume that the difference in throughput which the two congestion control algorithms TCP Reno and TCP Vegas achieve is statistically significant

Is this enough?

- Good: Very simple
- Bad: Quite restricted applicability
  - Only should be applied if the response has the same variance for the two levels – not often the case
    - Better: Modified or Welch two-sample t confidence intervals
  - Calculating the confidence interval for the response differences only can tell us if two levels of one factor make a difference
  - What if we want to analyse more than two levels for a given factor?
    - E.g., TCP Reno vs. TCP Vegas vs. TCP Cubic: 3 levels
  - What if we have more than one factor?
    - E.g., TCP congestion control algorithm, TCP window size, network delay, link bandwidth: 4 factors

Linear model and regression

- Have n samples $x_{1...n}$ and $y_{1...n}$ of two random variables x and y
- y is 'not really' a random variable: it's also dependent on x
- Linear model: $y = a \cdot x + b + e$
  - a: slope
  - b: intercept
  - e: error
- Idea: Chose a and b such that e is minimised
  - Calculate sum of squared errors:

$$SSE = \sum_{i=1}^{n} (y_i - b - ax_i)^2$$

  - Minimise SSE

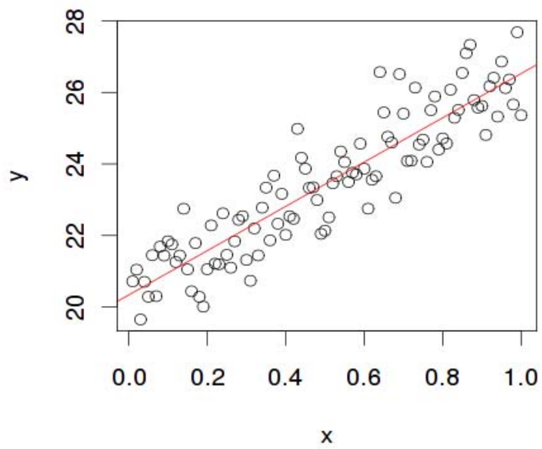
Calculating a and b

$$a = \frac{Cov(x, y)}{Var(x)}$$

$$b = mean(y) - a \cdot mean(x)$$

- N.B.: different, but equivalent formulae in literature (you can omit dividing by n—1 in var and cov)
- Usually built into statistical programs

- Graphical interpretation:
  Fit a straight line that goes through the points in the (x,y) scatterplot
  - b: intercept (Achsenabschnitt)
  - a: slope (Steigung)



How good is our regression?

- Correlation coefficient r:

$$r = \frac{\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)}{n-1}$$

- Coefficient of determination: r^2
  - i.e., simply square above result
  - Can be better compared than non-squared r, because it is proportional to the correlation, e.g.:
    $r^2 = 0.4$ provides double the correlation than $r^2 = 0.2$
  - Can be simply added up if multiple independent factors are combined
- Don’t confuse these two with the covariance!

Are we actually allowed to apply regression!?

- Warning:
  - The residuals e (as in y = a·x + b + e) must be normally distributed!
  - Exploit the central limit theorem: Calculate averages of multiple independent simulation runs with the same factor level
  - Check that it looks normal: QQ plots or some normality test
- N.B.: This check has, of course, nothing to do with the "quality" of the regression expressed as r^2
  - Normality check: Are we allowed to look for linearity?
  - r^2: How much linearity is there?

Regression and experiment planning

- In our nomenclature: y = response, x = factor level
- Regression can tell us how much the factor influences the response. Answers questions like:
  - Does it make sense to explore further factor levels in a given direction?
  - Does it make sense to check factor levels in between?
- Good:
  - We now can have multiple factor levels
- Bad:
  - We still have only one factor
  - It must be linearly proportional
  - The residuals must be normally distributed
    (but that constraint won't go away with ANOVA either)

Nonlinear regression 1/2

- Often, the relationship between x and y is not linear
- Solution: Try to find a suitable transformation
  - Let y be the simulation outcome (response)
  - Then apply the model $y^* = a \cdot x + b + e$
    where $y^* = f(y)$
  - Transformation function f can be, for example:
    - Logarithm
    - Exponentiation
    - Square root
    - Square
    - Some other polynomial (usually quadratic or cubic)
    - Logistic function (logistic regression)
    - Inverse (1/x)
    - …

Nonlinear regression 2/2

- Which transformation function is the right one?
  - Careful consideration of the system: You have to think!
  - Check if the y* are normally distributed – the y are probably not normally distributed in this case
- QQ plots can help
- Admittedly, a matter of experience
- Warning:
  - Overfitting, arbitrary curve fitting: “Just try around with some transformations and pick the one that matches best” – no, try to avoid that!
  - A correlation can be coincidence
  - Correlation does not imply causation
  - Example: Decreasing number of pirates leads to increasing global temperatures (Church of the Flying Spaghetti Monster)
  - Again: First *think* about the system, *then* postulate a meaningful transformation

Multiple [linear] regression

- We want to look at multiple factors
- For historic reasons, we relabel our 'old' values a and b from the regression formulae as $\beta_0 \ldots \beta_m$ and the error as ε
  - Linear model is now:
    $y = \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots + \beta_m \cdot x_m + \beta_0 + \varepsilon$
  - Warning about the indices:
    Now, x_1 means 'the first factor', not 'the first simulation run' (there may be many simulation runs for the same choice of the x_i)
- Will not go into detail here

ANOVA

- Short for ‘analysis of variances’
  - Historical term
  - Explained in next slides
- Be careful: “variance analysis” is a more general term!
  Often, that term describes a slightly different analysis:
  - Calculate variances of the responses for different levels of one (or several) factors
  - Analyse statistically if the variances are the same
  - Very similar to ANOVA, but slightly different!

ANOVA nomenclature

- Factor has a levels ('treatments' for historical reasons: ANOVA was developed in pharmaceutical research)
- Each level is replicated/observed *n* times
- Data:

level	*replication*		
	1	$\cdots$	*n*
1	y_{11}	$\cdots$	y_{1n}
$\vdots$	$\vdots$		$\vdots$
a	y_{a1}	$\cdots$	y_{an}

- Question we want to answer:
  - Is there an effect of factor levels on system responses?
  - If so: how much?

