



THE UNIVERSITY OF AUCKLAND

DOCTORAL THESIS

Computer generation of designs for
two-phase experiments with
applications to multiplex experiments
in proteomics

Kevin Chih-Tao Chang

*A thesis submitted in fulfilment of the requirements for
the degree of Doctor of Philosophy in School of Biological Sciences,
The University of Auckland, 2017.*

*This thesis is for examination purposes only and is confidential to the
examination process. November 1, 2017*

Abstract

Two-phase experiments arise when the treatment effects of interest in an experiment (Phase 1) cannot be measured directly and the material from the experiment requires further processing (Phase 2 experiment) in order for these effects to be evaluated. Two designs are needed for these experiments, one for each phase. Since the allocation of experimental units from the Phase 1 experiment to blocks in the Phase 2 experiment results in the block effects from the two phases interacting with one another, the Phase 2 design must be carried out with consideration to the Phase 1 design. Except for very small experiments, how this should be done in an optimal way is non-trivial.

Theoretical analysis of variance (ANOVA) tables, showing the decomposition of the total variability in the data space into its constituent components of known sources of variation, and their corresponding degrees of freedom, are shown to play an important role in assessing the properties of competing designs for two-phase experiments. However, generating these ANOVA tables is a laborious manual task, even for relatively small two-phase experiments. To automate this process, an R package called **infoDecompuTE** was developed and is available on the Comprehensive R Archive Network. All of the ANOVA tables presented throughout this thesis were generated by **infoDecompuTE**.

While the theoretical ANOVA tables are an important tool in assessing the properties of competing designs, the manual generation of optimal designs for two-phase experiments is non-trivial, particularly for non-orthogonal designs. Thus, a fundamental component of this thesis is the development methodologies for the computer generation of designs for two-phase experiments. A combination of theory, to derive multi-criterion objective functions, and computing, in which a modified simulated annealing algorithm is developed, are used to identify A-optimal designs for the Phase 2 when the Phase 1 experiment is arranged in either a completely randomised, a randomised complete block or a balanced incomplete block design. Optimal designs for range of design parameters for both the Phase 1 and Phase 2 experiments are catalogued in the appendices of this thesis, as are summary tables of their properties.

Data simulations were carried out to explore how well the variances of treatment effects are

estimated among competing Phase 2 designs. For this, the effective degrees of freedom (EDF) for estimating the error variance, using Satterthwaite's approximation, were calculated using two methods of variance component estimation, namely taking linear combinations of expected mean squares and restricted maximum likelihood. The two methods of variance component estimation were found to have little effect on the EDF. However, the simulation studies were shown to be informative with respect to preferred choice of two competing designs when the relative magnitudes of the variance components are known.

While the motivating examples in this thesis come from proteomics experiments, which have as their goal link the identities and abundances of proteins in a biological sample to different experimental conditions (treatments), the methods presented in this thesis apply more generally across a wide range of biological, and other, experiments.

Acknowledgments

First of all, I would like to express my sincere gratitude to my primary supervisor Dr. Kathy Ruggiero for her continual support and patience throughout my PhD study. As well as providing an enormous amount of ideas, time, support and guidance both on my project and the thesis.

Besides my primary supervisor, I would like to acknowledge to my co-supervisors Prof. Chris Triggs and Dr. Richard Jarrett, who have been influential in my statistical thinking, and provided crucial support, expertise and important considerations throughout my PhD study.

I would also like to thank all the friends and colleagues in Bioinformatic Institute, Statistical Consulting Centre and Centre of Methods and Policy Application in the Social Sciences for all the helps and encouragements.

Finally, I would like to thank my family for all the support throughout my PhD study to make it all possible.

Contents

Abstract	iii
Acknowledgement	iv
Table of Contents	ix
List of Tables	xxi
List of Figures	xxiv
1 Introduction	1
1.1 The introduction of the two-phase experiment	2
1.2 Analysis of Variance Tables Based on Experimental Structure	5
1.3 Sweeping operations for two-phase experiments	8
1.4 Work by Brien and Bailey	10
1.5 Systematic approach to design two-phase experiments	11
1.6 Experimental design for multi-plex proteomic experiments	11
1.7 Quantitative Proteomics using MudPIT coupled with iTRAQ TM	13
1.7.1 Proteins and the Proteome	13
1.7.2 Multi-dimensional Protein Identification Technology	14
1.7.3 iTRAQ TM for Protein Quantitation	15
1.7.4 MudPIT Coupled with iTRAQ TM : Laboratory Workflow	17
1.7.5 Data Generation from MudPIT-iTRAQ TM	20
1.8 Overview of thesis	22
2 infoDecompuTE: An R package for constructing theoretical ANOVA tables for two-phase experiments	23
2.1 Introduction	23
2.2 Information decomposition of designed experiments	26

2.2.1	The linear mixed-effects model	26
2.2.2	Null ANOVA using projection matrices	27
2.2.3	Computing the treatment SS	29
2.2.4	Computing the residual SS	33
2.3	Information decomposition for two-phase experiments	33
2.3.1	The linear mixed-effects model	34
2.3.2	Null ANOVA of Phase 2 block structure	35
2.3.3	Null ANOVA of Phase 1 block structure	35
2.3.4	Computing the treatment SS	37
2.4	Application to quantitative proteomics experiments	38
2.5	Example of two-phase proteomics experiment	39
2.5.1	Phase 1 experiment	39
2.5.2	Phase 2 experiment	40
2.6	InfoDecompuTE	43
2.6.1	Installation instructions	43
2.6.2	Functions	43
2.6.3	Phase 1 experiment	44
2.6.4	Phase 2 experiment	45
2.6.5	Crossed or nested	47
2.6.6	Artificial strata or pseudo-factors	48
2.6.7	Manually defined contrasts	49
2.6.8	Mean squares computation	50
2.6.9	Latex output	51
2.6.10	Shiny Application	52
2.7	Example	52
2.8	Conclusion	55
3	Optimal designs for two-phase experiments when the Phase 1 experiment is arranged in a completely randomised design	57
3.1	Introduction	57
3.2	Design parameters and model	59
3.2.1	Linear models	59
3.2.2	Linear models in matrix notation	60
3.3	Information matrix for the objective function	61

3.4	Optimality criteria	63
3.4.1	A-optimality	63
3.4.2	MS-optimality	65
3.5	An illustrative example	66
3.6	An objective function for identifying optimal designs for the Phase 2 experiment	68
3.6.1	Maximising the animal information	68
3.6.2	Maximising the treatment information	71
3.6.3	Maximising intact Treatment DF	74
3.7	Construction of the initial design	79
3.8	Simulated annealing	81
3.8.1	Acceptance probability	82
3.8.2	Initial and final temperatures	82
3.9	Nested simulated annealing	83
3.10	Modified simulated annealing	84
3.10.1	Swapping method for two or more technical replicates	84
3.10.2	Three-stage swapping method	84
3.11	An illustrative example using the four-plex iTRAQ TM system	85
3.12	An illustrative example using the eight-plex iTRAQ TM system	87
3.13	Optimal designs for experiments involving two to eight treatments and two technical replicates	91
3.13.1	Phase 1 Animal (Experimental unit) effects	92
3.13.2	Treatment effects	93
3.14	Extension to more than one treatment factor in the Phase 1 experiment	94
3.14.1	Model and design parameters	94
3.14.2	Information matrix	96
3.14.3	Objective function	96
3.15	Summary	97
4	Optimal designs for two-phase experiments when the Phase 1 experiment is arranged in blocks	99
4.1	Introduction	99
4.2	An illustrative example	100
4.3	Design parameters and model	103
4.4	Defining the information matrix	104

4.5	An objective function for identifying optimal designs for the Phase 2 experiment	105
4.5.1	Illustrative example applying the old objective function	106
4.5.2	New Objective function and optimisation procedure	108
4.5.3	Illustrative example applying new objective function	110
4.6	Construction of the initial design	110
4.6.1	An initial design when the Tray effects are intentionally confounded with Run effects (better than confounded with Tag effects)	112
4.6.2	Performance remains unchanged whether Tray effects are confounded with Tag or Run effects	116
4.7	Modified simulated annealing for searching for the optimal design	119
4.8	An illustrative example with six treatments	120
4.9	Optimal designs for experiments involving two to eight treatments and two tech- nical replicates	125
4.9.1	Phase 1 Block and Plot effects	125
4.9.2	Treatment effects	126
4.10	Extension when the Phase 1 experiment is arranged with BIBD	126
4.10.1	BIBD example with six treatments	127
4.10.2	Optimal designs when the Phase 1 experiment is a BIBD	131
4.11	Summary	131
5	Estimation of variance components and effective degrees of freedom in two- phase proteomic experiments	133
5.1	Introduction	133
5.2	An illustrative example	134
5.3	Estimation of variance components	136
5.3.1	Constructing the score function and Fisher information matrix	136
5.3.2	Transformation from expected values in ξ^2 to estimates in ς	138
5.3.3	Estimating VCs in ς^2	139
5.4	Satterthwaite's approximation in deriving the EDF	139
5.4.1	Comparing VCs between LC and REML methods	140
5.5	EDF when Phase 1 experiment is arrange in a CRD	141
5.5.1	Example 1: A CRD with 2 treatments and 12 animals	141
5.5.2	Example 2: A CRD with 8 treatments and 16 animals	143
5.5.3	Example 3: A CRD with 4 treatments and 24 animals	146

5.6	Comparing the EDF when Phase 1 experiment is arranged in a RCBD	149
5.6.1	Four-plex system	150
5.6.2	Eight-plex system	153
5.6.3	Four-plex versus Eight-plex system	157
5.7	Summary	157
6	Discussion	161
6.1	Summary	161
6.2	Future lines of research	164
6.2.1	Shiny application for generating optimal designs of Phase 2 experiments .	164
6.2.2	Effective degrees of freedom versus average treatment efficiency factor . .	165
6.2.3	Missing values	166
6.3	More general future research directions	169
	Appendices	171
A	Reference manual of infoDecompuTE package	173
B	Information matrix in more detail	207
C	Tables of optimal designs when Phase 1 is CRD	211
D	Tables of properties of optimal designs when Phase 1 is CRD	227
E	Tables of optimal designs when Phase 1 is RCBD	235
F	Tables of properties of optimal designs when Phase 1 is RCBD	265
G	Tables of optimal designs when Phase 1 is BIBD	277
H	Tables of properties of optimal designs when Phase 1 is BIBD	283
	Bibliography	289

List of Tables

1.1	Partial theoretical ANOVA table by McIntyre (1955)	4
1.2	Partial theoretical ANOVA table by Curnow (1959)	5
1.3	ANOVA table of the wine-evaluation experiment described by Brien (1983)	8
2.1	Theoretical null ANOVA for an experiment with m_1 blocks of size k	29
2.2	Notations of the treatment and block factors between single and two-phase experiments.	35
2.3	Design of Phase 1 experiment showing the assignment of the eight animals, labelled $A - H$, to disease status' group.	39
2.4	Theoretical ANOVA showing the decomposition of DF and EMS associated with each source of variation.	40
2.5	Design of Phase 2 experiment showing allocation of sub-samples from animals and disease statues to runs and tags. The letters denote the animal IDs.	41
2.6	Phase 2 ANOVA of design in Table 2.5 with the coefficients of variance components of EMS.	42
2.7	Description of <code>summaryAovOnePhase</code> and <code>summaryAovTwoPhase</code> functions' arguments.	44
2.8	Theoretical ANOVA table.	51
3.1	Theoretical ANOVA of (a) the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 4$ animals arranged in a CRD and (b) the Phase 2 experiment ignoring Phase 1 with $n_r = 4$ runs and $n_\gamma = 4$ tags.	66

3.2	Design of Phase 2 proteomics experiment showing allocation of sub-samples from animals and disease statues to runs and tags, when the Phase 1 experiment consisting of $\nu = 2$ treatments assigned to each of $n_a = 4$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 2$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	67
3.3	Theoretical ANOVA of the Phase 2 experiment in Table 3.2.	67
3.4	Theoretical ANOVA of the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 6$ animals.	69
3.5	The MS-optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consisting of $\nu = 2$ treatments assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	69
3.6	Theoretical ANOVA of the MS-optimal design for the Phase 2 experiment in Table 3.5.	70
3.7	The A-optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consisting of $\nu = 2$ treatments assigned to assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	70
3.8	Theoretical ANOVA of the A-optimal design for the Phase 2 experiment in Table 3.7.	71
3.9	Theoretical ANOVA table for the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 4$ animals.	72
3.10	Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags based on the objective function with equal weights for E_a and E_τ , when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 4$ animals, $n_s = 3$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	73

3.11 Theoretical ANOVA for the Phase 2 experiment in Table 3.10, where E_a and E_τ share the same weight in the objective function.	73
3.12 Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags based on the objective function with $w_a = 0.75$ and $w_\tau = 0.25$ for E_a and E_τ , when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 4$ animals, $n_s = 3$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	74
3.13 Theoretical ANOVA for the Phase 2 experiment in Table 3.12 where $w_a = 0.75$ and $w_\tau = 0.25$ for E_a and E_τ in the objective function.	74
3.14 Theoretical ANOVA for the Phase 1 experiment with $\nu = 3$ treatments assigned to $n_a = 6$ animals.	76
3.15 Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags without the maximised DF associated with the Treatment effects, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	77
3.16 Theoretical ANOVA for the Phase 2 experiment in Table 3.15 without maximization of the DF associated with the Treatment effects.	77
3.17 Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags with the maximization of the DF associated with the Treatment effects, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	78
3.18 Theoretical ANOVA for the Phase 2 experiment in Table 3.17 with maximization of the DF associated with the Treatment effects.	79

3.19	Experiments involving 4 animals in the Phase 1 experiment and (a) two runs and two technical replicates (b) 3 runs and 3 technical replicates and (c) 4 runs and 4 technical replicates in the Phase 2 proteomics experiment using the four-plex iTRAQ TM labelling system.	80
3.20	Initial design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_a = 8$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 4$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	81
3.21	Different initial design with the same design parameters as Table 3.20.	81
3.22	Theoretical ANOVA for the Phase 1 experiment with $\nu = 6$ treatments assigned to $n_a = 18$ animals.	86
3.23	Optimal design for Phase 2 experiment showing the allocation of sub-samples from treatments assigned to animals based on objective function value of 0.9932, when the Phase 1 experiment consists of $\nu = 6$ treatments assigned to $n_a = 18$ animals, $n_s = 2$ sub-samples are then taken from each animals and measured in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 9$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	86
3.24	Theoretical ANOVA from the optimal design for Phase 2 experiment in Table 3.23 based on objective function value of 0.9932.	87
3.25	Theoretical ANOVA for a Phase 1 experiment with $\nu = 8$ treatments assigned to $n_a = 16$ animals.	87
3.26	Initial design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 8$ treatments assigned to $n_a = 16$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 4$ runs and $n_\gamma = 8$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	88
3.27	Theoretical ANOVA of the initial design for the Phase 2 experiment in Table 3.26.	88

3.28	Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 8$ treatments assigned to $n_a = 16$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 4$ runs and $n_\gamma = 8$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	89
3.29	Theoretical ANOVA of the optimal design for the Phase 2 experiment in Table 3.28.	89
4.1	Experimental design of Phase 1 plant experiment with $\nu = 2$ treatments (labelled a and b) assigned to $n_p = 4$ plants (labelled A to D) in $n_b = 2$ trays (labelled 1 and 2).	101
4.2	Theoretical ANOVA for the Phase 1 experiment in Table 4.1 with $\nu = 2$ treatments assigned to $n_p = 4$ plants in $n_b = 2$ trays.	101
4.3	Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to each of $n_p = 4$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 2$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	102
4.4	Theoretical ANOVA table for the Phase 2 experiment in Table 4.3.	103
4.5	Phase 1 experimental design showing the assignment of treatments and plants to trays, with $\nu = 3$ treatments assigned to $n_p = 12$ plants in $n_b = 2$ trays.	106
4.6	Theoretical ANOVA of the Phase 1 experiment in Table 4.5.	106
4.7	Theoretical ANOVA for the Phase 2 experiment in Table 4.14b.	107
4.8	Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags based on objective function (4.6), when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	107
4.9	Theoretical ANOVA table of the Phase 2 experiment in Table 4.8 from objective function (4.6).	108

4.10	Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags based new objective function, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	110
4.11	Theoretical ANOVA of the Phase 2 experiment in Table 4.10 from the new objective function described in Section 4.5.2.	111
4.12	Initial design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in each of $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	112
4.13	Theoretical ANOVA table of Phase 1 experiment with $\nu = 4$ treatments assigned to each of $n_p = 16$ plants in $n_b = 4$ trays.	113
4.14	Initial and final optimal designs for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consisting of $\nu = 4$ treatments assigned to each of $n_p = 16$ plants in $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 8$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	113
4.15	Theoretical ANOVA for the Phase 2 experiment in Table 4.14b.	114

4.16	Initial and final optimal designs for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Run effects, when the Phase 1 experiment consisting of $\nu = 4$ treatments assigned to each of $n_p = 4$ plants in each of $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 8$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	115
4.17	Theoretical ANOVA for the Phase 2 experiment in Table 4.16b.	115
4.18	Theoretical ANOVA of the Phase 1 experiment with $\nu = 4$ treatments assigned to $n_p = 4$ plants in each of $n_b = 2$ trays.	116
4.19	Initial and final optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_p = 8$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 4$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	117
4.20	Theoretical ANOVA for the Phase 2 experiment in Table 4.19.	117
4.21	Initial and final optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Run effects, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_p = 8$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 4$ runs and $n_\gamma = 4$ tags.	118
4.22	Theoretical ANOVA for the Phase 2 experiment in Table 4.21.	118
4.23	Illustrate the swapping between Tray 1 of Plant A under Treatment a and Tray 1 of Plant E under Treatment b to preserve the assignments of the sub-samples between trays and tags on the initial design for the Phase 2 experiment in Table 4.12, when the Tray effects are intentionally confounded with the Tag effects. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	119
4.24	Theoretical ANOVA table of Phase 1 experiment with $\nu = 6$ treatments assigned to $n_p = 18$ plants in $n_b = 3$ trays.	120

4.25	Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consists of $\nu = 6$ treatments assigned to each of $n_p = 18$ plants in each of $n_b = 3$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 9$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	121
4.26	Theoretical ANOVA table for the Phase 2 experiment in Table 4.25.	122
4.27	Theoretical ANOVA for the Phase 2 experiment in Table 4.25 with five basic treatment contrasts showing which stratum estimates each of the five different basic treatment contrasts.	124
4.28	Phase 1 experimental design with $\nu = 6$ treatments assigned to $n_p = 30$ plants in $n_b = 6$ trays. Upper case letters denote plant IDs, while the lower case letters denote the treatments.	127
4.29	Theoretical ANOVA table of the Phase 1 experiment arranged in a BIBD with $\nu = 6$ treatments assigned to $n_p = 5$ plants within each of $n_b = 6$ trays.	128
4.30	Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 6$ treatments assigned to $n_p = 30$ plants in $n_b = 6$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 15$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments. . . .	129
4.31	Theoretical ANOVA for the Phase 2 experiment in Table 4.30.	130
4.32	Theoretical ANOVA with DF and average efficiency factors for the Phase 2 experiment in Table 4.30 which generated by treating Tray and Plant effects as the fixed effects.	130
5.1	Optimal design of Phase 2 proteomics experiment showing allocation of sub-samples from animals and treatment to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to each of $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal, and labelled by $n_\gamma = 4$ tags and analysed in $n_r = 3$ runs of the Phase 2 MudPIT-iTRAQ TM experiment.	134
5.2	Theoretical ANOVA table of the Phase 2 experiment in Table 5.1.	135

5.3	Optimal (a) four- and (b) eight-plex designs of Phase 2 proteomics experiment when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to each of $n_a = 12$ animals, with $n_s = 2$ sub-samples taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment. Animal IDs are denoted by upper case letters, while the lower case letters a and b denote the two treatments. . . .	142
5.4	Theoretical ANOVA table for the optimal design of the Phase 2 experiment in Table 5.3a.	143
5.5	Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.3b.	143
5.6	Optimal (a) four- and (b) eight-plex designs of Phase 2 proteomics experiment when the Phase 1 experiment consists of $\nu = 8$ treatments each assigned to $n_a = 16$ animals, with $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	144
5.7	Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.6a.	145
5.8	Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.6b.	145
5.9	Optimal (a) four- and (b) eight-plex designs of Phase 2 proteomics experiment, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to each of $n_a = 24$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQ TM experiment. Upper case letters denote animal IDs, while the lower case letters denote the treatments.	147
5.10	Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.9a.	148
5.11	Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.9b.	148
5.12	Magnitudes of the run-to-run variation over animal-to-animal variation when the EDF become higher for four-plex experiment than eight-plex experiments based on Figure 5.4.	149

5.13	Optimal design of Phase 2 proteomics experiment showing allocation of sub-samples from trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 16$ plants in $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment using $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	150
5.14	Theoretical ANOVA table from the Phase 2 experiment in Table 5.13a.	151
5.15	Theoretical ANOVA table from the Phase 2 experiment in Table 5.13b.	151
5.16	Magnitudes of the run-to-run variation over plant-to-plant variation when the EDF become higher for design when Tray effects are intentionally confounded with Tag effects than the design when Tray effects are intentionally confounded with Run effect based on Figure 5.5.	153
5.17	Optimal design of Phase 2 proteomics experiment showing allocation of sub-samples from trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 16$ plants in $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQ TM experiment using $n_\gamma = 8$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.	154
5.18	Theoretical ANOVA table of the Phase 2 experiment in Table 5.17a.	155
5.19	Theoretical ANOVA table of the Phase 2 experiment in Table 5.17b.	155
6.1	Theoretical ANOVA table from the Phase 1 experiment arranged in CRD with $\nu = 8$ and $r_b = 2$, and from the Phase 2 experiment using the four-plex system. .	165
6.2	Theoretical ANOVA table from the Phase 1 experiment arranged in CRD with $\nu = 8$ and $r_b = 2$ and the Phase 2 experiment using the eight-plex system. . . .	166
6.3	Optimal design for Phase 2 experiment showing the allocation of sub-samples from treatments assigned to animals, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_a = 12$ animals, $n_s = 2$ sub-samples are then taken from each animals and measured in the Phase 2 MudPIT-iTRAQ TM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags.	167
6.4	Theoretical ANOVA table of design in Table 6.3.	167
6.5	Theoretical ANOVA table of design in Table 6.3 with Run 6 missing.	168
6.6	Theoretical ANOVA table of design in Table 6.3 with Runs 5 and 6 missing. . .	168

- 6.7 Residual DF and treatment average efficiency factors from the optimal design with different number of lanes and barcodes for Next-Generation Sequencing technology. [170](#)

List of Figures

1.1	Basic biological processes of producing functional proteins from DNA.	14
1.2	Structure of eight-plex-iTRAQ TM tags showing the reporter and balance group masses measured using m/z (Choe et al., 2007).	16
1.3	The process of protein digestion by trypsin. a)Intact protein. b)Digested peptides showing free N-termini. c)Peptides labelled with iTRAQ TM tag.	17
1.4	Schematic of analysis of a protein digest in the QSTAR [®] qTOF-MS/MS (Sciex, 2004).	18
1.5	The intensity peak resulting from the mass analysis, of a peptide species with $m/z = 1823.23$. A single peak is generated for the intact peptide since the mass spectrometer cannot distinguish between the eight samples of origin due to the isobaric tags.	19
1.6	Example of MS (blue) and MS/MS (red) spectra.	20
1.7	Possible b- and y-ions generated by collision-induced dissociation.	21
1.8	a) MS/MS spectra showing the b- and y-ions, and b) zoom-in into 113 to 121 m/z reporter ion region.	21
5.1	EDF for the variance of the treatment effects for two-phase experiment, i.e. for the optimal design of the Phase 2 experiment which takes account of the design of the Phase 1 experiment involves $\nu = 2$ treatments assigned to $n_a = 6$ animals. Each sample is further split into $n_s = 2$ sub-samples labelled by $n_\gamma = 4$ tags and measured in $n_r = 3$ runs, based on REML estimates of the variance components and on a linear combination of the mean squares.	141
5.2	EDF plots for optimal designs shown in Tables 5.3a and 5.3b, where EDF is calculated using VCs estimated by both the REML and LC methods.	144
5.3	EDF plots for optimal designs shown in Tables 5.6a and 5.6b, where EDF is calculated using VCs estimated by both the REML and LC methods.	146

5.4	EDF plots for optimal designs shown in Tables 5.9a and 5.9b, where EDF is calculated using VCs estimated by both the REML and LC methods.	147
5.5	EDF plots for optimal designs shown in Tables 5.13a and 5.13b, where EDF is calculated using VCs estimated by both the REML and LC methods.	152
5.6	EDF plots for optimal designs shown in Tables 5.17a and 5.17b, where EDF is calculated using VCs estimated by both the REML and LC methods.	156
5.7	EDF for optimal designs shown in Tables 5.13a, 5.13b, 5.17a and 5.17b, with Tray effects are assumed to be fixed, where EDF is calculated using VCs estimated by both the REML and LC methods.	158

Chapter 1

Introduction

This thesis considers on the computer generation of optimal designs for two-phase experiments. Two-phase experiments have two distinct features. The first feature is that samples of interest cannot be directly measured from a single experiment, thus, a second experiment is required to process the sample and to obtain the measurements of the response variable(s) of interest. The second feature is that two-phase experiments require two distinct experimental designs, one for each phase. The first design is directly associated with the first, or *Phase 1 experiment*, while the second design, which is of most interest in this thesis, is concerned with how the units from the first phase should be allocated to the units of the second phase, or *Phase 2 experiment*.

This thesis comprises three components. The first component describes the method of information decomposition of the designs of single- and two-phase experiments (with automation of the construction of theoretical ANOVA tables). The second component develops a computational approach for finding optimal designs for the Phase 2 experiment given the Phase 1 experiment is arranged in either a completely randomised, a randomised complete block or a balanced incomplete block design. The last component examines two methods of estimating variance components and how these affect the estimated effective degrees of freedom and the implications this has in terms of how well the error variance for treatment effects is estimated in the Phase 2 experiment.

Our motivating examples come from proteomic experiments, which are used to identify and measure all of the protein species in a biological sample with the goal being to link changes in their abundances to, for example, the presence or severity of conditions of disease. Proteomics experiments are just one example of a wide range of “omics” experiments which utilise technologies to universally detect the different molecular species (e.g. gene transcripts, metabolites, lipids, etc.) in a cell. What they all have in common is that detection and measurement of the target species cannot be made *in-vivo*. Each requires a subsequent laboratory-based exper-

iment for these measurements to be made. Thus, while the applications in this thesis focus on proteomics experiments, the methods which are presented apply more generally to other multiplexing “omics” platforms, where the term multiplexing refers to the simultaneous assaying of multiple biological samples. Moreover, these experiments are very expensive to conduct. Thus, a carefully thought out experimental design becomes very important.

The typical work-flow of a proteomic experiment starts with experimental units - these may be cells, tissues, bio-fluids, or entire organisms - being first perturbed by the experimental conditions of interest, i.e. the *Phase 1 experiment*. Since protein abundance cannot be measured directly from the experimental unit, the biological material must first be harvested, and the proteins are extracted and then measured in a subsequent laboratory-based experiment, i.e. the *Phase 2 experiment*. Hence, proteomics experiments necessarily have a *two-phase* structure.

The biological technology that will be discussed in this thesis is Multi-dimensional Protein Identification Technology (MudPIT), which allows the researcher to identify and quantify the entire complement of proteins in a given condition within a single biological sample. Due to the high variation between different MudPIT experiments, or *runs*, a multiplexing technology is introduced that allows the simultaneous analysis of multiple samples, i.e. multiple biological samples are assayed under homogeneous conditions. One such technology is the isobaric Tags for Relative and Absolute Quantitation (iTRAQTM) (Ross et al., 2004). Another advantage of multiplexing is that it reduces overall experimental costs. However, complications arise in the design of the Phase 2 experiment applying the multiplexing technology, more specifically in how samples from the Phase 1 experiment should be differentially labelled by iTRAQTM *tags* within each MudPIT run of the Phase 2 experiment.

This Chapter establishes some general insights into two-phase experiments, in particular, the multiplex experiments in proteomics. Sections 1.1 to 1.5 describe the evolution of the methods surrounding the two-phase experiments over recent decades. Section 1.6 describes some recent work in the design of proteomics experiments. Moreover, Section 1.7 details the biological background of the MudPIT-iTRAQTM experiments. Finally, Section 1.8 presents a brief overview of this thesis.

1.1 The introduction of the two-phase experiment

Two-phase experiments were first introduced by McIntyre (1955), who investigated the effects of four light treatments on the synthesis of tobacco mosaic virus in tobacco leaves. The Phase 1 experiment comprised two 4×4 arrays consisting of four successive leaves taken at defined

positions on the stem of each of eight tobacco test plants infected by the virus. Four light treatments were then assigned to the plants and leaves, such that each treatment occurred only once within each row and column in each of two 4×4 square arrays. This assignment is also known as the Latin square design. This Phase 1 experiment thus yielded 32 samples. However, the virus content of each leaf could not be measured directly from the test plants used in the Phase 1 experiment. Therefore, to measure the disease severity, sap was first expressed from the test tobacco plants and then injected into the leaves of a different set of assay plants (Phase 2 experiment) on which lesions subsequently appeared and were counted.

The Phase 2 experiment was arranged in a Graeco-Latin square design involving 16 assay plants divided into 4 sets of 4 assay plants. Within each set of 4 assay plants, four consecutive leaves were taken from four distinct positions on each plant. Then, the expressed sap from the four leaves from a test plant in the first Latin square of the Phase 1 experiment was taken and injected into half-leaves in a Latin square arrangement. The process was repeated with the sap from a plant in the second Latin square in the Phase 1 experiment. The Latin square assignments of the sap from each pair of Phase 1 plants to a set of assay plants at Phase 2 were such that when superimposed on one another they formed a Graeco-Latin square. Thus, each of the 32 samples in the Phase 1 experiment was measured four times in the Phase 2 experiment thereby giving a total of 128 measurements. This two-phase experiment illustrates the requirement of two different experimental designs; the first design is concerned with the allocation of light treatments to test plants, while the second design is concerned with assigning samples (sap expressed from the leaves of the Phase 1 test plants) to the leaves of the Phase 2 test plants in a manner that enables the differences in disease severity between light treatments to be estimated as efficiently as possible.

In McIntyre's experiment, the sap expressed from a single leaf (from the same phase 1 test plant) was injected into a half-leaf of each of four different Phase 2 assay plants. These four half-leaves yield four *technical replicates* of the light treatment applied to the leaf in the Phase 1 experiment. Technical replication enables us to estimate the variation introduced by all of the steps in the laboratory processes used to prepare samples for measurement. To truly get at this source of variation, we need to take sub-samples (e.g. from sap expressed from the same leaf of a Phase 1 test plant) and independently apply to each laboratory step. This enables the separation of biological variation (e.g. variation between leaves from the same plant and variation between plants) from technical variation. But, if only one measurement is made on each sub-sample, then technical variation and measurement error are confounded with one another. To separate technical variation and measurement error, we must also make multiple measurements on each

sub-sample. The latter is commonly known as *pseudo-replication*. As McIntyre (1955) stated, technical replication is required if the Phase 2 experiment introduces large and uncontrollable variation. For the light treatment experiment, each sample (sap from a single leaf) from the Phase 1 experiment is assayed and measured four times in the Phase 2 experiment.

McIntyre (1955) made three observations regarding the error variance of the treatment effects based on the light treatment experiment. Foremost, he established that if the Phase 1 design remains unchanged while the Phase 2 design is modified, then the error variances may differ between the different Phase 2 designs. Second, if the design is modified to have more technical replicates, the error variance for treatment effects can be reduced, but the time required to complete the experiment is increased. Third, if the design is modified so that the biological replication is eliminated, the error variance for the treatment effects increases. This is because the error variance in this case includes the variation among the leaves from single plants from both the Phase 1 and 2 experiments.

McIntyre (1955), as stated by Curnow (1959), performed an unweighted estimation of the treatment MS in the Between and Within Leaves Within Assay Plants strata, as shown in the partial theoretical ANOVA in Table 1.1. The theoretical ANOVA table contains the degrees of freedom (DF) and expected mean squares (EMS) for each source of variation. The components in the EMSs are: σ_ϵ^2 , which denotes the variation between half-leaves within leaves and assay plants; σ_Δ^2 , which denotes the variation between leaves within assay plants; σ_L^2 , which denotes the variation between leaves within test plants, and θ which is the fixed effects component for the light treatments.

Table 1.1: Partial theoretical ANOVA table by McIntyre (1955).

Source of Variation	DF	EMS
Treatment	3	$\sigma_\epsilon^2 + \sigma_\Delta^2 + 4\sigma_L^2 + 32\theta$
Residual	15	$\sigma_\epsilon^2 + \sigma_\Delta^2 + 4\sigma_L^2$

Curnow (1959) revisited McIntyre's light treatment experiment and generated a new ANOVA table showing the decomposition of the EMS corresponding to each source of variation associated with each treatment or block factor. Table 1.2 presented only those strata in Curnow's ANOVA table relevant to Treatment effects showing that the Treatment effects are estimated across Between and Within Leaves Within Assay Plants strata. This is because some of the Treatment information is confounded with leaves within the Phase 2 Assay Plants. Thus, the Treatment EMS in the Between Leaves Within Assay Plants stratum contains σ_Δ^2 , whereas the Treatment EMS in the Within Leaves Within Assay Plants stratum does not. Curnow (1959) termed the

estimation of Treatment effects in the Between Leaves Within Assay Plants stratum the *sums* analysis, which is equivalent to an *inter*-block analysis; and the estimation of Treatment effects in the Within Leaves Within Assay Plants stratum the *differences* analysis, which is equivalent to an *intra*-block analysis (Yates, 1936). Curnow (1959) then showed how to combine the inter- and intra-block analyses for this experiment using the weights computed from the error variances from these two strata. This new ANOVA table by Curnow (1959) provided a first step towards developing a better analytical procedure for two-phase experiments.

Table 1.2: Partial theoretical ANOVA table by Curnow (1959).

Source of Variation	DF	EMS
Between Leaves in Assay Plants		
Treatment	3	$\sigma_\epsilon^2 + 2\sigma_\Delta^2 + 2\sigma_L^2 + 16\theta$
Residual	15	$\sigma_\epsilon^2 + 2\sigma_\Delta^2 + 2\sigma_L^2$
Within Leaves in Assay Plants		
Treatment	3	$\sigma_\epsilon^2 + 2\sigma_L^2 + 16\theta$
Residual	15	$\sigma_\epsilon^2 + 2\sigma_L^2$

Wood et al. (1988) demonstrated how the Treatment effects and the variances of the Treatment effects can be estimated using the generalized least squares method. Their equations contain the *efficiency factor*, which is the proportion of treatment information present in the intra-block stratum. The example they presented has the Phase 1 Block effects confounded with the Phase 2 Block effects, but they stated that the efficiency factor of the Phase 1 Block effects to Phase 2 Block effects was not in the expression for estimating the Treatment effects; this is because their Phase 1 experiment was arranged in a completely randomised design. Thus, the efficiency factor associated with the Treatment effect is related to the degree to which the Phase 1 Block effects are confounded with the Phase 2 Block effects. If the Phase 1 design has Treatment effects confounded with the Phase 1 Block effects, arranged in a balanced incomplete block design, then the proportion of treatment information will be further diluted. Lastly, they also mentioned that the best Phase 2 design has the property that the treatment information in the lowest stratum is maximised.

1.2 Analysis of Variance Tables Based on Experimental Structure

Neither McIntyre (1955) nor Curnow (1959) described how their ANOVA tables were derived. Brien (1983) thus presented a generalised procedure for deriving the ANOVA table, showing

only the decomposition of the total DF into each source of variation, for both single- and two-phase experiments. This derivation resulted from recognising that experimental factors formed groups, which [Brien \(1983\)](#) referred to as *tiers*, such that factors from one tier were randomised to factors of another. This section briefly describes the procedure presented by [Brien \(1983\)](#) and simultaneously introduces some basic terminology used in experimental design.

[Brien \(1983\)](#) noted that before constructing an ANOVA the overall structure of the experiment must be determined. The first step is to identify treatment and block factors in the experiment and the *observational unit*. The observational unit is the smallest measurable quantity of experimental material ([Bailey, 2008](#)). The second step is to divide the factors into different tiers based on the randomisation scheme. *Randomisation*, one of the important principles in experimental design presented by [Fisher \(1935\)](#), involves the random assignment of one set of objects to another. The main purpose of randomisation is to enable researchers to obtain a data set with minimal systematic bias and to make the observations behave in a way that they are independent of one another.

A single-phase experiment, in general, has two tiers of factors. The first tier consists of those factors that jointly identify the observational unit in the absence of randomisation. [Nelder \(1965a\)](#) also referred to first tier factors as *block* factors. The second tier factors are those whose factor-level combinations are directly associated with factors in the first tier via randomisation. The second tier factors thus are what [Nelder \(1965b\)](#) termed *treatment* factors. The smallest unit in the first tier to which a treatment can be assigned, is known as the *experimental unit*. The last step is to determine the relationships between the factors within each tier. [Wilkinson and Rogers \(1973\)](#) developed a symbolic syntax to represent the relationships between factors within tiers, which are commonly referred to as block and treatment structures. [Brien and Payne \(1999\)](#) referred to their symbolic representations as *structure formulae*.

[Wilkinson and Rogers's](#) syntax was originally developed to generate and analyse ANOVA models in the GenStat statistical analysis program, but is now widely used in many statistical packages. Two basic operations described by [Wilkinson and Rogers \(1973\)](#) are used to represent block and treatment structures, namely *crossing* denoted by an asterisk, *, and *nesting*, denoted by a slash, /. These two operators represent the type of joint effects between more than one factor.

A two-phase experiment, in general, involves three tiers of factors, two of block factors and one of treatment factors. Consequently, two-phase experiments are also known as *multi-tiered experiments*. Tiers 1 and 2 comprise block factors from the Phase 2 and 1 experiments, respectively. Tier 3 contains treatment factors of the two-phase experiments.

The remainder of this section will illustrate the procedure of deriving the ANOVA table using the wine-evaluation experiment described by [Brien \(1983\)](#), where each wine is evaluated and presented for scoring once to tasters at a sitting. The order of wine presented is randomised to each taster. Further, the wines are made from a field trial to test the effects of several viticultural treatments assigned to plots arranged in a randomised complete block design.

This experiment is then a two-phase experiment, where the Phase 1 experiment is a field trial, consisting of p treatments assigned to b blocks each containing p plots. The Phase 2 experiment is a wine evaluation experiment, consisting of t testers and bp sittings. The observational unit is the sample of wine given to a taster at a particular sitting. The factor-level combinations of Taster and Sitting factors cannot be randomised; thus, Taster and Sitting factors form the first tier. Block and Plot factor combinations are randomised to Sittings within each Taster; thus, Block and Plot factors form the second tier. Finally, levels of Treatment are randomised to the Plots within each Block; thus, the Treatment factor forms the third tier. The structure formulae of the Phase 2 and 1 block tiers are therefore written as

$$\text{Taster/Sitting,} \tag{1.1}$$

$$\text{Block/Plot} \tag{1.2}$$

and the treatment tier is simply expressed as

$$\text{Treatment.} \tag{1.3}$$

Once the structure formula for each tier of the two-phase experiment is determined, the first step in generating the ANOVA table is to expand each formula for each tier, the rules for which are described by [Wilkinson and Rogers \(1973\)](#). Thus, block structure formulae (1.1) and (1.2) are expanded to

$$\text{Taster} + \text{Taster.Sitting} \tag{1.4}$$

and

$$\text{Block} + \text{Block.Plot}, \tag{1.5}$$

respectively, where Taster.Sitting and Block.Plot are read as Between Sittings Within Tasters and Between Plots Within Blocks, respectively.

The second step of generating the ANOVA table is to examine every pair of terms between the two tiers of the expanded structure formula in (1.4) and (1.5) for the presence of confounding in the design. The terms in the block structure formula, given in (1.4), describe the relationship between first tier factors, which form the basic structure of the ANOVA table. The terms in the

structure formula, given in (1.5), associated with second tier block factors are inserted below the terms of the first tier, with indentation if the two terms of different tiers are confounded. The confounding between the terms can be elucidated using orthogonal contrasts generated from each term. The terms from tier 3, i.e. the treatment tier in (1.3), are grouped with those from tiers 1 or 2, with another indentation in the event of confounding between terms from different tiers. After going through these processes, the ANOVA table, with only the decomposition of the DF, can be derived (see Table 1.3). The indentation on the line with Treatment indicates that the effects of Treatment are confounded with the Block.Plot. A more detailed description of these procedures is available in Brien (1983).

Table 1.3: ANOVA table of the wine-evaluation experiment described by Brien (1983).

Source of Variation	DF
Between Tasters	$t - 1$
Between Sitzings Within Tasters	$t(bp - 1)$
Between Blocks	$b - 1$
Between Plots Within Blocks	$b(p - 1)$
Treatment	$p - 1$
Residual	$(b - 1)(p - 1)$
Residual	$(bp - 1)(t - 1)$

1.3 Sweeping operations for two-phase experiments

The sweeping operations for general ANOVA were introduced by Wilkinson (1970), and Payne and Wilkinson (1977). Every experiment generates n responses which can be viewed as a data vector, \mathbf{y} say, with length n in n -dimensional space (Payne and Wilkinson, 1977). The standard sweeping operation estimates the effects of a term from the expanded structure formula, and then subtracts the estimated effects from the current working vector, which then becomes the working vector for the next sweep (Brien and Payne, 1999).

The Phase 1 field experiment described in Section 1.2 is used to demonstrate the sweeping operations. The factors for this experiment consist of p treatments assigned to b blocks each containing p plots. The expanded structure formulae of the block and treatment tiers are given in (1.5) and (1.3), respectively. The first sweep is for the grand mean, i.e. removing the effects of the grand mean, which is implicit in the structure formulae. This is performed using the *reanalysis sweep*, which can be expressed by

$$\mathbf{I} - E_G \mathbf{P}_G, \quad (1.6)$$

where \mathbf{I} is the $n \times n$ identity matrix, and E_G and \mathbf{P}_G denote the efficiency factor and projection matrix, respectively, associated with the grand mean. The projection matrix \mathbf{P}_G is defined as

$$\mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}',$$

where \mathbf{K} denotes an $n \times n$ averaging matrix with all elements equal to n^{-1} . In general, the reanalysis sweep is used to sweep for non-orthogonal terms from the data vector. Thus, the sweep which removes the Block effects from the data vector is given by

$$\mathbf{I} - E_B\mathbf{P}_B,$$

where E_B and \mathbf{P}_B denote the efficiency factor and projection matrix, respectively, associated with the Block term in the block structure formula.

The second type of sweep operator is the *pivotal sweep* which involves projecting the data vector onto the vector subspace of the associated terms. This sweep is used to determine the effects and the SS for terms in the second structure formula, within terms of the first structure formula. Thus, the pivotal sweep for estimating the Treatment effect is given by

$$E_T\mathbf{P}_T,$$

where E_T and \mathbf{P}_T denote the efficiency factor and projection matrix, respectively, associated with the Treatment term in the treatment structure formula. If a given design has Treatment effects confounded with Block effects, the sweeping sequence involves two pivotal sweeps, first for the Block term and then for the Treatment term, i.e.

$$E_T\mathbf{P}_TE_BE_B\mathbf{P}_B.$$

In general, the pivotal sweep is required for the terms in the structure formula for the first (block) tier before performing the same sweep for the terms in the structure formula for the second (treatment) tier that are confounded with it. Thus, the sweeping sequence to estimate the effects of Treatment in the Between Blocks stratum for the field experiment is given by

$$E_T\mathbf{P}_TE_BE_B(\mathbf{I} - E_G\mathbf{P}_G)\mathbf{y},$$

which consists of a reanalysis sweep for the grand mean in (1.6), and then two pivotal sweeps for the Block and Treatment terms. The sweeping sequence for estimation of the other effects of the experiment can also be derived from a combination of reanalysis and pivotal sweeps. Thus, the ANOVA table can be constructed with the effects for each source of variation.

Brien and Payne (1999) extended the sweeping algorithm to work for two-phase experiments by performing additional sweeps for terms from the structure formula of the third tier. Using the example of the two-phase wine-evaluation experiment described in Section 1.2, Treatment effects estimated across the Between Plots Within Blocks and Between Sitzings Within Tasters strata are given by

$$E_T \mathbf{P}_T E_P \mathbf{P}_P (\mathbf{I} - E_B \mathbf{P}_B) E_S \mathbf{P}_S (\mathbf{I} - E_J \mathbf{P}_J) (\mathbf{I} - E_G \mathbf{P}_G) \mathbf{y},$$

where \mathbf{P}_P , \mathbf{P}_S and \mathbf{P}_J denote the projection matrices associated with Plots, Sitzings and Tasters, respectively, and E_P , E_S and E_J denote the average efficiency factors associated with Plots, Sitzings and Tasters, respectively.

To summarise, Brien and Payne (1999) demonstrated the use of the reanalysis and pivotal sweep operations to construct ANOVA tables. Chapter 2 further discusses this approach to constructing the ANOVA tables.

1.4 Work by Brien and Bailey

Brien and Bailey (2006) described how randomisation can be performed for a multi-tiered experiment, involving at least three tiers of factors. Thus, two-phase experiments are a special case of multi-tiered experiments. As Brien (1983) pointed out, randomisation can affect the structure of the experiment. In an experiment involving three tiers of factors, a randomisation typically must be performed twice: (1) for the allocation of treatments to experimental units in the Phase 1 experiment, and (2) for the allocation of experimental units from the Phase 1 experiment to the experimental units in the Phase 2 experiment. The randomisation procedure for multi-tiered experiments thus is termed *multiple randomisation* (Brien and Bailey, 2006).

Brien and Bailey (2006) compared and contrasted six types of multiple randomisation three-tiered experiments. Brien and Bailey (2009, 2010) discussed the aspects of orthogonal decomposition of the data space for multi-tiered experiments with respect to different types of multiple randomisation. Bailey et al. (2016) further discussed estimation theory for the randomisation-based model of multi-tiered experiments.

Different multiple randomisations can have different directions of randomisation between tiers; Brien and Bailey (2009, 2010) showed that this can affect the ordering of orthogonal decomposition between terms of different tiers. Additionally, knowing the initial randomisation procedure allows researchers to recognise how terms are confounded with one another between tiers, which can let researchers more easily determine the decomposition method required for the

experiment. However, identifying the types of multiple randomisation that fit a given multi-tiered experiment and then determining how the ordering of the decomposition should be achieved can be laborious. Given that the structure formulae for each tier and the design are known, a decomposition method presented in Chapter 2 can be used for all two-phase experiments. This method does not require the identification of the type of multiple randomisations that should be fitted. Note that the randomisation procedure should still be applied before the experiment to minimise systematic bias and to obtain a more accurate result.

1.5 Systematic approach to design two-phase experiments

More recently, [Brien et al. \(2011\)](#) discussed a systematic approach to design two-phase experiments. This paper only considered designs with an orthogonal structure. Some principles described in their paper are used to develop the methods presented in Chapters 3 and 4 of this thesis.

[Brien et al. \(2011\)](#) first presented a list of rules for calculating the EMS in ANOVA tables for two-phase experiments. However, these rules can be laborious to follow for experiments that involve numerous treatment and block factors. Chapter 2 of this thesis thus presents an R package that automates the generation of the ANOVA tables, including the decomposition of DF and EMS, for any single- or two-phase experiment.

[Brien et al. \(2011\)](#) then discussed some fundamental principles for designing two-phase experiments. First, the unit of the Phase 1 experiment with the highest variation, but orthogonal to the Treatment effects, should be confounded with that of the Phase 2 experiment containing the highest variation. If the experimental unit in the Phase 1 experiment has large variation, Phase 1 unit should be confounded with the Phase 2 unit with a smaller variation. This principle attempts to minimise the error variance of the Treatment effects ([Brien et al., 2011](#)).

1.6 Experimental design for multi-plex proteomic experiments

[Oberg and Vitek \(2009\)](#) discussed the principles of experimental design for mass spectrometry-based proteomic experiment, including the MudPIT-iTRAQTM experiments. The goal of an experimental design is to allocate individual samples to tags and runs in a way that avoids systematic bias and reduces the error variance of treatment comparisons.

Oberg and Vitek (2009) recognised that each MudPIT run is a block and, thus, showed how the treatment allocation to runs can be arranged as either a randomised complete block design (RCBD), or a balanced incomplete block design (BIBD). Two additional designs were also discussed: the reference design and loop design. The reference design is where each run contains a reference sample for comparison, while the loop design is where the samples are allocated such that they are cycled through the blocks systematically. The reference sample is not in itself of interest in the experiment, but used as method of increasing the robustness of the design to the loss or failure of runs. Oberg and Vitek (2009) concluded that the best designs are those which minimise the variance of the pairwise differences between treatment means. They found the RCBD to be the best amongst the designs they considered because it requires fewer runs. Additionally, if a run is lost, although entire set of biological replicates is lost, the resultant design still retains an orthogonal structure. However, if the loss of a run occurs in the BIBD or loop design, the resultant design may not preserve the balanced structure, i.e. the treatment replications and within block pairwise treatment concurrences become different. The reference design is most robust under conditions of run failure, but generally requires more MudPIT runs than the other designs to measure the additional reference samples. Given that the number of tags can be either four (Ross et al., 2004) or eight (Choe et al., 2007), and given the high cost of performing the experiment in each MudPIT run, biologists are likely to utilise all of the tags to measure protein abundances of samples.

Oberg and Mahoney (2012) recognised that the allocation of treatments to runs and tags is a randomised block design, where the MudPIT run is the block factor. Moreover, Oberg and Mahoney (2012) suggested that multiple MudPIT runs are required to avoid the confounding of tag effects with treatment effects. However, Oberg and Vitek (2009), and Oberg and Mahoney (2012) only considered the allocations of treatments to runs and tags. The allocation of any other block factors from the Phase 1 experiment, e.g. animals or plants, is equally important because it can also affect the error variance of treatment comparisons.

Any proteomics experiments, using MudPIT coupled with iTRAQTM to compare the proteomes between different conditions, requires two phases of experimentation to generate the protein abundance data. The first phase involves the preparation of the animal or plant models under different treatment conditions of interest (Phase 1). Since protein abundance cannot be measured directly from these tissue samples, a second phase of experimentation is needed. Thus, the Phase 1 experiment provides the physical material, which is analysed in a subsequent experiment using MudPIT coupled with iTRAQTM (Phase 2). This a two-phase experiment (Jarrett and Ruggiero, 2008). It will become clear throughout this thesis that designing such an

experiments, careful consideration must be given to sources of variation by each phase.

1.7 Quantitative Proteomics using MudPIT coupled with iTRAQTM

This section we discuss in detail how the data are generated. This type of proteomic study requires the use of a combination of technologies, coupled with database searching, to identify and quantify the proteins within a target cell, tissue or biofluid. This section aims to describe this identification and measurement in detail, which would help design proteomics experiments using the technologies. Section 1.7.1 thus provides a detailed introduction to proteins and some of their properties, and defines the term proteome. Section 1.7.2 describes the process of separating a complex protein mixture into smaller subunits to enable high resolution measurements of the constituent components of proteins. Section 1.7.3 describes a recent protein labelling technology that enables the simultaneous analysis of complex multi-protein mixtures. Section 1.7.4 describes the laboratory workflow where this protein labelling technology is coupled with mass spectrometry, and Section 1.7.5 describes the role of database searching in protein identification.

1.7.1 Proteins and the Proteome

Proteins are one of the major macromolecules within a biological system that contribute to every process in a living system and are considered the essential building blocks of life. They are constructed from chains of amino acids derived from the genes within a cell nucleus through a process known as transcription, where DNA is copied to messenger RNA (mRNA), followed by translation, where mRNA is decoded into proteins (see Figure 1.1). Furthermore, a pre-mRNA may undergo alternative splicing as part of the post-transcription process, and thus can produce multiple protein sequences. This makes the study of protein expression more functionally relevant than mRNA transcription, also known as *gene expression*.

A protein is made up of many smaller units, or *peptides*, that comprise subsequences of the amino acids in the intact protein. Proteins thus are also termed *polypeptides*, while peptides are produced by enzymatic digestion of a whole protein. An example of such an enzyme is *trypsin* (Eidhammer et al., 2007).

The *proteome* is the entire complement of proteins expressed by the genome (i.e. genetic material) in a cell, or in tissue or bio-fluid of an organism at a given time under specific conditions (Boehm et al., 2007). The function of a protein corresponds to the timing and location of its

expression. Proteomic research thus aims to identify, localise, and quantify many more unknown and known proteins, and better understand their functions.

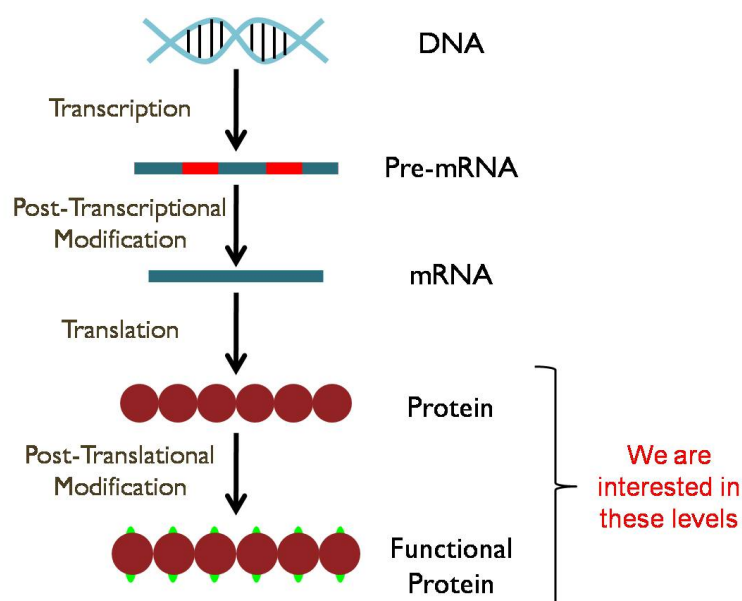


Figure 1.1: Basic biological processes of producing functional proteins from DNA.

There are many ways to study proteins. For example, physical structures of a protein can be studied by X-ray crystallography (Blow, 2002), protein-protein interactions can be studied using the yeast two-hybrid system (Fields and Song, 1989), and the abundance of an individual protein under a defined condition can be studied using isotopic-labelling. The latter, referred to as *quantitative proteomics*, is achieved via a process of separating a complex protein mixture into smaller subunits to enable high resolution measurements of the constituent components of proteins. The use of **M**ulti-dimensional **P**rotein **I**dentification **T**echnology (MudPIT) together with **i**sobaric **T**ags for **R**elative and **A**bsolute **Q**uantitation (iTRAQ™) is just one technology used to measure protein abundance.

1.7.2 Multi-dimensional Protein Identification Technology

Multi-dimensional Protein Identification Technology (MudPIT) is a chromatography-based method, which uses a suite of technologies to separate a peptide mixture in order to identify and quantify the constituent proteins in the original sample (Washburn et al., 2001). The process typically involves the separation of the peptide mixture into three orthogonal dimensions. The first separation of peptide species is by their charge, using *strong cation exchange chromatography* (SCX). This is followed by a second separation by hydrophobicity, using *reversed phase liquid chromatography* (RPLC). The third separation is by mass, and is carried out by *mass spectrometry* (MS). This reduces the complexity of the sample and enables high throughput protein analysis. Each

MudPIT *run* thus comprises these three steps of separation.

The SCX column contains immobilised negatively charged sulfonic acids that form an ionic interaction with the positively charged peptides. Hence, the charge separation by SCX divides a protein digest (i.e. a mixture of peptides) into many different fractions based on the strength of the charge interaction between the peptides and the sulfonic acids. Different peptide sequences have different affinities for the SCX resin. This allows for complex peptide mixtures to be fractionated by gradually increasing the concentration of a competing salt solution (for binding to the sulfonic acid groups in a gradient) in a step-wise manner. The salt concentration ranges between 10mM to 500mM. Hence, at each interval of salt concentration, a set of peptides is released into the next MudPIT phase. Each of these charge intervals is termed a *salt step*, and increasing the number of salt steps enables the detection of proteins with low abundance.

The set of peptides from each charge fraction is then separated by RPLC based on hydrophobicity. This is achieved with a separate column that contains silica beads with chains of 18 carbon atoms attached. The peptides are loaded into the column, which causes hydrophobic interactions with the carbon chains. An organic solvent is then added to the column in concentrations that increase over time, causing the peptides to emerge or *elute*, with the least hydrophobic peptides eluting first. Eluted peptides are then detected using the mass spectrometer.

The mass analysis is performed by MS, whereby each peptide's *mass-to-charge ratio* (m/z) is measured and then used to calculate its molecular mass. Peptide separation is described in more detail in [Eidhammer et al. \(2007\)](#).

Tandem mass spectrometry (MS/MS), which consists of two repeated phases of MS, was used for this study. Peptide fragmentation occurs between the two phases of MS, and identification and quantification of the peptides and proteins is thus based on these peptide fragments. MS/MS enables higher specificity in protein identification and enables more accurate quantification.

MudPIT has some limitations. For example, large variation in signal intensity between different MudPIT runs can make inter-sample comparisons of peptide or protein abundance difficult. This limitation has been resolved by iTRAQTM labelling, which enables the simultaneous analysis of up to eight distinct protein digests within a single MudPIT run. For this thesis, each MudPIT run is referred to as a *run*.

1.7.3 iTRAQTM for Protein Quantitation

In 2004, [Ross et al.](#) introduced a peptide labelling technology, namely isobaric Tags for Relative and Absolute Quantitation (iTRAQTM), enabling differential labelling of peptides in different biological samples. In its initial format, iTRAQTM comprised four isobaric tags, each consisting

of a reporter group, a balance group, and a peptide reactive group. The reactive group binds the N-terminus at the start of each peptide, and, for those peptides contain lysine residues (i.e. amino acid), then also on the lysine's side chain. The four reporter groups have m/z values 114, 115, 116 and 117, with corresponding balance group values of 31, 30, 29 and 28. Each of the four tags thus has an identical total m/z value of 145 Da, making them isobaric. This enables identical peptide species, differentially labelled with the four tags, to be indistinguishable with respect to the intact mass of the peptide when selected for MS/MS (Ross et al., 2004). For MS/MS, the relative abundances are determined using the reporter ion signals at m/z values of 114, 115, 116 and 117 on the *mass spectrum*, i.e. the graphical representation of the peptides and peptide fragments based on their m/z values and abundances. Mass spectra are generated for both phases of MS/MS, i.e. for the intact peptide during the first cycle, and for the fragmented peptides during the second fragmentation cycle. The four different labels thus allow the simultaneous analysis of four different samples (Ross et al., 2004).

	Reporter Group	Balance Group	Peptide
m/z ratio {	113	192	
	114	191	
	115	190	
	116	189	
	117	188	
	118	187	
	119	186	
	121	184	
	Isobaric Tag		
	Total Mass = 305		

Figure 1.2: Structure of eight-plex-iTRAQTM tags showing the reporter and balance group masses measured using m/z (Choe et al., 2007).

Choe et al. (2007) described a new multiplexing strategy, based on the same concept as the four-plex iTRAQTM system, allowing the simultaneous analysis of up to eight distinct protein samples (see Figure 1.2). This scheme involves reporter ion signals located at m/z values of 113, 114, 115, 116, 117, 118, 119 and 121. No label is used for an m/z of 120 because it has the same mass as the phenylalanine immonium ion (Pierce et al., 2008). For this thesis, each iTRAQTM tag is referred to as a *tag*.

1.7.4 MudPIT Coupled with iTRAQTM: Laboratory Workflow

Once the target cells or tissues are harvested, each sample is independently processed, including steps for reduction, alkylation, total protein quantification, and enzymatic digestion (usually with trypsin) into many smaller peptide fragments (Ross et al., 2004). Since proteins exist in a three-dimensional structure, held together by disulfide bonds, reduction breaks down these bonds to produce a two-dimensional linear structure. The alkylation step prevents the reformation of the disulfide bonds. A protein assay is used to measure the total protein content of each sample after reduction and alkylation. This ensures that the total amount of protein to be compared using iTRAQTM is approximately equal for all samples. Enzymatic digestion with trypsin specifically cleaves the protein immediately after every lysine (K) and arginine (R) residue (see Figure 1.3).

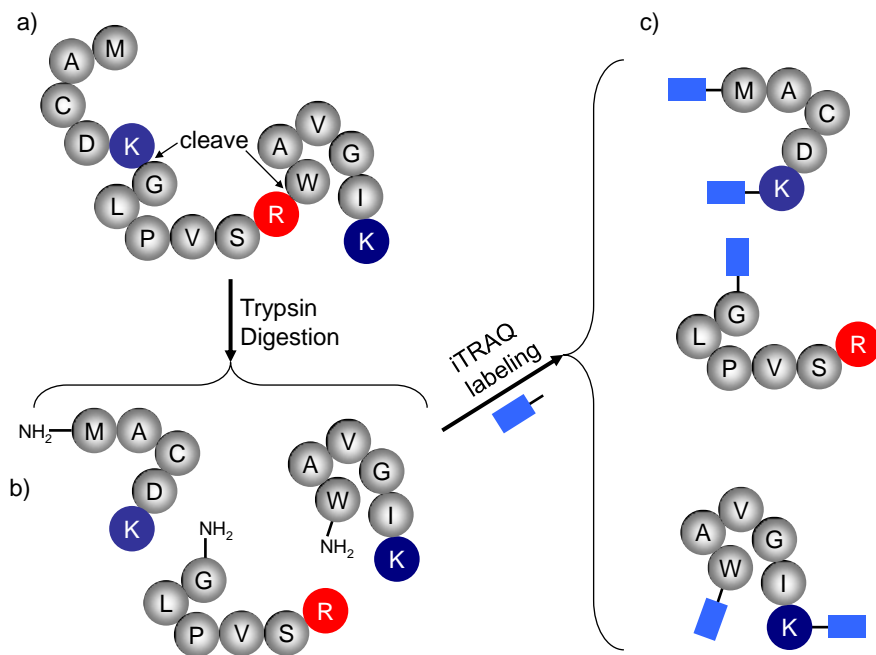


Figure 1.3: The process of protein digestion by trypsin. a) Intact protein. b) Digested peptides showing free N-termini. c) Peptides labelled with iTRAQTM tag.

The iTRAQTM labelling chemistry works by binding the tag to the free N-termini at the start of each peptide and on the K side chain. Thus, with fully efficient enzymatic cleavage, peptides containing one K residue are labelled twice, as the K residue's side-chain also contains a free N-terminus (Wiese et al., 2007). However, miss-cleavages do occur, resulting in some peptides containing more than one K residue. A consequence of this is that these additional K residues are also tagged resulting in a much higher reporter ion abundance due to the additional tags. For example, if a peptide has been labelled twice, the abundance of this peptide is shown as double

the true amount in the mass spectrum. To avoid this complication, a *relative quantification* is used instead of an *absolute quantification*. Absolute quantification is calculated as the ratio of a target protein's abundance to a protein of a known concentration either within the same sample, or in a different sample in the same MudPIT run (Thelen and Peck, 2007). The sample of known concentration is also referred as a *spike-in*. Relative quantification is the ratio of a particular protein's abundance in one sample to the same protein's abundance in another sample, within the same MudPIT run. Calculating the relative abundance overcomes the issue of the multiple tags on the same peptide because they cancel through this division process.

Next, approximately equal concentrations of the differentially labelled peptide samples are pooled and undergo the first two phases of MudPIT, i.e. separation by charge with SCX, and separation by hydrophobicity with RPLC (see Section 1.7.2). Finally, the peptide mixture is analysed according to molecular mass using MS/MS. There are many different types of mass spectrometer. An example presented here is the *ElectroSpray Ionisation Quadrupole-Time-Of-Flight tandem mass spectrometer* (ESI-qTOF-MS/MS), called the QSTAR[®] (Sciex) (see Figure 1.4).

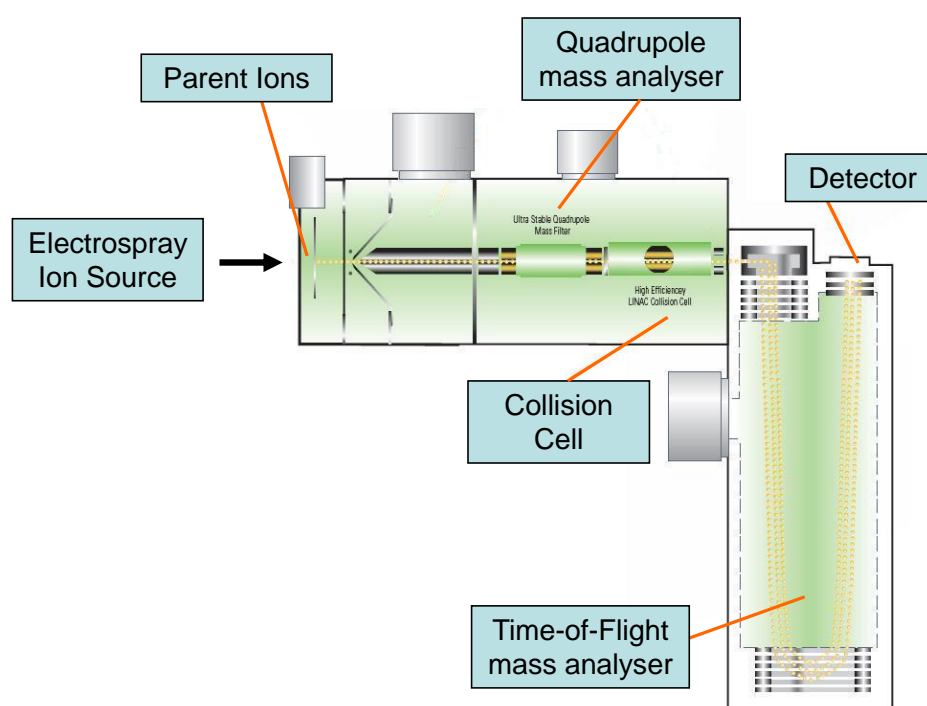


Figure 1.4: Schematic of analysis of a protein digest in the QSTAR[®] qTOF-MS/MS (Sciex, 2004).

The first step of MS/MS involves subjecting the peptide mixture to ElectroSpray Ionisation (ESI), which transforms each peptide into a positively charged ion, thereby allowing the peptides to be selected, analysed and detected by the mass spectrometer. The separated peptide ions (i.e. ionised peptides) eluting from the RPLC are sprayed continuously into the mass spectrometer

for up to 100 minutes for each salt step.

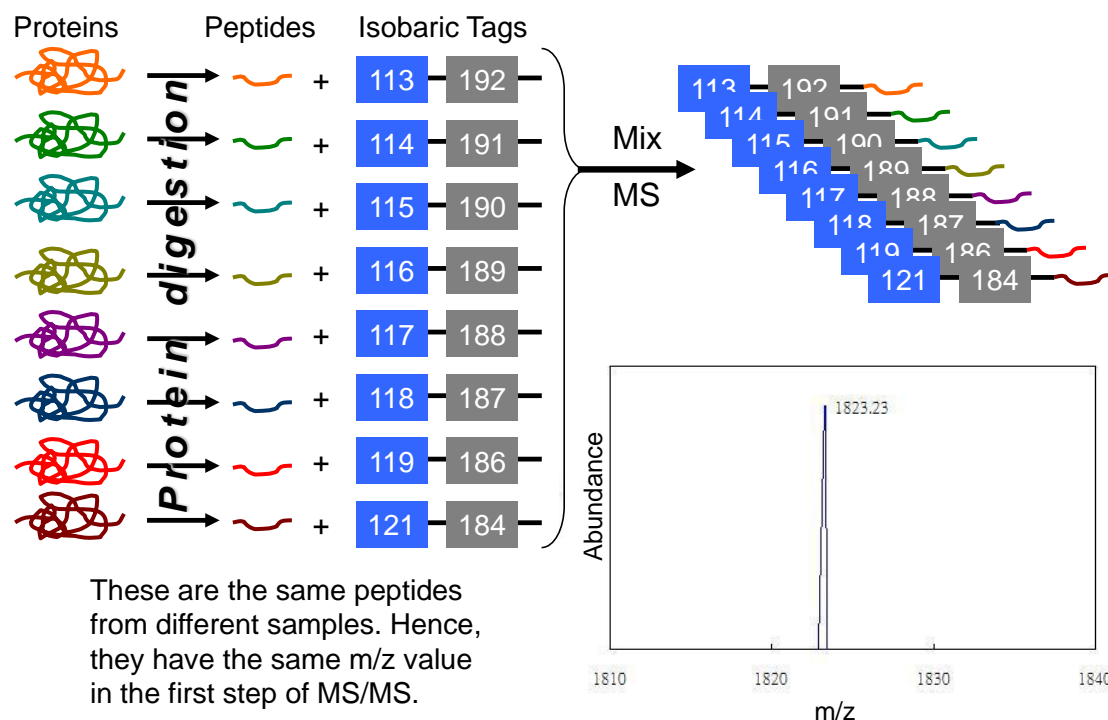


Figure 1.5: The intensity peak resulting from the mass analysis, of a peptide species with $m/z = 1823.23$. A single peak is generated for the intact peptide since the mass spectrometer cannot distinguish between the eight samples of origin due to the isobaric tags.

MS/MS consists of two repeated phases of mass analysis and both phases use the Quadrupole-Time-Of-Flight (qTOF) mass analyser. In the first phase, all the peptide ions are passed through the quadrupole mass analyser, and their molecular masses are analysed by the time-of-flight (TOF). The TOF measures the time taken for these ions to travel a known distance. From this, each ion's m/z and abundance is estimated and presented in a MS spectrum (see Figure 1.5). In the second phase, the peptide ions that have the highest abundance, as detected by the MS, are subsequently isolated by the quadrupole mass analyser. Normally up to three different peptide ions, also known as *precursor* or *parent* ions, are selected in one cycle. Note that these selected peptide ions for the second phase of MS are new peptide ions sprayed by the ESI, as once peptide ions are moved into the TOF they are never recovered. Further, the rate of flow is slow enough that software can instantly inform the quadrupole mass analyser to let a specific peptide pass through. Each of these cycles generally takes around 5.5 seconds: one second to acquire the first MS spectrum, then one, one-and-a-half, and two seconds to acquire the three subsequent MS/MS spectra. Figure 1.6 illustrates how the MS and MS/MS spectra are produced under the same time scale, where one MS spectrum will generate the three MS/MS spectra from the three most abundant peptides.

After a specific precursor ion is isolated for the second phase of MS/MS, it moves into a

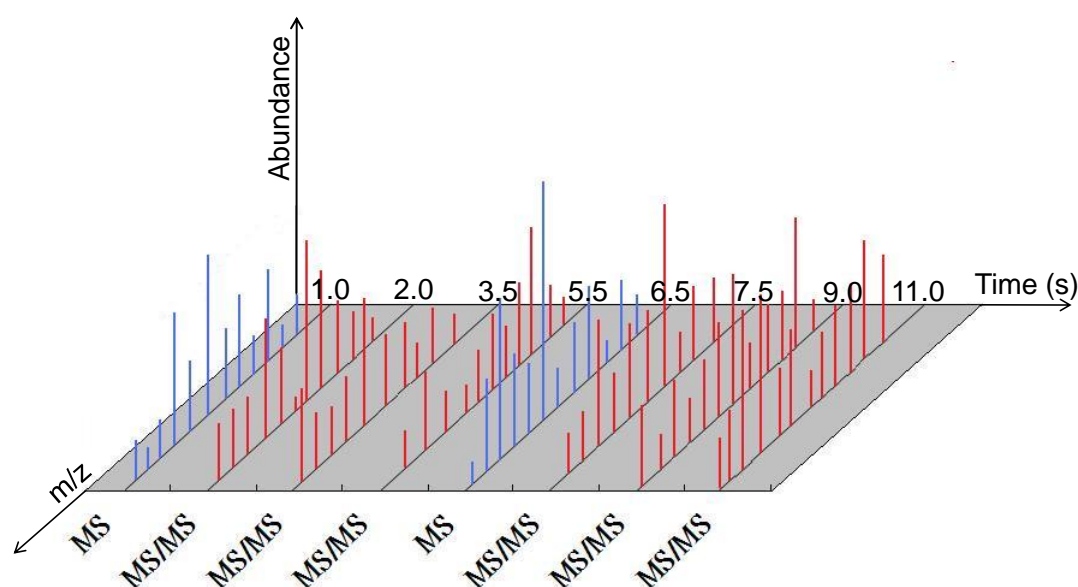


Figure 1.6: Example of MS (blue) and MS/MS (red) spectra.

collision cell where it is fragmented into smaller components by collisions with an inert gas. This process, called *collision-induced dissociation* (CID), cleaves the peptide bonds and the chemical bonds with the iTRAQTM tags. This produces b-ions, y-ions and iTRAQTM reporter ions (see Figure 1.7) (Ross et al., 2004). The molecular masses of these fragment ions are measured in the TOF. The abundances of the iTRAQTM labels are estimated from the peaks in the 113 to 121 m/z reporter ion regions. The area under each of these peaks serves as an approximation for the abundance of the peptide in a given sample. The rest of the peaks consist mainly of b- and y- ions, and their m/z are used for peptide identification by comparing the molecular weights of the experimental fragments (i.e. MS and MS/MS spectra) to theoretical fragments generated by in-silico fragmentation of peptides in an appropriate database (see Figure 1.8).

1.7.5 Data Generation from MudPIT-iTRAQTM

In every MudPIT run, the QSTAR[®] machine generates a WIFF file for each salt step that is performed. Each WIFF file contains all of the MS and MS/MS spectra from the fragmented peptides. There are a number of software packages which are available to perform both peptide-protein identification and the statistical analysis.

Peptide-to-protein identification involves database searching to identify the proteins that match a given peptide from the MudPIT-iTRAQTM experimental data. Examples of such software are MascotTM 2.6 (Matrix Science), SEQUEST[®] (Sadygov et al., 2004), and ProteinPilotTM 5.0 (Sciex). In general, the available software employ a *bottom-up* approach (Nunn and Timper-

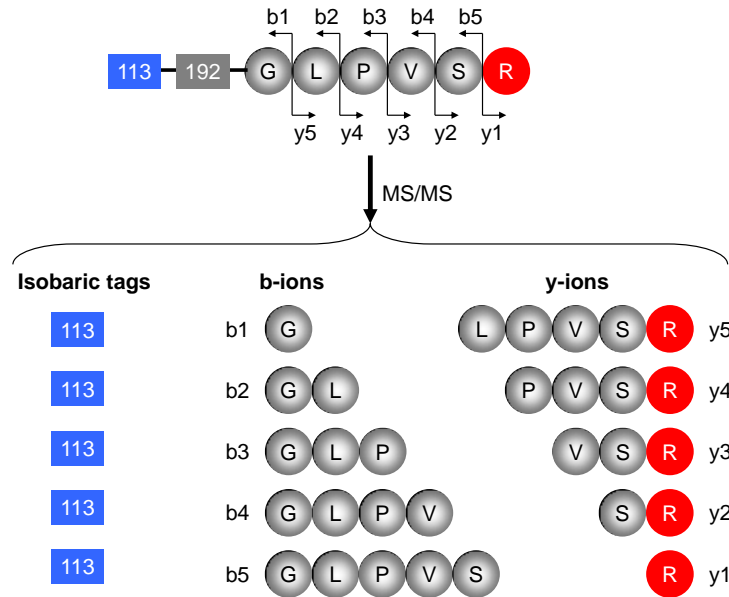


Figure 1.7: Possible b- and y-ions generated by collision-induced dissociation.

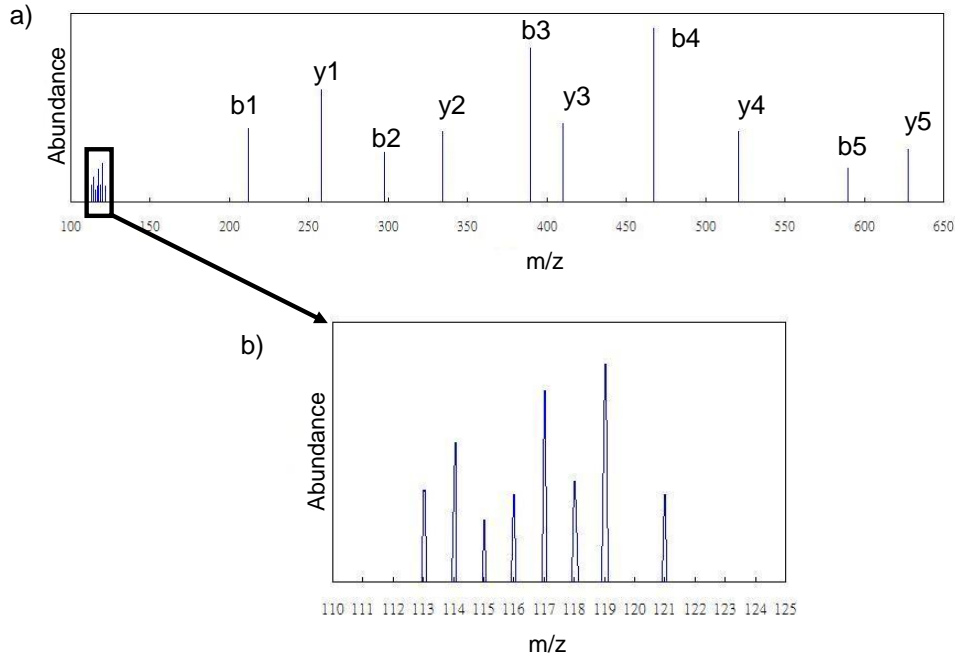


Figure 1.8: a) MS/MS spectra showing the b- and y-ions, and b) zoom-in into 113 to 121 m/z reporter ion region.

man, 2007), also known as *shotgun proteomics* (Nesvizhskii and Aebersold, 2005), for protein identification and quantification. The identification method involves a series of peptide matching steps. The mass spectra of the observed fragment ions that were detected in the MudPIT-iTRAQTM analysis are compared with the mass spectra of theoretical fragment ions, also known as *theoretical spectra*. These theoretical spectra are created from the in-silico tryptic digestion of proteins from an appropriate protein database (Sadygov et al., 2004). Then, the original intact proteins can be identified from the matched peptides (Nesvizhskii and Aebersold, 2005).

The database should contain all of the known proteins from a chosen species together with their sequences. Hence, choosing a comprehensive protein sequence database is essential to allow as many peptides as possible to be matched. There are many different protein sequence databases available. The largest and most complete database is Entrez Protein from the National Center for Biotechnology Information (NCBI), because this database is made up of several other databases. However, since this database is generated automatically without any manual correction, some peptide sequences may not be determined correctly. Furthermore, this database has a high degree of sequence redundancy and can slow down the searching process. For a higher quality of sequence annotation, it is more appropriate to use well-curated databases such as Swiss-Prot from the European Bioinformatics Institute (EBI), or Reference Sequence (RefSeq) from NCBI. If the database search cannot find a match for an observed spectrum, then manual peptide sequencing is possible using a process called *de novo peptide sequencing* (Colinge and Bennett, 2007). Rather than pattern matching the observed spectrum to a theoretical spectrum, it sequences incrementally using the m/z corresponding to the masses of amino acid residues to build the potential sequence of the unknown peptide (Colinge and Bennett, 2007).

1.8 Overview of thesis

The primary purpose of this thesis is to develop a method for the computer generation of optimal designs of two-phase multiplex proteomics experiments. This thesis has three parts. The first part, presented in Chapter 2, describes the method of information decomposition the design of any single- and two-phase experiment, and automates the construction of theoretical ANOVA tables. The second part is the main component of thesis where we describes a computational approach for finding optimal designs for Phase 2 proteomics experiments. Chapter 3 has the Phase 1 experiment arranged in a completely randomised design and Chapter 4 has the Phase 1 experiment arranged in a randomised complete block design (RCBD), or a balanced incomplete block design (BIBD). The third part, presented in Chapter 5, shows how to estimate the variance components and EDF using the restricted maximum likelihood (REML). The comparison of some optimal designs found in Chapters 3 and 4 using the EDF can help clarify the properties of different candidate designs. Chapter 6 summarises and concludes the entire thesis. Furthermore, it presents future directions in the design and analysis of two-phase experiments, in particular, for quantitative high-throughput biotechnology experiments with multiplexing technology. These technologies are currently useful and will remain so for the next decade.

Chapter 2

infoDecompuTE: An R package for constructing theoretical ANOVA tables for two-phase experiments

2.1 Introduction

A primary objective of comparative experiments is to contrast measurements made on experimental units of material (e.g. humans, animals, plants, tissues, cells, etc.) in response to interventions, or *treatments*, to which they are subjected. Many situations arise in practice in which the response variable(s) of interest cannot be measured directly from the experimental units in an experiment (Phase 1). Instead, the experimental units must be further processed in a subsequent experiment (Phase 2) before measurements can be made. Such *two-phase experiments* were introduced by [McIntyre \(1955\)](#) in the context of a study of the effects of four light treatments on the synthesis of tobacco mosaic virus in tobacco leaves. Healthy tobacco plants were inoculated with the virus, and then subjected to different light treatments (Phase 1 experiment). To measure the disease severity, sap was first expressed from the experimental tobacco plants, and then injected into the leaves of specific assay plants (Phase 2 experiment) on which lesions subsequently appeared and were counted.

Efforts have been made to develop a general theory for the design of two-phase experiments ([Brien, 1983](#); [Wood et al., 1988](#); [Brien and Payne, 1999](#); [Jarrett and Ruggiero, 2008](#)). Initial efforts by [Brien \(1983\)](#) yielded a set of rules for deriving the analysis of variance (ANOVA) tables for such designs. This resulted from recognising that experimental factors formed groups, or *tiers*, such that factors from one tier were randomised to factors of another ([Brien, 1983](#)).

Two-phase experiments generally involve three tiers of factors: two tiers of block factors and a tier of treatment factors. Consequently, two-phase experiments fall within the class of *multi-tiered experiments* (Brien, 1983). Tiers 1 and 2 comprise block factors from the Phase 2 and 1 experiments, respectively, while Tier 3 contains treatment factors from the Phase 1 experiment. Design construction comprises a two-step process: (1) the allocation of treatments to experimental units in the Phase 1 experiment, and (2) the allocation of experimental units from the Phase 1 experiment to the experimental or observational units in the Phase 2 experiment. Hence, randomisation generally must be performed twice; once for each allocation. Brien and Bailey (2006) thus named the randomisation procedure for multi-tiered experiments, including two-phase experiments, as *multiple randomisation*. Furthermore, Brien and Bailey (2009, 2010) defined the connection between the type of multiple randomisation associated with an experiment's design and the decomposition of the vector space spanned by the data vector from the experiment. In this Chapter, we refer to the process of separating the total variability in the data vector collected from a two-phase experiment into its constituent components of known sources of variation, and their corresponding degrees of freedom (DF), simply as *information decomposition*.

Jarrett and Ruggiero (2008) conducted a detailed comparative study of the properties of two competing designs – multiple dye-swap (MD) and alternating loop (AL) (Kerr and Churchill, 2001a,b) – for a two-colour microarray experiment at Phase 2, when the Phase 1 experiment involved two treatments arranged in a completely randomised design (CRD) in each case. Using information decomposition to construct the theoretical ANOVA tables of both the MD and AL designs, they demonstrated that the final analysis depended on the designs of both the Phase 1 and 2 experiments. More specifically, while the multiple dye-swap design could be analysed using a simple ANOVA, the alternating loop design required a more involved analysis to test for treatment effects. This was shown to be a consequence of the sources of variation introduced in the Phase 2 experiment interacting with those introduced at Phase 1. Thus, Jarrett and Ruggiero (2008) illustrated the importance of considering the sources of variation introduced at each phase when designing two-phase experiments, and showed that constructing the relevant ANOVA tables can achieve this end.

Construction of complete theoretical ANOVA tables is a laborious manual task, even for relatively small two-phase experiments. The availability of software to automate this task is much needed. There are a number of software packages that address various aspects of this task. For example, the commercial statistical software **GenStat** can be used to decompose known explanatory and structural sources of variation from an experiment into their associated DF in

an ANOVA, but only for generally balanced designs (Payne, 2003). The production of ANOVA tables via this approach requires knowledge of pseudofactors. The **AMTIER** procedure in **GenStat** eliminates the need for pseudofactors by enabling users to specify three separate model formulae: one for each of the block structures in the Phase 1 and 2 experiments and the other for the treatment structure. **AMTIER** performs a decomposition which generates ANOVA tables containing only the names of the sources of variation and their corresponding DF (Brien and Payne, 2006). These tables are, therefore, of limited utility in assessing competing designs since the Treatment and Residual expected mean squares (EMSs) that are needed for this purpose are not generated. While the **GLM** procedures in **JMP** and **SAS**, and the **ANOVA** command in **Minitab** can compute the EMSs, none of these programs offer a direct procedure to do this for two-phase experiments. The **dae** R package (Brien, 2011) performs information decomposition. However, it requires manual intervention by users to construct the projection matrices for (1) generating the strata of the ANOVA table and (2) performing the decomposition of treatment information across these strata. We are unaware of any existing statistical software packages that automate all of the required tasks in a straightforward manner.

We introduce the R package **infoDecompuTE** which we have developed for **information Decomposition under Two-phase Experiments**. **infoDecompuTE** quickly generates complete theoretical ANOVA tables (i.e. both DF and EMSs) for both single- and two-phase experiments, thereby facilitating researchers in their comparisons of the properties of competing designs. It requires minimal intervention by the user, requiring only three pieces of information: 1) a data frame containing the design, 2) the block structure formulae for Phase 1 and/or Phase 2 experiments, and 3) the treatment structure formula. Additionally, the ANOVA table can be constructed by fitting contrasts which allows more flexible analysis and can provide further insight into how the information separates across different strata.

In this Chapter, we present the theoretical concepts and methods underlying **infoDecompuTE**, as well as its use. Section 2.2 explains the information decomposition for a single-phase experiment. Section 2.3 then generalises this to two-phase experiments, and shows how it differs from that of a single-phase experiment. Section 2.4 gives a brief overview about the technology uses for proteomics experiments. Section 2.5 presents an example of a two-phase experiment with the theoretical ANOVA table. Section 2.6 then demonstrates the use of **infoDecompuTE**. Finally, Section 2.7 illustrates the application of **infoDecompuTE** to a two-phase viticulture-sensory evaluation experiment first described by Brien (1992).

2.2 Information decomposition of designed experiments

An experiment's design and the model describing the measurements that are collected from it are intimately connected. More specifically, the model describes the *block structure* (i.e. the relationships between the experimental and observational units), *treatment structure* (i.e. the relationships between the treatments) and the relationships between the treatments and experimental units. Section 2.2.1 considers a general design involving both block and treatment factors whose relationships can be represented mathematically by a linear *mixed-effects model*. This section also shows how the data vector from such an experiment is decomposed into its constituent components based on the experiment's block and treatment structures.