The model for ANOVA

- Model for system responses with one factor

$$y_{ij} = m + t_i + e_{ij}$$

average (!) response

treatment effect

random error (‘noise’, ‘residuals’)

- i: factor level (‘treatment’)
- j: simulation run
- Please note: We’re dealing with only one factor so far

One-way ANOVA (1)

- Similar to linear regression:
- One factor, multiple levels
- $y_{ij} = \mu + \alpha_i + \varepsilon_i$
- μ: population mean (of the total population, i.e., across all different factor levels – in other words, across all simulation runs, regardless of their parameters!), also called grand average
- α_i: the influence of the different factor levels (how much do they contribute to a diversion from the mean?)
- ε_i: errors, also called 'residuals' or 'noise'

One-way ANOVA (2)

Important things to note about the model:

- Factor levels α_i
  - We do not require them to have a linear relationship on the response y
  - They even can be categorical data, e.g.: {male, female} or {child, student, employed, unemployed, retired, other}
- Residuals ε_i
  - Any deviation from the model that cannot be explained
  - Usually, the index is dropped for the errors, as ε is an independent random variable that must not (!) depend on the factor level
  - If that is not the case, we do not have a truly random but a systematic error. That's bad – it violates our assumptions!

One-way ANOVA (3)

- We suspect that the α_i are different and influence the response variable y
- Formulate this as a statistical test

Digression:

Statistical tests, revisited

Statistical tests

- So far, we've seen the χ^2 distribution fitting test and the Kolmogorov-Smirnov test (KS)
- Both test if a given set of measurements is consistent with a theoretical distribution
  - Note the wording: „Consistent with“, but not „comes from“
- There are many, many other statistical tests for many, many other applications

Statistical tests = hypothesis tests

- We would like to „prove“ some statement, based on statistical calculations
  Examples:
  - Measurements x_i are consistent with a normal distribution
  - The mean of the measurements xi is greater than 5
- Call this statement our 'work hypothesis' or 'alternative hypothesis' (Arbeitshypothese) H_A
- Formulate the contrary: null hypothesis H_0
- H_A and H_0 need to be:
  - Exclusive: Either H_A is true or H_0 is true
  - Exhaustive: All possible results will satisfy one of the two

Test statistic

- Hope to find statistical evidence that H_0 is highly improbable
- Mathematically:
  - Input data = x_i (...rather arbitrary label)
  - Calculate a so-called test statistic: $TS(x_i)$
  - Usually: If test statistic is above some threshold, then refuse H_0
  - Test statistic depends on specific test
  - Threshold depends on specific test and on desired accuracy

Test accuracy: Error types

- As mentioned before: No test can give a 100% guarantee – we're talking about statistics here, and statistics always deals with the unknown
- Differentiate between two types of errors:

	Test rejects H_0	Test accepts H_0
In reality, H_0 is false	Correct decision	Type II error, β error, false negative
In reality, H_0 is true	Type I error, α error, false positive	Correct decision (albeit not the one that we wanted in most cases…)

Error types explained by example (1/2)

- Suppose you have developed a medical drug. Development has cost an enormous amount of money. Now you want to test if the drug is harmful to your patients
- Type I error (α error)
  - Probability that people get harmed
  - Can cost lives: Invest a lot of effort to avoid it.
- Type II error (β error)
  - Probability that you reject a drug that is actually perfectly safe
  - Can waste money: Unpleasant, but more acceptable.

Error types explained by example (2/2)

- Suppose you have developed a new network protocol. By applying a statistical test to the output of some network simulations, you hope to show that the protocol increases network performance (=H_A).
- Type I error (α error)
  - Probability that you claim that the protocol is great, whereas it is actually rubbish
  - If you don‘t specify your α error, or if it is too large (i.e., your confidence level is too low), then nobody will believe your results!
- Type II error (β error)
  - Probability that you wrongly assume that your great protocol does not help anything
  - Presumably interesting to you, but the reader of your paper does not care about the risk that you might have failed detecting the performance increase: Obviously, you did not fail, since otherwise the paper would not have been written…

Balancing error types

- Problem:
  - Reducing one error increases the other and vice versa. Damn.
  - Only solution to reduce both: Increase the sample size. Usually a superlinear factor (e.g., to reduce one error by 1/2 while keeping the other constant, we must increase sample size by 4)
- In the majority of the cases, keeping the α error low is more important
  - α = 5% has been accepted for years (although there has been some criticism), 1% is better, 0.1% is extremely good
  - β = 10% or 20% is usually acceptable; but usually, it's not calculated
  - Don't choose α too small if there are only few samples: Small sample size and small α both will increase β to unacceptable values – then you would almost always accept the null hypothesis and thus (wrongly) reject your work hypothesis

Power of a test

- 'Power' of a test := $(1 - \beta)$
  - Obviously the higher, the better
- Can be used to compare tests:
  - Fix an α and a number of measurements
  - The better test will feature a higher power for this input
- Rules of thumb:
  - Parametric tests (make assumptions about input distribution) are stronger than nonparametric tests (work with any distribution)
  - One-sided tests are stronger than two-sided tests (later slide).
  - The more general the test, the weaker it is.

Error types: summary

- Usually, Type-1 errors (α errors) are the more serious ones
- In order to minimise one type of error (e.g., Type 1 error), you only have the choice between…:
  - Increasing the Type 2 error
  - Increasing the sample size
  - Picking a different statistical test that has better error properties

An „alternative“: significance tests

P-value (R. A. Fisher): How likely is the result to happen?

- Test statistic is a dependent random variable that follows a specific distribution (test distribution, e.g., Student's t distribution or χ^2 distribution) if the null hypothesis holds
- Using the theoretical distribution, calculate the probability that our measurements attain our given values or even more extreme values if the null hypothesis holds:
  - This is defined as the p value
  - Note that the p value itself is uniformly distributed in [0...1] if the null hypothesis holds, and it is near 0 if it does not hold.
- Refuse H_0 if this seems unlikely: i.e., refuse if $p \leq \alpha$
- In other words: Our threshold for the test statistic is the point where its distribution „has no meat“, i.e., the p value gets too low

We have two types of tests?