2.2.1 The linear mixed-effects model

In general, consider a designed experiment involving $t = t_1 t_2 \dots t_\nu$ treatments arranged in a block design involving b block factors, where t_i denote the number of levels of the i -th treatment factor, F_i ($i = 1, 2, \dots, \nu$). The block size for the j -th block factor, B_j , is denoted by m_j ($j = 1, 2, \dots, b$). Letting \mathbf{y} be an $n \times 1$ vector of responses, the linear mixed-effects model for the experiment can be written in matrix notation as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (2.1)$$

where $\mathbf{1}$ is an $n \times 1$ vector whose elements are all unity, μ denotes the population mean, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ is an $n \times 1$ random vector of experimental errors. The matrix \mathbf{I}_n denotes the $n \times n$ identity matrix. The treatment parameter vector of length $t = t_1 t_2 \dots t_\nu$ is defined as

$$\boldsymbol{\alpha} = (\alpha_{11\dots 11}, \alpha_{11\dots 12}, \dots, \alpha_{11\dots 1t_\nu}, \dots, \alpha_{11\dots t_{\nu-1}t_\nu}, \dots, \alpha_{t_1 t_2 \dots t_{\nu-1} t_\nu})', \quad (2.2)$$

where $\alpha_{f_1 \dots f_\nu}$ denotes the effect of treatment $f_1 \dots f_\nu$, ($f_i = 1, \dots, t_i; i = 1, \dots, \nu$). Note that $\boldsymbol{\alpha}$ is defined based on the ordering of the treatment combination. The binary $n \times t$ treatment design matrix, \mathbf{X} , defines the allocation of treatments to experimental units. Thus, the row h and column j of \mathbf{X} is 1 if the h -th observation corresponds to an experimental unit assigned the j -th treatment combination, ($h = 1, 2, \dots, n; j = 1, 2, \dots, t$), and is 0 otherwise. The $m \times 1$ vector of block parameters

$$\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_b)'$$

where $m = \sum_{j=1}^b m_j$, and where

$$\mathbf{u}'_j = (u_{j1}, u_{j2}, \dots, u_{jm_j})$$

and $u_{jk} \sim \mathcal{N}(0, \sigma_{B_j}^2)$, ($j = 1, 2, \dots, b; k = 1, 2, \dots, m_j$). The $n \times m$ block design matrix, \mathbf{Z} , in (2.1) can be partitioned into $n \times m_j$ submatrices \mathbf{Z}_j , i.e.

$$\mathbf{Z} = [\mathbf{Z}_1 | \mathbf{Z}_2 | \dots | \mathbf{Z}_b],$$

where \mathbf{Z}_j has (h, k) -th element equal to 1 if the h -th observation corresponds to an experimental unit in the k -th block of B_j , and is zero otherwise, ($h = 1, 2, \dots, n; j = 1, 2, \dots, b; k = 1, 2, \dots, m_j$).

2.2.2 Null ANOVA using projection matrices

Once the block and treatment structures have been defined for the experiment, the first step is to construct the ANOVA for the *null* experiment, i.e. the experiment ignoring treatment effects (Nelder, 1965a). In this case, the total variability of the data is decomposed into its constituent components based on the block structure of the experiment, yielding the *null* ANOVA.

Consider the vector of responses, \mathbf{y} , in (2.1) spanning the n -dimensional Euclidean space

$$\mathbb{R}^n = \mathbb{V}_0 \oplus \mathbb{V}_1 \oplus \dots \oplus \mathbb{V}_q, \quad (2.3)$$

where \mathbb{V}_l is the l -th vector *subspace* of \mathbb{R}^n (i.e. $\mathbb{V}_l \subset \mathbb{R}^n$) and \oplus denotes the vector space addition operator. Thus, the information decomposition of \mathbf{y} is achieved by projecting \mathbf{y} from \mathbb{R}^n space onto each of its constituent $q+1$ vector subspaces. Then the l -th vector subspace, \mathbb{V}_l , corresponds to the l -th *stratum* of the ANOVA. The strata corresponding to the grand mean and intra-block vector subspaces, denoted by \mathbb{V}_0 and \mathbb{V}_q , respectively, are in general not displayed in the ANOVA table.

The variance of \mathbf{y} , in (2.1), is given by

$$\text{var}(\mathbf{y}) = \sigma^2 \mathbf{I}_n + \sum_{j=1}^b \sigma_j^2 \mathbf{Z}_j \mathbf{Z}_j', \quad (2.4)$$

and is expressed in spectral form as

$$\text{var}(\mathbf{y}) = \sum_{l=0}^q \xi_l \mathbf{Q}_l, \quad (2.5)$$

where ξ_l is a linear combination of the variance components, and \mathbf{Q}_l is an $n \times n$ *orthogonal projector* matrix which projects \mathbf{y} from \mathbb{R}^n onto \mathbb{V}_l . The \mathbf{Q}_l matrices are constructed by expressing \mathbf{y} in terms of a *yield identity* (Nelder, 1965a), i.e. an equation defining the partitioning of the overall variation in \mathbf{y} in terms of a set of orthogonal components based on the experimental design's block structure. Consequently, the \mathbf{Q}_l matrices are symmetric, (i.e. $\mathbf{Q}_l' = \mathbf{Q}_l$), ortho-

nal, (i.e. $\mathbf{Q}_l \mathbf{Q}_{l'}' = 0; l \neq l'$), and idempotent (i.e. $\mathbf{Q}_l^2 = \mathbf{Q}_l$) (Hadi, 1996). The remainder of this subsection describes how the orthogonal projector matrices, \mathbf{Q}_l , are defined and connects the block design matrices in (2.4) to the projector matrices in (2.5).

Consider a null experiment comprising m blocks of size k , i.e. each block comprises k experimental units under homogeneous condition. The linear mixed model in (2.1) simplifies to

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}_1 \mathbf{u}_1 + \boldsymbol{\epsilon}, \quad (2.6)$$

so that

$$\text{var}(\mathbf{y}) = \sigma^2 \mathbf{I}_n + \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1'.$$

For this experiment, the null yield identity for observation y_{ij} from the j -th experimental unit in the i -th block is given by

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.}). \quad (2.7)$$

where $\bar{y}_{..}$ denotes the grand mean, and $\bar{y}_{i.}$ denotes the mean of the observations in block i , ($i = 1, \dots, m_1; j = 1, \dots, k$). Writing the observation vector \mathbf{y} , in (2.6), in lexicographical order, i.e. $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_{m_1})'$, where $\mathbf{y}'_i = (y_{i1}, \dots, y_{ik})$, it follows that (2.7) may be written in matrix notation as

$$\mathbf{y} = \sum_{l=0}^2 \mathbf{Q}_l \mathbf{y}, \quad (2.8)$$

where $\mathbf{Q}_0 = \mathbf{K}_n$ is an $n \times n$ averaging matrix with all elements equal to n^{-1} , $\mathbf{Q}_1 = \mathbf{P}_1 - \mathbf{K}_n$ and $\mathbf{Q}_2 = \mathbf{I}_n - \mathbf{P}_1$, so that $\mathbf{Q}_0 + \mathbf{Q}_1 + \mathbf{Q}_2 = \mathbf{I}_n$, and where

$$\mathbf{P}_1 = \mathbf{Z}_1 (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' \quad (2.9)$$

is, by definition, the *projection matrix* of \mathbf{Z}_1 . Thus, pre-multiplying \mathbf{y} by \mathbf{P}_1 projects \mathbf{y} onto the column space of \mathbf{Z}_1 , which is also known as *pivotal sweep* (Brien and Payne, 1999). Pre-multiplying \mathbf{y} by the *orthogonal complement* of \mathbf{P}_1 , namely $\mathbf{I}_n - \mathbf{P}_1$, projects \mathbf{y} onto the space orthogonal to that spanned by the column space of \mathbf{Z}_1 , which is also known as *reanalysis sweeps* (Brien and Payne, 1999). Thus, the projector matrices \mathbf{Q}_0 , \mathbf{Q}_1 and \mathbf{Q}_2 defined in (2.8) project \mathbf{y} from \mathbb{R}^n onto the grand mean, between blocks and within blocks vector subspaces, respectively. It follows from (2.8) and (2.9) that the total SS is given by

$$\mathbf{y}' \mathbf{y} = \mathbf{y}' \left(\sum_{l=0}^2 \mathbf{Q}_l \right) \mathbf{y} \quad (2.10)$$

$$= \mathbf{y}' \mathbf{K}_n \mathbf{y} + \mathbf{y}' (\mathbf{P}_1 - \mathbf{K}_n) \mathbf{y} + \mathbf{y}' (\mathbf{I}_n - \mathbf{P}_1) \mathbf{y}, \quad (2.11)$$

showing its decomposition into the grand mean, between blocks and within blocks SS, respectively.

Following from (2.10), it can be shown that the expected sum of squares (ESS) in the l -th stratum of the null ANOVA is given by

$$\begin{aligned} E(\mathbf{y}'\mathbf{Q}_l\mathbf{y}) &= \text{trace}[\mathbf{Q}_l \text{var}(\mathbf{y})] \\ &= \sigma_1^2 \text{trace}(\mathbf{Z}_1'\mathbf{Q}_l\mathbf{Z}_1) + \sigma^2 \text{trace}(\mathbf{Q}_l), \end{aligned} \quad (2.12)$$

since the trace of the product of the matrices is invariant under any cyclic permutation of them (Searle, 1982). Further, the $\text{trace}(\mathbf{Q}_l)$ yields the total DF in the l -th stratum. It follows, therefore, that the EMS in the l -th stratum given by

$$\text{EMS}_l = \frac{E(\mathbf{y}'\mathbf{Q}_l\mathbf{y})}{\text{trace}(\mathbf{Q}_l)}. \quad (2.13)$$

Thus, from (2.11) – (2.13), we can generate the theoretical null ANOVA shown in Table 2.1.

Table 2.1: Theoretical null ANOVA for an experiment with m_1 blocks of size k .

Source of Variation	DF	SS	EMS
Between Blocks	$m_1 - 1$	$\mathbf{y}'(\mathbf{P}_1 - \mathbf{K}_n)\mathbf{y}$	$\sigma^2 + k\sigma_1^2$
Within Blocks	$m_1(k - 1)$	$\mathbf{y}'(\mathbf{I}_n - \mathbf{P}_1)\mathbf{y}$	σ^2
Grand Mean	1	$\mathbf{y}'\mathbf{K}_n\mathbf{y}$	
Total	n	$\mathbf{y}'\mathbf{y}$	

2.2.3 Computing the treatment SS

In Section 2.2.2 it was shown how, by defining a null yield identity we were able to generate orthogonal projector matrices, \mathbf{Q}_l , enabling block-information decomposition yielding the theoretical null ANOVA. We now consider the process of treatment-information decomposition in an ANOVA when treatments are applied to the experimental units (Nelder, 1965b).

Consider again the null experiment described in section 2.2.2 with m blocks of size k . We now overlay on this a two-factor factorial experiment with treatment factors, F_1 and F_2 at t_1 and t_2 levels, respectively, arranging the treatments in a randomized complete block design. We extend the linear model defined in (2.6) to include a $t \times 1$ vector of treatment parameters $\boldsymbol{\alpha}$, i.e.

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}_1\mathbf{u}_1 + \boldsymbol{\epsilon}, \quad (2.14)$$

where $\boldsymbol{\alpha} = (\alpha_{11}, \dots, \alpha_{1t_2}, \alpha_{21}, \dots, \alpha_{t_2t_2})$. The effect of the treatment combination having factor F_1 at level f_1 and factor F_2 at level f_2 is denoted by $\alpha_{f_1f_2}$, ($f_1 = 1, \dots, t_1; f_2 = 1, \dots, t_2$). The

yield identity for treatments, ignoring blocks, for this factorial experiment is given by

$$\alpha_{f_1 f_2} = \bar{\alpha}_{..} + (\bar{\alpha}_{f_1.} - \bar{\alpha}_{..}) + (\bar{\alpha}_{.f_2} - \bar{\alpha}_{..}) + (\alpha_{f_1 f_2} - \bar{\alpha}_{f_1.} - \bar{\alpha}_{.f_2} + \bar{\alpha}_{..}), \quad (2.15)$$

where $\bar{\alpha}_{..}$ denotes the overall mean, and where $\bar{\alpha}_{f_1.}$ and $\bar{\alpha}_{.f_2}$ are the means averaged over the levels of factor F_1 at level f_1 and Factor F_2 at level f_2 , respectively. Thus, $\bar{\alpha}_{f_1.} - \bar{\alpha}_{..}$ corresponds to the main effect of factor F_1 at level f_1 , $\bar{\alpha}_{.f_2} - \bar{\alpha}_{..}$ corresponds to the main effect of factor F_2 at level f_2 , and $\alpha_{f_1 f_2} + \bar{\alpha}_{f_1.} + \bar{\alpha}_{.f_2} - \bar{\alpha}_{..}$ corresponds to the interaction effect between factors F_1 and F_2 at levels f_1 and f_2 , respectively.

In matrix notation, the yield identity in (2.15) can be expressed as

$$\boldsymbol{\alpha} = \mathbf{C}_{00}\boldsymbol{\alpha} + \mathbf{C}_{10}\boldsymbol{\alpha} + \mathbf{C}_{01}\boldsymbol{\alpha} + \mathbf{C}_{11}\boldsymbol{\alpha}$$

where

$$\mathbf{C}_{00} = \mathbf{K}_{t_1} \otimes \mathbf{K}_{t_2}, \mathbf{C}_{01} = \mathbf{K}_{t_1} \otimes (\mathbf{I}_{t_2} - \mathbf{K}_{t_2}), \mathbf{C}_{10} = (\mathbf{I}_{t_1} - \mathbf{K}_{t_1}) \otimes \mathbf{K}_{t_2}, \mathbf{C}_{11} = (\mathbf{I}_{t_1} - \mathbf{K}_{t_1}) \otimes (\mathbf{I}_{t_2} - \mathbf{K}_{t_2}).$$

The matrix \mathbf{C}_x is the treatment projection matrix for a generalised interaction $x = x_1 x_2$, where $x_i = 1$ if factor F_i is present in the interaction and zero otherwise, given by

$$\mathbf{C}_{x_1 x_2} = \mathbf{C}_{x_1} \otimes \mathbf{C}_{x_2}$$

where

$$\mathbf{C}_{x_i} = \begin{cases} \mathbf{I}_{t_i} - \mathbf{K}_{t_i}, & \text{if } x_i = 1 \\ \mathbf{K}_{t_i}, & \text{if } x_i = 0 \end{cases}. \quad (2.16)$$

More generally, it can be shown that for a v -factor experiment with treatment factor F_i at t_i levels, ($i = 1, \dots, v$), the vector of treatment parameters is given by

$$\boldsymbol{\alpha} = \sum_x \mathbf{C}_x \boldsymbol{\alpha} \quad (2.17)$$

where

$$\mathbf{C}_x = \mathbf{C}_{x_1} \otimes \mathbf{C}_{x_2} \otimes \dots \otimes \mathbf{C}_{x_v},$$

is the treatment projection matrix for a generalised interaction $x = x_1 x_2 \dots x_v$ and where \mathbf{C}_{x_i} is given by (2.16). The generalised interaction is denoted by $F^x = F_1^{x_1} F_2^{x_2} \dots F_v^{x_v}$ ($x = x_1 x_2 \dots x_v$; $x_i = 1, 2$; $i = 1, 2, \dots, v$) (John and Williams, 1995).

The next step is to estimate the treatment main and interaction effects in $\boldsymbol{\alpha}$ from the reduced normal equations, which are derived by: (1) minimizing the error sum of squares $\boldsymbol{\epsilon}'\boldsymbol{\epsilon}$ in (2.6) with respect to the vector of treatment parameters, $\boldsymbol{\alpha}$, and (2) eliminating from the resultant

normal equations the mean and block parameters (John and Williams, 1995). Now consider the least squares estimator of $\boldsymbol{\alpha}$ decomposed in stratum l , denoted by $\hat{\boldsymbol{\alpha}}_l$, the reduced normal equations corresponding to the l -th stratum of the ANOVA are given by

$$\mathbf{A}_l \hat{\boldsymbol{\alpha}}_l = \mathbf{q}_l, \quad (2.18)$$

where

$$\mathbf{A}_l = \mathbf{L}_x \mathbf{Q}_l \mathbf{L}_x' \text{ and} \quad (2.19)$$

$$\mathbf{q}_l = \mathbf{L}_x \mathbf{Q}_l \mathbf{y}, \quad (2.20)$$

are the symmetrical treatment information matrix and the vector of the adjusted treatment totals, respectively. The matrix $\mathbf{L}_x = \mathbf{C}_x \mathbf{X}'$ is to simplify the notation in (2.19) and (2.20).

A solution to the normal equations, obtained by substituting (2.19) and (2.20) into (2.18) and solving for $\hat{\boldsymbol{\alpha}}_l$ yields the vector of estimated treatment effects

$$\hat{\boldsymbol{\alpha}}_l = \mathbf{A}_l^- \mathbf{q}_l = (\mathbf{L}_x \mathbf{Q}_l \mathbf{L}_x')^- \mathbf{L}_x \mathbf{Q}_l \mathbf{y}, \quad (2.21)$$

where \mathbf{A}_l^- is a generalised inverse of \mathbf{A}_l satisfying $\mathbf{A}_l \mathbf{A}_l^- \mathbf{A}_l = \mathbf{A}_l$. Thus, the treatment SS corresponding to the generalised interaction F^x in the l -th stratum is given by

$$\mathbf{q}_l' \mathbf{A}_l^- \mathbf{q}_l = \mathbf{y}' \mathbf{Q}_{x(l)} \mathbf{y}. \quad (2.22)$$

where $\mathbf{Q}_{x(l)} = \mathbf{Q}_l \mathbf{L}_x' \mathbf{A}_l^- \mathbf{L}_x \mathbf{Q}_l$ and denotes the projection matrix which projects \mathbf{y} onto the treatment vector subspace spanned by F^x in the l -th stratum. The treatment ESS is then given by

$$\mathbb{E}(\mathbf{y}' \mathbf{Q}_{x(l)} \mathbf{y}) = \text{trace}[\mathbf{Q}_{x(l)} \text{var}(\mathbf{y})] + \boldsymbol{\alpha}' \left(\sum_x \mathbf{L}_x \mathbf{Q}_{x(l)} \mathbf{L}_x' \right) \boldsymbol{\alpha}, \quad (2.23)$$

where $\text{trace}[\mathbf{Q}_{x(l)} \text{var}(\mathbf{y})]$ contains the coefficients of the variance components, denoted by σ^2 , and $\boldsymbol{\alpha}' \left(\sum_x \mathbf{L}_x \mathbf{Q}_{x(l)} \mathbf{L}_x' \right) \boldsymbol{\alpha}$ contains the fixed effects components, denoted by θ^2 , of the ESS in the l th stratum.

The quadratic form $\boldsymbol{\alpha}' \left(\sum_x \mathbf{L}_x \mathbf{Q}_{x(l)} \mathbf{L}_x' \right) \boldsymbol{\alpha}$ in (2.23) can be re-written as

$$\boldsymbol{\alpha}' \left(\sum_x \mathbf{L}_x \mathbf{Q}_{x(l)} \mathbf{L}_x' \right) \boldsymbol{\alpha} = \sum_{g=1}^{t_x-1} \lambda_g (\mathbf{p}_g' \boldsymbol{\alpha})^2,$$

where λ_g and \mathbf{p}_g denote the g -th eigenvalue and eigenvector of the information matrix in stratum l , denoted by $\mathbf{L}_x \mathbf{Q}_{x(l)} \mathbf{L}_x$, ($g = 1, 2, \dots, t_x - 1$). Thus, $\mathbf{p}_g' \boldsymbol{\alpha}$ is also known as g -th basic treatment contrast, which provides an orthogonal decomposition of the $t_x - 1$ DF for treatment effects into

single degree of freedom components (John and Williams, 1995). The fixed effects components of the EMS can then be expressed as

$$\frac{\boldsymbol{\alpha}'\mathbf{L}_x\mathbf{Q}_{x(l)}\mathbf{L}_x'\boldsymbol{\alpha}}{t_x - 1} = \frac{\sum_{g=1}^{t_x-1} \lambda_g (\mathbf{p}_g'\boldsymbol{\alpha})^2}{t_x - 1} = \theta_x \frac{\sum_{g=1}^{t_x-1} \lambda_g}{t_x - 1}, \quad (2.24)$$

where $\theta_x = \frac{\sum_{g=1}^{t_x-1} (\mathbf{p}_g'\boldsymbol{\alpha})^2}{t_x - 1}$, which denotes the fixed effect parameter for the generalised interaction F^x .

If treatment allocation is balanced, i.e. the treatment is equally replicated, the eigenvalues should be identical, i.e. all $\lambda_g = \lambda$, then from (2.24)

$$\theta_x \frac{\sum_{g=1}^{t_x-1} \lambda_g}{t_x - 1} = \lambda \theta_x. \quad (2.25)$$

Thus, the eigenvalue is the same as the coefficient of the fixed effect parameter θ_x .

The proportion of treatment information across different strata can be quantified using the *efficiency factor* (Yates, 1936). The *canonical efficiency factors*, denoted by e_g for the g -th basic treatment contrast, are calculated by dividing the eigenvalue by the treatment replication, r , i.e.

$$e_g = \frac{\lambda_g}{r}. \quad (2.26)$$

If every estimable treatment contrast is fully efficient, then the canonical efficiency factors all equal one; hence, the eigenvalue and the coefficient of θ_x must be the same as the treatment replication.

If the treatment allocation is not balanced, then the eigenvalues are not identical; so each coefficient needs to be presented separately for the fixed effect parameter of each basic treatment contrast, i.e.

$$\lambda_1 \theta_{x(1)} + \lambda_2 \theta_{x(2)} + \dots + \lambda_{t_x-1} \theta_{x(t_x-1)}.$$

Thus, the canonical efficiency factors are obtained as the eigenvalues of the information matrix for a block design with unequally replicated treatments, i.e.

$$\mathbf{r}^{-\delta/2} \mathbf{L}_x \mathbf{Q}_{x(l)} \mathbf{L}_x' \mathbf{r}^{-\delta/2} \quad (2.27)$$

where $\mathbf{r}^{-\delta/2}$ is a diagonal matrix with i -th diagonal element equal to $r_i^{1/2}$ and r_i is the replication of i -th treatment combination in $\boldsymbol{\alpha}$ (John and Williams, 1995). Since the canonical efficiency factors are not identical, the proportion of treatment information for the intra-block analysis can be computed, namely the *average efficiency factor* (Yates, 1936). The average proportion of the treatment information associated with generalised interaction F^x in stratum l , denoted

by E_x , is given by

$$E_x = \frac{t_x - 1}{\sum_{g=1}^{t_x-1} e_g^{-1}}, \quad (2.28)$$

which is the harmonic mean of the canonical efficiency factors (John and Williams, 1995).

2.2.4 Computing the residual SS

The residual SS in stratum l , denote by RSS_l , is computed by subtracting the sum over all the SS of all generalised interactions F^x , denote by $\mathbf{y}' \left(\sum_x \mathbf{Q}_{x(l)} \right) \mathbf{y}$, from the total SS, denote by $\mathbf{y}' \mathbf{Q}_l \mathbf{y}$, i.e.

$$RSS_l = \mathbf{y}' \mathbf{Q}_l \mathbf{y} - \mathbf{y}' \left(\sum_x \mathbf{Q}_{x(l)} \right) \mathbf{y} = \mathbf{y}' \left(\mathbf{Q}_l - \sum_x \mathbf{Q}_{x(l)} \right) \mathbf{y}. \quad (2.29)$$

In summary, this section described the information decomposition of a single-phase experiment that involved basic decomposition steps: adjusting for the grand mean, defining the strata based on the block structures and computing treatment SS based on the treatment structure within each defined strata. The next section extends the decomposition method to the two-phase experiment.

2.3 Information decomposition for two-phase experiments

Information decomposition of two-phase experiments involves an additional decomposition step, namely that of the block information from the Phase 2 experiment. Recall that in a single-phase experiment, the block-information decomposition process begins with the construction of the null ANOVA (see Section 2.2). In a single-phase experiment, this is straightforward as there is a single tier for block factors. In a two-phase experiment, however, there are two tiers of block factors, and the decomposition begins with the strata corresponding to the block structure in the Phase 2 experiment, followed by the decomposition into the strata corresponding to the Phase 1 experiment block structure.

The allocation of experimental units from the Phase 1 experiment to blocks in the Phase 2 experiment often results in the block effects from the two phases interacting with one another in such a way that they are non-orthogonal. When this happens, the Phase 1 experiment's block information is dispersed across multiple strata of the ANOVA arising from the Phase 2 experiment's strata (Wood et al., 1988), just as treatment effects in a balanced incomplete block design (BIBD) are dispersed across strata in a single-phase experiment. Consequently, the procedure for the Phase 1 block-information decomposition follows the method described in Section 2.2.3 by *regarding the Phase 1 block factors just as we would treatment factors*.

2.3.1 The linear mixed-effects model

Consider a two-phase experiment involving $t = t_1 t_2 \dots t_\nu$ treatments arranged in a block design with overall block size of $m = m_1 m_2 \dots m_b$ for the Phase 1 experiment, where t_i denotes the number of levels of the i -th treatment factor, F_i , and m_j is the block size for the j -th Phase 1 block factor, B_j ($i = 1, \dots, \nu; j = 1, \dots, b$). Furthermore, the experimental units of the Phase 1 experiment are further arranged in another block design involving p block factors in the Phase 2 experiment. The block size of the k -th Phase 2 block factor, H_r is denoted by s_r ($r = 1, \dots, p$). The linear mixed-effects model for a two-phase experiment can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}^*\mathbf{u}^* + \mathbf{W}\mathbf{v} + \boldsymbol{\epsilon}, \quad (2.30)$$

where the vector of fixed treatment effects parameters, $\boldsymbol{\alpha}$, and its corresponding treatment design matrix, \mathbf{X} , are defined as in (2.1). Since the Phase 1 block factors are regarded as treatment factors, the $m = m_1 m_2 \dots m_b$ Phase 1 block parameter \mathbf{u}^* is defined with the same structure as the treatment effects parameters, $\boldsymbol{\alpha}$, in (2.2), i.e.

$$\mathbf{u}^* = (u_{11\dots 11}, u_{11\dots 12}, \dots, u_{11\dots 1m_j}, \dots, u_{11\dots m_{b-1}m_b}, \dots, u_{m_1 m_2 \dots m_{b-1} m_b})', \quad (2.31)$$

where $u_{B_1 B_2 \dots B_b} \sim N(0, \sigma_{B_1 B_2 \dots B_b}^2)$ denotes the random effect of block B_b within the combination of blocks $B_1 B_2 \dots B_{b-1}$ in the Phase 1 experiment. The Phase 1 block parameter \mathbf{u}^* is defined based on the ordering of the combination of all block factors in the Phase 1 experiment, which is associated with the effects of plots within all the blocks of the Phase 1 experiments. In addition, the structure of the Phase 1 block design matrix, \mathbf{Z}^* , is a binary $n \times m$ described the allocation of the allocation of the Phase 1 experimental units, which is also defined differently to the block design matrix as in (2.1). The new layout of the information from the Phase 2 experiment consists of $m \times 1$ vector of Phase 2 block parameters

$$\mathbf{v} = (\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_p)'$$

where $s = \sum_{r=1}^p s_r$, and where

$$\mathbf{v}'_r = (v_{r1}, v_{r2}, \dots, v_{rs_r})$$

and $v_{ro} \sim \mathcal{N}(0, \sigma_{H_r}^2)$, ($r = 1, 2, \dots, p; o = 1, 2, \dots, s_r$). The $n \times s$ block design matrix, \mathbf{W} , in (2.1) can be partitioned into $n \times s_r$ sub-matrices \mathbf{W}_r , i.e.

$$\mathbf{W} = [\mathbf{W}_1 | \mathbf{W}_2 | \dots | \mathbf{W}_p],$$

where \mathbf{W}_r has (h, o) -th element equals to 1 if the h -th observation corresponds to an experimental unit in the o -th block of H_j , and is zero otherwise, ($h = 1, 2, \dots, n; r = 1, 2, \dots, p; o = 1, 2, \dots, s_k$). Table 2.2 shows the notations of design matrices and parameters of the treatment and block factors between the single-phase and two-phase experiments.

Table 2.2: Notations of the treatment and block factors between single and two-phase experiments.

Factors		Single-phase	Two-phase
Treatment	Notation	F_i	F_i
	Index	$i = 1, 2, \dots, \nu$	$i = 1, 2, \dots, \nu$
	Length	$t = t_1 t_2 \dots t_v$	$t = t_1 t_2 \dots t_v$
	Parameters	$\boldsymbol{\alpha} = (\alpha_{11\dots 1}, \dots, \alpha_{t_1 t_2 \dots t_v})'$	$\boldsymbol{\alpha} = (\alpha_{11\dots 1}, \dots, \alpha_{t_1 t_2 \dots t_v})'$
	Design matrices	\mathbf{X}	\mathbf{X}
(Phase 1) Block	Notation	B_j	B_j
	Index	$j = 1, 2, \dots, b$	$j = 1, 2, \dots, b$
	Length	$\sum_{j=1}^b m_j$	$m = m_1 m_2 \dots m_b$
	Parameters	$\mathbf{u} = (\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_b)'$ $\mathbf{u}'_j = (u_{j1}, u_{j2}, \dots, u_{jm_j})'$	$\mathbf{u}^* = (u_{11\dots 1}, \dots, u_{m_1 m_2 \dots m_b})'$
	Design matrices	$\mathbf{Z} = [\mathbf{Z}_1 \mathbf{Z}_2 \dots \mathbf{Z}_b]$	\mathbf{Z}^*
Phase 2 Block	Notation		H_r
	Index		$r = 1, 2, \dots, p$
	Length		$\sum_{r=1}^p s_r$
	Parameters		$\mathbf{v} = (\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_p)'$ $\mathbf{v}'_r = (v_{r1}, v_{r2}, \dots, v_{rs_r})$
	Design matrices		$\mathbf{W} = [\mathbf{W}_1 \mathbf{W}_2 \dots \mathbf{W}_p]$

2.3.2 Null ANOVA of Phase 2 block structure

The null ANOVA resulting from the decomposition of the Phase 2 block information is computed exactly as per Section 2.2.2. The first step is to construct a list of $n \times n$ orthogonal projector matrices which project \mathbf{y} onto each stratum, based on the Phase 2 block structure, where the orthogonal projector matrix for l -th stratum is denoted by \mathbf{Q}_l .

2.3.3 Null ANOVA of Phase 1 block structure

Once the orthogonal projector matrices for each stratum are defined based on the Phase 2 block structure, the next step is to construct the ANOVA of the null experiment by decomposing the Phase 1 block information within each of the l -th stratum. Consider a Phase 1 experiment arranged in a randomised complete block design with m_1 blocks of size m_2 , where the $m_1 m_2 \times 1$

vector \mathbf{u}^* corresponds to the random effects of plots within blocks as

$$(u_{11}, u_{12}, \dots, u_{m_1 m_2})'. \quad (2.32)$$

where $u_{m_1 m_2}$ denotes the effect from plot B_2 within block B_1 , ($m_j = 1, \dots, s_j; j = 1, 2$). The yield identity of $u_{h_1 h_2}$ can be written as,

$$u_{m_1 m_2} = \bar{u}_{..} + (\bar{u}_{m_1.} - \bar{u}_{..}) + (u_{m_1 m_2} - \bar{u}_{m_1.}), \quad (2.33)$$

where $\bar{u}_{..}$ denotes the mean overall the observations from all plots in all blocks, $\bar{u}_{m_1.} - \bar{u}_{..}$ denotes the effect from block B_1 at level m_1 , and $u_{m_1 m_2} - \bar{u}_{m_1.}$ denotes the effect from plot B_2 at level m_2 within block B_1 at level m_1 .

In matrix notation, the yield identity in (2.33) can be expressed as

$$\mathbf{u}^* = \mathbf{C}_{00}\mathbf{u}^* + \mathbf{C}_{10}\mathbf{u}^* + \mathbf{C}_{21}\mathbf{u}^*$$

where

$$\mathbf{C}_{00} = \mathbf{K}_{b_1} \otimes \mathbf{K}_{b_2}, \mathbf{C}_{10} = (\mathbf{I}_{b_1} - \mathbf{K}_{b_1}) \otimes \mathbf{K}_{b_2}, \mathbf{C}_{21} = \mathbf{I}_{b_1} \otimes (\mathbf{I}_{b_2} - \mathbf{K}_{b_2}).$$

The matrix \mathbf{C}_z is the block projection matrix for a generalised interaction $z = z_1 z_2$, given by

$$\mathbf{C}_z = \mathbf{C}_{z_1} \otimes \mathbf{C}_{z_2}$$

where

$$\mathbf{C}_{z_j} = \begin{cases} \mathbf{K}_{m_j}, & \text{if } z_j = 0 \\ \mathbf{I}_{m_j} - \mathbf{K}_{m_j}, & \text{if } z_j = 1 \\ \mathbf{I}_{m_j}, & \text{if } z_j = 2. \end{cases} \quad (2.34)$$

More generally, for a Phase 1 experiment with b block factors with B_j at m_j levels, ($j = 1, \dots, b$), the vector of Phase 1 block parameters then is given by

$$\mathbf{u}^* = \sum_z \mathbf{C}_z \mathbf{u}^* \quad (2.35)$$

where

$$\mathbf{C}_z = \mathbf{C}_{z_1} \otimes \mathbf{C}_{z_2} \otimes \dots \otimes \mathbf{C}_{z_b} = \bigotimes_{j=1}^b \mathbf{C}_{z_j}$$

is the block projection matrix for a generalised interaction $z = (z_1 z_2 \dots z_b)$ and where \mathbf{C}_{z_i} is given by (2.34).

The procedure for the Phase 1 block-information decomposition follows the method described in Section 2.2.3 by regarding the Phase 1 block factors just as we would treatment factors. Given

that $\hat{\mathbf{u}}_l^*$ denotes the least square estimator of \mathbf{u}^* in the l -th stratum from the Phase 2 experiment, reduced normal equations for the decomposition the Phase 1 block structure in the l -th stratum of the ANOVA is given by

$$\mathbf{A}_l \hat{\mathbf{u}}_l^* = \mathbf{q}_l, \quad (2.36)$$

where

$$\mathbf{A}_l = \mathbf{L}_z \mathbf{Q}_l \mathbf{L}_z', \quad (2.37)$$

$$\mathbf{q}_l = \mathbf{L}_z \mathbf{Q}_l \mathbf{y}, \quad (2.38)$$

are the symmetrical Phase 1 block information matrix and vector of adjusted Phase 1 block totals, respectively. The matrix $\mathbf{L}_z = \mathbf{C}_z \mathbf{Z}'$ is again used to simplify the notation.

Substituting (2.37) and (2.38) into (2.36), solving for $\hat{\mathbf{u}}_l$ yields the vectors of estimated block effects

$$\hat{\mathbf{u}}_l^* = \mathbf{A}_l^- \mathbf{q}_l.$$

Thus, the Phase 1 block SS for B_j in the l -th stratum of the Phase 2 experiment is given by

$$\mathbf{q}_l' \mathbf{A}_l^- \mathbf{q}_l = \mathbf{y}' \mathbf{Q}_l \mathbf{L}_z' \mathbf{A}_l^- \mathbf{L}_z \mathbf{Q}_l \mathbf{y}. \quad (2.39)$$

The residual Phase 1 block SS in the l -th stratum is then derived by subtraction which gives

$$RSS_l = \mathbf{y}' \mathbf{Q}_l \left(\mathbf{I} - \sum_z \mathbf{L}_z' \mathbf{A}_l^- \mathbf{L}_z \right) \mathbf{Q}_l \mathbf{y}.$$

In general, suppose the matrix \mathbf{C}_z in matrix \mathbf{L}_z represents the Phase 1 block projection matrix for the j -th Phase 1 block factor in l -th stratum of a two-phase experiment, then the orthogonal projector that projects the data vector, \mathbf{y} , onto the vector subspace j of the Phase 1 experiment within the vector subspace l of the Phase 2 experiment is given by

$$\mathbf{Q}_{j(l)} = \mathbf{Q}_l \mathbf{L}_z' \mathbf{A}_l^- \mathbf{L}_z \mathbf{Q}_l.$$

This orthogonal projector $\mathbf{Q}_{j(l)}$ can then be used to compute the Treatment SS in the following subsection.

2.3.4 Computing the treatment SS

The information decomposition of the treatment structure for the two-phase experiment is computed exactly as per Section 2.2.3, apart from replacing the orthogonal projector matrix from \mathbf{Q}_l by $\mathbf{Q}_{j(l)}$. This is because the overall treatment information is now decomposed in the j -th

stratum of the Phase 1 experiment within the l -th stratum of the Phase 2 experiment in the ANOVA.

This section summarised the information decomposition method for the two-phase experiments. The main difference from the single-phase experiments is the additional decomposition procedure of the Phase 1 block factors to the Phase 2 block factors. Furthermore, since the Phase 1 block factor/s can be confounded with multiple Phase 2 block factors, this additional decomposition procedure must be performed in the same way as the decomposition of the treatment information to block structure for the single-phase experiment. The vector of the block parameters and design matrix of the linear model for the Phase 1 experiment, thus are defined in the same way as the treatment parameters and design matrix of the linear model for the single-phase experiment.

2.4 Application to quantitative proteomics experiments

Quantitative proteomics uses analytical chemistry techniques to quantify the abundances of the complement of proteins in a biological sample at a cross-section of space and time. Many quantitative proteomics experiments have as their primary objective the identification of proteins that are differentially abundant between different experimental conditions or treatments. Such studies are intrinsically two-phase because protein identification and abundance cannot be measured directly from the experimental units to which the treatments are applied (Phase 1); rather, measurements are made in a subsequent laboratory-based experiment (Phase 2) which itself introduces additional sources of variation to those from the earlier phase.

Once the target cells or tissues are harvested, each sample is independently processed, including steps for reduction, alkylation, total protein quantification and enzymatic digestion (usually with trypsin) into many smaller peptide fragments (Ross et al., 2004). The peptide fragments are then separated using the different properties (Washburn et al., 2001). The first separation is by charge, using *strong cation exchange chromatography* (SCX), followed by hydrophobicity, using *reversed phase liquid chromatography* (RPLC). The third dimension of separation is by mass and is carried out by *mass spectrometry* (MS). These series of separations, namely *Multi-dimensional Protein Identification Technology* (MudPIT), reduces sample complexity and allow more accurate measurement of protein abundance. Hence, each MudPIT *run* is comprised of these three steps of separation. Since the measurement of abundances was performed on peptides, the corresponding proteins need to be identified using an appropriate bioinformatics software searched from protein database. Examples of such software are MascotTM 2.6 (Matrix

Science), SEQUEST[®] (Sadygov et al., 2004) and ProteinPilot[™] 5.0 (Sciex).

However, comparison of protein abundances between samples is difficult due to the variability between different MudPIT experiments. This limitation has been resolved by the introduction of *isobaric Tags for Relative and Absolute Quantitation* (iTRAQ[™]). The iTRAQ[™] labelling chemistry works by binding the tags to each peptide which enables the simultaneous analysis of up to eight distinct samples within a single MudPIT experiment (Ross et al., 2004; Choe et al., 2007).

2.5 Example of two-phase proteomics experiment

This section presents an example of a two-phase proteomics experiment when the Phase 1 experiment consists of eight animals randomly assigned to one of two groups in which four are injected with saline (Healthy) and the remaining four are injected with a disease-inducing drug (Disease). The Phase 2 experiment consists of four MudPIT runs and four iTRAQ[™] tags (i.e. four differentially labelled proteomic samples per run). The main objective of this experiment is to compare the protein abundances between healthy and diseased animals. Additionally, we will also show how the theoretical ANOVA tables are derived for the Phase 1 and 2 experiments.

2.5.1 Phase 1 experiment

The Phase 1 experiment is arranged in a completely randomised design (CRD) with eight animals, when four animals are randomly assigned to each of the healthy and diseased groups as shown in Table 2.3. The four animals in each treatment group are used to assess biological variation and are known as *biological replicates*. Thus, the Phase 1 experiment consists of a block structure of just Animal factor and a treatment structure of Disease status factor.

Table 2.3: Design of Phase 1 experiment showing the assignment of the eight animals, labelled A - H, to disease status' group.

Healthy	A	C	E	G
Diseased	B	D	F	H

Let y_{ij} denote the log abundance of a given protein in animal j under disease status i . Then, the linear model of the Phase 1 design is given by

$$y_{ij} = \mu + \tau_i + a_j, \quad (2.40)$$

where μ denotes the grand mean, τ_i denotes the fixed effect of disease status i , and $a_j \sim \mathcal{N}(0, \sigma_a^2)$

denotes the random effect of animal j , ($i = \text{healthy, diseased}$; $j = A, \dots, H$). Since the treatments are directly applied to the animals and no measurements are made in the Phase 1 experiment, there is no measurement error in linear model (2.40).

Table 2.4: Theoretical ANOVA showing the decomposition of DF and EMS associated with each source of variation.

Source of Variation	DF	EMS
Between Animals		
Disease status	1	$\sigma_a^2 + 4\theta_\tau$
Residual	6	σ_a^2
Total	7	

Using the information decomposition method described in Section 2.2, the theoretical ANOVA table for the Phase 1 experiment can be generated as shown in Table 2.4. The theoretical ANOVA table contains the DF and EMS of each source of variation. Since no measurements were made in the Phase 1 experiment, the Between Animals stratum captured the overall analysis. The DF associated with the Between Animals stratum is seven, because there is a total of eight animals. The Between Animals stratum is decomposed to 1 and 6 DF for the EMS of Disease status and Residual, respectively. Furthermore, Disease status EMS is $\sigma_a^2 + 4\theta_\tau$, where σ_a^2 denotes the variation between animals and θ_τ is the fixed effects component for Disease status. The coefficient of θ_τ is four, because each group of Disease status is replicated four times. Finally, the Residual EMS consists of only the variation between animals, denoted by σ_a^2 .

2.5.2 Phase 2 experiment

In practice, protein abundances cannot be measured directly from the animals and so Table 2.4 cannot be used to test for the Disease status effects. Obtaining measurements of protein abundances requires tissue to be harvested from the target organ in each animal. Each tissue sample requires laboratory processing for the proteins to be extracted and for their subsequent proteomic analysis in the Phase 2 MudPIT-iTRAQTM experiment. The aim is to assign samples in such a way that both Disease status and Animal effects are orthogonal to MudPIT Run and iTRAQTM Tag effects, which allows us to estimate Disease status effects in the Phase 2 experiment as precisely as we did in the Phase 1 experiment. Thus, the Phase 2 experiment consists of Phase 2 block structure of MudPIT Run factor, Phase 1 block structure of Animal factor, and treatment structure of Tag and Disease status factors.

Given there are eight animals from the Phase 1 experiment, two tissue samples are harvested from the target organ of each animal and independently processed in the laboratory. These

two tissue samples are referred to as *technical replicates*, because we have shown the technical variation is largely due to series of the laboratory processes which includes pipetting error, differences in digestion efficiency and recovery of proteins during database searching (Chang, 2008). There is thus a total of 16 samples to be measured in the Phase 2 MudPIT-iTRAQ™ experiment.