- In theory, distinguish:
  - Hypothesis test that we just explained:
    Fix an α, calculate the test statistic and accept or reject the null hypothesis
  - Fisher's probability test:
    For the given data, calculate the p value for the null hypothesis, and decide how likely the null hypothesis is
- In practice, combine both!
  - p value is more expressive
  - Fixed α is more commonly known/accepted; often allows better comparisons to other studies

How to combine both types of a test?

- With modern statistical programs, this is possible – in most cases, it‘s even done automatically!
- Good practice:
  - Tell the reader your p value (especially if null hypothesis sounds quite likely!)
  - Traditionally, the p value is judged with star symbols within braces:
    - [***] means: $P \leq 0.1\%$
    - [**] means: $0.1\% < P \leq 1\%$
    - [*] means: $1\% < P \leq 5\%$
- If possible, calculate the p value and derive statements about *α*
  - e.g.: „The null hypothesis could be refused at a confidence level of α=0.5, but not at a confidence level of α=0.1“

One-sided tests, two-sided tests

- One-sided test:
  $H_A: \mu < \mu_0$ (or $\mu > \mu_0$)
  - Example: „With the new routing protocol, network latency is significantly reduced from the old value“
- Two-sided test:
  $H_A: \mu \neq \mu_0$
  - Example: „With the new routing protocol, network throughput has significantly changed from the old value (either better or worse)“

- Which one to choose?
  - One-sided tests are stronger than two-sided tests
  - Two-sided tests are more expressive

Back to one-way ANOVA! (4)

- Recall our model: $y_{ij} = \mu + \alpha_i + \varepsilon_i$
- We suspect that the α_i are different and influence the response variable y
- Formulate this as a statistical test:
  - Hypothesis: At least one of the α_i influences y
  - Null hypothesis: $\alpha_1 = \alpha_2 = ... = 0$
  - Equivalent formulation of null hypothesis:
    The means of the factor levels are equal

One-way ANOVA (5)

- Analysis of variance: Analyse total sum of squares
- Introduce these variables (SS = sum of squares):
  - SS_{Total}
    - The total variation across all samples
    - I.e.: the total sum of squared deviations from the general mean μ
    - How much variability is in the general population?
  - $SS_{Between}$
    - The variation between the different sample groups (i.e., one group for each different factor level)
    - How much variability can be attributed to the different factor levels?
  - SS_{Within}
    - The variation between the samples of one factor group (i.e., all samples that hold for the same factor)
    - As we can see, we need to do multiple simulation runs for one factor level
    - How much variability can be attributed to the errors (‚noise‘)?

One-way ANOVA (6)

- Important observation: $SS_{Total} = SS_{Between} + SS_{Within}$
- Coarse idea:
  - If $SS_{Between}$ (the treatment variability) is much larger than SS_{Within} (the error variability), then the overall variability is likely to be caused by the factor
  - Otherwise, the overall variability is likely to be caused by ‚random‘ noise
    - Take care: The errors also can be unexplained effects
- More precisely: If H_0 holds, then $SS_{Between}$ and SS_{Within} have the same value
- Check this by applying the F test

The F test

- Developed by R. A. Fisher
- Input: two samples from two different populations
  - Populations have to be normally distributed (!)
- F test tells if the populations have a large difference in variance
- Test statistic: the F value

$$F = \frac{Var(X_1)}{Var(X_2)}$$

- If the null hypothesis holds, then the F value is F distributed
  - F distribution: a test distribution
  - As usual: degrees of freedom = #samples – 1

One-way ANOVA (8)

- Further mathematical details…?
  - Usually, the F test is built into statistical software
  - Usually, ANOVA is built into statistical software
  - We want to apply statistics, not learn any proofs of theorems → For more details, refer to literature

ANOVA: Caveats

Prerequisites similar to linear regression:

- The *measurements* have to be normally distributed
  - Easy if the response can be expected to be normally distributed (but that's generally not the case)
  - Easy if means are sampled from several (i.e., enough!) simulation runs: central limit theorem
- The *residuals* have to be normally distributed
  - Residuals: $e_{ij} = y_{ij} - \bar{y}_i$ (i.e., the deviation from the group mean)
  - Warning: You must ensure that this is really the case!
  - If not, the result is meaningless!
- The variances of the α_i need to be equal
  - F test
- How to check for normality?
  - QQ-plots
  - or some statistical test for normality

Two-way ANOVA (1)

- Two factor response analysis
  Factor A and B at levels *a* and *b*, *n* replications
- Change in quality of the results compared to one-way ANOVA?
  Yes!
  Both factor effects *and* effects from interacting factors
  - **main effect** of each factor
  - **interaction** of the two factors!
- **System model**: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$

grand average (μ), main effect A (α_i), main effect B (β_j), inter-action (γ_{ij}), error (ε_{ijk})

Two-way ANOVA (2)

- Data

		Factor A		
		1	$\cdots$	b
Factor B	1	$y_{111} \cdots y_{11n}$	$\cdots$	$y_{1b1} \cdots y_{1bn}$
	$\vdots$	$\vdots$		$\vdots$
	a	$y_{a11} \cdots y_{a1n}$	$\cdots$	$y_{ab1} \cdots y_{abn}$

Two-way ANOVA (3)

- Three null hypotheses:
  - $\alpha_i = 0$
  - $\beta_j = 0$
  - $\gamma_{ij} = 0$

- Sums and averages similar to one-way ANOVA:
  - $SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Within}$
- Usually built into statistical software packages

Two-way ANOVA (5)

- Two-way ANOVA table

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

Two-way ANOVA (6)

- Interpretation of the results:
  Check the p-values corresponding to the individual F_0 tests;
  if they are small, there are significant effects.

- Note: statistical significance does not tell anything about practical relevance! Decide yourself!

- Check model adequacy by analysis of residuals:
  - They should be consistent with a normal distribution
  - They should be free of structure (e.g., check that a higher response value does not usually imply higher error values)

Summary: ANOVA

- Generalisation: n-way ANOVA
- Usually performed using a statistical program
- Usually only two levels per factor.
  Examples:
  - Small window size, large window size
  - TCP Reno, TCP Cubic
- Tests if one or several factors have or have no influence on some response variable
  - E.g.: Does TCP window size affect TCP throughput?
- Can tell how much influence the individual factors have
- Can tell how much influence the interactions of the factors have
  - E.g.: Window size and congestion control algorithm taken together have significant influence

ANOVA and experiment planning

- Usually many factors
  - Example: TCP window size, TCP congestion control algorithm, network bandwidth, network delay, packet loss rate
- Which factor combinations should we try out? – ANOVA can give answers to these questions:
  - Which factors are interesting factors (i.e., have much influence), so we should try out more levels for them?
  - Which factors have interesting interactions, so we should try out more factor level combinations for them?
  - Which factors, which interactions can be left out?
- Structuring the experiments like this is called factorial design
  - Of course, not limited to simulation experiments

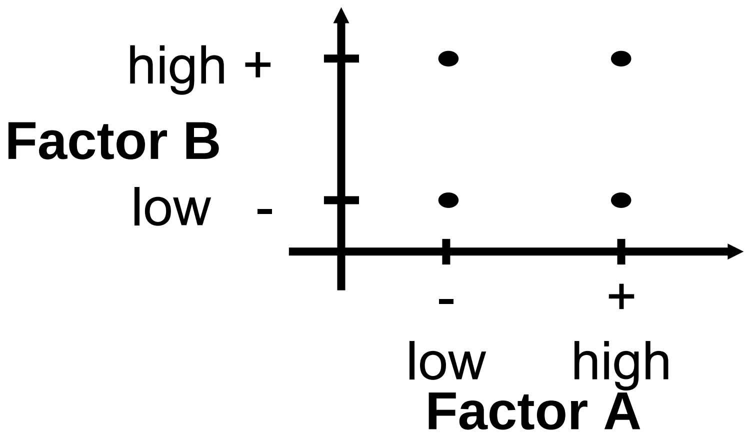
2^k factorial designs (1)

- Problem with general factorial designs:
  explosion of number of runs for multi-factor multi-level designs
- Solution:
  Two levels are often enough for detecting general trends and to screen out important factors
- k factors, each one with 2 levels: 2^k design points
- Underlying assumption: effects depend linearly on factors

2^k factorial designs (2)

- Example: 2 factors, i.e., a 2^2 design

- 4 design points:



- Design matrix:

Run	Factor A	Factor B	Response
1	-	-	r_1
2	+	-	r_2
3	-	+	r_3
4	+	+	r_4

2^k factorial designs (3)

- Construction of the "+/-" area of the design matrix:
  - Each row is the binary coding of the run number minus 1
  - with the least significant bit on the left side
  - where '-' represents 0 and '+' represents 1

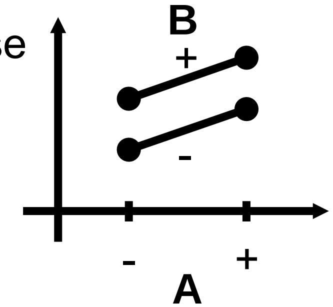
2^k factorial designs (4)

- Computation of the effects:
  - Main effect of factor A: how does the response change if A is changed while B is left constant?
    - $effect_A = ½ ((r_2 – r_1) + (r_4 –r_3))$
  - Main effect of factor B: how does the response change if B is changed while A is left constant?
    - $effect_B = ½ ((r_3 – r_1) + (r_4 –r_2))$
  - Main effect equations for other designs: Similar (Use factor column as signs for responses and sum up, then divide sum by $2^{k—1}$)
- Usually, the ANOVA module of a statistical program will help

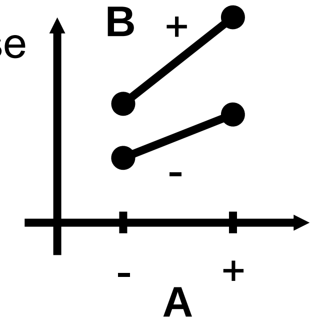
2^k factorial designs (5)

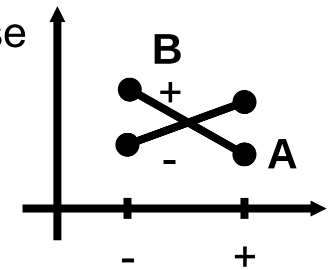
- Interaction of factors A and B: Is there a difference in the changes of the response if A is changed while B is kept either on level ‘+’ or ‘–’?

- no interaction, i.e.
  no (or small) difference in changes:

response


- interaction, difference in changes:

response


response


2^k factorial designs (6)

- How to find the interaction equations for other designs?
  - In theory: Multiply columns of factors of interest element by element and use the resulting column as signs for responses and sum up.
    Then divide sum by 2^{k-1} .
  - In practice: Built into statistical software…
  - In addition to main effects and interactions, the average response is usually computed

2^k factorial designs (7)

- Example: main effects and interactions of the 2^3 design

average	A	B	C	AB	AC	BC	ABC
+	-	-	-	+	+	+	-
+	+	-	-	-	-	+	+
+	-	+	-	-	+	-	+
+	+	+	-	+	-	-	-
+	-	-	+	+	-	-	+
+	+	-	+	-	+	-	-
+	-	+	+	-	-	+	-
+	+	+	+	+	+	+	+

Fractional factorial designs (1)

- Full factorial design can be costly for larger number of factors
- In most cases, we are only interested in main effects and two-way interactions
- Example: Full 2^7 design requires 128 times replications runs! (And each needs to be run multiple times.) Effects obtained:

Avg.	Main effects	2-way	3-way	4-way	5-way	6-way	7-way
1	7	21	35	35	21	7	1

- More than 75% of the effects are 3-way interactions and higher
- Obtain the main effects and two-way interactions with less runs? Yes, by using fractional factorial designs!