If a four-plex labelling system is used, then four runs are required. However, since the run size is four, it is not possible to allocate samples from each of the eight animals into a single run. The next best option is to divide the runs into two pairs: runs 1 and 2 in the first and runs 3 and 4 in the second. Then the samples from all eight animals can be allocated to each of these two pairs of runs. The next step is to assign the Disease status groups. Since there are two Disease status groups, the four runs and four tags arrangement is split into four 2-by-2 arrays and the disease status groups are then assigned to each array with the Latin square arrangement. The 16 samples from 8 animals are labelled A - H, with Animals A, C, E and G assigned to the Healthy group and B, D, F and H assigned to the Diseased group. This allocation of animals to runs and tags for the Phase 2 design is shown in Table 2.5.

Table 2.5: Design of Phase 2 experiment showing allocation of sub-samples from animals and disease statuses to runs and tags. The letters denote the animal IDs.

Run	Tag			
	114	115	116	117
1	<i>A:Healthy</i>	<i>B:Diseased</i>	<i>C:Healthy</i>	<i>D:Diseased</i>
2	<i>F:Diseased</i>	<i>E:Healthy</i>	<i>H:Diseased</i>	<i>G:Healthy</i>
3	<i>C:Healthy</i>	<i>D:Diseased</i>	<i>A:Healthy</i>	<i>B:Diseased</i>
4	<i>H:Diseased</i>	<i>G:Healthy</i>	<i>F:Diseased</i>	<i>E:Healthy</i>

Let y_{kls}^{ij} denote the log abundance of *one given protein* in sub-sample s from animal j under disease status i and measured from the l -th MudPIT run with iTRAQ™ tag k . Note that the MudPIT-iTRAQ™ experiment measures the protein abundance values of all the proteins in the collected sample. The subscript and superscript indices of the observation, y_{kls}^{ij} , correspond to the indices from the Phase 2 and Phase 1 experiments, respectively. The linear model of the Phase 2 experiment is then given by

$$y_{kls}^{ij} = \mu + \tau_i + a_j + \gamma_k + r_l + \epsilon_{kls}^{ij}, \quad (2.41)$$

where $r_l \sim \mathcal{N}(0, \sigma_r^2)$ denotes the random effects from MudPIT run l , γ_k denotes the fixed effects of tag k , $\epsilon_{kls}^{ij} \sim \mathcal{N}(0, \sigma^2)$ denotes the effects from sub-sample s in animal j from run l under disease status j and tag k , ($k = 114, \dots, 117$, $l = 1, \dots, 4$, $s = 1, 2$). Additionally, ϵ_{kls}^{ij} corresponds to the experimental error, because the sample is the smallest unit of the Phase 2 experiment. The

remaining terms are defined as in (2.40). Furthermore, the Disease status and Tag effects are assumed not to interact.

Using the information decomposition method described in Section 2.3, the theoretical ANOVA table of the Phase 2 experiment can be generated and shown in Table 2.6. Since there are 16 observations, there are total of 15 DF to be decomposed. These 15 DF are broken down into 3 and 12 DF for the Between and Within Runs strata, respectively. The Between Runs stratum is further separated into 1 and 2 DF corresponding to the Between Animals and Residual (Within Animals Between Runs) strata, respectively. The Within Runs stratum is decomposed into two sets of 4 DF for the Between Animals and Between Samples Within Animals strata, respectively. The Between Animals Within Runs stratum is further decomposed to the EMS of the Disease status (1 DF), Tag (1 DF) and Residual (4 DF). Notably, all of the Disease status effect is estimated in the Between Animals Within Runs stratum. Finally, the Between Sample Within Animals Within Runs stratum is decomposed to the EMS of Tag (2 DF) and Residual (4 DF).

Comparing this ANOVA table (see Table 2.6) to that of the Phase 1 experiment (see Table 2.4), the animals originally have 7 DF; however, 1 DF is now in the Between Runs stratum and another 1 DF is confounded with Tag effects. A valid F-test for the Disease status effect can still be conducted for the Between Animals Within Runs stratum, but the DF associated with the Residual EMS is reduced from 6 DF to 4 DF, leading to a reduction in the precision in which the Disease status effects are estimated.

Table 2.6: Phase 2 ANOVA of design in Table 2.5 with the coefficients of variance components of EMS.

Source of Variation	DF	EMS
Between Runs		
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$
Residual	2	$\sigma^2 + 4\sigma_r^2$
Within runs		
Between Animals		
Tag	1	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma$
Disease status	1	$\sigma^2 + 2\sigma_a^2 + 8\theta_\tau$
Residual	4	$\sigma^2 + 2\sigma_a^2$
Between Samples Within Animals		
Tag	2	$\sigma^2 + 4\theta_\gamma$
Residual	4	σ^2
Total	15	

2.6 An R package: InfoDecompuTE

InfoDecompuTE is written in the R programming language (R Core Team, 2017). It automates the information decomposition procedures described in Sections 2.2 and 2.3, producing theoretical ANOVA tables such those shown in Tables 2.4 and 2.6. **InfoDecompuTE** comprises a suite of functions (see Appendix A), that performs different stages of the information decomposition and which are used by the two primary functions: named `summaryAovOnePhase` and `summaryAovTwoPhase`, to generate the theoretical ANOVA tables for single-and two-phase experiments. The remainder of this Chapter consider these two main functions.

2.6.1 Installation instructions

InfoDecompuTE requires at least version 3.3.0 of the R statistical programming environment which is available from the Comprehensive R Archive Network at <http://CRAN.R-project.org/> (R Core Team, 2017). The system requirements for this package depend on the size of the experimental design, for example, number of blocks and block sizes, number of treatment factors and number of replicates of each treatment. This is because the number of factors and observations is directly related to the dimensions of the matrices which can affect the speed of computation.

Providing that the user has an internet connection, **infoDecompuTE** can be installed and initiated by typing the following two commands in a new R session:

```
> install.packages("infoDecompuTE")
> library("infoDecompuTE")
```

The package can also be downloaded from <http://cran.r-project.org/web/packages/infoDecompuTE/index.html>.

2.6.2 Functions

This section explains the arguments for the two main functions in the **infoDecompuTE** package.

These two functions and their arguments are:

```
summaryAovOnePhase(design.df, blk.str, trt.str, var.comp = NA,
                    trt.contr = NA, table.legend = FALSE,
                    response = NA, latex = FALSE, fixed.names = NA,
                    decimal = FALSE, digits = 2, list.sep = TRUE)

summaryAovTwoPhase(design.df, blk.str1, blk.str2, trt.str, var.comp = NA,
                    blk.contr = NA, trt.contr = NA, table.legend = FALSE,
                    response = NA, latex = FALSE, fixed.names = NA,
                    decimal = FALSE, digits = 2, list.sep = TRUE)
```

where the description of each argument is shown in Table 2.7. The rest of this section first shows how to generate the theoretical ANOVA tables from the two-phase experiment example described in Section 2.5 and then explains the usage of each argument which provides additional functionalities.

Table 2.7: Description of `summaryAovOnePhase` and `summaryAovTwoPhase` functions' arguments.

Argument	Description
<code>design.df</code>	data frame containing the experimental design.
<code>blk.str</code>	structure formula for the block factors.
<code>blk.str1</code>	structure formula for the block factors of Phase 1 experiment.
<code>blk.str2</code>	structure formula for the block factors of Phase 2 experiment.
<code>trt.str</code>	structure formula for the treatment factors.
<code>var.comp</code>	specifies the variance components to be shown in the ANOVA table.
<code>trt.contr</code>	list of treatment contrast vectors.
<code>table.legend</code>	legend for the names of variance components for large designs.
<code>response</code>	experimental data of the experiment.
<code>latex</code>	allows output of the Latex script to Latex table.
<code>fixed.names</code>	symbols for the fixed effects in the Latex outputs.
<code>decimal</code>	allows display of the coefficients as decimals.
<code>digits</code>	number of decimal places.
<code>list.sep</code>	shows the efficiency factors and coefficients of the fixed effects together.

2.6.3 Phase 1 experiment

We first show how to use the `summaryAovOnePhase` function which generates the theoretical ANOVA table of the Phase 1 experiment. The first argument of the function, `design.df`, consists of the experimental design in a data frame format. The class of each vector in the data frame should be a factor. Variable names must not contain any punctuation characters or symbols such as parentheses. The Phase 1 design in Table 2.3 can be presented as a data frame shown below,

```
> design1
```

```

  Ani      Trt
1  A healthy
2  B diseased
3  C healthy
4  D diseased
5  E diseased
6  F healthy
7  G diseased
8  H healthy

```

where **Ani** denotes animal ID and **Trt** denotes disease status.

The `blk.str` and `trt.str` in `summaryAovOnePhase` allow the user to input the block and treatment structures, respectively, using [Wilkinson and Rogers](#)' syntax ([Wilkinson and Rogers, 1973](#)). The user can also refer to the documentation file for R's `formula` function for further information on structure formulae. For the Phase 1 experiment (Section 2.5.1), the block structure formula is `Ani`, whereas the treatment structure formula contains a single term for the disease status, `Trt`. The output from `summaryAovOnePhase` is

```
> summaryAovOnePhase(design1, blk.str = "Ani", trt.str = "Trt")
```

```
$ANOVA
      DF Ani
Between Ani
  Trt    1  1
Residual 6  1

$Fixed
$Fixed$Coef
      Trt
Between Ani
  Trt    4

$Fixed$EF
      eff.Trt
Between Ani
  Trt    1
```

where each part has the decomposition of the sources of variation. The first sub-table, denoted by `ANOVA`, is the random effects table which lists the DF associated with each source of variation. The EMSs are split between the two sub-tables with the first sub-table giving coefficients of the variance components for each source of variation. The second sub-table, denoted by `Fixed`, is further split into two sub-tables where the first sub-table, denoted by `Coef`, containing the coefficients for the fixed effects part of the EMS, and the second sub-table, denoted by `EF`, contains the average efficiency factors for each fixed effect.

2.6.4 Phase 2 experiment

The function `summaryAovTwoPhase` is used to generate the theoretical ANOVA for two-phase experiments. The data frame of the Phase 2 design in Table 2.5 is shown below

```
> design2
```

	Run	Ani	Sam	Tag	Trt
1	1	A	1	114	healthy
2	1	B	1	115	diseased
3	1	C	1	116	healthy
4	1	D	1	117	diseased
5	2	E	1	114	diseased
6	2	F	1	115	healthy

```

7    2    G    1 116 diseased
8    2    H    1 117 healthy
9    3    C    2 114 healthy
10   3    D    2 115 diseased
11   3    A    2 116 healthy
12   3    B    2 117 diseased
13   4    G    2 114 diseased
14   4    H    2 115 healthy
15   4    E    2 116 diseased
16   4    F    2 117 healthy

```

where **Run** denotes MudPIT runs, **Sam** denotes sub-samples and **Tag** denotes iTRAQTM tags.

Recall the two-phase experiment from Section 2.5.2. The Phase 1 block structure consists of Animals and the Phase 2 block structure is MudPIT run. The Phase 1 and 2 block structures are represented by the arguments `blk.str1` and `blk.str2`, and their structure formulae are `Ani` and `Run`, respectively. The treatment structure is defined in the argument `trt.str`. Since Disease status and Tag effects do not interact, the structure formula is `Tag + Trt`.

The output from `summaryAovTwoPhase` of the two-phase experiment in Section 2.5.2 is

```

> summaryAovTwoPhase(design2, blk.str1 = "Ani", blk.str2 = "Run",
+ trt.str = "Tag + Trt")

```

```

$ANOVA
      DF e Ani Run
Between Run
  Between Ani 1  1 2  4
  Within Ani  2  1 0  4
Within Run
  Between Ani
    Tag      1  1 2  0
    Trt      1  1 2  0
  Residual  4  1 2  0
  Within Ani
    Tag      2  1 0  0
  Residual  4  1 0  0

$Fixed
$Fixed$Coef
      Tag Trt
Between Run
  Between Ani
  Within Ani
Within Run
  Between Ani
    Tag      4
    Trt      8
  Within Ani
    Tag      4
$Fixed$EF
      eff.Tag eff.Trt
Between Run
  Between Ani
  Within Ani
Within Run
  Between Ani
    Tag      1
    Trt      1

```



```
Within Ani
Tag      1
```

where the layout of this output is the same as that of the output from the `summaryAovOnePhase` function. The only main difference is the column `e`, of sub-table ANOVA, which denotes the variance component of experimental error.

2.6.5 Crossed or nested

Constructing these ANOVA tables requires first expanding each structure formula; R has provided a `terms` function to perform this task. However, when the structure formula comprises two or more terms, these factors may be nested, e.g. `Animal/Sample`, then the output from the `terms` gives

```
[1] "Animal"      "Animal:Sample"
```

where `Animal` denote the effects of different animals and `Animal:Sample` denotes the effects of different samples within animals. If the sample is assumed to be crossed with the animals, then the output from the `terms` function gives

```
[1] "Animal"      "Sample"      "Animal:Sample"
```

where `Animal` and `Sample` denote the effects of different animals and samples, respectively. However, `Animal:Sample` is identical to the previous output, but denotes the interaction effect between the animals and samples. Thus, it can become difficult to interpret whether the `Animal:Sample` means crossed or nested in the final ANOVA table for complex single or two-phase experiment. To overcome this confusion, we re-express the interaction effect of Between Animals and Samples by `Animal*Sample` and the effects of Between Samples Within Animals by `Animal(Sample)`.

This additional functionality can be useful for the user to examine complex structure formulae such as $((A/B)*C)/D$, where the `terms` function will generate

```
[1] "A"          "A:B"        "C"          "A:C"        "A:B:C"      "A:B:C:D"
```

whereas, the function in **infoDecompuTE** generates

```
[1] "A"          "A(B) "      "C"          "A*C"        "A(B)*C"     "A.B.C(D) "
```

whereby the user can easily identify that the relationship between A and B is nested and that between A and C is crossed. In addition, the `A(B)*C` denotes the effect of Between B Within A crossed with C, and `A.B.C(D)` denotes the effect of Between D Within A, B and C.

2.6.6 Artificial strata or pseudo-factors

In the two-phase proteomics experiment described in Section 2.5, the four runs can be viewed as two sets of two runs, namely Runs 1 and 3, and Runs 2 and 4. Samples of the same four animals can be assigned to each of these two sets of runs. The Set of Runs factor does not have its own random effects, as this artificial construct can partition the sources of variation to allow us to examine the ANOVA table more easily. Since the construct is artificial, there are no variance components associated with it. The `var.comp` argument, of `summaryAovOnePhase` and `summaryAovTwoPhase` functions, can be used to hide the variance components of the artificial construct. The new block factor, `Set`, is included in the block tier and the block structure formula to enable the additional decomposition of the artificial strata. The first and second Run sets are denoted by 1 and 2, respectively. This new vector of animal set, denoted by `Set`, is given by

```
> Set
```

```
[1] 1 1 1 1 2 2 2 2 1 1 1 1 2 2 2 2
Levels: 1 2
```

The Phase 2 block structure is then written as `Set/Run`, which means MudPIT runs are nested within sets of runs. Since the Set of Runs factor is an artificial construct, it contributes no additional variation and, therefore, does not have variance components associated with it. By default, **infoDecompuTE** associates variance components with each block factor. A mechanism is needed whereby actual random effect terms can have their variance components explicitly specified. The `var.comp` argument is supplied with a vector of character strings containing the names of the variance components which should be retained and appear as column names in the ANOVA table. This argument can also be used to define the order in which the variance components appear in the output table, i.e.

```
> summaryAovTwoPhase(design2, blk.str1 = "Ani", blk.str2 = "Set/Run",
+                      trt.str = "Tag + Trt", var.comp = c("Ani", "Set(Run)"))
```

```
$ANOVA
```

	DF	e	Ani	Set(Run)
Between Set				
Between Ani	1	1	2	4
Between Set(Run)	2	1	0	4
Within Set.Run				
Between Ani				
Tag	1	1	2	0
Trt	1	1	2	0
Residual	4	1	2	0
Within Ani				
Tag	2	1	0	0
Residual	4	1	0	0

Note `Within Set.Run` denotes Within Sets and Runs.

2.6.7 Manually defined contrasts

There can be situations when the researcher needs to split up the information further by re-expressing the treatment factor in terms of a set of 1 DF mutually orthogonal contrasts. In Table 2.6, the 3 DF associated with Tags are split across two strata, namely Between Animals Within Runs and Within Animals Within Runs, and this approach allows us to quickly identify which contrast is confounded with which stratum. The example here illustrates the treatment contrasts, where four iTRAQTM tags can be represented by three orthogonal contrasts from a classical 2^k design, as shown below

```
> Tag = list(Tag1 = Tag1, Tag2 = Tag2, Tag3 = Tag3)
> Tag
```

```
$Tag1
[1] 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1
$Tag2
[1] 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1 1 -1
$Tag3
[1] 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1 1 -1 -1 1
```

Note it is important that each contrast is uniquely named so that where its information lies is easily identifiable in the ANOVA table. The argument `table.legend` allows the use of letters to represent column names, and then inserts a legend at the bottom of the table. In this example, argument `table.legend` of function `summaryAovTwoPhase` is set to `TRUE`, because once the treatment contrasts are fitted separately, the larger number of columns in the table of fixed components make the table difficult to read. The output is shown as follows,

```
> summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
+ trt.str = "Tag + Trt",
+ trt.contr = list(Tag = list(Tag1 = Tag$Tag1, Tag2 = Tag$Tag2, Tag3 = Tag$Tag3),
+ Trt = Trt), table.legend = TRUE)
```

```
$ANOVA
$ANOVA$VC
          DF a b c
Between Run
  Between Ani      1  1 2 4
  Between Ani(Sam) 2  1 0 4
Within Run
  Between Ani
    Tag.Tag2      1  1 2 0
    Trt           1  1 2 0
  Residual        4  1 2 0
  Between Ani(Sam)
    Tag.Tag1      1  1 0 0
    Tag.Tag3      1  1 0 0
  Residual        4  1 0 0
```

```
$ANOVA$Legend
[1] "a = Ani(Sam)" "b = Ani"      "c = Run"
$Fixed
```

```

$Fixed$Coef
              a b c d
Between Run
  Between Ani
    Between Ani(Sam)
Within Run
  Between Ani
    Tag.Tag2      4
    Trt           8
  Between Ani(Sam)
    Tag.Tag1      4
    Tag.Tag3      4

$Fixed$Legend.Coeff
[1] "a = Tag.Tag1" "b = Tag.Tag2" "c = Tag.Tag3" "d = Trt"
$Fixed$EF
              a b c d
Between Run
  Between Ani
    Between Ani(Sam)
Within Run
  Between Ani
    Tag.Tag2      1
    Trt           1
  Between Ani(Sam)
    Tag.Tag1      1
    Tag.Tag3      1
$Fixed$Legend.EF
[1] "a = eff.Tag.Tag1" "b = eff.Tag.Tag2" "c = eff.Tag.Tag3" "d = eff.Trt"

```

Having broken down the tag contrasts, the random effects table shows that the Tag2 contrast is estimated in the Between Animals Within Runs stratum, and Tag1 and Tag3 contrasts are estimated in the Between Samples Within Animals Within Runs stratum. By breaking the Treatment effects into multiple orthogonal contrasts, it is possible to more closely examine how these contrasts contribute to each source of variation.

2.6.8 Mean squares computation

The `summary.aov` function in R can only compute the mean squares (MS) of single-phase experiments using experimental data. For two-phase experiments, the researcher can input their experimental data in the `response` argument to obtain the MS for each source of variation. When this argument is used, an additional column of MS will appear in the last column of the random effect table shown as,

```

> summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
+   trt.str = "Tag + Trt", response = rnorm(16))$ANOVA
$

```

	DF	Ani(Sam)	Ani	Run	MS
Between Run					
Between Ani	1	1	2	4	6.175
Between Ani(Sam)	2	1	0	4	0.667

Within Run					
Between Ani					
Tag	1	1	2	0	0.025
Trt	1	1	2	0	1.733
Residual	4	1	2	0	1.086
Between Ani(Sam)					
Tag	2	1	0	0	3.287
Residual	4	1	0	0	0.445

2.6.9 Latex output

The output from R is not always easy to read on the screen. The argument `latex` allows the user to transform R output into L^AT_EX script. Using the example from Section 2.5.2, the `latex` argument of the `summaryAovTwoPhase` function is set to `TRUE`, i.e.

```
> summaryAovTwoPhase(design, blk.str1 = "Ani", blk.str2 = "Run",
+ trt.str = "Tag + Trt", latex = TRUE, fixed.names = c("\\gamma", "\\tau") )
```

Table 2.8 shows the ANOVA table which results from the compilation of the L^AT_EX script, that can be generated by the `summaryAovOnePhase` and `summaryAovTwoPhase` functions. Two L^AT_EX packages `bm` and `booktabs` are required to compile the L^AT_EX script.

Table 2.8: Theoretical ANOVA table.

Source of Variation	DF	EMS	E_γ	E_τ
Between Run				
Between Ani	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Between Ani(Sam)	2	$\sigma^2 + 4\sigma_r^2$		
Within Run				
Between Ani				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma$	1	
Trt	1	$\sigma^2 + 2\sigma_a^2 + 8\theta_\tau$		1
Residual	4	$\sigma^2 + 2\sigma_a^2$		
Between Ani(Sam)				
Tag	2	$\sigma^2 + 4\theta_\gamma$	1	
Residual	4	σ^2		

Production of the L^AT_EX script requires Greek letters to define the fixed effects, i.e. θ_γ and θ_τ . Since different experiments can have different sets of Greek letters, users can choose their own sets of Greek letters to be displayed using the argument `fixed.names`. The subscripts of Between Runs and Between Animals variance components, σ_r^2 and σ_a^2 , come from the lower case of the first letter of the factor names in the design of the data frame and the Phase 1 and 2 block structure formulae in `blk.str1` and `blk.str2`, respectively. If the user need to use a different subscript, then they can change the first letter of the factor names both in the design of the data frame and the block structure formulae. Note that this table not only contains the DF and

EMS, but also the average efficiency factors for all the treatment effects in the last two columns. Further modification of this table may still be required, such as adjusting the names in the sources of variation, e.g. from Between Ani to Between Animals and from Between Ani(Sam) to Between Samples Within Animals. Nonetheless, this additional functionality allows rapid generation of the theoretical ANOVA table from the L^AT_EX scripts.

2.6.10 Shiny Application

Not all researchers can apply this R package, especially for those who have no statistical background. Hence, a Shiny application has been developed for more friendly use. Shiny is an R package which allows the construction of interactive web applications straight from R. This interactive web application can be deployed onto the web with its own URL, so the user can visit the application and not need to worry about installing the R program and coding. The Shiny application of this package is being hosted at the following link: https://kcha193.shinyapps.io/infoDecompuTE_Shiny/. One limitation of web applications is the requirement of stable internet connection.

There are three type of outputs that can be generated from this Shiny application: 1) output from the R console as a text file, 2) latex code as a text file, and 3) latex compiled portable document format (pdf) file.

2.7 Two-phase viticultural experiment using InfoDecompuTE

This section is to demonstrate the degree of complexity in a two-phase experiment which **InfoDecompuTE** can deal with by generating the theoretical ANOVA table for the viticultural experiment described by [Brien \(1992\)](#) and [Brien and Payne \(1999\)](#). The Phase 1 field trial involved a viticultural experiment comparing four different types of trellising and two pruning methods. The Phase 2 experiment involved the evaluation of the wines made from the produce of the viticultural experiment.

The Phase 1 viticultural experiment was arranged into two adjacent Youden square designs, each with three rows and four column blocks. The four trellising methods were assigned to the row blocks as a randomised complete block design and to the column blocks as a BIBD. Furthermore, each main plot was halved, and one of two different pruning methods was randomly assigned to each half-plot.

The Phase 2 experiment consisted of six judges evaluating the wines made from the grapes grown in the viticultural experiment. The wines were evaluated on two separate occasions, with wines made from grapes grown within the same square at Phase 1 being evaluated on the same occasion at Phase 2. Each occasion was divided into three intervals, with four sittings per interval. At each sitting, each judge was presented with four glasses of wine, two replicate wines from each half-plot within a main plot. [Brien and Payne \(1999\)](#) referred to these glasses as the positions. The two-phase experiment yielded a total of 576 measurements.

The block structure formulae of Phase 2 and 1 experiments are

$$((\text{Occasions}/\text{Intervals}/\text{Sittings}) * \text{Judges})/\text{Positions} \text{ and} \quad (2.42)$$

$$(\text{Rows} * (\text{Squares}/\text{Columns}))/\text{Halfplots, respectively.} \quad (2.43)$$

The treatment structure formula is

$$\text{Trellis} * \text{Method.} \quad (2.44)$$

The block structure (2.42) indicates that Sittings are nested within Intervals which are nested within Occasions. However, since all Judges were present at every Sitting, Judges is crossed with Sittings within Intervals within Occasions. Finally, Positions are nested within Judges and Sittings because each Judge evaluated four glasses of wine at each Sitting. The block structure in (2.43) for the Phase 1 experiment indicates that the main plots, to which Trellising methods are assigned, are defined as Rows crossed with Columns nested within Squares, with Half-plots being nested within main plots. The treatment structure defined in (2.44) is a 2×2 factorial experiment, and thus the Trellising and pruning Methods are crossed.

The structure formulae in (2.42), (2.43) and (2.44) are input into the function `summaryAovTwoPhase`. The block and treatment factor names are replaced by their first three characters, so the table can be fitted on the page. The output can be shown as follows:

```
> summaryAovTwoPhase(designPhase2, blk.str1 = "(Row*(Squ/Col))/Hal",
+                      blk.str2 = "((Oc/In/St)*Ju)/Pos",
+                      trt.str = "Tre*Met", decimal = TRUE,
+                      digits = 2, table.legend = TRUE)
```

Note: Complete confounding between Squ and Occ!

\$ANOVA																
\$ANOVA\$VC																
	DF	a	b	c	d	e	f	g	h	i	j	k	l	m	n	
Between Occ																
Between SquCCW	1	12	24	96	72	288	0	1	4	16	48	0	24	96	288	
Between Occ(Int)	4	0	0	0	0	0	0	1	4	16	0	0	24	96	0	
Between Occ.Int(Sit)																
Between SquCCW(Col)																
Tre	3	4	8	0	24	0	0	1	4	0	0	0	24	0	0	

Residual	3	4	8	0	24	0	0	1	4	0	0	0	24	0	0
Within Row.SquCCW.Col.Hal	12	0	0	0	0	0	0	1	4	0	0	0	24	0	0
Between Jud	5	0	0	0	0	0	0	1	4	16	48	96	0	0	0
Between Occ*Jud	5	0	0	0	0	0	0	1	4	16	48	0	0	0	0
Between Occ(Int)*Jud															
Between Row	2	12	24	96	0	0	192	1	4	16	0	0	0	0	0
Between Row*SquCCW	2	12	24	96	0	0	0	1	4	16	0	0	0	0	0
Within Row.SquCCW.Col.Hal	16	0	0	0	0	0	0	1	4	16	0	0	0	0	0
Between Occ.Int(Sit)*Jud															
Between SquCCW(Col)															
Tre	3	8	16	0	48	0	0	1	4	0	0	0	0	0	0
Residual	3	8	16	0	48	0	0	1	4	0	0	0	0	0	0
Between Row*SquCCW(Col)															
Tre	3	12	24	0	0	0	0	1	4	0	0	0	0	0	0
Residual	9	12	24	0	0	0	0	1	4	0	0	0	0	0	0
Within Row.SquCCW.Col.Hal	72	0	0	0	0	0	0	1	4	0	0	0	0	0	0
Between Occ.Int.Sit.Jud(Pos)															
Between Row.SquCCW.Col(Hal)															
Met	1	12	0	0	0	0	0	1	0	0	0	0	0	0	0
Tre*Met	3	12	0	0	0	0	0	1	0	0	0	0	0	0	0
Residual	20	12	0	0	0	0	0	1	0	0	0	0	0	0	0
Within Row.SquCCW.Col.Hal	408	0	0	0	0	0	0	1	0	0	0	0	0	0	0

\$ANOVA\$Legend

[1] "a = Row.SquCCW.Col(Hal)"	"b = Row*SquCCW(Col)"	"c = Row*SquCCW"
[4] "d = SquCCW(Col)"	"e = SquCCW"	"f = Row"
[7] "g = Occ.Int.Sit.Jud(Pos)"	"h = Occ.Int(Sit)*Jud"	"i = Occ(Int)*Jud"
[10] "j = Occ*Jud"	"k = Jud"	"l = Occ.Int(Sit)"
[13] "m = Occ(Int)"	"n = Occ"	

\$Fixed

\$Fixed\$Coef

	a	b	c
Between Occ			
Between SquCCW			
Between Occ(Int)			
Between Occ.Int(Sit)			
Between SquCCW(Col)			
Tre	5.33		
Within Row.SquCCW.Col.Hal			
Between Jud			
Between Occ*Jud			
Between Occ(Int)*Jud			
Between Row			
Between Row*SquCCW			
Within Row.SquCCW.Col.Hal			
Between Occ.Int(Sit)*Jud			
Between SquCCW(Col)			
Tre	10.67		
Between Row*SquCCW(Col)			
Tre	128		
Within Row.SquCCW.Col.Hal			
Between Occ.Int.Sit.Jud(Pos)			
Between Row.SquCCW.Col(Hal)			
Met		288	
Tre*Met			72
Within Row.SquCCW.Col.Hal			

\$Fixed\$Legend.Coeff

[1] "a = Tre"	"b = Met"	"c = Tre*Met"
---------------	-----------	---------------

\$Fixed\$EF

a	b	c
---	---	---


```

Between Occ
  Between SquCCW
Between Occ(Int)
Between Occ.Int(Sit)
  Between SquCCW(Col)
    Tre                                0.04
  Within Row.SquCCW.Col.Hal
Between Jud
Between Occ*Jud
Between Occ(Int)*Jud
  Between Row
  Between Row*SquCCW
  Within Row.SquCCW.Col.Hal
Between Occ.Int(Sit)*Jud
  Between SquCCW(Col)
    Tre                                0.07
  Between Row*SquCCW(Col)
    Tre                                0.89
  Within Row.SquCCW.Col.Hal
Between Occ.Int.Sit.Jud(Pos)
  Between Row.SquCCW.Col(Hal)
    Met                                1
    Tre*Met                            1
  Within Row.SquCCW.Col.Hal

$Fixed$Legend.EF
[1] "a = eff.Tre"      "b = eff.Met"      "c = eff.Tre*Met"

```

The above output takes about a minute to generate using eight-gigabytes of RAM with a Quad Core 2.5GHz machine running R 3.3.3 and Microsoft Windows 7.

There are several additional features that can be observed from this new output table. Firstly, recall that wines made from grapes grown within the same square in the Phase 1 experiment were evaluated on the same occasion in the Phase 2 experiment. Thus, Square and Occasion are completely confounded with each other. The function is capable of detecting such confounding and alerts the user to its presence by printing the warning

Note: Complete confounding between Squ and Occ!

in the first row of output. Further, the `decimal` argument allows the user to choose between displaying the coefficients in the VC and fixed tables, and the efficiency factors in the EF table, as proper fractions (`decimal = FALSE` default) or in decimal format (`decimal = TRUE`). Lastly, the `digits` argument is used to specify the number of decimal places to be printed when users set `decimal` to `TRUE`.

2.8 Conclusion

InfoDecompuTE, a freely available R package, allows researchers to study any complex single- or two-phase experimental design by generating the theoretical ANOVA table with the coefficients

of variance components of the EMS, as shown in Section 2.7. This package allows researchers to study how data vector space spanned by the data vector is decomposed across different strata and sources of variation.

This package can also analyse designs with treatment or block factors that are non-orthogonal to multiple block factors, and can produce average efficiency factors, as shown in Section 2.7. The researcher thus can identify how much treatment information remains when conducting the test for the Treatment effects. Additionally, users can fit each block or treatment contrast separately, allowing more flexible analysis, and can further clarify on how the block or treatment information is split across different strata.

However, this package has its limitations. Currently it can only analyse single- and two-phase experiments. If another phase was added, it would increase the computation time from n^2 to n^3 . This is due to an additional for-loop being required to define the block structure of the additional phase. The best solution would be to re-implement the matrix calculation in another programming language such as C to accelerate the computation.

Additionally, users need some understanding of how to build the model using the structure formulae as described by [Wilkinson and Rogers \(1973\)](#) for block and treatment structures of the two-phase experiments. Nonetheless, **infoDecompuTE** gives statistical researchers an additional tool to help them better understand experimental designs and construct better experiments.

Chapter 3

Optimal designs for two-phase experiments when the Phase 1 experiment is arranged in a completely randomised design

3.1 Introduction

This Chapter discusses the search for optimal designs for Phase 2 proteomics experiments, when the Phase 1 experiment is a completely randomised design (CRD), while the Phase 2 experiment employs a multiplexing technique such as iTRAQTM. The optimal design of experiments is a branch of experimental design theory. The aim of optimal design is to plan an experiment to assign treatments to experimental units in such a way that we maximise the amount of treatment information in the intra-block stratum. In general, an experiment is made up of a set of design parameters, which are the number of treatments, replication number, number of blocks and block sizes. A group of designs can be generated with the same design parameters, namely *competing designs*. The challenge is to find an optimal arrangement of treatments across blocks, i.e. to achieve maximum efficiency across a competing set of designs of the same size.

When the Phase 1 experiment is arranged in a CRD, the treatments are randomly allocated to each experimental unit of the experiment. Since animals are used as examples in this Chapter, the experimental units are animals. The biological material harvested from each Phase 1 experimental unit is divided into multiple aliquots, namely technical replicates, which will undergo MudPIT analysis in the Phase 2 experiment. Thus, the Phase 1 experimental units form blocks

in the Phase 2 experiment. The Phase 2 experiment involves a second block factor, namely MudPIT runs, with Phase 2 treatment factor, namely iTRAQTM tags, in which multiple biological samples from the Phase 1 experiment are analysed under homogeneous conditions.

Determining whether a treatment arrangement of experimental units is optimal is based on the definition of the optimality criterion. Numerous optimality criteria have been presented in the literature (Goos, 2012). Many of them belong to the class of alphabetic optimality criteria. This Chapter considers A- and MS-optimality criteria. If the harmonic mean of the canonical efficiency factors, or the average efficiency factor, of a design is the largest compared to the other competing designs, then the design is said to be A-optimal. The MS-optimality criterion is a two-stage criterion where (1) a class of designs has the highest mean of the efficiency factors, and (2) within this class, the design having the lowest spread of the efficiency factors is said to be the MS-optimal design.

Once a specific optimality criterion is chosen, an *objective function* can be derived. The objective function is a mathematical expression describing the relationship between variables that correspond to the competing designs. Thus, it plays an important role in helping us to locate the optimal design. A well-chosen objective function produces a maximum or minimum value given an optimal input design. The main goal of this Chapter is to derive an objective function specifically for the finding the optimal design of Phase 2 experiments.

The method of finding optimal designs for different design classes (such as block, row-column and α_n designs) has previously been discussed (Whitaker et al., 1990; Williams and John, 1996; John et al., 2002). The method aims to find the design with the most treatment information in the intra-block stratum among the set of competing designs. Finding the optimal design of Phase 2 experiment requires us to also choose the best allocation of the experimental units from the Phase 1 experiment to the experimental units of the Phase 2 experiment. This is because any confounding of the block effects of the Phase 1 experiment with the block effects of the Phase 2 experiments will impact the estimation of the treatment effects and the number of Residual DF needed to estimate their variance, and hence perform a valid statistical test. Given that a suitable optimality criterion has not been defined, the method of finding such optimal designs of Phase 2 experiments has yet to be described.

This Chapter discusses the search for optimal designs using the two-phase proteomics experiment, when the Phase 1 experiment is arranged in a CRD, while the Phase 2 experiment uses a multiplexing technique such as iTRAQTM. This Chapter is laid out as follows: Sections 3.2, 3.3 and 3.4 describe the linear models of the Phase 1 experiment arranged in a CRD and Phase 2 MudPIT-iTRAQTM experiments, the information matrix and optimality criteria, respectively.

The theoretical ANOVA tables of the Phase 1 and 2 experiments are described in Section 3.5. The construction of the objective function is presented in Section 3.6. Section 3.7 discusses how the initial designs are generated. Sections 3.8 and 3.9 describe the *simulated annealing algorithm* (SA) and nested SA algorithm used to search for optimal designs. Section 3.10 discusses improvements that we have made to the nested SA algorithm to find optimal designs of Phase 2 experiments. Section 3.11 and 3.12 set forth two examples to illustrate the objective function and the SA algorithm. Section 3.13 summarises the properties of the optimal design that are found using the method described. Section 3.14 briefly demonstrates how to modify the objective function in situations when the Phase 1 experiment is a 2×2 factorial experiment. The **infoDecompuTE** package, introduced in Chapter 2, is employed consistently to construct the theoretical ANOVA tables using the treatment average efficiency factor to compare the competing designs. Even though we have used four- and eight-plex labelling systems as the motivating example, the method are more general and can be applied all two-phase designs when the Phase 1 experiment is arranged in a CRD.

3.2 Design parameters and model

This section defines the linear models of the Phase 1 and 2 proteomics experiments described in this Chapter. This experiment is a generalised version of the one in Section 2.5 with some adjustments that allow us to better describe the experimental design structure and mathematical derivation of the information matrix.

3.2.1 Linear models

The Phase 1 experiment is arranged in a CRD which consists of ν treatments assigned to n_a experimental units. Note that we will refer to animals as experimental units for the remainder of this Chapter. Let y_{ij} denote the abundance of a given protein from animal j under treatment i . The linear model of this Phase 1 experiment can be written as

$$y_{ij} = \mu + \tau_i + a_j, \quad (3.1)$$

where μ denotes the overall mean log protein abundance across all samples, τ_i denotes the fixed effect of treatment i , and $a_j \sim \mathcal{N}(0, \sigma_a^2)$ denotes the random effect from animals j , ($i = 1, \dots, \nu$; $j = 1, \dots, n_a$). Note that this model in (3.1) is a generalised version of the linear model in (2.40).

The Phase 2 experiment is arranged in blocks, where the numbers of MudPIT run, n_r , and iTRAQTM tags, n_γ , correspond to the number of blocks and block size, respectively. Each sample

of the Phase 1 experiment is further split into n_s sub-samples (technical replicates) for estimating measurement errors in the Phase 2 experiment. Hence, there are a total of $n = n_a n_s = n_r n_\gamma$ sub-samples to be analysed in the Phase 2 proteomics experiment. Now let y_{kls}^{ij} denote the abundance of the same protein as in (3.1) from sub-sample s of treatment i assigned to animal j which is then differentially labelled by tag k and analysed in run l . The linear model of the Phase 2 experiment can be written as

$$y_{kls}^{ij} = \mu + \tau_i + a_j + \gamma_k + r_l + \epsilon_{kls}^{ij}, \quad (3.2)$$

where γ_m denotes the fixed effect of tag k , $r_l \sim \mathcal{N}(0, \sigma_r^2)$ represents the random effect of run l , and $\epsilon_{kls}^{ij} \sim \mathcal{N}(0, \sigma^2)$ denotes the experimental error, ($k = 1 \dots n_\gamma$; $l = 1 \dots n_r$; $s = 1 \dots n_s$).

3.2.2 Linear models in matrix notation

Based on the definition in Section 2.3.1, model (3.2) is then expressed as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \boldsymbol{\epsilon}, \quad (3.3)$$

where $\mathbf{1}$ is an $n \times 1$ vector whose elements are all unity, μ denotes the grand mean, and $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ is an $n \times 1$ random vector of experimental errors. The matrix \mathbf{I}_n denotes the $n \times n$ identity matrix. The treatment parameter vector of length νn_γ is defined as

$$\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \dots, \alpha_{1n_\gamma}, \alpha_{21}, \dots, \alpha_{\nu n_\gamma})', \quad (3.4)$$

where α_{im} denotes the effect of treatment i and tag k , ($i = 1, \dots, \nu$; $m = 1 \dots n_\gamma$). The $n \times \nu n_\gamma$ treatment design matrix, \mathbf{X} , in (3.3) is a binary matrix indicating the allocation of treatments to the experimental units. The Phase 1 block design matrix is denoted by $\mathbf{Z} = \mathbf{Z}_a$ which is a $n \times \nu n_b$ design matrix of animals. The Phase 1 block parameter is denoted by $\mathbf{u} = (a_{11}, a_{12}, \dots, a_{\nu n_b})$. The Phase 2 block design matrix is denoted by $\mathbf{W} = \mathbf{W}_r$ which is a $n \times n_r$ design matrix of runs. The Phase 2 block parameter is denoted by $\mathbf{v} = (m_1, \dots, m_{n_r})$.

For the purposes of finding optimal designs for Phase 2 experiments, the linear model in matrix notation in (3.3) needs to be reformulated. This is because the main consideration is in the *how the sub-samples from treatments assigned to each animal in the Phase 1 experiment are labelled by which tag and measure in which run of the Phase 2 experiment*. Hence, during the search for the optimal design, the design matrices of Phase 2 Run and Tag factors stay the same, and the goal is to find the optimal design based on the design matrices of Phase 1 Treatment and Animal factors. Furthermore, the Tag effects belong to the Phase 2 experiment and the effects

of treatment and tag are assumed not to interact. Thus, the overall treatment design matrix is split into treatment and tag design matrices, denoted by X_τ and X_γ , respectively. The model in (3.3) can be rewritten as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_\tau\boldsymbol{\alpha}_\tau + \mathbf{Z}_a\mathbf{u} + \mathbf{X}_\gamma\boldsymbol{\alpha}_\gamma + \mathbf{W}_r\mathbf{v} + \boldsymbol{\epsilon}, \quad (3.5)$$

where $\boldsymbol{\alpha}_\tau$ and $\boldsymbol{\alpha}_\gamma$ are the vectors of treatment and tag parameters

$$\boldsymbol{\alpha}_\tau = (\tau_1, \dots, \tau_\nu)' \text{ and } \boldsymbol{\alpha}_\gamma = (\gamma_1, \dots, \gamma_{n_\gamma})',$$

respectively.

3.3 Information matrix for the objective function

The *information matrix* is essential for evaluating the objective function. This section shows how the animal and treatment information matrices are defined for two-phase MudPIT-iTRAQTM experiments.

In finding the optimal design for a MudPIT-iTRAQTM experiment, we need to consider the variation between runs and differences between tags which are introduced from the Phase 2 experiment. The best allocation of the sub-samples from treatments assigned to each animal, which are then labelled by a tag and measured in a run occurs when the confounding of the Phase 1 Animal and Treatment effects with the Phase 2 Run and Tag effects is minimised. We define an orthogonal projection matrix which projects \mathbf{y} onto Within Runs and Tags vector subspace as

$$\mathbf{Q}_{r\gamma} = (\mathbf{I} - \mathbf{P}_r)(\mathbf{I} - \mathbf{P}_\gamma) \quad (3.6)$$

where $\mathbf{P}_r = \mathbf{W}_r(\mathbf{W}_r'\mathbf{W}_r)^{-1}\mathbf{W}_r'$ denotes the projection matrix which projects \mathbf{y} onto the Between Runs vector subspace, and $\mathbf{P}_\gamma = \mathbf{X}_\gamma(\mathbf{X}_\gamma'\mathbf{X}_\gamma)^{-1}\mathbf{X}_\gamma'$ denotes the projection matrix which projects \mathbf{y} onto the Between Tags vector subspace. Note that, the Tag effects should still be estimated as the fixed effects in the ANOVA.