Fractional factorial designs (2)

- Example: Full 2^3 design requires 8 runs

Average	Main effects	2-way inter.	3-way inter.
1	3	3	1

- Only interested in main effects – let's do only 4 runs and ignore the interactions
- 2^{3-1} design requires 4 runs, but:
  how to accommodate 3 factors with a 2 factor design?

Factor A	Factor B	Factor C
-	-	+
+	-	-
-	+	-
+	+	+

C = AB

Fractional factorial designs (3)

- Did we get information for free?
  Half the runs to obtain the same result?

- NO! There are confounded (or aliased) effects!

- Main effects and two-way interactions are confounded, i.e.:
  - Influence of C indistinguishable from influence of interaction AB
  - Influence of B indistinguishable from influence of interaction AC
  - Influence of A indistinguishable from influence of interaction BC
- What does this mean?
  - Main effect of factor C is only useful if interaction of A and B is small, i.e., 2^{3-1} design is a bad choice if two-way interactions are significant.
- N.B. There also is a graphical explanation for this (→later slides)

Fractional factorial designs (4)

- Resolution of a fractional design (denoted in Roman numbers)
  - III: only main effects are not confounded
  - IV: main effects/two-way interactions not confounded
  - V: main effects/two-way interactions and two-way/two-way interactions not confounded
  - Higher order effects are confounded!
- Practical advice:
  - Use resolution III designs only in complete desperation!
  - Interactions of more than 3 factors are rarely relevant
- Notation: $2^{k-p}_{resolution}$, e.g., 2^{4-1}_{IV}
  - Examples:
    - III: $2^{3-1}_{III}, 2^{5-2}_{III}, 2^{6-3}_{III}, \ldots$
    - IV: $2^{4-1}_{IV}, 2^{6-2}_{IV}, 2^{7-3}_{IV}, \ldots$
    - V: $2^{5-1}_{V}, 2^{8-2}_{V}, 2^{10-3}_{V}, \ldots$

Fractional factorial designs (5)

- Construction of the design matrix
  - Basis is always full factorial design for k—p factors, e.g., a 2^3 matrix for a fractional 2^{5-2}_{III} design
  - Missing columns are computed from existing ones by rules from DOE text books. These rules guarantee fractional designs of maximum resolution.
    Example: for 2^{5-2}_{III} design, columns D and E missing
    rules: D = AB or -AB, E = AC or -AC
    (AB: multiply signs of columns A and B)
  - Resolution and construction of design matrix for fractional designs from DOE text books
  - Often already built in run controllers of simulation tools or statistical programs

Fractional factorial design, graphically explained

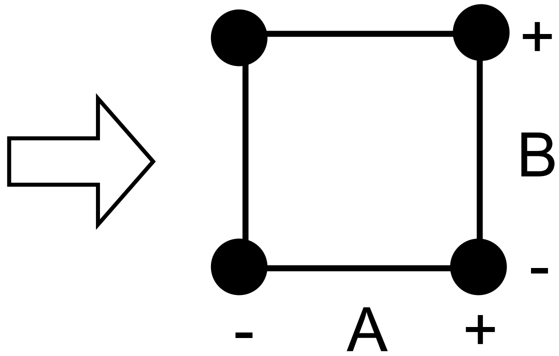
Motivation for the graphical approach:

- Successful application of graphical methods in other areas of statistics, in particular, for data analysis and data mining
- Application of the creative potential of the right brain half
- Intuitive understanding of “good” characteristics of DOE
- Approach was used for the development of DOE methods, but no longer in the application phase
- Straightforward approach, often even without use of computers

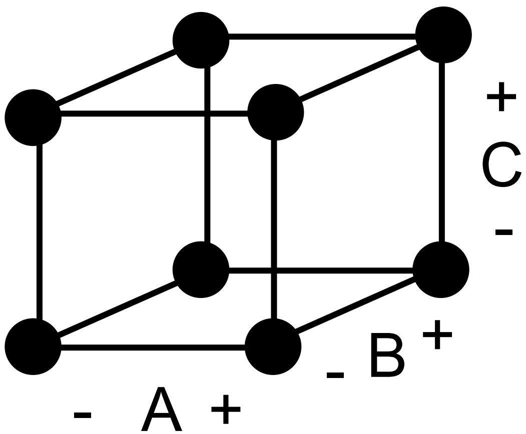
From the design matrix to the design graph

- Approach: Transform the design matrix into an appropriate and equivalent graphical representation
- 2 factors:

Run	Factor A	Factor B
1	-	-
2	+	-
3	-	+
4	+	+



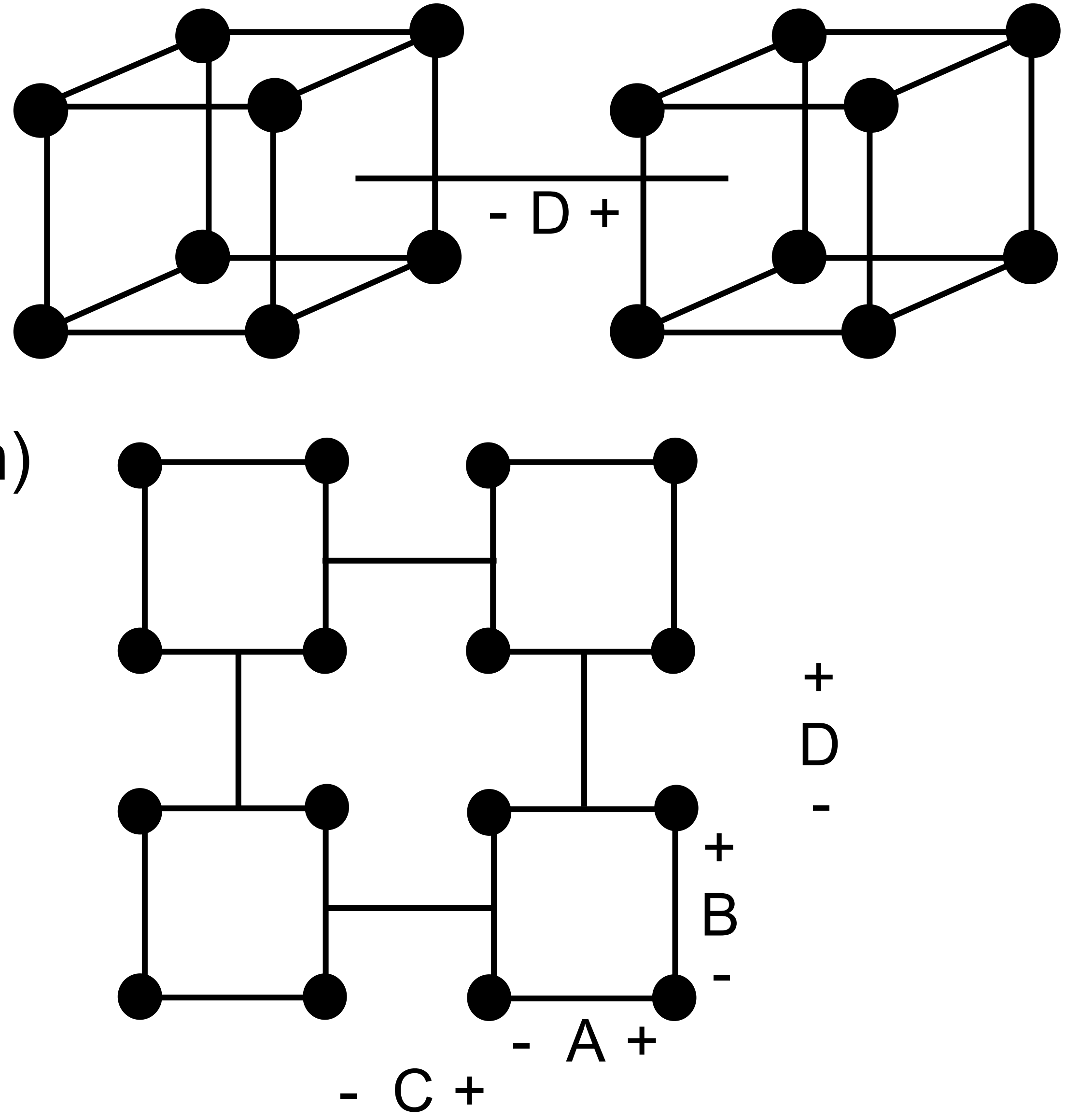
- 3 factors:



Graphical representation of designs

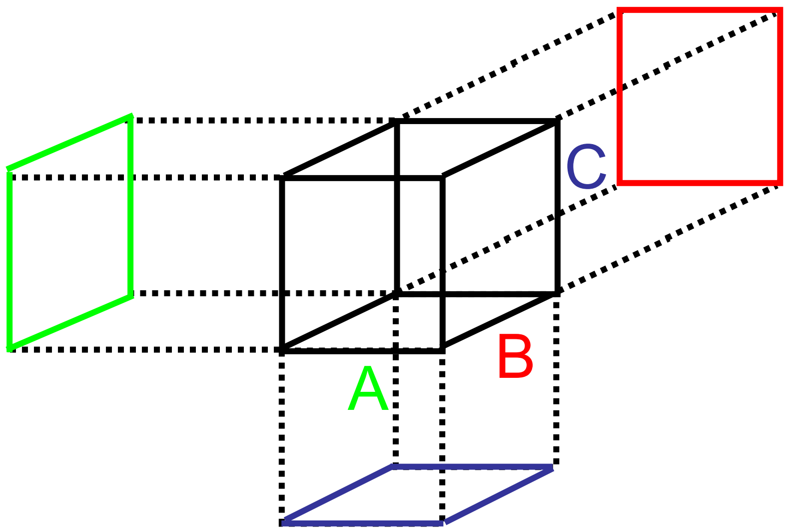
- 4 factors:
  (hypercube)

  (Problem:
  Humans don't
  have 4-D vision)



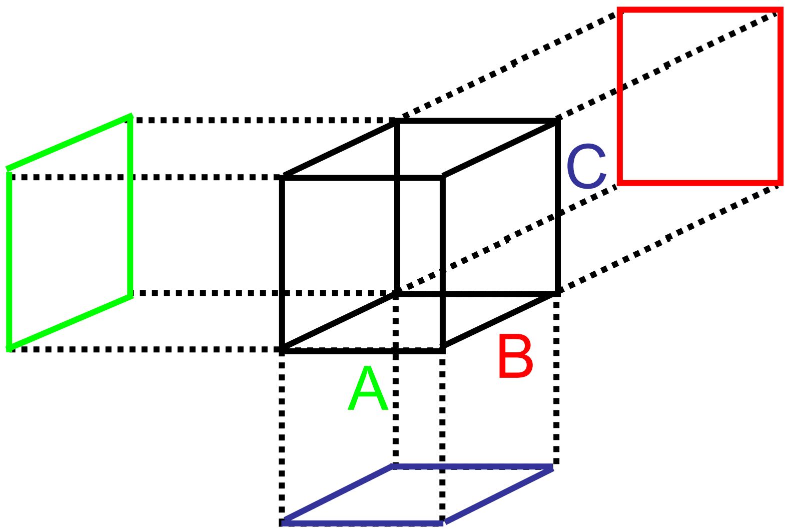
We're talking about an optimisation problem

- Goal: Minimize information loss of a fractional factorial design reduced by *p* factors
- Graphically: Projections of the design graph where p dimensions disappear (graph collapses)
- Example: 1 factor of a 2^3 design disappears