The matrix $\mathbf{Q}_{r\gamma}$ in (3.6), an orthogonal projection matrix, has to be symmetric. To show this property, we first expand the matrix $\mathbf{Q}_{r\gamma}$ in (3.6) as

$$\mathbf{Q}_{r\gamma} = (\mathbf{I} - \mathbf{P}_r)(\mathbf{I} - \mathbf{P}_\gamma) = \mathbf{I} - \mathbf{P}_r - \mathbf{P}_\gamma + \mathbf{P}_r\mathbf{P}_\gamma \quad (3.7)$$

where matrices \mathbf{I} , \mathbf{P}_r and \mathbf{P}_γ are symmetric. To show $\mathbf{P}_r\mathbf{P}_\gamma$ in (3.7) is also symmetric, we will expand matrices \mathbf{P}_r and \mathbf{P}_γ separately. The projection matrix which projects \mathbf{y} onto the Between

Runs vector subspace is

$$\mathbf{P}_r = \mathbf{W}_r(\mathbf{W}_r' \mathbf{W}_r)^{-1} \mathbf{W}_r'$$

where \mathbf{W}_r is the design matrix for Run factor of the Phase 2 experiment and can be expressed as

$$\mathbf{W}_r = \mathbf{1}_{n_\gamma} \otimes \mathbf{I}_{n_r}.$$

Thus, the projection matrix for Between Runs can be re-written as

$$\mathbf{P}_r = \frac{1}{n_\gamma} \mathbf{1}_{n_\gamma} \mathbf{1}_{n_\gamma}' \otimes \mathbf{I}_{n_r}.$$

The projection matrix which projects \mathbf{y} onto the Between Tags vector subspace is

$$\mathbf{P}_\gamma = \mathbf{X}_\gamma(\mathbf{X}_\gamma' \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma$$

where \mathbf{X}_γ is design matrix for Tag factor of the Phase 2 experiment and can be expressed as

$$\mathbf{X}_\gamma = \mathbf{I}_{n_\gamma} \otimes \mathbf{1}_{n_r}$$

Thus, projection matrix for Between Tags can be re-written as

$$\mathbf{P}_\gamma = \frac{1}{n_r} \mathbf{I}_{n_\gamma} \otimes \mathbf{1}_{n_r} \mathbf{1}_{n_r}'.$$

Then, $\mathbf{P}_r \mathbf{P}_\gamma$ in (3.7) becomes

$$\begin{aligned} \mathbf{P}_r \mathbf{P}_\gamma &= \frac{1}{n_r n_\gamma} \mathbf{1}_{n_\gamma} \mathbf{1}_{n_\gamma}' \otimes \mathbf{1}_{n_r} \mathbf{1}_{n_r}' \\ &= \frac{1}{n} \mathbf{1}_{n_\gamma n_r} \mathbf{1}_{n_\gamma n_r}' \\ &= \frac{1}{n} \mathbf{1}_n \mathbf{1}_n' \\ &= \mathbf{K}_n. \end{aligned} \tag{3.8}$$

Thus, $\mathbf{P}_r \mathbf{P}_\gamma$ is also symmetric because it is an averaging matrix.

The animal information matrix in the Within Runs and Tags vector subspace, denoted by \mathbf{A}_a , can be written as

$$\mathbf{A}_a = \mathbf{Z}_a' \mathbf{Q}_{r\gamma} \mathbf{Z}_a \tag{3.9}$$

where \mathbf{Z}_a denotes the Phase 2 design matrix of Phase 1 Animal factor (experimental units).

The treatment information matrix in the Within Runs and Tags vector subspace, denoted

by \mathbf{A}_τ , is defined in the same way as the animal information matrix, which is given by

$$\mathbf{A}_\tau = \mathbf{X}'_\tau \mathbf{Q}_{r\gamma} \mathbf{X}_\tau, \quad (3.10)$$

where \mathbf{X}_τ denotes the Phase 2 treatment design matrix.

3.4 Optimality criteria

A specific optimality criterion is used to construct the objective function and search for the optimal design. The value from the objective function design at the i -th iteration is denoted by $O(D_i)$, where D_i is the design at i -th iteration during the search. Different optimality criteria may result in designs with different properties. This section explains the A- and MS-optimality criteria.

3.4.1 A-optimality

The *A-optimal design* has the highest average efficiency factor among the competing designs (John and Williams, 1995). The efficiency factors provide a measure of the amount of treatment information obtained from intra-block analysis and thus can be useful in comparing different competing designs. In order to understand the basis of the A-optimal design, we need to show how the average efficiency factor is defined using pair-wise treatment contrasts.

The variance of the pairwise treatment comparison, $\tau_j - \tau_{j'}$, can be expressed as

$$\text{var}(\widehat{\tau}_j - \widehat{\tau}_{j'}) = (\omega_{jj} + \omega_{j'j'} - 2\omega_{jj'})\sigma^2 \quad (j \neq j'),$$

where $\omega_{jj'}$ is the (jj') -th element in generalised inverse of the treatment information matrix, denoted by \mathbf{A}^- .

If all the treatment information can be estimated in the intra-block analysis, i.e. orthogonal block design, the variance of any pairwise treatment comparison is

$$(1/r + 1/r - 0)\sigma^2 = 2\sigma^2/r,$$

where σ^2 denotes the variance components associated with the measurement error of the experiment, and the term r is the treatment replication.

For a non-orthogonal block design, the efficiency factor for treatment contrast $\tau_j - \tau_{j'}$ is then defined to be

$$e_{ij} = \frac{2\sigma^2/r}{\text{var}(\widehat{\tau}_j - \widehat{\tau}_{j'})}. \quad (3.11)$$

The variance over pairwise difference can be re-written as

$$\text{var}(\hat{\tau}_j - \hat{\tau}_{j'}) = \text{var}(\mathbf{c}'\hat{\boldsymbol{\tau}}) = \mathbf{c}'_{jj'}\mathbf{A}^+\mathbf{c}_{jj'}\sigma^2 = (\sigma^2/r) \sum_{j=1}^{\nu-1} e_i^{-1}(\mathbf{c}_{jj'}\mathbf{p}_i)^2$$

where $\mathbf{c}_{jj'}$ denotes a vector of length ν with j -th element equal to 1, the j' -th element equal to -1 , and the remaining $\nu - 2$ elements equal to zero, and \mathbf{A}^+ denotes the Moore-Penrose generalised inverse of the information matrix. The average efficiency factor over all pairwise comparisons, denoted by E , can be defined as $2\sigma^2/r$ divided by the average variance of all estimated pairwise comparisons, i.e.

$$\begin{aligned} E &= \frac{2\sigma^2/r}{\text{var}(\hat{\tau}_i - \hat{\tau}_j)/[\nu(\nu - 1)]} \\ &= \frac{2\sigma^2/r}{(\sigma^2/r[\nu(\nu - 1)]) \sum_{i=1}^{\nu-1} e_i^{-1}(\mathbf{c}'_{jj'}\mathbf{p}_i)^2} \\ &= \frac{2\sigma^2/r}{(2\sigma^2/r) \sum_{i=1}^{\nu-1} e_i^{-1}/(\nu - 1)} \\ &= \frac{\nu - 1}{\sum_{i=1}^{\nu-1} e_i^{-1}}, \end{aligned} \tag{3.12}$$

namely the harmonic mean of the pairwise efficiency factors.

Since the treatment information matrix is symmetric, it has a complete set of orthogonal eigenvectors. These orthogonal eigenvectors, denoted by $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{\nu-1}$, can be used as treatment contrasts to compute the average efficiency factor (John and Williams, 1995). Let $\lambda_1, \lambda_2, \dots, \lambda_{\nu-1}$ denote the eigenvalues of the treatment information matrix, then the canonical form of \mathbf{A}^+ can be written in spectral form, i.e.

$$\sum_{i=1}^{\nu-1} \lambda_i^{-1} \mathbf{p}_i \mathbf{p}_i'. \tag{3.13}$$

The variance of the treatment contrast, $\mathbf{p}'_i \hat{\boldsymbol{\tau}}$, is given by

$$\text{var}(\mathbf{p}'_i \hat{\boldsymbol{\tau}}) = \mathbf{p}'_i \mathbf{A}^+ \mathbf{p}_i \sigma^2 = \frac{\sigma^2}{\lambda_i}.$$

Thus, the variance of the Treatment effect is still σ^2/r for an orthogonal block design.

For a non-orthogonal block design, the efficiency factor is

$$e_i = \frac{\sigma^2/r}{\text{var}(\mathbf{p}'_i \hat{\boldsymbol{\tau}})} = \frac{\sigma^2/r}{\sigma^2/\lambda_i} = \frac{\lambda_i}{r}, \quad (i = 1, \dots, \nu - 1),$$

namely the *canonical efficiency factors* of the i -th *basic contrast*, denoted by $\mathbf{p}'_i \boldsymbol{\tau}$ (James and Wilkinson, 1971). Following (3.12), the average efficiency factor is then given by the harmonic mean of the canonical efficiency factors.

The objective function for finding the A-optimal design is the computation of the average efficiency factor associated with the Animal effects, i.e.

$$E_a = \frac{n_a - 1}{\sum_{i=1}^{n_a-1} e_i^{-1}}, \quad (3.14)$$

where e_i denotes the canonical efficiency factor computed from the animal information matrix defined in (3.9).

3.4.2 MS-optimality

The *MS-optimality criterion*, introduced by Shah (1960) and Eccleston and Hedayat (1974), is a two-step criterion, i.e. an M-step and an S-step. The M-step involves finding the designs which maximise the trace of the information matrix, where the trace of the information matrix is the sum of the eigenvalues. From the set of designs found in the M-step, the S-step identifies the design that minimises the trace of the square of the information matrix. Thus, the S-step attempts to find the design where the spread of the canonical efficiency factors is the smallest. The resultant design is said to be *variance-balanced* in the sense that the variances of all estimated pairwise treatment differences are the same. MS-optimal designs can be found using two separate objective functions, one for the M-step and another for the S-step. Alternatively, these two objective functions can be combined into a single objective function (Williams and John, 1996).

The information matrix associated with the Animal effects, $\mathbf{A}_a = \mathbf{Z}'_a \mathbf{Q}_{r\gamma} \mathbf{Z}_a$, can be re-written as

$$\mathbf{A}_a = n_s \left[\mathbf{I} - \mathbf{K} - \left(\frac{\mathbf{N}_{ar} \mathbf{N}'_{ar}}{n_s n_\gamma} + \frac{\mathbf{N}_{a\gamma} \mathbf{N}'_{a\gamma}}{n_s n_r} - 2\mathbf{K} \right) \right], \quad (3.15)$$

where $\mathbf{N}_{am} = \mathbf{Z}'_a \mathbf{W}_r$ and $\mathbf{N}_{a\gamma} = \mathbf{Z}'_a \mathbf{X}_\gamma$ denotes the $n_a \times n_r$ animal incidence matrices with runs and tags, respectively. Since the number of sub-samples from each animal, n_s , identity matrix, \mathbf{I} , and averaging matrix, \mathbf{K} , are always the same, the equation (3.15) is equivalent to that given by Williams and John (1996) as

$$\mathbf{S} = \frac{\mathbf{N}_{am} \mathbf{N}'_{am}}{n_s n_\gamma} + \frac{\mathbf{N}_{a\gamma} \mathbf{N}'_{a\gamma}}{n_s n_r} - 2\mathbf{K}.$$

MS-optimal designs can then be found by minimising the following objective function

$$O(D_i) = \text{tr}(\mathbf{S}) + \text{tr}(\mathbf{S}^2), \quad (3.16)$$

where minimising $\text{tr}(\mathbf{S})$ and $\text{tr}(\mathbf{S}^2)$ is the same as the M-step and S-step, respectively, (Williams and John, 1996).

3.5 An illustrative example

This illustrative example is based on the linear models defined in Section 3.2. This section uses the most trivial case to show the theoretical ANOVA tables of the Phase 1 experiment, Phase 2 ignoring Phase 1 experiment, and the optimal design of the Phase 2 experiment.

Consider a two-phase experiment when the Phase 1 experiment involves $\nu = 2$ treatments (labelled by lower case letters, i.e. a and b), assigned to $n_a = 4$ animals (labelled by upper case letters, i.e. A to D). The treatments are arranged so that Treatment a is assigned to Animals A and C , and Treatment b is assigned to Animals B and D . The theoretical ANOVA table of these two experiments can then be generated using the R package **infoDecompuTE**, as shown in Table 3.1a. The content of these theoretical ANOVA tables has been previously described in Section 2.5, which shows that 3 DF associated with the Animal effects can be partitioned into 1 DF of Treatment effects and 2 DF of Residual EMS.

The Phase 2 ignoring Phase 1 experiment consists of $n_r = 2$ runs with samples from each run labelled by $n_\gamma = 4$ tags. The theoretical ANOVA table of the Phase 2 experiment ignoring the Phase 1 experiment is shown in Table 3.1b. This theoretical ANOVA table shows the total of 7 DF is separated into 1 DF associated with the Between Runs stratum and 6 DF with the Within Runs stratum. These 6 DF of the Within Runs stratum are further partitioned into 3 DF of Tag effects and 3 DF of Residual EMS.

Table 3.1: Theoretical ANOVA of (a) the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 4$ animals arranged in a CRD and (b) the Phase 2 experiment ignoring Phase 1 with $n_r = 4$ runs and $n_\gamma = 4$ tags.

(a)				(b)			
Source of Variation	DF	EMS	E_τ	Source of Variation	DF	EMS	E_γ
Between Animals				Between Runs	1	$\sigma^2 + 4\sigma_r^2$	
Treatments	1	$\sigma_a^2 + 2\theta_\tau$	1	Within Runs			
Residual	2	σ_a^2		Tag	3	$\sigma^2 + 2\theta_\gamma$	1
				Residual	3	σ^2	

The optimal design for the Phase 2 experiment, given the Phase 1 experiment is arranged in a CRD, is presented in Tables 3.2. Each of four samples from each animal are split into $n_s = 2$ sub-samples, thus there is a total of eight sub-samples to be measured in the Phase 2 experiment. Each run comprises sub-samples from all four animals, so that the effects of runs are orthogonal to the effects of animals. However, since sub-samples from Animals A and B are differentially labelled with Tags 114 and 115, and sub-samples from Animals C and D are differentially labelled with Tags 116 and 117, the Tag effects are partially confounded with

Animal effects. Since each treatment is assigned twice in each run and differentially labelled once with each tag, Treatment effects are orthogonal to both Run and Tag effects.

Table 3.2: Design of Phase 2 proteomics experiment showing allocation of sub-samples from animals and disease statuses to runs and tags, when the Phase 1 experiment consisting of $\nu = 2$ treatments assigned to each of $n_a = 4$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 2$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Aa</i>	<i>Bb</i>	<i>Ca</i>	<i>Db</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Db</i>	<i>Ca</i>

The theoretical ANOVA for the optimal design of the Phase 2 proteomics experiment is shown in Table 3.3. As expected, the theoretical ANOVA shows that all 3 DF associated with the Between Animal stratum are in the Within Runs stratum. Residual DF in the Between Animals Within Runs stratum is reduced from 2 DF to 1 DF, because 1 of the 3 DF associated with Tag effects is in the Between Animals stratum. There is only 1 DF left to estimate variance of the Treatment effects, so the estimation will not be as precise as compared to the Phase 1 experiment. However, the amount of treatment information is preserved from the Phase 1 experiment, because the treatment average efficiency factor, E_τ , is 100% in the Within Runs stratum. Furthermore, there is a valid F-test for the Treatment effects in the Between Animals Within Runs stratum, because the variance components of Treatment EMS and Residual EMS are identical. Thus, the design for the Phase 2 experiment shown in Table 3.2 is optimal in this case.

Table 3.3: Theoretical ANOVA of the Phase 2 experiment in Table 3.2.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs	1	$\sigma^2 + 4\sigma_r^2$		
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 2\theta_\gamma$	1	
Treatment	1	$\sigma^2 + 2\sigma_a^2 + 4\theta_\tau$		1
Residual	1	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 2\theta_\gamma$	1	
Residual	1	σ^2		

3.6 An objective function for identifying optimal designs for the Phase 2 experiment

An objective function is a mathematical expression that generates a value used to compare competing designs. Having observed the example in Section 3.5, a well-chosen objective function produces a maximum or minimum value given an optimal input design. In this section, we will derive an objective function to identify the optimal design for a Phase 2 proteomics experiment where the material for this experiment comes from a Phase 1 experiment arranged in a CRD. Here, an optimal design is defined as one in which:

- The animal information is maximised in the Within Runs stratum.
- The treatment information is maximised in the Between Animals Within Runs stratum.
- DF of the Treatment effects must still be intact in the Between Animals Within Runs stratum.

These three properties form the three components of the objective function, that is a compound-criterion objective function which maximises both the animal and treatment information in the Between Animal Within Runs stratum of the analysis for the Phase 2 experiment. We will show, using three different examples, how as we build this function the optimal designs, of the Phase 2 experiment given the Phase 1 experiment is a CRD, that are identified improve in terms of meeting the criteria that we consider yield an improved design on the previous step.

3.6.1 Maximising the animal information

The first step in constructing the objective function is to maximise the animal information in the Within Runs stratum. Specifically, the first step is to compare the MS- and A-optimal designs. These two optimality criteria and the objective function were described in Section 3.4.

The MS- and A-optimal designs are compared using a two-phase experiment with Phase 1 experiment involving $\nu = 2$ treatments assigned to $n_a = 6$ animals. Samples of each animal are further split into $n_s = 2$ sub-samples which are then differentially labelled by $n_\gamma = 4$ tags and analysed in $n_r = 3$ runs. In the Phase 1 experiment, Treatment a is assigned to Animals A , C and E and Treatment b is assigned to Animals B , D and F . The ANOVA for this Phase 1 experiment is presented in Table 3.4, and shows there is a total of 5 DF partitioned into 1 DF for estimating Treatment effects and 4 DF for the Residual EMS. The aim is to find an allocation of

differentially labelled samples to runs which maximises the amount of animal information within runs.

Table 3.4: Theoretical ANOVA of the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 6$ animals.

Source of Variation	DF	EMS	E_τ
Between Animals			
Treatments	1	$\sigma_a^2 + 3\theta_\tau$	1
Residual	4	σ_a^2	

Two allocations of sub-samples from animals and treatments labelled by tags and measured in runs are compared based on the MS- and A-optimality criteria. Using the objective function based on MS-optimality criterion, given in (3.16), the optimal allocation of sub-samples from animals and treatments labelled by tags and measured in runs is shown in Table 3.5. The Animal effects are confounded with Run effects, because each run comprises sub-samples from different combinations of animals. The Animals effects are also confounded with Tag effects for the same reason. The sub-samples from treatments labelled by tags and measured in runs are also non-orthogonal, because each run and tag comprised sub-samples of different combinations of the treatment groups.

Table 3.5: The MS-optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consisting of $\nu = 2$ treatments assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Ea</i>	<i>Db</i>	<i>Ca</i>	<i>Aa</i>
2	<i>Bb</i>	<i>Ea</i>	<i>Aa</i>	<i>Fb</i>
3	<i>Ca</i>	<i>Fb</i>	<i>Db</i>	<i>Bb</i>

Based on the MS-optimal design of the Phase 2 experiment, the theoretical ANOVA in Table 3.6 shows the total of 11 DF is first partitioned into 2 DF of Between Runs and 9 DF of Within Runs strata. Since the Treatment and Animal effects are confounded with Run effects, some Treatment and Animal effects can be estimated in the Between Runs stratum. In the Between Animals Within Runs stratum, there is 1 Residual DF compared to the 4 DF from the Phase 1 experiment. This is due to the confounding of animal effects with the tag effects; thus, the tag EMS used up 3 DF. Moreover, there is no valid F-test for the treatment effects, because the coefficients of the variance components in the Between Animals Within Runs stratum are not the same. Five canonical efficiency factors for Animal effects in the Within Runs stratum

are 1, 1, 1, 0.75 and 0.75, which means 2 of 5 DF for animals have only 0.25 of the information in the Between Runs stratum. Hence, the *variance is not balanced* in the Between Animals Within Runs stratum. In addition, there is no valid F-test for comparing between the treatment groups, because the coefficients of the variance component in the Treatment EMS and Residual EMS are not the same.

Table 3.6: Theoretical ANOVA of the MS-optimal design for the Phase 2 experiment in Table 3.5.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals				
Treatments	1	$\sigma^2 + 0.5\sigma_a^2 + 4\sigma_r^2 + \theta_\tau$		0.17
Residual	1	$\sigma^2 + 0.5\sigma_a^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 1.8\sigma_a^2 + 1.1\theta_\gamma + 1.8\theta_\tau$	0.36	0.29
Treatments	1	$\sigma^2 + 1.7\sigma_a^2 + 3.3\theta_\tau$		0.54
Residual	1	$\sigma^2 + 1.9\sigma_a^2$		
Within Animals				
Tag	3	$\sigma^2 + 1.9\theta_\gamma$	0.63	
Residual	1	σ^2		

Using the objective function for the A-optimality criterion in 3.14, the optimal allocation of sub-samples from animals and treatments labelled by tags and measured in runs is shown in Table 3.7. Notice that Run 1 comprised sub-samples from Animals *C* and *F*, while Runs 2 and 3 comprised sub-samples from Animals *A*, *B*, *D* and *E*. Tags 114 and 117 label sub-samples from Animals *C*, *D* and *E*, and Tags 115 and 116 label sub-samples from Animals *A*, *B* and *F*. The Treatment effects are orthogonal to Run effects, because each run comprises sub-samples from two of each treatment. However, the Treatment effects are still confounded with Tags effects, because the allocation of treatments to tags has a same structure to the optimal design based on the MS-optimality criterion.

Table 3.7: The A-optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consisting of $\nu = 2$ treatments assigned to assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Ca</i>	<i>Fb</i>	<i>Fb</i>	<i>Ca</i>
2	<i>Ea</i>	<i>Aa</i>	<i>Bb</i>	<i>Db</i>
3	<i>Db</i>	<i>Bb</i>	<i>Aa</i>	<i>Ea</i>

The theoretical ANOVA of the A-optimal design is presented in Table 3.8, which again shows a total of 11 DF are first partitioned into 2 DF of Between Runs and 9 DF of Within Runs strata. However, this optimal design only has Animal effects confounded with Run effects, so the Treatment effects can be fully estimated in the Within Run stratum. Notice that there are 2 Residual DF in the Between Animals Within Runs stratum, which means this design can estimate the variances of the Treatment effects more precisely than the optimal design found using the objective function based on MS-optimality. A valid F-test for the Treatment effects can also be performed, because both the Treatment and Residual EMS in the Between Animals Within Runs stratum contain identical coefficients of the variance components. Therefore, the A-optimal design has been shown to be preferable to the MS-optimal design in terms of the allocation of animals to runs and tags.

Table 3.8: Theoretical ANOVA of the A-optimal design for the Phase 2 experiment in Table 3.7.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Residual	1	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 3\theta_\gamma + 0.67\theta_\tau$	1	0.11
Treatments	1	$\sigma^2 + 2\sigma_a^2 + 5.3\theta_\tau$		0.89
Residual	2	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 3\theta_\gamma$	1	
Residual	3	σ^2		

3.6.2 Maximising the treatment information

The objective function based only on the average efficiency factor from the animal information, while optimising the allocation of sub-samples from treatments to runs and tags, does not take into account treatment information. The reason for using E_a is that the treatments are assigned to animals in Phase 1, so some the treatment information is carried with it when samples are differentially labelled and assigned to runs in the Phase 2 experiment. However, on its own, this does not directly result in the optimal allocation of sub-samples from treatments labelled by tags and analysed in runs, so E_τ is added to the objective function to do this. The treatment average efficiency factor is included in the objective function using the weighted average of the

average efficiency factor for Animals and that for Treatments, i.e.

$$O(D_i) = w_a E_a + w_\tau E_\tau, \quad (3.17)$$

where $w_a + w_\tau = 1$.

This subsection compares the optimal designs found using an objective function with two different weighting schemes. The first scheme equally weights the average efficiency factors, i.e. $w_a = w_\tau = 0.5$. The second scheme gives greater weight to E_a , with $w_a = 0.75$ and $w_\tau = 0.25$. To illustrate how these two weighting schemes behave in terms of the optimal designs they generate, consider a two-phase experiment when there are still $\nu = 2$ treatments assigned to $n_a = 4$ animals in the Phase 1 experiment. The design of this Phase 1 experiment consists of Treatment a assigned to Animals A and C , and Treatment b assigned to Animals B and D . The theoretical ANOVA of the Phase 1 experiment shows that all of the treatment information is in the Between Animals stratum (in Table 3.9). Each sample from the Phase 1 experiment is split into $n_s = 3$ sub-samples, and assigned to $n_r = 3$ runs and $n_\gamma = 4$ tags of the Phase 2 experiment.

Table 3.9: Theoretical ANOVA table for the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 4$ animals.

Source of Variation	DF	EMS	E_τ
Between Animals			
Treatment	1	$\sigma_a^2 + 2\theta_\tau$	1
Residual	2	σ_a^2	

Table 3.10 shows the optimal design found by using the objective function which equally weights E_a and E_τ . While the resulting design is satisfactory in the sense that Animal effects are orthogonal to Run effects, i.e. all runs contain a sample from each animal, the allocation of tag labelling to samples is not entirely satisfactory because each tag labels a sample with a different set of animals. Thus, the Animal effects are confounded with Tag effects. This is an example of balanced incomplete block design with the arrangement of sub-samples of animals to tag labels.

The theoretical ANOVA for the design from Table 3.10, given in Table 3.11, shows the total of 11 DF partitioned to 2 DF and 9 DF from Between Runs and Within Runs strata, respectively. Since the Animal and Treatment effects are orthogonal to Run effects, all the animal and treatment information is in the Within Runs stratum. However, it appears all 3 DF with Tag effects are confounded with the 3 DF from the Between Animals Within Runs stratum. Since the Treatment effects are estimated in the Between Animals stratum, the Tag effects are also confounded with Treatment effects. Thus, Treatment effects are not estimable in

Table 3.10: Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags based on the objective function with equal weights for E_a and E_τ , when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 4$ animals, $n_s = 3$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Bb</i>	<i>Aa</i>	<i>Ca</i>	<i>Db</i>
2	<i>Ca</i>	<i>Db</i>	<i>Aa</i>	<i>Bb</i>
3	<i>Aa</i>	<i>Bb</i>	<i>Db</i>	<i>Ca</i>

this design.

Table 3.11: Theoretical ANOVA for the Phase 2 experiment in Table 3.10, where E_a and E_τ share the same weight in the objective function.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs	2	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 3\sigma_a^2 + 0.3\theta_\gamma + 6\theta_\tau$	0.11	1
Within Animals				
Tag	3	$\sigma^2 + 2.7\theta_\gamma$	0.89	
Residual	3	σ^2		

Table 3.12 shows the optimal design found by the objective function with weights $w_a = 0.75$ and $w_\tau = 0.25$. As per the previous scheme, each run comprises a sub-sample from each of the four animals and thus two of each treatment group. The unequal weighting scheme, however, partitions the tag labelling into two parts: each sub-sample from Animals *A*, *B* and *C* is differentially labelled with Tags 114, 115 and 116, while all sub-samples from Animal *D* are differentially labelled with Tag 117. Thus, the contrast Tags (114, 115, 117) versus Tag 116 is confounded with the contrasts Animals (*A*, *B*, *C*) versus Animal *D*, i.e. showing that 1 DF associated with Animal effects is confounded with 1 DF with Tag effects. Further, note that the tag labelling of sub-samples from treatment also has the same structure of partitioning, when Tag 116 only labels the sub-sample for Treatment *b*. Thus, the Treatment effects are also confounded with the Tag effects.

The theoretical ANOVA from the design in Table 3.12, given in Table 3.13, again shows the total of 11 DF partitioned to 2 DF and 9 DF from Between Runs and Within Runs strata, respectively, with all animal and treatment information being estimated in the Within Runs stratum. In the Between Animals Within Runs stratum, there is only 1 DF associated with the

Table 3.12: Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags based on the objective function with $w_a = 0.75$ and $w_\tau = 0.25$ for E_a and E_τ , when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 4$ animals, $n_s = 3$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Ca</i>	<i>Aa</i>	<i>Db</i>	<i>Bb</i>
2	<i>Aa</i>	<i>Bb</i>	<i>Db</i>	<i>Ca</i>
3	<i>Bb</i>	<i>Ca</i>	<i>Db</i>	<i>Aa</i>

Tag effects, due to the confounding of the contrast Tags (114, 115, 117) versus Tag 116 with the contrasts Animals (*A*, *B*, *C*) versus Animal *D*. The Tag effects are still not orthogonal with treatment effects, because 0.33 of the treatment information is confounded with Tag effects, which means there is still 0.67 of pure treatment effects that can be estimated in the Between Animals Within Runs stratum. Finally, a valid F-test for the treatment effects can also be conducted. Therefore, this example shows that the objective function needs to have greater weight to E_a , and we have chosen $w_a = 0.75$ and $w_\tau = 0.25$. The current objective function is

$$O(D_i) = 0.75E_a + 0.25E_\tau. \quad (3.18)$$

Table 3.13: Theoretical ANOVA for the Phase 2 experiment in Table 3.12 where $w_a = 0.75$ and $w_\tau = 0.25$ for E_a and E_τ in the objective function.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs	2	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 3\sigma_a^2 + 3\theta_\gamma + 2\theta_\tau$	1	0.33
Treatments	1	$\sigma^2 + 3\sigma_a^2 + 4\theta_\tau$		0.67
Residual	1	$\sigma^2 + 3\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 3\theta_\gamma$	1	
Residual	4	σ^2		

3.6.3 Maximising intact Treatment DF

The final desirable property we want to incorporate in the objective function is to keep the total DF associated with treatments as intact as possible. This issue can arise when there is *complete* confounding of DF associated with Treatment effects with DF associated with either

Between Runs stratum or DF associated with Tag effects, which results in some DF associated with Treatment effects that cannot be estimated in the Between Animals Within Runs stratum.

This subsection is a continuation of derivation of the objective function. So far in this section, we have put together an objective function in (3.18), which first targets designs that maximise the amount of animal information, E_a , in the Within Runs stratum, and then maximise the amount of treatment information, E_τ , in the Between Animals Within Runs stratum. Finally, to ensure all of the treatment contrasts are, as much as possible, estimated in the Between Animals Within Runs stratum, we add a final component to the objective function, namely the ratio of DF associated with Treatment effects between the Phase 1 and Phase 2 experiments, denoted by $\frac{\nu_2}{\nu_1}$, where ν_1 and ν_2 denote the DF associated with Treatment effects in the Phase 1 and Phase 2 experiments, respectively. The DF associated with Treatment effects in the Phase 1 experiment provides an upper bound, and main purpose of this component is to find a design which allocates treatments in such a way that $\nu_2/\nu_1 = 1$. Thus, the objective function in (3.18) can be extended to

$$O(D_i) = w_a E_a + w_\tau E_\tau + w_\nu \frac{\nu_2}{\nu_1}. \quad (3.19)$$

where w_ν denotes the weight of the proportion of treatment DF in Phase 1 and 2 experiments. The objective function in (3.19) can be viewed as having two parts: animal component ($w_a E_a$) and treatment component ($w_\tau E_\tau + w_\nu \frac{\nu_2}{\nu_1}$). The objective function in (3.19) is first re-written as

$$O(D_i) = 0.75 E_a + 0.25 \left(w_\tau E_\tau + w_\nu \frac{\nu_2}{\nu_1} \right),$$

where the weight for the animal component is maintained as 0.75 of the objective function in (3.18); the weight for the treatment component and treatment DF proportion is then 0.25.

As the number of treatment increases in the experiment, the chance of all contrasts (or DF) associated with the Treatment effects being estimated in the Between Animals Within Runs stratum decreases. Since $w_\tau + w_\nu = 1$, we choose weights of $w_\tau = 1/\nu$ and $w_\nu = (\nu - 1)/\nu$, which suggests as the number of treatment increases, we give greater weight to the part of the objective function that keeps all Treatment DF intact, so the estimation of the effects is more efficient. The objective function then becomes

$$O(D_i) = 0.75 E_a + 0.25 \left(\frac{1}{\nu} E_\tau + \frac{\nu - 1}{\nu} \frac{\nu_2}{\nu_1} \right).$$

Since the Phase 1 experiment is arranged with CRD, the ν_1 is the upper bound of the DF associated with the Treatment effects and thus is always the same as $\nu - 1$, and the objective

function becomes

$$O(D_i) = 0.75E_a + 0.25 \left(\frac{E_\tau + \nu_2}{\nu} \right). \quad (3.20)$$

To show this objective function is effective, we consider a two-phase experiment when the Phase 1 experiment involves $\nu = 3$ treatments assigned to $n_a = 6$ animals. The design of the Phase 1 experiment comprises Treatment a assigned to Animals A and D , Treatment b assigned to Animals B and E , and Treatment c assigned to Animals C and F . The Phase 1 theoretical ANOVA, see Table 3.14, shows there are now 2 DF associated with the Treatment effects.

Table 3.14: Theoretical ANOVA for the Phase 1 experiment with $\nu = 3$ treatments assigned to $n_a = 6$ animals.

Source of Variation	DF	EMS	E_τ
Between Animals			
Treatment	2	$\sigma_a^2 + 2\theta_\tau$	1
Residual	3	σ_a^2	

The samples from the Phase 1 experiment are further split into $n_s = 2$ sub-samples, and each of these 12 sub-samples is measured in the $n_r = 3$ runs of four-plex experiments. The aims of the optimal design of the Phase 2 experiment are to maximise the animal and treatment information, and preserve 2 DF associated with Treatment effects in the Between Animals Within Runs stratum.

Table 3.15 shows the optimal design for the Phase 2 experiment found by using the objective function defined in (3.18). Three groups can be observed in the allocation of animals to runs, when Run 1 contains the sub-samples from Animals B and D , Run 2 contains sub-samples from Animals C and F , and Run 3 contains sub-samples from Animals A and E . In addition, two groups can be observed with respect to the tags when sub-samples from Animals A , B and C are differentially labelled by Tags 114 and 115 while the sub-samples from Animals D , E and F are differentially labelled by Tags 116 and 117. Thus, Animal effects are confounded with both Run and Tag effects. The treatment allocation to runs and tags is unsatisfactory. Although three sub-samples from each treatment group are differentially labelled with the four different tags, thereby yielding orthogonal Treatment and Tag effects, the Treatment effects, which are of greatest interest to researchers, are confounded with runs since Runs 1 and 3 contain Treatments a and b , and Run 2 contains Treatment c .

The theoretical ANOVA for the design in Table 3.15 is shown in Table 3.16. The total of 11 DF is first separated into 2 DF in Between Runs and with 9 DF in Within Runs strata. The 2 DF associated with the Between Runs stratum are further partitioned to 1 DF associated with Treatment effects with 100% of treatment information. This 1 Treatment DF in the Between

Table 3.15: Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags without the maximised DF associated with the Treatment effects, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Bb</i>	<i>Bb</i>	<i>Da</i>	<i>Da</i>
2	<i>Cc</i>	<i>Cc</i>	<i>Fc</i>	<i>Fc</i>
3	<i>Aa</i>	<i>Aa</i>	<i>Eb</i>	<i>Eb</i>

Animals Between Runs stratum is confounding the treatment contrast of Treatments *a* and *b* versus Treatment *c* with Run effects. Thus, the estimation of Treatment effects in this stratum will not be very precise, because the run-to-run variation tend to be large. The other 1 DF associated with the Treatment effects in the Between Animals Within Runs stratum and the estimation of the Treatment effects in this stratum will not include the differences between Treatments *a* and *c* versus Treatment *b*.

Table 3.16: Theoretical ANOVA for the Phase 2 experiment in Table 3.15 without maximization of the DF associated with the Treatment effects.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals				
Treatment	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2 + 4\theta_\tau$		1
Residual	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 3\theta_\gamma$	1	
Treatment	1	$\sigma^2 + 2\sigma_a^2 + 4\theta_\tau$		1
Residual	1	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 3\theta_\gamma$	1	
Residual	4	σ^2		

Table 3.17 shows the optimal design for the Phase 2 experiment based on the three-criterion objective function defined in (3.20). Two groups can be observed in the allocation of animals to runs, when Run 1 contains the sub-samples from Animals *C* and *D*, and Runs 2 and 3 contain the sub-samples from Animals *A*, *B*, *E* and *F*. There are also groups of animals assigned to tags in the same way as the allocation in Table 3.15. The allocation of sub-samples from animals differentially labelled by tags are also in two groups, when Tags 114 and 115 label sub-samples from Animal *A* - *C*, while Tags 116 and 117 label sub-samples from Animal *D* - *F*. The effects

of Animals are still confounded with both Run and Tag effects for the new design. For the treatment allocation to runs and tags (see Table 3.15), there are no run that contains one of each treatment, but each of the four tags does have each of the three treatments. Hence, while the Treatment effects are still confounded with the Run effects, they are orthogonal to the Tag effects.

Table 3.17: Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags with the maximization of the DF associated with the Treatment effects, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 3$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Cc</i>	<i>Cc</i>	<i>Da</i>	<i>Da</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Eb</i>	<i>Fc</i>
3	<i>Aa</i>	<i>Bb</i>	<i>Fc</i>	<i>Eb</i>

The theoretical ANOVA of the new design in Table 3.15 shows that both of the 2 DF associated with the Treatment effects are in the Between Animals Within Runs stratum. However, there is still 1 DF associated with the Treatment effects with 0.25 of the treatment information in the Between Animals Between Runs stratum. The 2 DF associated with the Treatment effects in the Between Animals Within Runs stratum have the canonical efficiency factors of 1 and 0.75, so that $E_\tau = 0.857$. The treatment contrasts that are fully estimated in the Between Animals Within Runs stratum is for Treatment *a* versus *b*, while we only obtain 0.75 of the information from the second contrast of Treatments *a* and *b* versus Treatment *c*. Since the 2 DF for the Treatment effects can be estimated in the Within Runs stratum, the estimation of Treatment effects is more precise than the design with 1 DF associated with Treatment effects. Note that there are valid F-tests for both design, but that in the previous design of this subsection one of the treatment contrasts is likely to be estimated very poorly due to the high run-to-run variation.

The allocation of treatments to runs and tags in Table 3.17 is *connected*, because we cannot split runs or tags into groups in such a way that the treatment in any one group of runs or tags is distinct from the treatment in the other groups of runs or tags. Having a connected design is important, because every treatment contrast is estimable from comparing within runs and not confounded with any one of the tag contrasts. Thus, all the DF associated with the Treatment effects are preserved for estimation and performing the F-test.

Table 3.18: Theoretical ANOVA for the Phase 2 experiment in Table 3.17 with maximization of the DF associated with the Treatment effects.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals				
Treatment	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2 + \theta_\tau$		0.250
Residual	1	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 3\theta_\gamma$	1	
Treatment	2	$\sigma^2 + 2\sigma_a^2 + 3.43\theta_\tau$		0.857
Residual	1	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 3\theta_\gamma$	1	
Residual	3	σ^2		

3.7 Construction of the initial design

The optimal designs found in Section 3.6 were generated using random initial designs, i.e. the animal were randomly assigned to the runs and tags for the initial Phase 2 design. Additionally, the search method used involved swapping random pairs of observational units of the Phase 2 experiments until the objective function value was maximised. However, as the size of the experiment increases, so too does the search space, and, thus, the computational effort needed to find the optimal design. One way to speed up the search time is to start with an initial design having properties that are close to those we anticipate the optimal designs will have, i.e. near-optimal initial designs.

The optimal design resulting from the objective function in (3.20) showed a pattern in the allocation of Animal samples to runs and tags that was common across a range of designs we explored with this objective function. Based on this pattern, we can provide a systematic approach to generate a near-optimal initial designs. Table 3.19 shows allocations of sub-samples of animals to runs and tags, where rows correspond to runs and columns to tags. We have presented generic patterns in the allocation of these samples for three cases, when each case consists of four animals in the Phase 1 experiment and then where the number of runs is equal to the number of sub-samples from each animal in the Phase 2 experiment, i.e. $n_r = n_s = 2, 3$ and 4. The design in Table 3.19 (a) shows the allocation of animals is split into two arrays each comprising two runs and two tags, and then sub-samples from each of one pair of animals are arranged in a Latin square design in the first array, and the remaining sub-samples from the second pair of animals are similarly arranged in a Latin square design in the second array. In

Table 3.19 (b), the three runs by four tags design is partitioned into one array comprised of three runs and three tags, and sub-samples from all 3 animals arranged in a Latin square. The remaining three sub-samples from Animal D are assigned to the three runs by one tag array. The design in Table 3.19 (c) shows the four runs and four tags are partitioned into four 2-run-and-2-tag arrays. The sub-samples from the first pair of animals are arranged in a Latin square design in the two arrays of the first two runs and first two tags, and the last two runs and last two tags. The sub-samples from the second pair of animals are also arranged in a Latin square design in the two arrays of first two runs and last two tags, and the last two runs and first two tags.

Table 3.19: Experiments involving 4 animals in the Phase 1 experiment and (a) two runs and two technical replicates (b) 3 runs and 3 technical replicates and (c) 4 runs and 4 technical replicates in the Phase 2 proteomics experiment using the four-plex iTRAQTM labelling system.

(a)				(b)				(c)			
A	B	C	D	A	B	C	D	A	B	C	D
B	A	D	C	B	C	A	D	B	A	D	C
				C	A	B	D	C	D	A	B
								D	C	B	A

Using the patterns of animal allocation to runs and tags as described in Table 3.19, treatment allocation also needs to be considered when setting up the initial design, with the aim of minimising the confounding Animals and Treatment effects within Run and Tag effects. Consider a Phase 1 experiment that consists of $\nu = 4$ treatments assigned to $n_a = 8$ animals, when Treatment a is assigned to Animals A and E , Treatment b is assigned to Animals B and F , Treatment c is assigned to Animals C and G , and Treatment d is assigned to Animals D and H . Each sample is then further split into $n_s = 2$ sub-samples, and is to be measured in the Phase 2 MudPIT-iTRAQTM experiment of $n_r = 4$ runs and $n_\gamma = 4$ tags.

Two possible initial designs are presented in Tables 3.20 and 3.21. The initial design in Table 3.21 appears to be better than the one in Table 3.20, because the sub-samples from Treatments a and b are only labelled by Tags 114 and 115 and sub-samples from Treatments c and d are only labelled by Tags 116 and 117. Thus, the Treatment effects are confounded with the Tag effects in the initial design shown in Table 3.20, while in the design shown in Table 3.21, sub-samples each of four treatments are labelled by all four tags and four runs, so that the Treatment effects are orthogonal to both Run and Tag effects.

If we examine the allocation of treatments in Tables 3.20 and 3.21 more closely, we can see that sub-samples from each treatment group are labelled twice by each tag in the design of Table 3.20, while sub-samples from each treatment group are labelled once in each tag in the

better design of Table 3.21. Based on this structure, we come up with a systematic approach for constructing an initial design for the Phase 2 experiment on the assignment of animals and treatments to runs and tags: the same treatment groups cannot be allocated to the same runs and tags more than ν/n_γ and ν/n_r times, respectively, while the same animals cannot be allocated to the same runs and tags more than n_a/n_γ and n_a/n_r times, respectively.

Table 3.20: Initial design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_a = 8$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 4$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Dd</i>	<i>Cc</i>
3	<i>Ea</i>	<i>Fb</i>	<i>Gc</i>	<i>Hd</i>
4	<i>Fb</i>	<i>Ea</i>	<i>Hd</i>	<i>Gc</i>

Table 3.21: Different initial design with the same design parameters as Table 3.20.

Run	Tag			
	114	115	116	117
1	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Dd</i>	<i>Cc</i>
3	<i>Gc</i>	<i>Hd</i>	<i>Ea</i>	<i>Fb</i>
4	<i>Hd</i>	<i>Gc</i>	<i>Fb</i>	<i>Ea</i>

3.8 Simulated annealing

Simulated annealing (SA) algorithm is a well-known heuristic method for finding the variables that maximise or minimise a numerical value obtained from an objective function with a suitable optimality criterion (Kirkpatrick et al., 1983). SA algorithm is thus important for continuously comparing the numerical values generated from the objective function of competing designs. Search algorithms have been introduced to resolve the problem of design optimization in a large search space. This Chapter considers the simulated annealing (SA) algorithm, which is inspired by the heating and cooling of metal to alter its physical properties (Kirkpatrick et al., 1983). This section describes SA algorithm in detail and how we use it to find the optimal design of Phase 2 experiment.

SA algorithm was inspired by the effects temperature changes have on metal: as the temperature cools down, the physical properties of the metal become fixed and cannot be changed.

In SA, the temperature is initially set to high and gradually decreases as the algorithm runs. At high temperatures the algorithm accepts solutions that are worse than the current solution, thereby avoiding the situation of being stuck at local optima. As the temperature decreases so does the chance of accepting worse solutions; thus, the algorithm will gradually close in on a globally optimum solution. This temperature change is what makes SA algorithm effective in finding an optimal solution across a large search space.

3.8.1 Acceptance probability

Acceptance probability is the probability of accepting a current design during the searching procedure. Consider a new design, D_1 , which can be generated by swapping any two animal IDs from the initial design, D_0 . A single swap is considered as one iteration. The current design at the i -th iteration is denoted by D_i and the previous design at the $(i - 1)$ -th iteration is denoted by D_{i-1} . Whenever the current design is better than the previous one, i.e.

$$O(D_i) > O(D_{i-1}),$$

where $O(D_i)$ denotes the value of the objective function for the design at the i -th iteration, the swap is accepted and in the next iteration the swapping process is carried out on the accepted design. Worse designs may also be accepted, with acceptance probability of the current design, D_i , given by

$$P(D_i, D_{i-1}, t_i) = \exp \left[\frac{O(D_i) - O(D_{i-1})}{t_i} \right], \quad (3.21)$$

where t_i denotes the annealing temperature at the i -th iteration, with higher temperatures allowing the algorithm to more frequently accept worse designs than the current solution. This probability is then compared to a randomly generated value between zero and one, and if it is lower than the random number, the current design is rejected and the previous design is retained for the next iteration. If the acceptance probability exceeds the random number, the current design is accepted for the next iteration. Since the algorithm sometimes accepts worse designs, the best overall design found is also stored. The overall best design is returned at the end of the search.

3.8.2 Initial and final temperatures

Temperature plays an important role in determining the acceptance probability as shown in (3.21), a poorly defined temperature will decrease the efficiency of using the SA algorithm. We can show that the SA algorithm's temperature is directly related to the difference between the

values of the objective function for the previous and current designs. If at the i -th iteration, this difference equals the negative value of the temperature, i.e.

$$O(D_i) - O(D_{i-1}) = -t_i,$$

then the acceptance probability given by

$$\exp \left[\frac{O(D_i) - O(D_{i-1})}{t_i} \right] = \exp \left(\frac{-t_i}{t_i} \right) = \exp(-1) = 0.3679.$$

There still exists a 0.3679 probability that D_i will be accepted during the search. Thus, the temperature can be determined based on the range of values calculated for the objective function from a set of designs (Whitaker et al., 1990).

Whitaker et al. (1990) described a procedure for determining the initial and final temperatures for the SA algorithm. Their method is to first generate a set of random designs and calculate values from the objective function. The initial temperature is then determined from the range of these values, i.e. the difference between the largest and smallest values. The final temperature is calculated as the difference between the largest and second largest values. Finally, the design with the highest value from the objective function is used as the initial design for SA.

3.9 Nested simulated annealing

John and Whitaker (1993) mentioned that the convergence of the SA algorithm can be very slow and the design found far from optimal. They describe a method of separating the entire process of SA algorithm into 10 levels, with each level performed one after the other. The process of how the temperature is rapidly reducing in one level after another is called *accelerated cooling* and applying SA algorithm is called *nested simulated annealing*.

The nested SA algorithm starts a random walk across a search space with a high temperature to diversify the search in order to escape local optima. The accelerated cooling allows the search to intensify as it becomes more local. As the random walk gradually becomes more confined, following the contours of the search space, the chance of accepting worse designs decreases. The nested SA algorithm for finding a good design requires quickly reducing the temperature across all levels sequentially while still carrying out SA algorithm at each temperature level. Let $t_{(l)(i)}$ denote the temperature at i -th iteration of level l , then the initial temperature at level l is denoted by $t_{(l)(0)}$, which we have shown in Section 3.8.2. The initial temperature at the next level is then given by

$$t_{(l+1)(0)} = \frac{t_{(l)(0)}}{c}, \tag{3.22}$$

where c denotes the factor by which the temperature is reduced from one temperature level to the next. [Whitaker et al. \(1990\)](#) set the value of c between 4 and 10, to achieve substantial temperature reductions.

Since the overall initial and final temperatures have been calculated as described in Section 3.8.2, we apply this overall initial temperature as the initial temperature of the second level, denoted by $t_{2(0)}$, and the overall final temperature as the initial temperature of the final level, denoted by $t_{10(0)}$. We can then calculate c in (3.22) as

$$c = \exp \left[\frac{\log(t_{2(0)}/t_{10(0)})}{10 - 2} \right].$$

Note the nested SA algorithm process only applies from level 2 to 10, because the first level is to compute the overall initial and final temperatures.

3.10 Modified simulated annealing

While the nested SA algorithm approach described in Section 3.9 works well, i.e. identifies designs with good properties, it is still slow. This section describes two further improvements in the swapping procedures of the nested SA algorithm.

3.10.1 Swapping method for two or more technical replicates

The current swapping method is applied to any two random observational units of the Phase 1 experiment in the allocation to runs and tags of the Phase 2 experiment. However, the initial design proposed in Section 3.7 for two or more technical replicates may not improve the design when swapping just two animal IDs, because the structure of the initial design is close to the optimal design. Hence, the swapping method is adjusted by swapping any two random *sets* of animal IDs of identical technical replicates, i.e. the experimental units of the Phase 1 experiment. The purpose of this adjustment is to preserve the structure of the initial design, defined in Section 3.7, while generating a new candidate design for comparison.

3.10.2 Three-stage swapping method

The swapping method of SA algorithm can be improved further still by using a three-stage swapping method, which is to divide a single large search space into three smaller ones. [Williams and John \(1996\)](#) describe the use of a swapping procedure employed in two stages in order to identify optimal row-column designs. For the MudPIT-iTRAQTM experiment, the runs and tags

are considered to be the rows and columns, respectively. The first stage is to apply swapping of animals and treatments within the runs and find the optimal design using the nested SA algorithm, followed by the second stage involving the swapping of animals and treatments within the same tags and finding the optimal design again using the nested SA algorithm.

This two-stage procedure increases coverage of finding the design with better properties than those found by using the single-step swapping procedure. However, it is clear that a two-stage procedure still has limited coverage of the search space. For example, a better design may be identified from a single swap of animals and treatments that do not belong to the same runs and tags, instead of performing two swaps of animals and treatments first within runs and then within tags. Thus, an additional swapping procedure was trialled, namely swapping the sub-samples of animals and treatments that do not belong to the same runs and tags, which proved to yield better coverage than both the one- and two-stage procedures. The differences between the procedures used in the three swapping steps mean that their corresponding search spaces are also different, and each thereby requires separate computation of its first-level starting and final temperatures. From level 2 onwards, the best design found from the one stage is used as the initial design of the next stage within each level of SA. The best design found after the final nested annealing level is considered as the overall optimal design.

3.11 An illustrative example using the four-plex iTRAQTM system

This section shows the optimal design when the Phase 2 experiment uses the four-plex iTRAQTM system with the objective function defined in (3.20), and the modified nested SA algorithm describe in Section 3.10. Consider a Phase 1 experiment consisting of $\nu = 6$ treatments assigned to $n_a = 18$ animals. The theoretical ANOVA of the Phase 1 experiment shows that all treatment information is in the Between Animals stratum (see Table 3.22). There are 5 DF associated with the Treatment effects, and so 12 DF remain associated with the Residual EMS in the Between Animals stratum. The aim is to find a design for the Phase 2 experiment that preserves, as much as possible, the amount of treatment information and the DF associated with Residual EMS from the Phase 1 experiment.

In the Phase 2 proteomics experiment, $n_s = 2$ sub-samples are taken from each sample of Phase 1 experiment, and these sub-samples are measured in the MudPIT-iTRAQTM experiment comprising $n_r = 9$ runs and $n_\gamma = 4$ tags. Using the objective function defined in (3.20) and the

Table 3.22: Theoretical ANOVA for the Phase 1 experiment with $\nu = 6$ treatments assigned to $n_a = 18$ animals.

Source of Variation	DF	EMS	E_τ
Between Animals			
Treatment	5	$\sigma_a^2 + 6\theta_\tau$	1
Residual	12	σ_a^2	

modified SA algorithm described in Section 3.10, the optimal design is given in Table 3.23. The final value of the objective function is 0.9932. This final design, given in Table 3.23, consists of eight 2-run-and-2-tag arrays from Run 1 to 8 and Tags 114 to 117, when each array has a Latin square arrangement for the animals and treatments. In addition, the samples from last pair of animals with the treatment, i.e. Animals J and Q and Treatments d and e , are assigned to Run 9. This structure has been described under the method for constructing an initial design in Section 3.7.

Table 3.23: Optimal design for Phase 2 experiment showing the allocation of sub-samples from treatments assigned to animals based on objective function value of 0.9932, when the Phase 1 experiment consists of $\nu = 6$ treatments assigned to $n_a = 18$ animals, $n_s = 2$ sub-samples are then taken from each animals and measured in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 9$ runs and $n_\gamma = 4$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	<i>Rf</i>	<i>Pd</i>	<i>Oc</i>	<i>Aa</i>
2	<i>Pd</i>	<i>Rf</i>	<i>Aa</i>	<i>Oc</i>
3	<i>Ic</i>	<i>Ke</i>	<i>Bb</i>	<i>Ma</i>
4	<i>Ke</i>	<i>Ic</i>	<i>Ma</i>	<i>Bb</i>
5	<i>Hb</i>	<i>Cc</i>	<i>Ee</i>	<i>Ff</i>
6	<i>Cc</i>	<i>Hb</i>	<i>Ff</i>	<i>Ee</i>
7	<i>Ga</i>	<i>Na</i>	<i>Lf</i>	<i>Dd</i>
8	<i>Nb</i>	<i>Ga</i>	<i>Dd</i>	<i>Lf</i>
9	<i>Jd</i>	<i>Jd</i>	<i>Qe</i>	<i>Qe</i>

Table 3.24 shows the theoretical ANOVA for this optimal design. The total of 35 DF is separated into 8 DF from Between Runs stratum and 27 DF from Within Runs stratum. The average efficiency factor for Treatment effects $E_\tau = 0.8370$ is computed from the five treatment canonical efficiency factors of 11/12, 11/12, 8/9, 3/4 and 3/4 for each of the canonical treatment contrasts in the Between Animals Within Runs stratum. Thus, the allocation of sub-samples from treatment to runs is not balanced, because the canonical efficiency factors are not identical. However, most of the treatment information is in the targeted stratum of Between Animals Within Runs. Additionally, the Residual DF decreases from 12 DF to 7 DF. The 5 DF associated

with Animal effects are lost due to the 4 DF now being in the Between Runs stratum, and 1 DF is confounded with the Tag effects. However, a valid F-test for detecting the treatment differences is still available in the Between Animals Within Runs stratum.

Table 3.24: Theoretical ANOVA from the optimal design for Phase 2 experiment in Table 3.23 based on objective function value of 0.9932.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals				
Treatment	4	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2 + 0.75\theta_\tau$		0.125
Residual	4	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 9\theta_\gamma + 0.67\theta_\tau$	1	0.111
Treatment	5	$\sigma^2 + 2\sigma_a^2 + 5.02\theta_\tau$		0.837
Residual	7	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 9\theta_\gamma$	1	
Residual	12	σ^2		

3.12 An illustrative example using the eight-plex iTRAQTM system

This section shows how the objective function defined in (3.20) and the modified nested SA algorithm described in Section 3.10 can also look for the optimal design when the Phase 2 experiment uses the eight-plex iTRAQTM system. This section will also compare the theoretical ANOVA of the Phase 2 experiment of the initial design and the optimal design.

Consider a Phase 1 experiment that consists of $\nu = 8$ treatments assigned to $n_a = 16$ animals. The theoretical ANOVA for the Phase 1 experiment, in Table 3.25, shows the total of 15 DF is separated to 7 DF associated with the Treatment effects and 8 DF for Residual EMS.

Table 3.25: Theoretical ANOVA for a Phase 1 experiment with $\nu = 8$ treatments assigned to $n_a = 16$ animals.

Source of Variation	DF	EMS	E_τ
Between Animals			
Treatment	7	$\sigma_a^2 + 2\theta_\tau$	1
Residual	8	σ_a^2	

Since each sample from the Phase 1 experiment is split into $n_s = 2$ sub-samples, there are a

total of 32 sub-samples that can be measured with four runs of the eight-flex MudPIT-iTRAQTM experiment. Table 3.26 shows the allocation of sub-samples of animals and treatments differentially labelled by tags and measured in runs in the initial design using the method described in Section 3.7. This initial design consists of eight 2-run-and-2-tag arrays, which has a Latin square arrangement of animals and treatments to runs and tags.

Table 3.26: Initial design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 8$ treatments assigned to $n_a = 16$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 4$ runs and $n_\gamma = 8$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>	<i>Ee</i>	<i>Ff</i>	<i>Gg</i>	<i>Hh</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Dd</i>	<i>Cc</i>	<i>Ff</i>	<i>Ee</i>	<i>Hh</i>	<i>Gg</i>
3	<i>Kc</i>	<i>Ld</i>	<i>Ia</i>	<i>Jb</i>	<i>Og</i>	<i>Ph</i>	<i>Me</i>	<i>Nf</i>
4	<i>Ld</i>	<i>Kc</i>	<i>Jb</i>	<i>Ia</i>	<i>Ph</i>	<i>Og</i>	<i>Nf</i>	<i>Me</i>

The theoretical ANOVA of the initial design shows there is 100% of the treatment information in the Between Animals Within Runs stratum (see Table 3.27). However, this is based on 6 DF compared to 7 DF in Table 3.25. For the analysis of the initial design, one of the 7 DF associated with the Treatment effects is completely confounded with the Tag effects. Therefore, the allocation of treatment to tags is considered as disconnected, because one treatment contrast cannot be estimated.

Table 3.27: Theoretical ANOVA of the initial design for the Phase 2 experiment in Table 3.26.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 8\sigma_r^2$		
Residual	2	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma + 4\theta_\tau$	1	1
Treatment	6	$\sigma^2 + 2\sigma_a^2 + 4\theta_\tau$		1
Residual	5	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	4	$\sigma^2 + 4\theta_\gamma$	1	
Residual	10	σ^2		

Using the objective function defined in (3.20) and the modified nested SA algorithm describe in Section 3.10, the optimal design is found and presented in Table 3.28. This optimal design still preserved the structure of the initial design in Table 3.26 with the eight 2- run-and-2-tag

arrays.

Table 3.28: Optimal design for the Phase 2 experiment showing assignment of animals and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 8$ treatments assigned to $n_a = 16$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 4$ runs and $n_\gamma = 8$ tags. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Og</i>	<i>Kc</i>	<i>Jb</i>	<i>Ph</i>	<i>Ia</i>	<i>Ld</i>	<i>Nf</i>	<i>Me</i>
2	<i>Kc</i>	<i>Og</i>	<i>Ph</i>	<i>Jb</i>	<i>Ld</i>	<i>Ia</i>	<i>Me</i>	<i>Nf</i>
3	<i>Hh</i>	<i>Ff</i>	<i>Cc</i>	<i>Aa</i>	<i>Ee</i>	<i>Gg</i>	<i>Bb</i>	<i>Dd</i>
4	<i>Ff</i>	<i>Hh</i>	<i>Aa</i>	<i>Cc</i>	<i>Gg</i>	<i>Ee</i>	<i>Dd</i>	<i>Bb</i>

Table 3.29 shows the theoretical ANOVA of the optimal design where all 7 DF associated with the Treatment effects are in the Between Animals Within Runs stratum. Thus, all treatment contrasts can be estimated in the stratum with minimum error. However, the DF associated with the Residual EMS of the Between Animals Within Runs stratum are reduced to 4 (from 5 of the initial design). The E_τ is also reduced to 0.8077, which is computed from seven canonical efficiency factors of 1, 1, 1, 0.75, 0.75 and 0.5.

Table 3.29: Theoretical ANOVA of the optimal design for the Phase 2 experiment in Table 3.28.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 8\sigma_r^2$		
Residual	2	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma + 1.2\theta_\tau$	1	0.3
Treatment	7	$\sigma^2 + 2\sigma_a^2 + 3.23\theta_\tau$		0.8077
Residual	4	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	4	$\sigma^2 + 4\theta_\gamma$	1	
Residual	10	σ^2		

Comparing the amount of treatment information between these two designs, the initial design provides 100% of the treatment information for the 6 DF associated with the Treatment effects, while the optimal design provides 7 DF associated with the Treatment effects. Out of these 7 DF, only 4 DF contain 100% of the treatment information, and the remaining 1 and 2 DF consist of 50% and 75% of the treatment information, respectively.

The next step is to further examine how the treatment groups compare, by extracting a set of basic contrasts which are the eigenvectors of the information matrix. For the initial design, a

matrix of six basic contrasts, denoted by $T1 - T6$, – where columns correspond to 6 DF associated with the Treatment effects and rows correspond to Treatments a to h – can be expressed as

$$\begin{array}{c} \begin{matrix} & T1 & T2 & T3 & T4 & T5 & T6 \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{matrix} \left(\begin{array}{cccccc} 0.0000 & 0.0000 & 0.0000 & 0.8660 & 0.0000 & 0.0000 \\ 0.0000 & -0.2677 & 0.0000 & -0.2887 & 0.0000 & 0.7714 \\ 0.0000 & -0.5342 & 0.0000 & -0.2887 & 0.0000 & -0.6175 \\ 0.0000 & 0.8019 & 0.0000 & -0.2887 & 0.0000 & -0.1538 \\ - & - & - & - & - & - \\ 0.0000 & 0.0000 & 0.8660 & 0.0000 & 0.0000 & 0.0000 \\ -0.2677 & 0.0000 & -0.2887 & 0.0000 & 0.7714 & 0.0000 \\ -0.5342 & 0.0000 & -0.2887 & 0.0000 & -0.6175 & 0.0000 \\ 0.8019 & 0.0000 & -0.2887 & 0.0000 & -0.1538 & 0.0000 \end{array} \right) \end{array}.$$

Note that all six contrasts have 100% of the treatment information. The treatment comparisons show two distinct groups of Treatments a, b, c and d , and Treatments e, f and g . This is because the comparison between these two groups of treatments is completely confounded with the Tag effects. For the optimal design, a matrix of seven basic contrasts, denoted by $T1 - T7$, – where columns corresponds to 7 DF associated with the Treatment effects and rows corresponds to Treatments a to h – can be expressed as

$$\begin{array}{c} \begin{matrix} & T1 & T2 & T3 & T4 & T5 & T6 & T7 \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \\ g \\ h \end{matrix} \left(\begin{array}{cccccc} 0.0000 & 0.0000 & 0.0000 & 0.6124 & 0.0000 & 0.7071 & 0.0000 \\ -0.3800 & 0.4095 & 0.1460 & -0.2041 & 0.7071 & 0.0000 & 0.0000 \\ 0.4712 & 0.2285 & -0.5559 & -0.2041 & 0.0000 & 0.0000 & 0.5000 \\ 0.6420 & 0.0420 & 0.4116 & -0.2041 & 0.0000 & 0.0000 & -0.5000 \\ -0.2621 & -0.4514 & -0.5575 & -0.2041 & 0.0000 & 0.0000 & -0.5000 \\ 0.0000 & 0.0000 & 0.0000 & 0.6124 & 0.0000 & -0.7071 & 0.0000 \\ -0.3800 & 0.4095 & 0.1460 & -0.2041 & -0.7071 & 0.0000 & 0.0000 \\ -0.0913 & -0.6379 & 0.4100 & -0.2041 & 0.0000 & 0.0000 & 0.5000 \end{array} \right) \end{array}.$$

The first four contrasts contain 100% of the treatment information, while Contrasts $T5$ and $T6$ contain 75% of the treatment information, and Contrast $T7$ has 50% of the treatment information. However, no distinct groups of treatments can be identified from these basic contrasts. Hence, this optimal design appears to be better, because the treatment allocation to runs and tags is connected, and all the treatment contrast are compared in the Between Animals Within

Runs stratum.

3.13 Optimal designs for experiments involving two to eight treatments and two technical replicates

A table of optimal designs for a range of design parameters are generated using the objective function described in (3.19) and the simulated annealing algorithm described in Section 3.10. This set of optimal designs is for experiments with $\nu = 2, \dots, 8$ treatments, $n_a = \nu r_b$ animals, $n_s = 2$ sub-samples, $n_\gamma = 4, 8$ tags, and $n_r = n/n_\gamma$ runs, where r_b denotes the number of biological replicates and n denotes total number of sub-samples, ($r_b = 2, \dots, 8$). This set of optimal designs is presented in Appendix C. If a biologist has an experiment with a specific set of design parameters, they can use the given design for their experiment. A set of tables, summarising the properties of each optimal design of the Phase 2 experiment, is presented in Appendix D. This section discusses the properties of the optimal designs found.

The general layout of each design is in the form of a two-way table comprising n/n_γ rows and n_γ columns, where n_γ can be four for the four-plex experiment or eight for the eight-plex experiment. We assume that one sub-sample is differentially labelled exactly once with one tag and analysed in only one run. As previously discussed, iTRAQTM is available in four-plex or eight-plex reagent kits. So, it is possible to analyse as few as two and up to eight differentially labelled proteomics samples per MudPIT run. Generally, however, because the cost of these kits is substantially more than the cost of obtaining biological samples, biologists will run experiments utilising each kit in its entirety. For this reason, here we consider only Phase 2 experiments with runs of size 4 and 8; although, the methods presented in this Chapter may also be used to generate optimal designs for runs of size 2 through 8. It follows from this assumption that the four-plex experiment can only measure samples from the Phase 1 experiment where the number of animals (experimental units) is even, without leaving any reagent unused. For example, consider an experiment with $n_a = 9$ animals, and taking two sub-samples from each animal yielding a total of $n = 6$ sub-samples. If a four-plex iTRAQTM reagent kit is used for the Phase 2 proteomics experiment, five runs will be required to measure all 18 sub-samples. This means that two reagents across the requisite five kits will remain unused and potentially be wasted. There are other possible ways to use these two remaining reagents by taking additional sub-sample from two animal-treatment group combinations of the Phase 1 experiment. Alternatively, two additional animals each treated with one each of the three treatments at Phase 1 could be used

without taking technical replicates in the Phase 2 experiment. These are interesting possibilities, but are beyond the scope of this thesis.

As for the eight-plex experiment, the Phase 2 experiment can only measure samples from the Phase 1 experiment where the number of animals (experimental units) is divisible by four. If we again consider the same experiment with $n_a = 9$ animals and n_s technical replicates, the 18 sub-samples from the Phase 1 experiment can be measured by three runs of the eight-plex experiment with six iTRAQTM reagents unused. Therefore, the number of animals (experimental units) from the Phase 1 experiment, assuming the biologists will utilise all iTRAQTM reagents, can influence the choice between the four-plex and eight-plex experiments.

3.13.1 Phase 1 Animal (Experimental unit) effects

The Between Animals Within Runs stratum is an important stratum to examine, because treatments are assigned to animals (experimental units) in the Phase 1 experiment, i.e. the animals are the experimental units. Thus, maximising the animal information for the Phase 2 experiment will allow us to have more precise estimation of the Treatment effects.

As the number of animals becomes larger than the number of tags used in the Phase 2 experiment, Animal effects will be confounded with Run effects. This means that some DF associated with the Between Animals stratum will be in the Between Runs stratum. Thus, as the number of animals used increases, there will be more DF associated with Animal effects in the Between Runs stratum. However, since the weight given to the animal average efficiency factor, E_a , in the objective function in (3.20) is high, *all* the identified optimal designs equal 1. Thus, every DF associated with the Animal effect that is still in the Within Runs stratum contains 100% of the animal information. This means the variances are balanced in the Between Animals Within Runs stratum, and allows us to have precise estimation of the Treatment effects and with a valid F-test.

In the four-plex experiment, there are 3 DF associated with Tag effects. Due to the structure of the initial design described in Section 3.7, these 3 DF are always split between two strata: with 1 DF going into the Between Animals Within Runs stratum and 2 DF going into the Within Animals Within Runs stratum. Using the example in Section 3.11, sub-samples from one half of the animals are labelled by the first two tags, and sub-samples from the other half of the animals are labelled by the last two tags. Thus, one basic contrast associated with Tag effects will always be confounded with Animal effects in the Phase 2 experiments using four-plex labelling system. In Phase 2 proteomics experiments using the eight-plex labelling system, there are 7 DF associated with Tag effects. Again due to the structure of the initial design described

in Section 3.7, these 7 DF are always split between two strata: with 3 DF going into the Between Animals Within Runs stratum and 4 DF going into the Within Animals Within Runs stratum. Using the example in Section 3.12, we can see that two tags always label eight sub-samples from four of 12 animals, i.e. Tags 113 and 114 label the sub-samples from Animals *F*, *H*, *K* and *O*, and so on. We can see that the animals are evenly split into four groups, when sub-samples from each group of animals are labelled by two of eight tags of the experiments. Thus, three basic contrasts associated with Tag effects are always confounded with Animal effects in the Phase 2 experiments using the eight-plex labelling system.

From reviewing tables of properties of each optimal design of the Phase 2 experiment in Appendix D. We can see that, given a same Phase 1 experiment is used, if the Phase 1 experiment consists of fewer than 16 animals (experimental units), the optimal design of four-plex system has higher Residual DF in the Between Animals Within Runs stratum than the eight-plex system. Furthermore, given a same Phase 1 experiment is used, if the Phase 1 experiment consists of more than 24 animals (experimental units), the optimal design of eight-plex system has higher Residual DF in the Between Animals Within Runs stratum than the four-plex system. Therefore, for the Phase 1 experiment with low numbers of animals (experimental units), it is more preferred to use the four-plex system instead of the eight-plex system, due to the two extra DF available in the Between Animals Within Runs stratum. However, when more Phase 1 animals (experimental units) are used, the degrees of confounded between the Animal effects and Run effects increases in the Phase 2 experiment. Thus, it becomes more preferable to use eight-plex system over the four-plex system.

3.13.2 Treatment effects

Confounding of Treatment effects with either Tag or Run effects, or both, can occur in optimal designs found using the objective function and the modified nested SA algorithm. Optimal designs that are found such that Treatment effects have some degree of confounding with Tag effects have a pattern of the number of runs not being divisible by the number of treatments. Such cases can be seen in the optimal designs found in Section 3.12 and 3.11. Optimal designs that are found such that Treatment effects have some degree of confounding with Run effects have a pattern of the number of tags not being divisible by the number of treatments. Such case can be seen in the optimal design found in Section 3.11.

Optimal designs found in Section 3.11 and 3.12 also demonstrate examples when the allocations of sub-samples from treatments to tags and runs can be unbalanced. Thus, for these optimal designs, the treatment canonical efficiency factors are not identical, and thus some treat-

ment contrasts contain more treatment information than other treatment contrasts. It becomes important to study the basic treatment contrasts on how the comparison is made. Appendix D lists both the canonical and average efficiency factors of the treatment effects for every optimal design.

From reviewing tables of properties of each optimal design of the Phase 2 experiment in Appendix D. We can see that, in general, optimal designs from the eight-plex system consists of higher treatment average efficiency factor than the four-plex system. There are some exceptions of cases when Phase 1 experiment consists of $\nu = 2$ and $\nu = 4$ treatments, their optimal designs are better with four-plex system than eight-plex system, due to the confounding between the Treatment and Tag effects.

3.14 Extension to more than one treatment factor in the Phase 1 experiment

In practice, biologists tend to be interested in more than one treatment factor in their proteomics experiments. For example, interest may lie in whether proteins are differentially abundant between healthy and diseased (disease status) animals when they are on normal versus high fat diets. This experiment is also known as a 2×2 factorial experiment. The methods described earlier in this Chapter for finding the optimal design can be extended to when the Phase 1 experiment has a factorial treatment structure. This section shows how to modify these methods to search for an optimal designs for the Phase 2 proteomics experiment featuring a 2×2 factorial Phase 1 experiment.

The main adjustments in extending the methods described for single-factor experiments involve the modification of the objective function to take account of the factorial treatment structure. Since the Phase 1 experiment involves additional treatment factors, the information matrix associated with each of the treatment factors needs to be defined. These information matrices are essential in calculating the treatment average efficiency factors for the objective function. The main focus of this section thus is on the generation of these new treatment information matrices.

3.14.1 Model and design parameters

Consider the Phase 1 experiment with “diseased and healthy” disease statuses and “normal and high fat” diet types which are randomly assigned to n_a animals; thus, there are a total of $n_a = 4r_b$

animals from the Phase 1 experiment. Let y_{ihj} denote the abundance level of protein from animal j under disease status i and diet type h . The linear model for the Phase 1 experiment can be written as

$$y_{ihj} = \mu + \tau_i + \rho_h + (\tau\rho)_{ih} + a_{ihj}, \quad (3.23)$$

where μ denotes the grand mean of all observations, and τ_i , ρ_h and $(\tau\rho)_{ih}$ represent the fixed effects of disease status i , diet type j and interaction, respectively. The $a_{ihj} \sim N(0, \sigma_a^2)$ denotes the random effect from animal ihj , ($i = \text{diseased, health}$; $h = \text{normal, high fat}$; $j = 1, \dots, r_b$).

Each sample from the Phase 1 experiment is further split into n_s sub-samples, when each sub-sample is to be differentially labelled by n_γ tags, and analysed in the Phase 2 MudPIT-iTRAQTM experiment. Now let y_{kls}^{ihj} denote the abundance level of the same protein as in (3.23) from s sub-sample of animal j under disease status i and diet type h , which then differentially labelled by tag k and analysed in run l . The linear model of the Phase 2 experiment can be written as

$$y_{kls}^{ihj} = \mu + \tau_i + \rho_h + (\tau\rho)_{ih} + a_{ihj} + \gamma_k + r_l + \epsilon_{kls}^{ihj} \quad (3.24)$$

where γ_k denotes the fixed effect of tag k , $r_l \sim N(0, \sigma_r^2)$ denotes the run effect, and ϵ_{kls}^{ihj} denotes the experimental error, ($k = 1 \dots n_\gamma$; $l = 1 \dots n_r$; $s = 1, \dots, n_s$).

The model in (3.24) can then be rewritten as matrix notation as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_{\tau\rho}\boldsymbol{\alpha}_{\tau\rho} + \mathbf{Z}_a\mathbf{u} + \mathbf{X}_\gamma\boldsymbol{\alpha}_\gamma + \mathbf{W}_r\mathbf{v} + \boldsymbol{\epsilon}, \quad (3.25)$$

where $\boldsymbol{\alpha}_{\tau\rho}$ and $\boldsymbol{\alpha}_\gamma$ are the vectors Phase 1 treatment and tag fixed effects parameters, respectively. The vector $\boldsymbol{\alpha}_{\delta\rho}$ can be expressed as

$$(\tau_{\text{diseased}} : \rho_{\text{normal}}, \tau_{\text{diseased}} : \rho_{\text{highfat}}, \tau_{\text{healthy}} : \rho_{\text{normal}}, \tau_{\text{healthy}} : \rho_{\text{highfat}}). \quad (3.26)$$

The matrices $\mathbf{X}_{\tau\rho}$ and \mathbf{X}_γ denote $n \times 4$ and $n \times n_\gamma$ treatment and tag design matrices, respectively.

The treatment parameter $\boldsymbol{\alpha}_{\tau\rho}$ can also be written in matrix notation as

$$\boldsymbol{\alpha}_{\tau\rho} = \mathbf{C}_{00}\boldsymbol{\alpha}_{\tau\rho} + \mathbf{C}_{10}\boldsymbol{\alpha}_{\tau\rho} + \mathbf{C}_{01}\boldsymbol{\alpha}_{\tau\rho} + \mathbf{C}_{11}\boldsymbol{\alpha}_{\tau\rho} \quad (3.27)$$

where $\mathbf{C}_{00} = \mathbf{K}_4$ is an averaging matrix. The remaining matrices $\mathbf{C}_{10} = (\mathbf{I}_2 - \mathbf{K}_2) \otimes \mathbf{K}_4$, $\mathbf{C}_{01} = \mathbf{K}_2 \otimes (\mathbf{I}_2 - \mathbf{K}_2)$ and $\mathbf{C}_{11} = (\mathbf{I}_2 - \mathbf{K}_2) \otimes (\mathbf{I}_2 - \mathbf{K}_2)$ are the treatment projection matrices for Disease status effects, Diet effects and Interaction effects. These projection matrices, \mathbf{C}_{10} , \mathbf{C}_{01} and \mathbf{C}_{11} , are then used to partition the overall treatment information matrix to three information matrices for each of the treatment effects.

3.14.2 Information matrix

The information matrix associated with the Animal effects, \mathbf{A}_a , for computing the average efficiency factors remains the same as in (3.9), i.e.

$$\mathbf{A}_a = \mathbf{Z}'_a \mathbf{Q}_{r\gamma} \mathbf{Z}_a.$$

Using the treatment projection matrices defined in (3.27), we can partition the overall treatment information matrix $\mathbf{X}'_{\tau\rho} \mathbf{Q}_{r\gamma} \mathbf{X}_{\tau\rho}$ for each of three treatment effects, namely, Disease status, Diet, and Interaction effects as

$$\mathbf{A}_\delta = \mathbf{C}_{10} \mathbf{X}'_{\tau\rho} \mathbf{Q}_{r\gamma} \mathbf{X}_{\tau\rho} \mathbf{C}_{10}, \quad (3.28)$$

$$\mathbf{A}_\rho = \mathbf{C}_{01} \mathbf{X}'_{\tau\rho} \mathbf{Q}_{r\gamma} \mathbf{X}_{\tau\rho} \mathbf{C}_{01}, \quad (3.29)$$

$$\mathbf{A}_{\delta\rho} = \mathbf{C}_{11} \mathbf{X}'_{\tau\rho} \mathbf{Q}_{r\gamma} \mathbf{X}_{\tau\rho} \mathbf{C}_{11}. \quad (3.30)$$

3.14.3 Objective function

The average efficiency factors for the Disease status, Diet and Interaction effects associated with the Between Animals Within Runs stratum can be calculated from the harmonic mean of the canonical efficiency factors. These canonical efficiency factors are derived using the eigenvalue decomposition of information matrices method defined in (3.28), (3.29) and (3.30). Additionally, the DF associated with each of these treatment effects are maximised in the objective function.

The objective function can then be constructed based on the importance of each treatment effect. If all treatment effects are equally important, the same weights can be assigned to each treatment average efficiency factor. The weights of all treatment components are thus identical, whereas the weights for the Disease status, Diet and Interaction effects are each set at 1/3. The objective function can then be expressed as

$$\frac{3}{4}E_a + \frac{1}{4} \left[\frac{1}{3} \left(\frac{\nu_\tau + E_\tau}{2} \right) + \frac{1}{3} \left(\frac{\nu_\rho + E_\rho}{2} \right) + \frac{1}{3} \left(\frac{\nu_{\tau\rho} + E_{\tau\rho}}{4} \right) \right], \quad (3.31)$$

where E_τ , E_ρ and $E_{\tau\rho}$, and ν_τ , ν_ρ and $\nu_{\tau\rho}$ are the average efficiency factors and DF associated with Disease status, Diet, and Interaction effects in the Between Animals Within Runs stratum, respectively. Therefore, with some slight modifications in the objective function presented in (3.31), the optimal design of Phase 2 experiments can be identified when the Phase 1 experiment is a factorial experiment.

3.15 Summary

This Chapter describes a computational approach for finding optimal designs for Phase 2 proteomics experiments using MudPIT-iTRAQTM technologies when the Phase 1 experiment is arranged in a CRD. A three-criterion objective function is derived for generating the optimal design with three properties: 1) animal information is maximised in the Within Runs stratum, 2) treatment information is maximised in the Between Animals Within Runs stratum, and 3) DF of the Treatment effects must still be intact in the Between Animals Within Runs stratum. Using this objective function, a catalogue of designs is generated for different combinations of design parameters.

The modified SA algorithm presented in Section 3.10 is able to optimise the objective function and obtain the optimal design. The modified SA algorithm consists of two further improvements in the swapping procedures of the nested SA algorithm in Section 3.9. The first improvement is applying the swapping method to only the two experimental units of the Phase 1 experiment instead of the observational units. The purpose of this improvement is to preserve the structure of the initial design, as defined in Section 3.7. The second improvement is the three-stage swapping procedure, which divides a single large search space into three smaller search spaces: this involves swapping the experiment units of animals and treatments: 1) within the same runs, 2) within the same tags, and 3) not within the same runs and tags.

Furthermore, researchers may carry out experiments in which the Phase 1 experiment is arranged in blocks, for example, plants can be arranged into different trays where each tray is referred to as a *block*. Thus, in the next Chapter we consider precisely this situation, and develop an objective function and further modify the SA algorithm described in this Chapter to enable the search for and identification of optimal designs.

Chapter 4

Optimal designs for two-phase experiments when the Phase 1 experiment is arranged in blocks

4.1 Introduction

This Chapter further develops the method for finding optimal designs of Phase 2 proteomics experiments using multiplexing technology. While Chapter 3 presented a methodology used to search for optimal designs when the Phase 1 experiment is arranged in a completely randomised design (CRD), this Chapter considers Phase 1 experiments arranged in a randomised complete block design (RCBD) or a balanced incomplete block design (BIBD).

Where the Phase 1 experiment is arranged in blocks, then the block structure consists of factors of Block and Plot. The treatments are randomly allocated in plots within each block, thus the plot is the experimental unit. The biological material harvested from each Phase 1 experimental unit is also divided into multiple aliquots, namely technical replicates, which will undergo MudPIT analysis in the Phase 2 experiment. Even though the additional Phase 1 Block factor is not the experimental unit, we still need to consider on how the allocation of the Phase 1 Block factor can affects the method in searching for the optimal design.

Due to ethical reasons in any biological experiment, researchers would never assign different treatments to different animals within the same block. Hence, an experiment on plants is used, instead of animals, as an example that is described throughout this Chapter. In the plant example, the Phase 1 experiment is arranged in blocks by trays of plants when each plant can be treated by a different type of fungicide. Thus, each tray is referred to as a block, and each

plant in the tray is referred to as a plot or an experimental unit. Sub-samples extracted and processed from each plant of the Phase 1 experiment are to be used to examine their protein content using the MudPIT-iTRAQTM experiment in Phase 2 experiments.

In this Chapter, Section 4.2 introduces an small experiment on plants with the design and theoretical ANOVA tables. Section 4.3 describes the linear models of the Phase 1 and 2 experiments. Section 4.4 defines the information matrix needed to derive the objective functions discussed in Section 4.5. Section 4.6 describes the construction of the initial design considering the additional block component. Section 4.7 discusses minor changes to the nested simulated annealing (SA) algorithm. Section 4.8 describes an example of finding optimal designs with six treatments. Section 4.9 summarises the properties of optimal designs found. Section 4.10 describes the situation when the Phase 1 experiment is arranged in a BIBD and presents another summary of the properties of optimal designs found. The **infoDecompuTE** package, introduced in Chapter 2, is used consistently throughout to construct the theoretical ANOVA tables for comparing the competing designs. Even though we have used four- and eight-plex labelling systems as the motivating example, the method are more general and can be applied all two-phase designs when the Phase 1 experiment is arranged in a RCBD or a BIBD.

4.2 An illustrative example

We first set the scene with a two-phase proteomics experiment on plants. The Phase 1 experiment comprises $\nu = 2$ treatments (labelled by lower case letter, i.e. a and b) on plants in a controlled environmental cabinet. This controlled environment consists of $n_b = 2$ trays (labelled by numbers, i.e. 1 and 2) when each tray can accommodate two plants, hence there are a total of $n_p = 4$ plants (labelled by upper cases letters, i.e. A, B, C and D). Let y_{ijh} denote the abundance of a given protein from plant h under treatment i in tray j , then the linear model of the Phase 1 experiment can be written as

$$y_{ijh} = \mu + \tau_i + b_j + p_h, \quad (4.1)$$

where μ denotes the grand mean of protein abundance from all observations, τ_i denotes the fixed effects of treatment i , $b_j \sim \mathcal{N}(0, \sigma_b^2)$ denotes the random effects from tray j , and $p_h \sim \mathcal{N}(0, \sigma_p^2)$ denotes the random effects from plant h in tray j ($i = a, b; j = 1, 2; h = A, \dots, D$).

The layout of the Phase 1 experimental design is presented in Table 4.1, which shows Tray 1 contains Plants A and B , respectively, whereas Tray 2 contains Plants C and D assigned by Treatment a and b , respectively. Treatment a is assigned to Plants A and C , and Treatment b

is assigned to Plants B and D . This arrangement is a RCBD, because all treatments are applied to each plant (plot) within a tray (block).

Table 4.1: Experimental design of Phase 1 plant experiment with $\nu = 2$ treatments (labelled a and b) assigned to $n_p = 4$ plants (labelled A to D) in $n_b = 2$ trays (labelled 1 and 2).

Tray 1	Aa	Bb
Tray 2	Ca	Db

The theoretical ANOVA of this Phase 1 experiment consists of Between Trays and Between Plants Within Trays strata (see Table 4.2). The total of 3 DF is partitioned into 1 DF for the Between Trays stratum and 2 DF for the Between Plants Within Trays stratum. Since the experimental unit is the plant, the Treatment effects are estimated in the Between Plants Within Trays stratum with 1 DF. This theoretical ANOVA also contains the variance components of between plants and trays denoted σ_p^2 and σ_b^2 , respectively, with the fixed effects component denoted by θ_τ . Finally, the treatment average efficiency factor, E_τ shows that all the treatment information can be estimated with the highest precision in the Between Plants Within Trays stratum. Therefore, the optimal design for Phase 2 experiment should be arrange with iTRAQTM labelling of sub-samples and their assignment to MudPIT runs, such that it retain all the DF associated with the Treatment effects and the Residual EMS of the Between Plants Within Trays stratum in the Within Runs stratum, as well as maximising the treatment information able to be estimated in the Within Runs stratum.

Table 4.2: Theoretical ANOVA for the Phase 1 experiment in Table 4.1 with $\nu = 2$ treatments assigned to $n_p = 4$ plants in $n_b = 2$ trays.