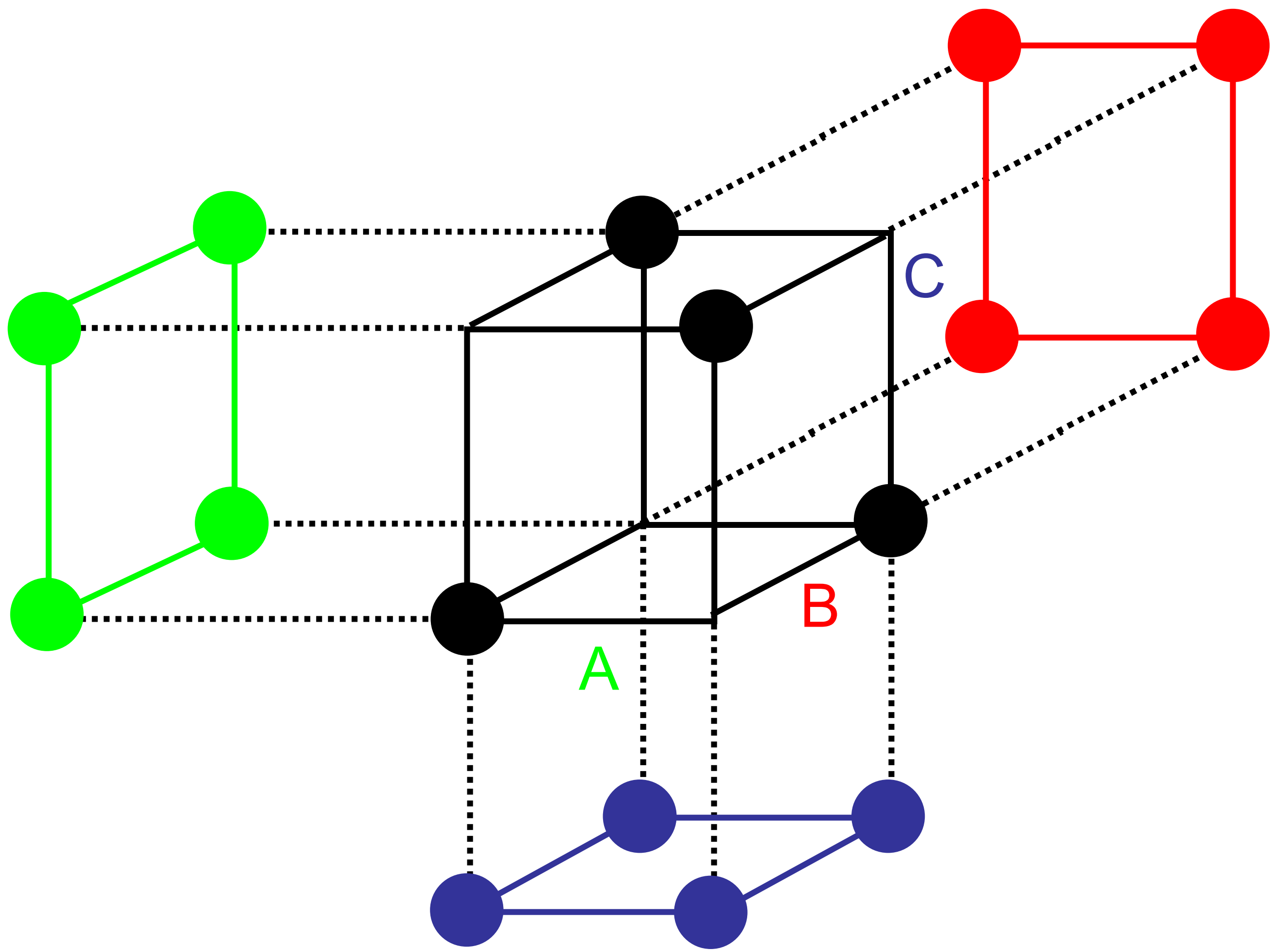


Optimum location of the design points

- Important graphical optimisation criteria for maximizing the information content in fractional designs:
  - Each projection must be a complete design graph
  - No multiple design points at the corners of the graph
- Example: Reduction of a 2^3 design to a 2^{3-1} design, i.e., from 8 to 4 design points $\rightarrow$ optimum location?