Source of Variation	DF	EMS	E_τ
Between Trays	1	$\sigma_p^2 + 4\sigma_b^2$	
Between Plants Within Trays			
Treatment	1	$\sigma_p^2 + 4\theta_\tau$	1
Residual	1	σ_p^2	

Suppose that we randomly select $n_s = 2$ sub-samples from each plant of the Phase 1 experiment and analyse all the proteins within these sub-samples using the MudPIT-iTRAQTM technology for the Phase 2 experiment. Since there are a total of $n_p n_s = 8$ sub-samples from the Phase 1 experiment, if the reagent kit of $n_\gamma = 4$ iTRAQTM tags is used, then $n_r = 2$ MudPIT runs is needed to analyse all eight sub-samples. Now let y_{kls}^{ijh} denote the log-protein abundance level of the same protein in (4.1) from sub-sample s of plant h in tray j under treatment i and analysed in run l with tag m , the linear model of this Phase 2 experiment can then be written

as

$$y_{kls}^{ijh} = \mu + \tau_i + b_j + p_h + \gamma_k + r_l + \epsilon_{kls}^{ijh} \quad (4.2)$$

where γ_k denotes the fixed effects of tag k , $r_l \sim \mathcal{N}(0, \sigma_r^2)$ denotes the random run effects, and $\epsilon_{kls}^{ijh} \sim \mathcal{N}(0, \sigma^2)$ denotes the experimental error, ($k = 1 \dots n_\gamma; l = 1 \dots n_r; s = 1, \dots, n_s$).

Table 4.3 shows one way to allocate eight sub-samples from each plant assigned by treatment in each tray of the Phase 1 experiment to be differentially labelled by iTRAQTM tags and MudPIT runs. We can observe that Tags 114 and 115 label sub-samples from Tray 2 and Tags 116 and 117 label sub-samples from Tray 2; thus, 1 DF associated with Tag effects is confounded with the Tray effects. This 1 DF of Tag effects is associated with Tag contrast of Tags 114 and 115 versus Tags 116 and 117. However, each run contains each of two trays, so the Tray effects are orthogonal to Run effects. The allocation of the plants and treatments follows a 2-by-2 Latin square arrangement in a two 2-run-and-2-tag arrays. Hence, the Plant and Treatment effects are orthogonal to Run and Tag effects.

Table 4.3: Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to each of $n_p = 4$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 2$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	1Bb	1Aa	2Ca	2Db
2	1Aa	1Bb	2Db	2Ca

The theoretical ANOVA of the Phase 2 experiment is shown in Table 4.4. The total of 7 DF is partitioned into 1 DF for Between Runs stratum and 6 DF for the Within Runs stratum. Since this Between Runs stratum only contains the variance components of measurement error and between runs are denoted by σ^2 and σ_r^2 , the Phase 2 design in Table 4.3 has preserved all the Phase 1 information in the Within Runs stratum. The Within Runs stratum (6 DF) are further separated into Between Trays (1 DF), Between Plants Within Trays (2 DF), and Within Plants and Within Trays (3 DF) strata. The 1 DF of the Between Trays stratum is completely confounded with 1 DF associated with the Tag effects, but that does not affect how we estimate the Treatment effects as it is estimated in the Between Plants Within Trays stratum. The Residual DF in the Between Plants Within Trays stratum remains 1 DF, and is unchanged from the Phase 1 experiment. Further, all treatment information is preserved from the Phase 1 experiment, with a valid F-test because the coefficients of σ^2 and σ_p^2 are identical in the Between

Plants Within Trays stratum. Therefore, this Phase 2 design can be considered as the optimal design for this case.

Table 4.4: Theoretical ANOVA table for the Phase 2 experiment in Table 4.3.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs	1	$\sigma^2 + 4\sigma_r^2$		
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 4\sigma_b^2 + 2\theta_\gamma$	1	
Between Plants Within Trays				
Treatment	1	$\sigma^2 + 2\sigma_p^2 + 4\theta_\tau$		1
Residual	1	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 2\theta_\gamma$	1	
Residual	1	σ^2		

4.3 Design parameters and model

For any two-phase experiment when the Phase 1 is arranged in a RCBD, we need to consider the block to which each experimental unit is allocated, and how the treatments are applied to each unit in the Phase 1 experiment. More generally, the Phase 1 experiment comprises ν treatments applied to n_p plots in each of n_b blocks. Each sample from the Phase 1 experiment is then split into n_s sub-samples and analysed in the Phase 2 experiment with n_r MudPIT runs, when each sub-sample is differentially labelled by n_γ iTRAQTM tags. Let \mathbf{y} be a vector of $n_p n_b n_s = n_r n_\gamma = n$ responses, the model (4.2) can then be expressed in matrix notation as

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_\tau \boldsymbol{\alpha}_\tau + \mathbf{X}_\gamma \boldsymbol{\alpha}_\gamma + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{v} + \boldsymbol{\epsilon}, \quad (4.3)$$

where $\mathbf{1}$ is also an $n \times 1$ vector for which all elements are unity, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of measurement error terms. The vectors of treatment and tag parameters are $\boldsymbol{\alpha}_\tau$ and $\boldsymbol{\alpha}_\gamma$, respectively, and can be further expressed as

$$\boldsymbol{\alpha}_\tau = (\tau_1, \dots, \tau_\nu), \boldsymbol{\alpha}_\gamma = (\gamma_1, \dots, \gamma_{n_\gamma}).$$

Matrices \mathbf{X}_τ and \mathbf{X}_γ are $n \times \nu$ and $n \times n_\gamma$ design matrices of treatment and tag, respectively. Furthermore, the vector of the Phase 1 block parameter, \mathbf{u} , is the combination of block and plot, which is given by the allocation of plots in blocks. For the example in Section 4.2, when Tray 1 contains Plants *A* and *B* and Tray 2 contains Plants *C* and *D*, the Phase 1 block parameter is

defined as

$$\mathbf{u} = (b_1 : p_A, b_1 : p_B, b_2 : p_C, b_2 : p_D).$$

In general, the Phase 1 block parameter, \mathbf{u} , has a length of n_p , when the Phase 1 experiment is arranged in a RCBD. Thus, the Phase 1 block design matrix is denoted by \mathbf{Z} , which is an $n \times n_p$ design matrix of block-plot combinations based on the Phase 1 block parameter. Lastly, the Phase 2 block design matrix is denoted by $\mathbf{W} = \mathbf{W}_r$, which is an $n \times n_r$ design matrix of runs. The Phase 2 block parameter is denoted by $\mathbf{v} = (r_1, \dots, r_{n_r})$.

4.4 Defining the information matrix

This section shows how to define the information matrices for evaluating the objective function. There are two information matrices of interest: plots within blocks and the treatment information matrices. The construction of these two information matrices is less straightforward compared to the Phase 1 experiments arranged in a CRD. This is because the block structure for the CRD includes only the Phase 1 experimental units, whereas for the RCBD it includes a Phase 1 Block factor on top of the Phase 1 experimental units (or the Phase 1 Plot factor).

Decomposition of the Phase 1 block structure is performed after the strata of the Phase 2 block structure are defined. Thus, the Phase 1 block design matrix, \mathbf{Z} , and the projection matrices for each factor in the Phase 1 block structure need to be computed. The Phase 1 block parameter \mathbf{u} in (4.3) can also be written in matrix notation as

$$\mathbf{u} = \mathbf{C}_{00}\mathbf{u} + \mathbf{C}_{01}\mathbf{u} + \mathbf{C}_{21}\mathbf{u}$$

where $\mathbf{C}_{00} = \mathbf{K}_{n_b} \otimes \mathbf{K}_{n_k}$ is the averaging matrix, where n_k denotes block size and is given by $n_k = n_p/n_b$. The remaining matrices $\mathbf{C}_{10} = (\mathbf{I}_{n_b} - \mathbf{K}_{n_b}) \otimes \mathbf{K}_{n_k}$ and $\mathbf{C}_{21} = \mathbf{I}_{n_b} \otimes (\mathbf{I}_{n_k} - \mathbf{K}_{n_k})$ are the projection matrices for Block and Plot Within Blocks effects.

The goal is to find the best allocation of the sub-samples from Phase 1 experimental unit assigned by treatment to be differentially labelled by a tag and analysed in a run of the Phase 2 experiment. This occurs when the confounding of the Phase 1 Plot Within Blocks and Treatment effects with the Phase 2 Run and Tag effects is minimised. To achieve this, we again use the orthogonal projection matrix which projects \mathbf{y} onto Within Runs and Tags vector subspace, $\mathbf{Q}_{r\gamma} = (\mathbf{I} - \mathbf{P}_r)(\mathbf{I} - \mathbf{P}_\gamma)$, as defined in (3.6) of Chapter 3.

Considering the estimation of Phase 1 Block effects in the Within Runs and Tags stratum,

the reduced normal equations for the \mathbf{u}_b can then be defined as

$$\mathbf{u}_b = \mathbf{A}_b^- \mathbf{q}_b,$$

where

$$\begin{aligned}\mathbf{A}_b &= \mathbf{L}_b \mathbf{Q}_{r\gamma} \mathbf{L}_b', \\ \mathbf{q}_b &= \mathbf{L}_b \mathbf{Q}_{r\gamma} \mathbf{y},\end{aligned}$$

where $\mathbf{L}_b = \mathbf{C}_{01} \mathbf{Z}'$, \mathbf{A}_b denotes the information matrix of block, vector \mathbf{q}_b denotes the adjusted Phase 1 Block total, and $\mathbf{Q}_{r\gamma}$ denotes the orthogonal projection matrix that projects \mathbf{y} onto Within Runs and Tags vector subspace. Thus, the orthogonal projection matrix that projects \mathbf{y} onto Between Blocks Within Runs and Tags vector subspace, denoted by \mathbf{Q}_b , is

$$\mathbf{Q}_b = \mathbf{Q}_{r\gamma} \mathbf{L}_b' \mathbf{A}_b^- \mathbf{L}_b \mathbf{Q}_{r\gamma}.$$

The Phase 1 Block information is then swept by deriving the orthogonal projection matrix that projects \mathbf{y} onto Within Blocks Within Runs and Tags vector subspace, i.e. $(\mathbf{I} - \mathbf{Q}_b)$. Thus, the information matrix for the Plots Within Blocks effects, denoted by \mathbf{A}_p , can be expressed as

$$\mathbf{A}_p = \mathbf{L}_p' (\mathbf{I} - \mathbf{Q}_b) \mathbf{L}_p, \quad (4.4)$$

where $\mathbf{L}_p = \mathbf{C}_{21} \mathbf{Z}'$.

Since the Treatment effects are to be maximised in the Between Plots Within Blocks Within Runs and Tags vector subspace, the treatment information matrix can be expressed as

$$\mathbf{A}_\tau = \mathbf{X}_\tau' \mathbf{Q}_p \mathbf{X}_\tau, \quad (4.5)$$

where \mathbf{Q}_p is the projection matrix that projects \mathbf{y} onto Between Plots Within Blocks in Within Runs and Tags vector subspace and is given by

$$\mathbf{Q}_p = (\mathbf{I} - \mathbf{Q}_b) \mathbf{L}_p' \mathbf{A}_p^- \mathbf{L}_p (\mathbf{I} - \mathbf{Q}_b).$$

4.5 An objective function for identifying optimal designs for the Phase 2 experiment

The next step is to derive the objective function for finding the optimal design of the Phase 2 experiment. The objective function used to find optimal designs for a Phase 2 multi-plexing

proteomics experiment when the biological material was derived from a Phase 1 experiment is arranged in a CRD is given as a single equation in (3.20). For the Phase 1 experiment arranged in a RCBD, the objective function is re-written as

$$0.75E_p + 0.25 \left(\frac{E_\tau + \nu_2}{\nu} \right) \quad (4.6)$$

where E_p and E_τ are the Phase 1 plot and treatment average efficiency factors in the Within Blocks Within Runs and Tags vector subspace. The optimisation procedure for the objective function uses the modified nested SA algorithm described in Section (3.10).

4.5.1 Illustrative example applying the old objective function

This section first uses the same objective function as in Chapter 3 on the plant example to search for the optimal design of Phase 2 experiment. For this example, the Phase 1 experiment consists of $\nu = 3$ treatments assigned to $n_p = 6$ plants in $n_b = 2$ trays. The layout of the Phase 1 experimental design is presented in Table 4.5. Tray 1 contains Plants A to F , and Tray 2 contains Plants G to L . In addition, Treatment a is assigned to Plants A , D , G and J , Treatment b is assigned to Plants B , E , H and K , and Treatment c is assigned to Plants C , F , I and L .

Table 4.5: Phase 1 experimental design showing the assignment of treatments and plants to trays, with $\nu = 3$ treatments assigned to $n_p = 12$ plants in $n_b = 2$ trays.

Tray 1	Aa	Bb	Cc
	Da	Eb	Fc
Tray 2	Ga	Hb	Ic
	Ja	Kb	Lc

The Phase 1 theoretical ANOVA (see Table 4.6), shows the total of 11 DF is separated into 1 DF for Between Trays and 10 DF Between Plants Within Trays strata. The Between Plants Within Trays stratum is further partitioned into 2 DF associated with Treatment effects and 8 Residual DF for estimating the variances of Treatment effects.

Table 4.6: Theoretical ANOVA of the Phase 1 experiment in Table 4.5.

Source of Variation	DF	EMS	E_τ
Between Trays	1	$\sigma_p^2 + 6\sigma_b^2$	
Between Plants Within Trays			
Treatment	2	$\sigma_p^2 + 4\theta_\tau$	1
Residual	8	σ_p^2	

The next step is first to split $n_s = 2$ sub-samples from each plant of the Phase 1 experiment, and each sub-sample is then further processed for the Phase 2 proteomics experiment. Since

Table 4.7: Theoretical ANOVA for the Phase 2 experiment in Table 4.14b.

Source of Variation	DF	EMS	E_γ
Between Runs	5	$\sigma^2 + 4\sigma_r^2$	
Within Runs			
Tag	3	$\sigma^2 + 6\theta_\gamma$	1
Residual	15	σ^2	

there are $n_p n_s = 24$ sub-samples from the Phase 1 experiment, the Phase 2 experiment uses $n_r = 6$ runs when each sub-sample is differentially labelled with each of $n_\gamma = 4$ tags.

Using the objective function in (3.20), we can find one allocation of sub-samples from treatments, plants and trays differentially labelled by tags and analysed in runs, which is given in Table 4.8. The effects of trays is orthogonal to both the effects of runs and tags because the sub-samples from Tray 1 and 2 are equally replicated in each run and tag. The Plant effects are confounded with Tag effects, because Plant C is only labelled by Tag 114 and 115, and Plant F is only labelled by Tag 116 and 117. The Plant effects are also confounded with Run effects, because different pairs of runs comprise different combinations of plants. However, the Treatment effects are orthogonal to the Tag effects, because each tag labels sub-sample assigned by two of the three treatments. Like Plant effects, the effects of treatment are also confounded with Run effects, as different pairs of runs comprise sub-samples from different combinations of treatments.

Table 4.8: Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags based on objective function (4.6), when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	2Ic	1Aa	2Ga	1Eb
2	1Aa	2Ic	1Eb	2Ga
3	1Cc	2Hb	2Ja	1Fc
4	2Hb	1Cc	1Fc	2Ja
5	2Kb	1Da	2Lc	1Bb
6	1Da	2Kb	1Bb	2Lc

The theoretical ANOVA (see Table 4.9) shows the total of 23 DF is separated into 5 DF in the Between Runs stratum and 18 DF in the Within Runs stratum. There are 2 DF associated with Treatment effects with 0.0625 of the treatment information able to be estimated in the Between Runs stratum. The Within Runs stratum (18 DF) is further separated into Between

Trays (1 DF), Between Plants Within Trays (8 DF) and Within Plants Within Trays (9 DF) strata. The estimation of the Treatment effects is from the Between Plants Within Trays Within Runs stratum based on 0.9375 of treatment information. In addition, the Residual DF of the Between Plants Within Trays Within Runs stratum is reduced to 5 DF from 8 DF of Phase 1 experiment. Thus, the design in Table 4.8 is very close to being as precise as the Phase 1 experiment in Table 4.5 with two minor deficiencies: lost of 0.0625 of the treatment information and 3 Residual DF that were from the Phase 1 experiment as shown in the theoretical ANOVA in Table 4.6. This means that the objective function in (3.20) of Chapter 3 is already effective, however, we believe that we can obtain an even better design with a different objective function.

Table 4.9: Theoretical ANOVA table of the Phase 2 experiment in Table 4.8 from objective function (4.6).

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Plants Within Trays				
Treatment	2	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + 0.5\theta_\tau$		0.0625
Residual	3	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 12\sigma_b^2$		
Between Plants Within Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 6\theta_\gamma$	1	
Treatment	2	$\sigma^2 + 2\sigma_p^2 + 7.5\theta_\tau$		0.9375
Residual	5	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 6\theta_\gamma$	1	
Residual	7	σ^2		

4.5.2 New Objective function and optimisation procedure

This newly developed objective function consists of four criteria:

1. The Phase 1 Plots average efficiency factor in the Within Blocks Within Runs and Tags vector subspace, denoted by E_p , must be 1. This allows us to obtain designs where the variances are balanced in the Between Plots Within Blocks Within Runs stratum, thus a valid F-test can be performed. The Phase 1 Plots average efficiency factor is computed from the information matrix of the Plots Within Blocks Within Runs and Tags vector subspace in (4.4),
2. The DF associated with Treatment effects in the Between Plots Within Blocks Within Runs stratum must be intact. This allows us to obtain designs when the treatment alloca-

tion is always connected to the runs and tags, so every pairwise treatment comparison is estimable in the Between Plots Within Blocks Within Runs stratum. The DF associated with treatment effects is computed from the trace of the treatment information matrix of Plots Within Blocks Within Runs and Tags vector subspace defined in (4.5).

3. The Residual DF in the Between Plots Within Blocks Within Runs stratum is maximised. This allows us to obtain design with higher precision in estimating the variances of the Treatment effects. This Residual DF are computed from the trace of the information matrix of Plots Within Blocks Within Runs and Tags vector subspace, defined in (4.4), minus the trace of the treatment information matrix of Plots Within Blocks Within Runs and Tags vector subspace, defined in (4.5), i.e. $\text{trace}(\mathbf{A}_p) - \text{trace}(\mathbf{A}_\tau)$.
4. The treatment average efficiency factor in the Between Plots Within Blocks Within Runs and Tags vector subspace, denoted by E_τ , is maximised. This is computed from the treatment information matrix of Plots Within Blocks Within Runs and Tags vector subspace (4.5),

Note that only the third component is newly introduced compared to the objective function given in (4.6).

The next issue is decide how to optimise this four-criterion objective function all at the same time. After trying different combinations of weights, we decided to use a more robust method to optimise this new objective function. The new method to optimise this four-criterion objective function consists of three incremental steps. The first step is to locate designs which satisfy the first two components of the objective function, i.e. E_p must be 1, and DF associated with treatment effects in the Between Plots Within Blocks Within Runs stratum must be intact. Then from among the designs located in the first step, the second step uses the modified nested SA algorithm to find optimal designs based on the third component of the objective function, i.e. Residual DF in the Between Plots Within Blocks Within Runs stratum is maximised. Finally, from among the designs found in the second step, the third step is to find the optimal design where the fourth component of the objective function is satisfied, i.e. treatment average efficiency factor in the Between Plots Within Blocks Within Runs and Tags vector subspace is maximised. Note that the last two steps in optimising the objective function are inspired by the method for finding the MS-optimal design. We construct the objective function in this manner to avoid the need to determine the weights of the four components in the objective function. However, this method suffers the drawback of requiring that the nested SA algorithm be performed twice in order to optimise two different objective functions.

4.5.3 Illustrative example applying new objective function

Returning to the plant example, this optimisation procedure with the newly developed four-criterion objective function, found a different allocation of sub-samples from treatments, plants and trays differentially labelled by tags and analysed in runs, which is given in Table 4.10. One noticeable difference of this design compared to the design in Table 4.8 is that the tray effects are completely confounded with tag effects, as Tags 114 and 115 only label the sub-samples from Tray 1 and Tags 116 and 117 only label the sub-samples from Tray 2.

Table 4.10: Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags based new objective function, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	2Lc	2Hb
2	1Bb	1Aa	2Hb	2Lc
3	1Cc	1Da	2Ga	2Kb
4	1Da	1Cc	2Kb	2Ga
5	1Eb	1Fc	2Ic	2Ja
6	1Fc	1Eb	2Ja	2Ic

The theoretical ANOVA of the optimal design from the new objective function is shown in Table 4.11. Comparing this theoretical ANOVA to the one in Table 4.9, there is still 0.0625 of the treatment information estimated in the Between Runs stratum. What makes this design better is that there is 1 DF associated with the Tag effects in the Between Trays stratum, which allows us to have 6 Residual DF in the Between Plants Within Trays Within Runs stratum. Thus, here the Residual DF is one higher than that from the design found using the objective function of the previous Chapter, offering us better precision in estimating the Treatment effects. We are able to generate this design due to having the third component of the objective function, which maximises the Residual DF in the Between Plants Within Trays Within Runs stratum.

4.6 Construction of the initial design

Section 4.5 shows that the newly developed objective function is better than the objective function from Chapter 3 in searching for the optimal design of the Phase 2 experiment when the Phase 1 experiment is arranged in a RCBD. Although the estimation of the Treatment effects occurs in the Between Plots Within Blocks Within Runs stratum, from observing the design in

Table 4.11: Theoretical ANOVA of the Phase 2 experiment in Table 4.10 from the new objective function described in Section 4.5.2.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Plants Within Trays				
Treatment	2	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + 0.5\theta_\tau$		0.0625
Residual	3	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 12\sigma_b^2 + 6\theta_\gamma$	1	
Between Plants Within Trays				
Treatment	2	$\sigma^2 + 2\sigma_p^2 + 7.5\theta_\tau$		0.9375
Residual	6	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 6\theta_\gamma$	1	
Residual	7	σ^2		

Table 4.10, we need to take extra caution in how the Block factor of the Phase 1 experiment is allocated for the Phase 2 experiment. This is because having a design when Phase 1 Block effects are confounded with Tag effects can increase the Residual DF in the Between Plots Within Blocks Within Runs stratum, which in turn can increase the precision of estimation for Treatment effects in the Between Plots Within Blocks Within Runs stratum.

Consider the plant example in Section 4.5 when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to $n_p = 12$ plants within each of $n_b = 2$ trays from Phase 1 and then each plant is split into $n_s = 2$ sub-samples differentially labelled by $n_\gamma = 4$ tags and analysed in $n_r = 6$ runs of the Phase 2 experiment. The optimal design of this Phase 2 experiment shows Tags 114 and 115 label sub-samples from Tray 1, and Tags 116 and 117 label sub-samples from Tray 2. Thus, if an initial allocation already has this structure between trays and tags, then we will be able to locate the optimal design quickly. An example of the initial allocation is shown in Table 4.12, when we already have the allocation that Tags 114 and 115 label sub-samples from Tray 1, and Tags 116 and 117 label sub-samples from Tray 2. We call this initial design as Tray (Phase 1 Block) effects are intentionally confounded with the Tag effects.

However, the initial design, when Tray effects are intentionally confounded with Tag effects, does not always generate a design with higher Residual DF in the Between Plots Within Blocks Within Runs stratum. This section shows two different examples of initial designs that perform better where (1) Tray effects are intentionally confounded with Run effects, as well as a situation where (2) both types of initial designs are equally effective.

Note that since the number of tags of MudPIT-iTRAQTM experiments of four or eight are

Table 4.12: Initial design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consists of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in each of $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	2Ga	2Hb
2	1Bb	1Aa	2Hb	2Ga
3	1Cc	1Cc	2Ic	2Ja
4	1Da	1Da	2Ja	2Ic
5	1Eb	1Fc	2Kb	2Lc
6	1Fc	1Eb	2Lc	2Kb

considered, we can only allocate sub-samples with equal number of each tray to tags where the number of trays is even. Therefore, for odd numbers of trays, the initial design will be constructed without considering the confounding between the Tray effects with the Run or Tag effects.

4.6.1 An initial design when the Tray effects are intentionally confounded with Run effects (better than confounded with Tag effects)

This example shows an exception to the rule that an initial design when the Tray effects are intentionally confounded with Run effects can generate a better design than the initial design when the Tray effects are confounded with Tag effects. This example considers the Phase 1 experiment involving $\nu = 4$ treatments assigned to each of $n_p = 16$ plants in $n_b = 4$ trays. The layout of the design for the Phase 1 experiment consists of Plants *A* to *D* in Tray 1, Plants *E* to *H* in Tray 2, Plants *I* to *L* in Tray 3, and Plants *M* to *P* in Tray 4. Furthermore, Treatment *a* assigned to Plants *A*, *E*, *I* and *M*, Treatment *b* assigned to Plants *B*, *F*, *J* and *N*, Treatment *c* assigned to Plants *C*, *G*, *K* and *O*, and Treatment *d* is assigned to Plants *D*, *H*, *L* and *P*. The theoretical ANOVA of the Phase 1 experiment is presented in Table 4.13, which shows the total of 15 DF is separated into 3 DF for Between Trays stratum and 12 DF for Between Plants Within Trays stratum. Since there are 3 DF associated with the Treatment effects, there are 9 Residual DF in the Between Plants Within Trays stratum.

The next step is to obtain $n_s = 2$ sub-samples from each plant and to make measurements in the Phase 2 proteomics experiment using $n_r = 8$ runs and $n_\gamma = 4$ tags. We first use the

Table 4.13: Theoretical ANOVA table of Phase 1 experiment with $\nu = 4$ treatments assigned to each of $n_p = 16$ plants in $n_b = 4$ trays.

Source of Variation	DF	EMS	E_τ
Between Trays	3	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2$	
Between Plants Within Trays			
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$	1
Residual	9	$\sigma^2 + 2\sigma_p^2$	

initial design when Tray effects are intentionally confounded with Tag effects. This first initial allocation of the sub-samples from trays, plants and treatments to runs and tags is shown in Table 4.14a, when Tags 114 and 115 label sub-samples from Trays 1 and 2, while Tags 116 and 117 label sub-samples from Trays 3 and 4. Notice that in this design, the Tray effects are also confounded with the Run effects as Runs 1 to 4 contain sub-samples from Trays 1 and 3, and Runs 5 to 8 contain sub-samples from Trays 2 and 4. The optimal design found is presented in Table 4.14b, as the assignment of sub-samples from trays to runs and tags is still the same as the initial design in Table 4.14a.

Table 4.14: Initial and final optimal designs for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consisting of $\nu = 4$ treatments assigned to each of $n_p = 16$ plants in $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 8$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

(a) The initial allocation.					(b) The final optimal design from using the initial design in Table 4.14a.				
Run	Tag				Run	Tag			
	114	115	116	117		114	115	116	117
1	1Aa	1Bb	3Ia	3Jb	1	1Aa	1Bb	3Kc	3Ld
2	1Bb	1Aa	3Jb	3Ia	2	1Bb	1Aa	3Ld	3Kc
3	1Cc	1Dd	3Kc	3Ld	3	1Cc	1Dd	3Jb	3Ia
4	1Dd	1Cc	3Ld	3Kc	4	1Dd	1Cc	3Ia	3Jb
5	2Ea	2Fb	4Ma	4Nb	5	2Ea	2Hd	4Oc	4Nb
6	2Fb	2Ea	4Nb	4Ma	6	2Hd	2Ea	4Nb	4Oc
7	2Gc	2Hd	4Oc	4Pd	7	2Gc	2Fb	4Ma	4Pd
8	2Hd	2Gc	4Pd	4Oc	8	2Fb	2Gc	4Pd	4Ma

The theoretical ANOVA of the optimal design in Table 4.14b is shown in Table 4.15. The total of 23 DF are separated into 7 DF and 16 DF of Between Runs and Within Runs strata, respectively. In the Between Runs stratum, there are 1 DF associated with Between Trays stratum and 2 DF associated with Between Plants Within Trays stratum. Due to the confounding

of Tray effects with Tag effects, there is 1 DF associated with the Tag effects in the Between Trays Within Runs stratum. In the Between Plants Within Trays Within Runs stratum, there are 7 Residual DF, which has been reduced from 9 Residual DF of the Phase 1 experiment in Table 4.13. In addition, the Treatment effects can be estimated with 100% of the treatment information with a valid F-test in the Between Plants Within Trays Within Runs stratum using the design in Table 4.14b.

Table 4.15: Theoretical ANOVA for the Phase 2 experiment in Table 4.14b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\sigma_r^2$		
Between Plants Within Trays	2	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2$		
Residual	4	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 8\theta_\gamma$	1	
Residual	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2$		
Between Plants Within Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$		1
Residual	7	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 8\theta_\gamma$	1	
Residual	10	σ^2		

Now consider a different initial design when Tray effects are intentionally confounded with the Run effects. An example of this type of initial design is presented in Table 4.16a, in which Runs 1 and 2 contain sub-samples from Tray 1, Runs 3 and 4 contain sub-samples of Tray 2, Runs 5 and 6 contain sub-samples of Tray 3 and Runs 7 and 8 contain sub-samples of Tray 4. Note that the Tray effects are orthogonal to Tag effects, because each tag labels sub-samples from two of all four trays. The optimal design found is presented in Table 4.16b, which shows the assignment of sub-samples from trays to runs and tags remains the same.

The theoretical ANOVA (in Table 4.17) again shows a total of 23 DF are separated into 7 DF and 16 DF of Between Runs and Within Runs strata, respectively. For this case, all 3 DF associated with the Between Trays stratum are in the Between Runs stratum, due to the confounding between the Tray effects and Run effects. Thus, the Tray effects can not be estimated in the Within Runs stratum. In the Between Plants Within Trays Within Runs stratum, there is 1 DF associated with Tag effects, but there are still 8 Residual DF remains, which is also reduced from 9 Residual DF in the Phase 1 experiment. Finally, the Treatment effects can be estimated with 100% of the treatment information, and there is a valid F-test in

Table 4.16: Initial and final optimal designs for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Run effects, when the Phase 1 experiment consisting of $\nu = 4$ treatments assigned to each of $n_p = 4$ plants in each of $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 8$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

(a) The initial allocation.					(b) The final optimal design from using the initial design in Table 4.16a.				
Run	Tag				Run	Tag			
1	114	115	116	117	1	114	115	116	117
2	1Aa	1Bb	1Cc	1Dd	2	1Cc	1Dd	1Bb	1Aa
3	2Ea	2Fb	2Gc	2Hd	3	1Dd	1Cc	1Aa	1Bb
4	2Fb	2Ea	2Hd	2Gc	4	2Hd	2Fb	2Ea	2Gc
5	3Ic	3Jd	3Ka	3Lb	5	2Fb	2Hd	2Gc	2Ea
6	3Jd	3Ic	3Lb	3Ka	6	3Ia	3Jb	3Kc	3Ld
7	4Mc	4Nd	4Oa	4Pb	7	3Jb	3Ia	3Ld	3Kc
8	4Nd	4Mc	4Pb	4Oa	8	4Oc	4Ma	4Nb	4Pd

the Between Plants Within Trays Within Runs stratum using the design in Table 4.16b.

Table 4.17: Theoretical ANOVA for the Phase 2 experiment in Table 4.16b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	3	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\sigma_r^2$		
Residual	4	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Plants Within Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 8\theta_\gamma$	1	
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$		1
Residual	8	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 8\theta_\gamma$	1	
Residual	10	σ^2		

We now compare these two ANOVA in Tables 4.15 and 4.17 which uses two different initial designs in Tables 4.14a and 4.16a, respectively. Despite both designs yielded ANOVA which 100% of treatment information with a valid F-test, we can see that using the initial design in Table 4.16a, when Tray effects are intentionally confounded with the Run effects, is better, because there is 1 more Residual DF in the Between Plants Within Trays Within Runs stratum for estimating the variances of Treatment effects. However, the 2 DF of the Between Plants Within Trays Between Runs stratum shown in Tables 4.15 from the initial design in Table 4.14a

can be recovered to obtain a F-test that effectively has higher Residual DF, if the Between Runs variances is lower than the Between Plants variances. This issue will be covered in Chapter 5.

4.6.2 Performance remains unchanged whether Tray effects are confounded with Tag or Run effects

There can also be a situation when the optimal design found has the same precision to that from either type of initial design. One such case involves $\nu = 4$ treatments assigned to $n_p = 8$ plants in $n_b = 2$ trays for the Phase 1 experiment. For the layout of the Phase 1 design, Tray 1 contains Plants A to D , and Tray 2 contains Plants E to H . Furthermore, Treatment a is assigned to Plants A and E , Treatment b is assigned to Plants B and F , Treatment c is assigned to Plants C and G , and Treatment d is assigned to Plants D and H . The theoretical ANOVA of the Phase 1 experiment is presented in Table 4.13, which shows the total of 7 DF is separated into 1 DF for the Between Trays stratum and 6 DF for the Between Plants Within Trays stratum. Since there are 3 DF associated with the Treatment effects, there are 3 Residual DF in the Between Plants Within Trays stratum.

Table 4.18: Theoretical ANOVA of the Phase 1 experiment with $\nu = 4$ treatments assigned to $n_p = 4$ plants in each of $n_b = 2$ trays.

Source of Variation	DF	EMS	E_τ
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2$	
Between Plants Within Trays			
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 4\theta_\tau$	1
Residual	3	$\sigma^2 + 2\sigma_p^2$	

In the Phase 2 proteomics experiment, $n_s = 2$ sub-samples from each plants of the Phase 1 experiment are analysed using $n_\gamma = 4$ tags with $n_r = 4$ runs. Table 4.19 shows the initial design when Tray effects are intentionally confounded with Tag effects, as Tags 114 and 115 labels sub-samples from Tray 1 and Tags 116 and 117 labels sub-samples from Tray 2. Run effects are orthogonal to Tray effects as each run contains sub-samples from two of each of both trays. The design in Table 4.19 is also the optimal design for this case.

The theoretical ANOVA in Table 4.20 shows that the total of 15 DF is separated into 3 DF for the Between Runs stratum and 12 DF for the Within Runs stratum. In the Between Runs stratum, there is 1 DF associated with the Between Plants Within Trays stratum. As for the Within Runs stratum, there is 1 DF associated with Tag effects in the Between Trays stratum due to the confounding between the Tray effects with Tag effects. In the Between Plants Within Trays Within Runs stratum, there is 2 Residual DF which is reduced from 3 Residual DF of the

Table 4.19: Initial and final optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_p = 8$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 4$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	2Hd	2Gc
2	1Bb	1Aa	2Gc	2Hd
3	1Cc	1Dd	2Fb	2Ea
4	1Dd	1Cc	2Ea	2Fb

Phase 1 experiment, as shown in Table 4.18. Further, there is still a valid F-test given from this design with 100% of the treatment information available.

Table 4.20: Theoretical ANOVA for the Phase 2 experiment in Table 4.19.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Plants Within Trays	1	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2$		
Residual	2	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\theta_\gamma$	1	
Between Plants Within Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 4\theta_\tau$		1
Residual	2	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 4\theta_\gamma$	1	
Residual	4	σ^2		

Another initial allocation is presented in Table 4.21 when Tray effects are intentionally confounded with Run effects. Runs 1 and 2 contain sub-samples from Tray 1, and Runs 3 and 4 contain sub-samples from Tray 2. Tag effects are orthogonal to Tray effects as each tag labels sub-samples from two of each of two trays. The design in Table 4.21 is also the optimal design for this case.

The theoretical ANOVA (in Table 4.22) again shows the total of 15 DF are separated into 3 DF for Between Runs stratum and 12 DF for Within Runs stratum. There is 1 DF associated with the Between Trays stratum in the Between Runs stratum due to the confounding between the Run effects with Tray effects. In the Between Plants Within Trays Within Runs stratum, 1 DF and 3 DF are associated with the Tag effects and Treatment effects, respectively. Thus,

Table 4.21: Initial and final optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Run effects, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_p = 8$ plants in $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 4$ runs and $n_\gamma = 4$ tags.

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	1Cc	1Dd
2	1Bb	1Aa	1Dd	1Cc
3	2Gc	2Hd	2Ea	2Fb
4	2Hd	2Gc	2Fb	2Ea

there are still 2 Residual DF given from this optimal design. Therefore, this situation provides an example where either type of initial design can generate optimal designs with identical precision.

Table 4.22: Theoretical ANOVA for the Phase 2 experiment in Table 4.21.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\sigma_r^2$		
Residual	2	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Plants Within Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 4\theta_\gamma$	1	
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 4\theta_\tau$		1
Residual	2	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 4\theta_\gamma$	1	
Residual	4	σ^2		

In summary, this section has shown that an initial design in which Phase 1 Tray (Block) effects are intentionally confounded with Run or Tag effects can affect the Residual DF of the ANOVA for estimating the variance of the Treatment effects. Specifically, in the initial design when Tray effects are confounded with Tag effects, the DF associated with the Tag effects can be pushed from the Between Plants Within Trays Within Runs stratum into the Between Trays Within Runs stratum. Consequently, the Residual DF in the Between Plants Within Trays Within Runs stratum increase, which also increases the precision in estimating the Treatment effects. Thus, an initial design when Tray effects are confounded with Tag effects should be used in most cases. However, some exceptions do exist, as shown in the Section 4.6.1, when the initial design, in which Phase 1 Block effects are intentionally confounded with Run effects generates an optimal design with higher Residual DF. Therefore, we suggest testing both types of initial designs, and comparing the optimal designs that they yield.

4.7 Modified simulated annealing for searching for the optimal design

The modified nested simulated annealing (SA) algorithm, as described in Chapter 3, can still be used to find optimal designs of the Phase 2 experiment when the Phase 1 experiment is arranged in a RCBD. The only consideration is when the initial design is constructed with Block effects of the Phase 1 experiment being intentionally confounded with Run or Tag effects (as discussed in Section 4.6). The three-stage swapping procedure becomes a single-stage swapping procedure to preserve the structure of the initial design between the Phase 1 Block factors and the Phase 2 Run and Tag factors.

Table 4.23: Illustrate the swapping between Tray 1 of Plant *A* under Treatment *a* and Tray 1 of Plant *E* under Treatment *b* to preserve the assignments of the sub-samples between trays and tags on the initial design for the Phase 2 experiment in Table 4.12, when the Tray effects are intentionally confounded with the Tag effects. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	2Ga	2Hb
2	1Bb	1Aa	2Hb	2Ga
3	1Cc	1Cc	2Ic	2Ja
4	1Da	1Da	2Ja	2Ic
5	1Eb	1Fc	2Kb	2Lc
6	1Fc	1Eb	2Lc	2Kb

Consider a two-phase experiment when the Phase 1 experiment consisting of $\nu = 3$ treatments assigned to each of $n_p = 12$ plants in each of $n_b = 2$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags, the initial design of the Phase 2 experiment is shown in Table 4.12. Table 4.23 demonstrates one possible swap between Plant *A* in Tray 1 with Plant *E* in Tray 1, which still results in a design when Tags 114 and 115 label sub-samples from Trays 1, and Tags 116 and 117 label sub-samples from Trays 2. Thus, the single-stage swapping is to be used instead of three-stage swapping, i.e. swapping of the sub-samples with labels in the only with same tags for case of the initial design when the Tray effects are intentionally confounded with Tag effects.

4.8 An illustrative example with six treatments

This section presents an example finding the optimal design using the objective function described in Sections 4.5. Consider the Phase 1 experiment arranged in a RCBD, each of $n_b = 3$ trays contains $n_p = 18$ plants with $\nu = 6$ different treatments assigned. The layout of the Phase 1 experiment consists of Tray 1 containing Plants A to F , Tray 2 containing Plants G to L and Tray 3 containing Plants M to R . Furthermore, Treatment a is assigned to Plants A , G and M , Treatment b is assigned to Plants B , H and N , Treatment c is assigned to Plants C , I and O , Treatment d is assigned to Plants D , J and P , Treatment e is assigned to Plants E , K and Q , and Treatment f is assigned to Plants F , L and R . The theoretical ANOVA of the Phase 1 experiment (in Table 4.24) shows a total of 17 DF are separated into 2 DF for the Between Trays stratum and 15 DF for the Between Plants Within Trays stratum. Since there are 5 DF associated with Treatment effects, there are still 10 Residual DF in the Between Plants Within Trays stratum.

Table 4.24: Theoretical ANOVA table of Phase 1 experiment with $\nu = 6$ treatments assigned to $n_p = 18$ plants in $n_b = 3$ trays.

Source of Variation	DF	EMS	E_τ
Between Trays	2	$\sigma_p^2 + 6\sigma_b^2$	
Between Plants Within Trays			
Treatment	5	$\sigma_p^2 + 3\theta_\tau$	1
Residual	10	σ_p^2	

Since the number of trays is not even, we cannot assign the sub-samples having the tray that are evenly allocated to tags of the Phase 2 experiment. Thus, we cannot consider the initial design when the Tray effects are intentionally confounded with Tag effects. Using the objective function and SA algorithm mentioned in Sections 4.5 and 4.7, respectively, an allocation of the sub-samples from trays, plants and treatments to runs and tags is shown in Table 4.25. Tray effects are confounded with run effects, because Runs 3, 4, 5, 6, 7 and 8 contain sub-samples from Trays 2 and 3, whereas Runs 1, 2 and 9 contain sub-samples from only Tray 1. There appears to be some confounding between the Plant and Treatment effects with both Run and Tag effects, as different runs and tags contain different combinations of plants and treatments.

The theoretical ANOVA of the Phase 2 experiment is presented in Table 4.26. A total of 35 DF are separated to 8 DF for Between Runs stratum and 27 DF for Within Runs stratum. In the Between Runs stratum, there are 1 DF for the Between Trays stratum and 3 DF for the Between Plants Within Trays stratum. The 3 DF associated with the Between Plants Within Trays Between Runs stratum are confounded with the 3 DF associated with Treatment effects

Table 4.25: Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags with the Tray effects are intentionally confounded with the Tag effects, when the Phase 1 experiment consists of $\nu = 6$ treatments assigned to each of $n_p = 18$ plants in each of $n_b = 3$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 9$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	2Jd	1Bb	1Ee	2Ic
2	1Bb	2Jd	2Ic	1Ee
3	1Cc	2Ke	2Lf	1Aa
4	2Ke	1Cc	1Aa	2Lf
5	1Ff	2Ga	2Hb	1Dd
6	2Ga	1Ff	1Dd	2Hb
7	3Pd	3Ma	3Oc	3Nb
8	3Ma	3Pd	3Nb	3Oc
9	3Qe	3Qe	3Rf	3Rf

based on 0.1667 of treatment information. In the Within Runs stratum, there are 1 DF associated with the Between Trays stratum, 12 DF with the Between Plants Within Trays stratum, and 14 DF with the Within Plants Within Trays stratum. Since there is 1 DF associated with Tag effects and 5 DF associated with the Treatment effects in the Between Plants Within Runs stratum, there are still 6 Residual DF for estimating the variance of Treatments effects, which is reduced from 10 DF of the Phase 1 experiment. The 4 DF associated with Plants Within Trays stratum are lost to the 3 DF in the Between Runs stratum and 1 DF associated with Tag effects. In addition, the amount of treatment information remaining is 0.8204, compared to 100% in the Phase 1 experiment.