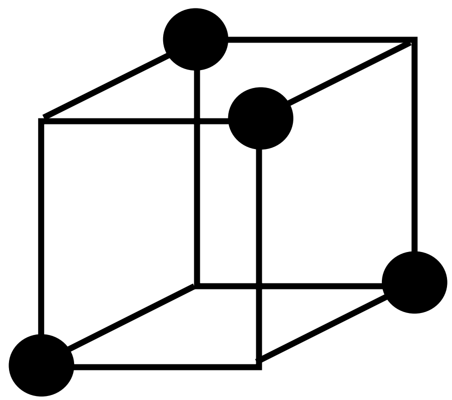


Optimum location of the design points



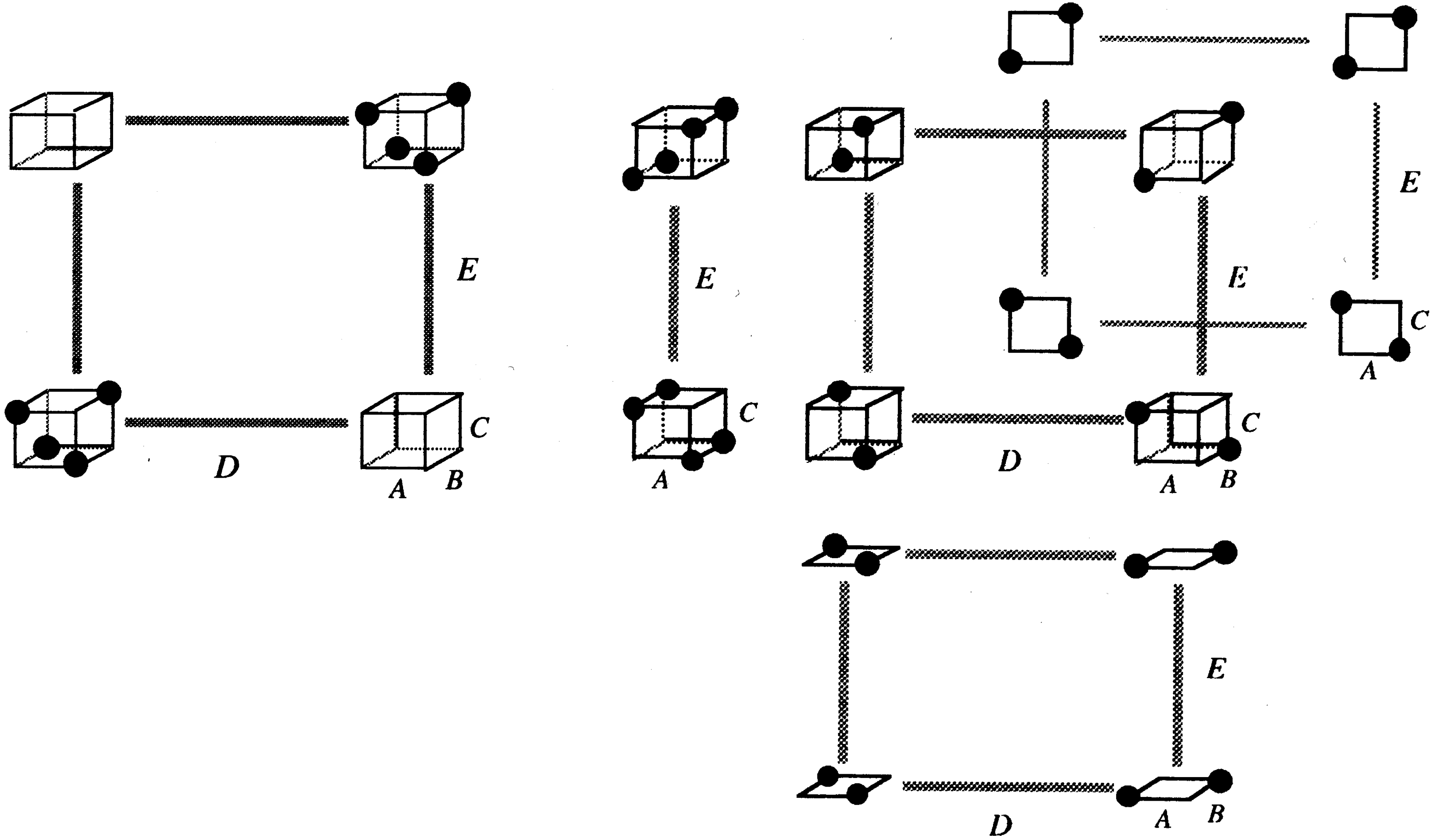
Criteria for larger 2^{k-p} designs

- Optimum 2^{3-1} design as as basic building block ("DOE lego")

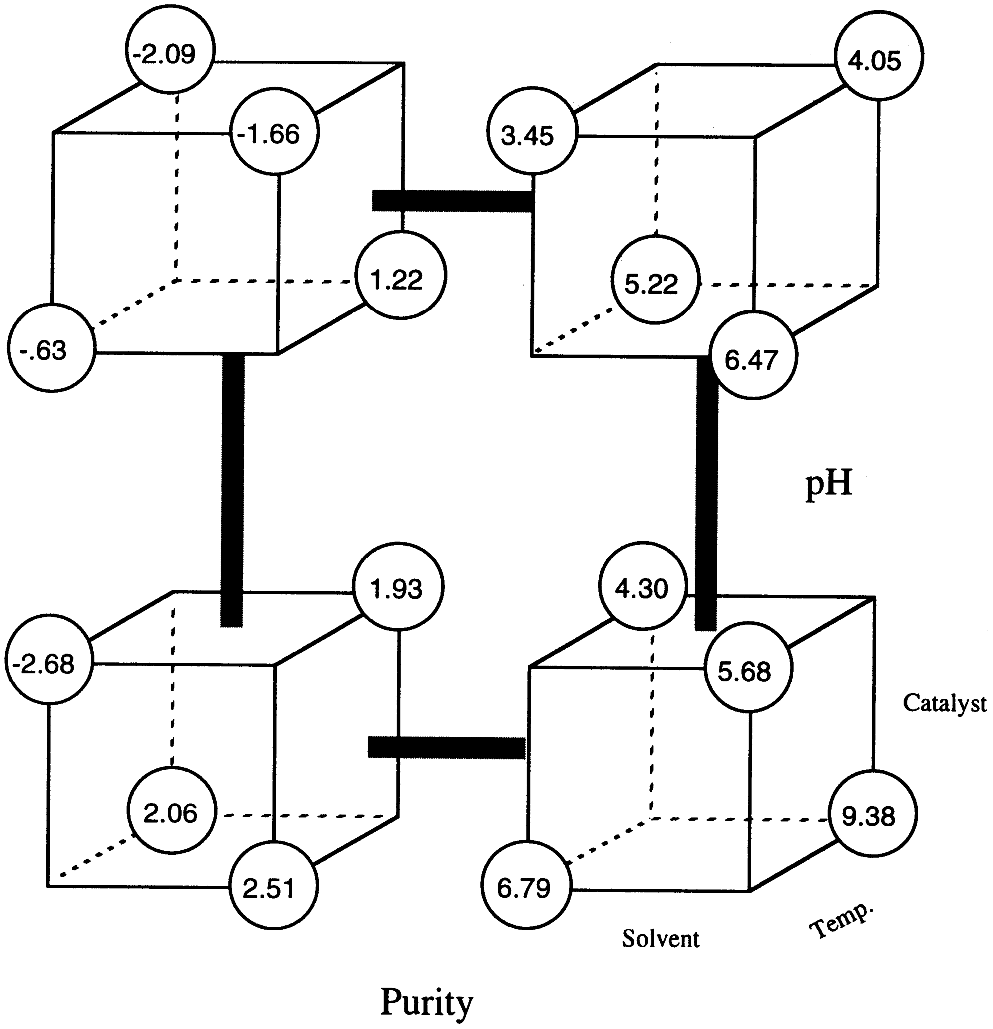


- Projections as complete as possible, but with single design points at the corners
- Maximising the minimum distance of the design points (even distribution of points)

Alternatives for a 2^{5-2} design



Design graphs for presenting results



Design graphs for presenting results