Notice that the treatment average efficiency factor in Between Plants Within Trays Between Runs (0.1667) and Within Runs (0.8204) does not add up to 1. This is because the treatment average efficiency factor in the Between Plants Within Trays Within Runs stratum of 0.8204 is computed from the harmonic mean of the five treatment canonical efficiency factors of 1, 0.894, 0.889, 0.833 and 0.606. Since these five treatment canonical efficiency factors are not identical, the optimal design of the Phase 2 experiment is not balanced. To examine how the treatment comparison are made for each of these treatment canonical efficiency factors, we

Table 4.26: Theoretical ANOVA table for the Phase 2 experiment in Table 4.25.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 12\sigma_b^2 + 4\sigma_r^2$		
Between Plants Within Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + \theta_\tau$		0.1667
Residual	4	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 12\sigma_b^2$		
Between Plants Within Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 9\theta_\gamma + 0.67\theta_\tau$	1	0.1111
Treatment	5	$\sigma^2 + 2\sigma_p^2 + 4.923\theta_\tau$		0.8204
Residual	6	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 9\theta_\gamma$	1	
Residual	12	σ^2		

generate the five basic contrasts corresponding to the five treatment canonical efficiency factors

$$\begin{matrix} & 1 & 2 & 3 & 4 & 5 \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{pmatrix} 0.2887 & -0.5577 & 0.4082 & -0.5 & -0.1494 \\ 0.2887 & 0.1494 & 0.4082 & 0.5 & 0.5577 \\ -0.2887 & -0.5577 & -0.4082 & 0.5 & -0.1494 \\ 0.5774 & 0.4082 & -0.4082 & 0 & -0.4082 \\ -0.2887 & 0.1494 & -0.4082 & -0.5 & 0.5577 \\ -0.5774 & 0.4082 & 0.4082 & 0 & -0.4082 \end{pmatrix} \end{matrix}.$$

The first contrast with 100% of treatment information, compares Treatment a , b and d with Treatments c , e , f ; the second contrast with 0.894 of treatment information, compares Treatments b , d , e and f with Treatments a and c ; the third contrast with 0.889 of treatment information compares, Treatments a , b and f with Treatments c , d and e ; the fourth contrast with 0.833 of treatment information, compares Treatments b and c with Treatments a and e ; and the fifth contrast with 0.606 of treatment information, compares Treatments b and e with Treatments a , c , d and f .

The theoretical ANOVA given in Table 4.27 is constructed using these five basic treatment contrasts to ascertain which stratum for each of these five basic treatment contrasts are estimated. Treatment contrast 1 has all its information in the Between Plants Within Trays Within Runs stratum. The Tag effects in the Between Plants Within Trays Within Runs stratum contains 0.111 of treatment information from treatment contrast 3, which means there is still

0.889 of treatment information remaining. Treatment contrasts 2, 4 and 5 have 0.1057, 0.1667 and 0.3943, respectively, of treatment information in the Between Plants Within Trays Between Runs stratum. Note that the harmonic mean of 0.1057, 0.1667 and 0.3943 is 0.1667, which represents the 0.167 of treatment information in the Between Plants Within Trays Between Runs stratum in the theoretical ANOVA in Table 4.26. The amount of treatment information remaining in the Between Plants Within Trays Between Runs stratum for treatment contrasts 1, 2, 3, 4 and 5 are 1, 0.894, 0.889, 0.833 and 0.606, respectively, giving us a harmonic mean of 0.8204 for the amount of treatment information in Between Plants Within Trays Within Runs stratum.

Table 4.27: Theoretical ANOVA for the Phase 2 experiment in Table 4.25 with five basic treatment contrasts showing which stratum estimates each of the five different basic treatment contrasts.

Source of Variation	DF	EMS	E_{γ}	E_{τ_1}	E_{τ_2}	E_{τ_3}	E_{τ_4}	E_{τ_5}
Between Runs								
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 12\sigma_b^2 + 4\sigma_r^2$						
Between Plants Within Trays								
Treatment contrast 2	1	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + 265/418\theta_{\tau_2}$			0.1057			
Treatment contrast 4	1	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + \theta_{\tau_4}$					0.1667	
Treatment contrast 5	1	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + 1351/571\theta_{\tau_5}$						0.3943
Residual	4	$\sigma^2 + 4\sigma_r^2$						
Within Runs								
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 12\sigma_b^2$		9				
Between Plants Within Trays								
Tag	1	$\sigma^2 + 2\sigma_p^2 + 9\theta_{\gamma} + 2/3\theta_{\tau_3}$	1			0.1111		
Treatment contrast 1	1	$\sigma^2 + 2\sigma_p^2 + 6\theta_{\tau_1}$		1				
Treatment contrast 2	1	$\sigma^2 + 2\sigma_p^2 + 3064/571\theta_{\tau_2}$			0.8943			
Treatment contrast 3	1	$\sigma^2 + 2\sigma_p^2 + 16/3\theta_{\tau_3}$				0.8889		
Treatment contrast 4	1	$\sigma^2 + 2\sigma_p^2 + 5\theta_{\tau_4}$					0.8333	
Treatment contrast 5	1	$\sigma^2 + 2\sigma_p^2 + 1519/418\theta_{\tau_5}$						0.6057
Residual	6	$\sigma^2 + 2\sigma_p^2$						
Within Plants Within Trays								
Tag	2	$\sigma^2 + 9\theta_{\gamma}$	1					
Residual	12	σ^2						

4.9 Optimal designs for experiments involving two to eight treatments and two technical replicates

A table of optimal designs for a range of design parameters is generated using the objective function described in Section 4.5, and the SA algorithm described in Section 4.7. This set of optimal designs is for Phase 1 experiments featuring $\nu = 2, \dots, 8$ treatments, $n_p = \nu r_b$ plots (plants), $n_b = 2, \dots, 10$ blocks (trays), $n_s = 2$ sub-samples, $n_\gamma = 4, 8$ tags, and $n_r = n/n_\gamma$ runs, where r_b denotes the number of biological replicates and n denotes total number of sub-samples, ($r_b = 2, \dots, 8$). This set of optimal designs is presented in Appendix E. This catalogue of designs enables any biologist without any experience in experimental design, to undertake a two-phase experiment that provides maximum precision for any given set of design parameters. A set of tables, summarising the properties of the theoretical ANOVA table for each optimal design of the Phase 2 experiment, is presented in the Appendix F. This section discusses the properties of the optimal designs found.

The general layout of the design is the same as what have been described in Chapter 3 when the Phase 1 experiment is arranged in a CRD. The design form of a two-way table comprising n/n_γ rows and n_γ columns, where n_γ can be four for the four-plex experiment or eight for the eight-plex experiment.

4.9.1 Phase 1 Block and Plot effects

With the addition of Block component in the Phase 1 experiment, the Residual DF in the Between Plots Within Blocks stratum will be decreased to accommodate the additional DF for estimating the Between Blocks effect at Phase 1. Thus, the Residual DF in the Between Plots Within Blocks Within Runs of the Phase 2 experiment will decrease as well. Therefore, in general, more number of Blocks, n_b , is used in the Phase 1 experiment, the lower the Residual DF in the Between Plots Within Blocks Within Runs stratum.

As shown in the example in Section 4.5, using the design when the Tag effects are intentionally confounded with the Block effects, the resultant design can have higher Residual DF in the Between Plots Within Blocks Within Runs stratum. This is because, based on the structure that we setup the initial design, the optimal designs of the Phase 2 experiment always have a property that 1 and 3 DF associated with Tag effects in the Between Plots Within Blocks stratum for the four-plex and eight-plex experiments, respectively. Having a design when the Block effects are intentionally confounded with Tag effects, the DF associated with Tag effects

will move from the Between Plots Within Blocks Within Runs stratum to Between Blocks Within Runs stratum, which free up the Residual DF in the Between Plots Within Blocks Within Runs stratum. This cases only applies where the number of blocks is even, because it is not possible to allocate the odd number of blocks equally across four or eight tags of the Phase 2 experiment. However, there can also be the case when the Block effects are intentionally confounded with the Run effects, which can obtain a better optimal design with higher Residual DF in the Between Plots Within Blocks Within Runs stratum. Such case is described in Section 4.6.1.

Comparing between the four-plex and eight-plex experiments, similar to the case when the Phase 1 experiment is arranged in a CRD, with a lower number of experimental units are used in the Phase 1 experiment, it is still recommended to use the four-plex system compare to the eight-plex system. The eight-plex system should be used instead of the four-plex system, where the number of the experimental units used in the Phase 1 experiment increases. However, there is not a clear cut-off number of experimental units, that the eight-plex system become better than the four-plex system. This is because having the an additional Block component can generate designs with higher Residual DF when the Blocks effects that confounded with the Tag effects.

4.9.2 Treatment effects

In general, the Treatment average efficiency factors from the optimal designs found are the same as for the Phase 1 experiment is arranged in CRD. Treatment effects are confounded with the Tag effects if the number of runs is not divisible by the number of treatments. Moreover, Treatment effects can also be confounded with Run effects where the number of tags is not divisible by the number of treatments. In most situations, both types of confounding results in optimal design with the treatment canonical efficiency factors not being identical. An example of the canonical efficiency factors not being identical is presented in Section 4.8. The Appendix F presents the treatment average and canonical efficiency factors of every optimal design found.

4.10 Extension when the Phase 1 experiment is arranged with BIBD

When the number of treatments exceeds the number of plots in a block, the design is known as an incomplete block design. A particular type of incomplete block design occurs when all blocks are of equal size and all treatments are equally replicated, where each pair of treatments occurs in two blocks together equally often. Since the treatment information of all treatment contrasts

is split evenly across the Between Blocks and Within Blocks strata, the amount of treatment information in the Within Blocks stratum is decreased.

From the optimal design found, when the Phase 1 experiment is arranged in either CRD or RCBD, it has been shown that the amount of treatment information can be diluted in the Phase 2 experiment. Thus, when allocating the samples from the Phase 1 experiment arranged in a BIBD to be analysed in the Phase 2 experiment, the amount of treatment information is likely to be further diluted.

Since the Phase 1 experiment still comprises a block structure of Blocks and Plots within Blocks, the aim of finding the optimal design is the same as when the Phase 1 experiment is arranged in a RCBD. Hence, we can still use the same objective function defined in Section 4.5.2 to find the optimal design of the Phase 2 experiment.

4.10.1 BIBD example with six treatments

Consider the Phase 1 experiment involving $\nu = 6$ treatments (labelled a, \dots, f) assigned to $n_p = 30$ plants (labelled AA, \dots, BE) in $n_b = 6$ trays (labelled $1, \dots, 6$). Since the Phase 1 experiment involves $n_p = 30$ plants, the labelling system requires two upper case letters instead of one. Table 4.28 illustrates the layout of the Phase 1 experiment, in which each treatment is replicated five times and is assigned to plants in five of six trays. In addition, any pair of treatments is present together in four trays.

Table 4.28: Phase 1 experimental design with $\nu = 6$ treatments assigned to $n_p = 30$ plants in $n_b = 6$ trays. Upper case letters denote plant IDs, while the lower case letters denote the treatments.

Tray 1	<i>AAb</i>	<i>ABc</i>	<i>ACd</i>	<i>ADe</i>	<i>AEf</i>
Tray 2	<i>AFc</i>	<i>AGd</i>	<i>AHe</i>	<i>AI f</i>	<i>AJa</i>
Tray 3	<i>AKd</i>	<i>ALe</i>	<i>AMf</i>	<i>ANa</i>	<i>AO b</i>
Tray 4	<i>APe</i>	<i>AQf</i>	<i>ARa</i>	<i>ASb</i>	<i>ATc</i>
Tray 5	<i>AUf</i>	<i>AVa</i>	<i>AWb</i>	<i>AXc</i>	<i>AYd</i>
Tray 6	<i>AZa</i>	<i>BA b</i>	<i>BBc</i>	<i>BCd</i>	<i>BDe</i>

Table 4.29 shows the theoretical ANOVA of the Phase 1 experiment in Table 4.28. The total of 29 DF are separated into Between Trays (5 DF) and Between Plants Within Trays (24 DF) strata. All 5 DF associated with the Between Trays stratum are confounded with the 5 DF associated with the Treatment effects. In the Within Trays stratum, since there are 5 DF associated with Treatment effects, there are 19 Residual DF in the Between Plants Within Trays stratum. As for the amount of treatment information, it is split between 0.04 and 0.96 in the Between Trays and Between Plants Within Trays strata, respectively. Since we know the number

of treatments is 6 and the size of each tray is 5, the amount of treatment information in the Between Plants Within Trays stratum can be calculated directly by

$$E_\tau = \frac{6(5-1)}{(6-1)5} = 0.96.$$

Table 4.29: Theoretical ANOVA table of the Phase 1 experiment arranged in a BIBD with $\nu = 6$ treatments assigned to $n_p = 5$ plants within each of $n_b = 6$ trays.

Source of Variation	DF	EMS	E_γ
Between Trays			
Treatments	5	$\sigma_p^2 + 5\sigma_b^2 + 0.2\theta_\gamma$	0.04
Between Plants Within Trays			
Treatments	5	$\sigma_p^2 + 4.8\theta_\gamma$	0.96
Residual	19	σ_p^2	

Each plant of the Phase 1 experiment is split into $n_s = 2$ sub-samples and analysed in $n_r = 15$ runs and $n_\gamma = 4$ tags of Phase 2 experiment. Since the number of trays is even, we can use an initial design when the Tray effects are intentionally confounded with Tag effects to start the search. The optimal design of the Phase 2 experiment is shown in Table 4.30, which shows the sub-samples from Trays 1, 2 and 3 are differentially labelled by Tags 114 and 115, while sub-samples from Trays 4, 5 and 6 are differentially labelled by Tags 116 and 117. Additionally, Runs 1 to 10 contain Trays 1, 2, 4 and 5, while Runs 11 to 15 contain only Trays 3 and 6. Hence, the Tray effects are confounded with both Run and Tag effects. Since Treatment effects are confounded with Tray effects in the Phase 1 experiment, the Treatment effects are also confounded with both Run and Tag effects in the Phase 2 experiment.

Table 4.31 shows the theoretical ANOVA of the Phase 2 experiment. The total of 59 DF is separated into 14 DF in the Between Runs stratum and 45 DF in the Within Runs stratum. The Between Runs stratum (14 DF) is further partitioned into the Between Trays (3 DF) and the Between Plants Within Trays (4 DF) strata, where 3 DF associated with the Treatment effects are present in both Between Trays and Between Plants Within Trays strata. Within Runs stratum (45 DF) is partitioned into Between Trays (4 DF), Between Plants Within Trays (18 DF) and Within Plants Within Trays (23 DF) strata. Since this design is constructed from the initial design when Tray effects are confounded more with Tag effects, 1 DF associated with Tag effects is in the Between Trays stratum. The Treatment effects in the Between Plants Within Trays Within Runs stratum can be estimated with 0.8606 of treatment average efficiency factor, which is computed from five treatment canonical efficiency factors of 0.9383, 0.9, 0.8736, 0.8217 and 0.7864. This amount of treatment information is smaller than that which can be yielded

Table 4.30: Optimal design for the Phase 2 experiment showing assignment of trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 6$ treatments assigned to $n_p = 30$ plants in $n_b = 6$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 15$ runs and $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

Run	Tag			
	114	115	116	117
1	2AJa	2AIf	5AXc	5AYd
2	2AIf	2AJa	5AYd	5AXc
3	1ADe	2AGd	4ASb	5AVa
4	2AGd	1ADe	5AVa	4ASb
5	1ABc	1ACd	5AWb	5AUf
6	1ACd	1ABc	5AUf	5AWb
7	2AFc	2AHe	4ARa	4AQf
8	2AHe	2AFc	4AQf	4ARa
9	1AAb	1AEf	4APe	4ATc
10	1AEf	1AAb	4ATc	4APe
11	3ANa	3AMf	6BDe	6BAb
12	3AMf	3ANa	6BAb	6BDe
13	3AOB	3AKd	6AZa	6BBc
14	3AKd	3AOB	6BBc	6AZa
15	3ALe	3ALe	6BCd	6BCd

from the Phase 1 experiment of 0.96 as shown in Table 4.29. The Residual DF in the Between Plants Within Trays Within Runs stratum is 13 DF which is reduced from 19 DF in the Phase 1 experiment as shown in Table 4.29. However, there is still valid F-test for testing the Treatment effects.

Further dissecting theoretical ANOVA in Table 4.31 reveals 3 and 4 DF associated with the Between Trays stratum in the Between and Within Runs strata, respectively. Given that the Phase 1 experiment only involves six trays, the total DF associated with Tray effects should be 5 DF. Additionally, the coefficients of the Between Trays variance component, σ_b^2 , are not identical across the theoretical ANOVA, which suggests the Tray effects are confounded with Run effects with an unbalanced structure.

The confounding between Tray effects and Run effects can be confirmed by using a single-phase theoretical ANOVA in Table 4.32 by fitting the Tray and Plant effects as the fixed effects. The 0.5 of the tray information in the Between Runs stratum is derived from the three canonical efficiency factors of 1, 0.4 and 0.4. The 0.75 of the tray information in the Within Runs stratum is derived from the four canonical efficiency factors of 1, 1, 0.6 and 0.6.

The consequence of this confounding causes the Plant effects to be confounded with Tray effects given 0.6 and 0.4 of plant information being estimated with Tray effects of the Between

Table 4.31: Theoretical ANOVA for the Phase 2 experiment in Table 4.30.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 6\sigma_b^2 + 4\sigma_r^2 + 0.0667\theta_\tau$		0.0667
Between Plants Within Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2 + 1.098\theta_\tau$		0.1098
Residual	1	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2$		
Residual	7	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 10\sigma_b^2 + 15\theta_\gamma + 0.4\theta_\tau$	1	0.04
Treatment	1	$\sigma^2 + 2\sigma_p^2 + 10\sigma_b^2 + 0.4\theta_\tau$		0.04
Residual	2	$\sigma^2 + 2\sigma_p^2 + 6\sigma_b^2$		
Between Plants Within Trays				
Treatment	5	$\sigma^2 + 2\sigma_p^2 + 8.606\theta_\tau$		0.8606
Residual	13	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 15\theta_\gamma$	1	
Residual	21	σ^2		

Runs and Within Runs strata, respectively. Notice that the theoretical ANOVA of Phase 2 experiment contains 13 Residual DF in the Between Plants Within Trays Within Runs stratum as shown in Table 4.31. The theoretical ANOVA of the Phase 1 experiment in Table 4.29 shows the Residual DF of Between Plants Within Trays stratum is 19 DF. This means there are 6 Residual DF lost moving from the Phase 1 to the Phase 2 design. There are 4 DF of the Between Plants Within Trays are being estimated in the Between Run stratum of the Phase 2 experiment. The remaining missing 2 Residual DF are the 2 DF of Tray effects that are confounded with Run and Plant effects.

Table 4.32: Theoretical ANOVA with DF and average efficiency factors for the Phase 2 experiment in Table 4.30 which generated by treating Tray and Plant effects as the fixed effects.

Source of Variation	DF	E_b	E_p
Between Runs			
Tray	3	0.5	0.6
Plant Within Trays	4		1
Residual	7		
Within Runs			
Tray	4	0.75	0.4
Plant Within Trays	18		1
Residual	23		

4.10.2 Optimal designs when the Phase 1 experiment is a BIBD

Using the objective function derived in Section 4.5, and the SA algorithm described in Section 4.7, a set of optimal designs were found and are presented in Appendix G. This set of optimal designs is for experiments with $\nu = 4, \dots, 8$ treatment, $n_b = \nu r_b$ plots, $n_b = \nu$ trays, $n_s = 2$ sub-samples, $n_\gamma = 4, 8$ tags and $n_r = n/n_\gamma$ runs ($r_b = 3, \dots, 7$). The researcher can again obtain a design for their experiment with a given set of design parameters. A table, summarising the properties of the theoretical ANOVA for each optimal Phase 2 design, is presented in the Appendix H. This section discusses the properties of the optimal designs found.

A four-plex system can only be used for experiments with $\nu = 6$ treatments and $n_p = 30$ plots, and $\nu = 7$ treatments and $n_p = 42$ plots. Due to the confounding of Block effects with both Run and Treatment effects, both the treatment average efficiency factors, E_τ , and Residual DF are further diluted from the Phase 1 experiment.

Both four-plex and eight-plex systems can be used for the Phase 2 experiment when Phase 1 experiments involves $\nu = 4$ treatments with $n_p = 12$ plots, $\nu = 5$ treatments with $n_p = 20$ plots, $\nu = 7$ treatments with $n_p = 28$ plots, and $\nu = 8$ treatments with $n_p = 56$ plots. For Phase 1 experiments involving $\nu = 4$, $\nu = 5$ and $\nu = 7$ treatments, the treatment average efficiency factors, E_τ , are the same between four-plex and eight-plex systems and preserved from the Phase 1 experiment. However, their Residual DF have all been reduced from the Phase 1 experiment, with the four-plex experiment yielding higher Residual DF for experiments involving $\nu = 5$ and $\nu = 7$ treatments and the eight-plex experiment yielding higher Residual DF for experiments involving $\nu = 4$ treatments. For the experiment with $\nu = 8$ treatments, the eight-plex experiment must be used as it preserves the treatment average efficiency factors from the Phase 1 experiment and yields higher Residual DF compared to the four-plex experiment.

4.11 Summary

This Chapter presented a method for finding optimal designs of Phase 2 experiments when the Phase 1 experiment is arranged in a RCBD or a BIBD. First step was to derived a new four-criterion objective function for finding optimal designs for Phase 2 experiments, when the Phase 1 experiment is arranged in a RCBD or a BIBD. These four criteria are: 1) Phase 1 Plots average efficiency factor in the Within Runs and Tags vector subspace must be 1, 2) the DF associated with treatment effects in the Between Plots Within Blocks Within Runs stratum must be intact. 3) the Residual DF in the Between Plots Within Blocks Within Runs stratum is maximised and 4) the treatment average efficiency factor in the Between Plots Within Blocks Within Runs

and Tags vector subspace is maximised. We have shown that this new four-criterion objective function can find optimal designs with different combinations of the design parameters.

Even though Treatment effects are estimated in the Within Blocks stratum, the construction of the initial design must take Phase 1 Blocks into account. This is because Phase 1 block can be allocated such that the Phase 1 Block effects are intentionally confounded with either Runs effects or Tags effects, which has shown to increase the Residual DF in the Between Plots Within Blocks stratum.

Finally, this Chapter also showed that the same method can be used when the Phase 1 experiment is arranged in a BIBD. Since the block structures are identical in such a design, the components of the objective function and the SA algorithm do not require adjustment. The main issue is that the treatment information has already been diluted since some of the treatment information is in the Between Blocks stratum of the Phase 1 experiment; hence, the degree to which treatment information can be further diluted in some optimal designs of Phase 2 experiments is problematic.

Chapter 5

Estimation of variance components and effective degrees of freedom in two-phase proteomic experiments

5.1 Introduction

In Chapter 2, we described the methodology needed to construct theoretical ANOVA tables for two-phase experiments. Chapters 3 and 4 then developed methods for constructing and searching for optimal designs for Phase 2 proteomics experiments when the Phase 1 experiments is arranged in a completely randomised design (CRD), a randomised complete block design (RCBD), or a balanced incomplete block design (BIBD). Theoretical ANOVA tables were shown to be a very useful tool for investigating and comparing the properties of optimal designs of different Phase 2 experiments for a given Phase 1 design. This Chapter presents a third component in estimating variance components (VCs) based on expected mean squares (EMS) of the theoretical ANOVA table, specifically the Residual mean squares (MS) of the same stratum for testing the Treatment effects, an example can be in the Between Animals Within Runs stratum. In addition, the assumption that the Phase 2 Block (Run) effects are random may allow us to obtain a test that effectively has higher degrees of freedom (DF) for the Residual MS, namely *effective degrees of freedom* (EDF).

Jarrett and Ruggiero (2008) demonstrated that given the same design at Phase 1, the choice of design at Phase 2 can affect the analysis of a micro-array experiment. MudPIT-iTRAQ™ experiments have their own unique set of problems. Either a four-plex or eight-plex labelling system can be used for the Phase 2 proteomics experiment, which allows researchers to measure

either four or eight biological samples simultaneously. The EDF provide us with another property to compare designs using four- and eight-plex systems, which we will apply to some of the optimal designs of the Phase 2 experiments found in Chapters 3 and 4.

This Chapter first uses an optimal design, described in Section 5.2, to illustrate the estimation of VCs in Section 5.3. Based on the VCs estimates, the approximation method for the EDF is then shown in Section 5.4. Section 5.5 compares the EDF between the Phase 2 design with four-plex and eight-plex systems given the same Phase 1 experiment arranged in a CRD. Section 5.6 compares the EDF between four Phase 2 designs, given the same Phase 1 experiment is arranged in a RCBD, using two different confounding schemes when the Phase 1 Block effects are intentionally confounded with Tag effects or Phase 1 Block effects are intentionally confounded with Run effects, and between the four-plex and eight-plex systems.

5.2 An illustrative example

This section presents the most trivial example of a two-phase experiment, in which the Phase 1 experiment is arranged in a CRD with $\nu = 2$ treatments assigned to $n_a = 6$ animals. The layout of the Phase 1 design consists of Treatment a assigned to Animals A , C and E , and Treatment b assigned to Animals B , D and F . The samples from each animal are further split into n_s sub-samples which are differentially labelled by n_γ tags and analysed in n_r runs of the Phase 2 experiment. The linear model of this Phase 2 experiment has been previously described in (3.2).

Table 5.1: Optimal design of Phase 2 proteomics experiment showing allocation of sub-samples from animals and treatment to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to each of $n_a = 6$ animals, $n_s = 2$ sub-samples are then taken from each animal, and labelled by $n_\gamma = 4$ tags and analysed in $n_r = 3$ runs of the Phase 2 MudPIT-iTRAQTM experiment.

Run	Tag			
	114	115	116	117
1	Db	Ca	Fb	Ea
2	Ca	Db	Ea	Fb
3	Bb	Bb	Aa	Aa

An optimal design found using the methods described in Chapter 3 is shown in Table 5.1. There are several characteristics of this design: Runs 1 and 2 contains sub-samples from Animals C , D , E and F , while Run 3 contains sub-samples from Animals A and B . Thus, 1 of 4 DF associated with the Animal effects is confounded with 1 DF of the Between Runs stratum. Furthermore, Animals B , C and D are assigned to Tags 114 and 115, and Animals A , E and

F are assigned to Tags 116 and 117. Thus, 1 DF associated with Tag effects is in the Between Animals stratum. Treatment effects are orthogonal to Run effects, as each run contains two of each treatment. There are unequal numbers of sub-samples from Treatment a and b labelled with each Tag resulting in some confounding between Treatment and Tag effects.

Table 5.2: Theoretical ANOVA table of the Phase 2 experiment in Table 5.1.

Source of Variation	DF	MS	EMS	E_γ	E_τ
Between Runs					
Between Animals	1	s_1^2	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Within Animals	1	s_2^2	$\sigma^2 + 4\sigma_r^2$		
Within Runs					
Between Animals					
Tag	1		$\sigma^2 + 2\sigma_a^2 + 3\theta_\gamma + 0.67\theta_\tau$	1	0.111
Treatment	1		$\sigma^2 + 2\sigma_a^2 + 5.33\theta_\tau$		0.889
Residual	2	s_3^2	$\sigma^2 + 2\sigma_a^2$		
Within Animals					
Tag	2		$\sigma^2 + 3\theta_\gamma$	1	
Residual	3	s_4^2	σ^2		

The theoretical ANOVA of the optimal design of the Phase 2 experiment in Table 5.1 is presented in Table 5.2. Notice that an additional column is present in this table, namely the mean square (MS) column, i.e. the estimated value of its corresponding EMS being computed from experimental data. There are four MS of interest, which contain only the random error: the Between Animals Between Runs MS (s_1^2), the Within Animals Between Runs MS (s_2^2), the Residual MS of the Between Animals Within Runs stratum (s_3^2) and the Residual MS of Within Animals Within Runs stratum (s_4^2). These MS, also known as residual variances, related to their associated EMS as follows

$$s_1^2 = \hat{\sigma}^2 + 2\hat{\sigma}_a^2 + 4\hat{\sigma}_r^2, \quad (5.1)$$

$$s_2^2 = \hat{\sigma}^2 + 4\hat{\sigma}_r^2, \quad (5.2)$$

$$s_3^2 = \hat{\sigma}^2 + 2\hat{\sigma}_a^2, \quad (5.3)$$

$$s_4^2 = \hat{\sigma}^2, \quad (5.4)$$

where $\hat{\sigma}^2$, $\hat{\sigma}_a^2$, and $\hat{\sigma}_r^2$ are the estimated between sub-samples, between animals, and between runs VCs, respectively. The $\hat{\sigma}^2$ is the residual variance from the Within Animals Within Runs stratum as shown in (5.4). Subtracting (5.4) from (5.3) giving $\hat{\sigma}_a^2 = (s_3^2 - s_4^2)/2$. We then subtract (5.4) from (5.2) to solve for $\hat{\sigma}_r^2$, which gives $\hat{\sigma}_r^2 = (s_2^2 - s_4^2)/4$. Following Jarrett and Ruggiero (2008), this method of solving a system of linear equations to obtain estimates of the variance components is, from this point forward, referred to as the linear combination (LC) method.

A purpose of constructing the theoretical ANOVA table is it enables us to determine variances of Treatment effects and, therefore, a valid F-test for the Treatment effects can be obtained. The estimated Treatment differences between Treatment a and b are given by $(\bar{y}_{...}^{a..} - \bar{y}_{...}^{b..})$, which, for the Phase 2 design presented in Table 5.1, has variance

$$\text{Var}(\bar{y}_{...}^{a..} - \bar{y}_{...}^{b..}) = \frac{\sigma^2 + 2\sigma_a^2}{3}, \quad (5.5)$$

where $\bar{y}_{...}^{i..}$ denotes the overall mean log-protein abundance values overall observations of Treatment i . The denominator of (5.5) is 3, because each treatment group is replicated six times (with two sub-samples from each of three animals). The numerator of (5.5), i.e. $\sigma^2 + 2\sigma_a^2$, can be estimated directly from the residual variance in the Within Animals Within Runs stratum, the same stratum in which the Treatment effects are being estimated.

If Run effects are regarded as fixed effects, the row of Table 5.2 containing σ_r^2 is not available to be used, i.e. s_1^2 and s_2^2 , which implies that the estimate of variance of the Treatment effects will be based solely on s_3^2 with 2 DF. If the Run effects are assumed to be random effects, we can recover extra information about σ_a^2 from s_1^2 , defined in (5.1), as well as information about σ^2 in the other residual variances. Thus, we can then improve the estimate of $\hat{\sigma}^2 + 2\hat{\sigma}_a^2$ either the LC method, as illustrated above, or by using restricted maximum likelihood (REML) approach to estimate the VC. How this is may be achieved is discussed in Section 5.3.

5.3 Estimation of variance components

The theoretical ANOVA in Table 5.2 identified four MS, with expected values ξ_i^2 and v_i DF ($i = 1, \dots, 4$), which are available for estimation of the VCs defined by $\boldsymbol{\varsigma} = (\hat{\sigma}^2, \hat{\sigma}_a^2, \hat{\sigma}_r^2)'$. This section shows the REML approach in estimating the VCs developed by Jarrett and Ruggiero (2008).

5.3.1 Constructing the score function and Fisher information matrix

The mean squares s_i^2 are assumed to have a χ^2 distribution, i.e.

$$s_i^2 \sim \frac{\xi_i^2}{v_i} \chi_{v_i}^2, \quad (i = 1, \dots, 4), \quad (5.6)$$

where v_i denotes the DF corresponding to s_i^2 . The log-likelihood of s_i^2 can be shown to be

$$L(\xi_i^2; s_i^2) = \text{constant} - \sum_{i=1}^4 \left[\frac{v_i \log(\xi_i^2)}{2} + \frac{v_i s_i^2}{2\xi_i^2} \right], \quad (i = 1, \dots, 4). \quad (5.7)$$

The score is the first derivative of the log-likelihood function with respect to the i th element of EMS, ξ_i^2 , i.e.

$$\frac{\partial L(\xi_i^2; s_i^2)}{\partial \xi_i^2} = \frac{v_i(s_i^2 - \xi_i^2)}{2\xi_i^4}, \quad (i = 1, \dots, 4).$$

Since the expected values ξ^2 is a vector containing $(\xi_1^2, \xi_2^2, \xi_3^2, \xi_4^2)'$, the score function with respect to ξ^2 can be re-written in vector form as follows

$$S(\xi^2) = \frac{\partial L(\xi^2; s^2)}{\partial \xi^2} = \begin{pmatrix} \frac{v_1(s_1^2 - \xi_1^2)}{2\xi_1^4} \\ \frac{v_2(s_2^2 - \xi_2^2)}{2\xi_2^4} \\ \frac{v_3(s_3^2 - \xi_3^2)}{2\xi_3^4} \\ \frac{v_4(s_4^2 - \xi_4^2)}{2\xi_4^4} \end{pmatrix}. \quad (5.8)$$

The Fisher information is defined as the variance of the score, which is computed from the negative of the expectation of the second derivative of the log-likelihood function with respect to ξ_i^2 . This is also known as the expected Fisher information.

Since the expected values ξ^2 is a vector containing $(\xi_1^2, \xi_2^2, \xi_3^2, \xi_4^2)'$, the negative expectation of the second partial derivative of the log-likelihood function gives a 4×4 Fisher information matrix, i.e.

$$E\left(-\frac{\partial^2 L(\xi^2; s^2)}{\partial \xi^4}\right) = E\left(-\begin{bmatrix} \frac{\partial^2 L}{\partial \xi_1^4} & \frac{\partial^2 L}{\partial \xi_1^2 \partial \xi_2^2} & \frac{\partial^2 L}{\partial \xi_1^2 \partial \xi_3^2} & \frac{\partial^2 L}{\partial \xi_1^2 \partial \xi_4^2} \\ \frac{\partial^2 L}{\partial \xi_2^2 \partial \xi_1^2} & \frac{\partial^2 L}{\partial \xi_2^4} & \frac{\partial^2 L}{\partial \xi_2^2 \partial \xi_3^2} & \frac{\partial^2 L}{\partial \xi_2^2 \partial \xi_4^2} \\ \frac{\partial^2 L}{\partial \xi_3^2 \partial \xi_1^2} & \frac{\partial^2 L}{\partial \xi_3^2 \partial \xi_2^2} & \frac{\partial^2 L}{\partial \xi_3^4} & \frac{\partial^2 L}{\partial \xi_3^2 \partial \xi_4^2} \\ \frac{\partial^2 L}{\partial \xi_4^2 \partial \xi_1^2} & \frac{\partial^2 L}{\partial \xi_4^2 \partial \xi_2^2} & \frac{\partial^2 L}{\partial \xi_4^2 \partial \xi_3^2} & \frac{\partial^2 L}{\partial \xi_4^4} \end{bmatrix}\right),$$

where L denotes $L(\xi_i^2; s_i^2)$ with the i th diagonal element given by

$$\frac{\partial^2 L}{\partial \xi_i^4} = -\frac{v_i s_i^2}{\xi_i^6} + \frac{v_i}{2\xi_i^4}, \quad (i = 1, \dots, 4).$$

Its negative has expectation

$$E\left(-\frac{\partial^2 L}{\partial \xi_i^4}\right) = E\left(\frac{v_i s_i^2}{\xi_i^6} - \frac{v_i}{2\xi_i^4}\right) = \frac{v_i E(s_i^2)}{\xi_i^6} - \frac{v_i}{2\xi_i^4} = \frac{v_i \xi_i^2}{\xi_i^6} - \frac{v_i}{2\xi_i^4} = \frac{v_i}{2\xi_i^4}.$$

The off-diagonal elements of the Fisher information matrix, $\frac{\partial^2 L}{\partial \xi_i^2 \partial \xi_j^2}$, ($i \neq j$), are all zero. Thus,

it follows that the Fisher information matrix with respect to ξ^2 is given by

$$\mathbf{A}_{\xi^2} = \mathbf{E} \left(-\frac{\partial^2 L(\xi^2; \mathbf{s}^2)}{\partial \xi^4} \right) = \text{diag} \left(\frac{v_i}{2\xi_i^4} \right), \quad (i = 1, \dots, 4). \quad (5.9)$$

5.3.2 Transformation from expected values in ξ^2 to estimates in ς

The score function and Fisher information matrix, as defined in (5.8) and (5.9), respectively, are functions of ξ^2 . However, the goal here is to estimate the VCs, which this formulation will not allow. Thus, we must first transform the score function and Fisher information matrix to functions of ς .

The relationship between the vector of expected values in ξ^2 and the vector of VCs estimates in ς can be written as

$$\xi^2 = \mathbf{G}\varsigma,$$

where the matrix \mathbf{G} , has (i, j) -th element equal to $\partial \xi_i^2 / \partial \varsigma_j$. From the theoretical ANOVA in Table 5.2, it follows that the row elements of the \mathbf{G} matrix are the coefficients of the VCs each Residual EMS, i.e.

$$\begin{bmatrix} 1 & 2 & 4 \\ 1 & 0 & 4 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Based on the product rule for differentiation, it follows that

$$\frac{\partial \xi^2}{\partial \varsigma} = \mathbf{G}. \quad (5.10)$$

Since the log-likelihood in (5.7) is a function of ξ^2 containing four elements, the change of variable technique can be implemented by using the multi-variable chain rule to calculate the score function with respect to ς , i.e.

$$S(\varsigma) = \frac{\partial L(\xi^2)}{\partial \varsigma} = \frac{\partial L(\xi^2)}{\partial \xi^2} \frac{\partial \xi^2}{\partial \varsigma}, \quad (5.11)$$

where $\frac{\partial \xi^2}{\partial \varsigma}$ is the matrix \mathbf{G} defined in (5.10). It follows, therefore, the score function with respect to ς is

$$S(\varsigma) = \mathbf{G}' S(\xi^2) = \mathbf{G}' \frac{\partial L(\xi^2)}{\partial \xi^2} \quad (5.12)$$

and the Fisher information matrix with respect to ς becomes

$$\mathbf{A}_{\varsigma} = \mathbf{G}' \mathbf{A}_{\xi^2} \mathbf{G} = \mathbf{G}' \left[\text{diag} \left(\frac{v_i}{2\xi_i^4} \right) \right] \mathbf{G}, \quad (i = 1, \dots, 4). \quad (5.13)$$

5.3.3 Estimating VCs in ς^2

Using the score function and the Fisher information matrix defined in (5.12) and (5.13), the vector of VCs, ς , can be estimated using Fisher's scoring algorithm, which is an iterative procedure that can be used to solve maximum likelihood equations. The formula for Fisher's scoring algorithm can be written as

$$\varsigma_{t+1} = \varsigma_t + \mathbf{A}_{\varsigma_t}^{-1} S(\varsigma_t), \quad (5.14)$$

where ς_t and ς_{t+1} are vectors of VCs estimates at the t -th and $(t+1)$ -th iterations, respectively. The initial estimates can be any value. The iterations stop when the differences between the VCs estimates in two consecutive iterations are less than 1×10^{-7} (Patterson and Thompson, 1971).

5.4 Satterthwaite's approximation in deriving the EDF

Assessment on how well the estimation of the variance, i.e. $\hat{\sigma}^2 + 2\hat{\sigma}_a^2$, has been performed is by examining the DF of Residual MS associated with the Between Animals Within Runs stratum. Once the VCs are estimated from the experimental data using either the LC or REML method, a higher DF, or EDF, may be approximated. The higher the EDF the better this variance is estimated, and also the higher the for the F-test of the Treatment effects.

Using the estimated VCs, the EDF are approximated as twice the square of the expected mean divided by the variance (Satterthwaite, 1941). We can show this based on the mean squares, s_i^2 , are assumed to have a $\chi_{v_i}^2$ distribution, from (5.6), then its expectation is given by

$$E(s_i^2) = E\left(\frac{\xi_i^2}{v_i} \chi_{v_i}^2\right) = \frac{\xi_i^2}{v_i} E(\chi_{v_i}^2) = \frac{\xi_i^2}{v_i} v_i = \xi_i^2, \quad (5.15)$$

and its variance is given by

$$\text{Var}(s_i^2) = \text{Var}\left(\frac{\xi_i^2}{v_i} \chi_{v_i}^2\right) = \frac{\xi_i^4}{v_i^2} \text{Var}(\chi_{v_i}^2) = 2 \frac{\xi_i^4}{v_i^2} v_i = 2 \frac{\xi_i^4}{v_i}. \quad (5.16)$$

From (5.15) and (5.16), the EDF are approximated from the twice the square of the expected mean divided by the variance, i.e.

$$\text{EDF} = \frac{2[E(s_i^2)]^2}{\text{Var}(s_i^2)} = \frac{2\xi_i^4}{\frac{2\xi_i^4}{v_i}} = v_i.$$

From Table 5.2, the MS of interest is the Residual MS in the Between Animals Within Runs

stratum, i.e. s_3^2 , thus the EDF associated with this MS are computed as

$$\text{EDF} = \frac{2(s_3^2)^2}{\text{Var}(s_3^2)} = \frac{2(\hat{\sigma}^2 + 2\hat{\sigma}_a^2)^2}{\text{Var}(\hat{\sigma}^2 + 2\hat{\sigma}_a^2)}. \quad (5.17)$$

The asymptotic variance of the estimates of ς is given by the inverse of the Fisher information matrix, which can be expressed as

$$\mathbf{A}_\varsigma^{-1} = \begin{pmatrix} \text{Var}(\sigma^2) & \text{Cov}(\sigma^2, \sigma_a^2) & \text{Cov}(\sigma^2, \sigma_r^2) \\ \text{Cov}(\sigma^2, \sigma_a^2) & \text{Var}(\sigma_a^2) & \text{Cov}(\sigma_a^2, \sigma_r^2) \\ \text{Cov}(\sigma^2, \sigma_r^2) & \text{Cov}(\sigma_a^2, \sigma_r^2) & \text{Var}(\sigma_r^2) \end{pmatrix}. \quad (5.18)$$

The estimated variance of $\sigma^2 + 2\sigma_a^2$ in (5.17), i.e. $\text{Var}(\hat{\sigma}^2 + 2\hat{\sigma}_a^2)$ is given by the sum of the four elements in the top left 2×2 submatrix in (5.18) and can be written as

$$\text{Var}(\hat{\sigma}^2 + 2\hat{\sigma}_a^2) = \text{Var}(\hat{\sigma}^2) + 4 \text{Var}(\hat{\sigma}_a^2) + 4 \text{Cov}(\hat{\sigma}^2, \hat{\sigma}_a^2).$$

From (5.17), the EDF of Residual MS in the Between Animals Within Runs stratum in Section 5.2 can be computed from

$$\text{EDF} = \frac{2(\hat{\sigma}^2 + 2\hat{\sigma}_a^2)^2}{\text{Var}(\hat{\sigma}^2) + 4 \text{Var}(\hat{\sigma}_a^2) + 4 \text{Cov}(\hat{\sigma}^2, \hat{\sigma}_a^2)}.$$

5.4.1 Comparing VCs between LC and REML methods

In this section, a simulation study, is used to compare the effects of the two VC estimation methods - LC and REML - on the estimated EDF of the variance, i.e. $\hat{\sigma}^2 + 2\hat{\sigma}_a^2$, for the design in the example discussed in Section 5.2. The simulation datasets were generated on the basis that the residual MS have a chi-square distribution, with the ratio of the Between Animals VC to measurement error VC, denoted by σ_a^2/σ^2 , set with 17 values ranging from 10^{-4} to 10^4 , and the ratio of Between Runs VC to measurement error, denoted by σ_r^2/σ^2 , set to 0, 0.25, 1, 5, 100. The case when run effects of run are fixed is also considered (effectively when $\sigma_r^2 = \infty$).

Figure 5.1 shows the EDF for the range of values of the ratio σ_r^2/σ^2 and σ_a^2/σ^2 , including a line for the fixed Run effects case. It shows that the EDF can be as low as 2 DF when there are large Run effects, but that the EDF approach 3 when the between-animal variation dominates. Note that, when the run-to-run variation is substantially larger than the between animal variation we see that the EDF based on VCs estimated using the LC method are slightly higher than those using REML. This suggests that the optimal design presented in Section 5.2 may be robust to the method of VCs estimation. We will show EDF approximated from both

LC and REML methods for the remaining of this Chapter to confirm this.

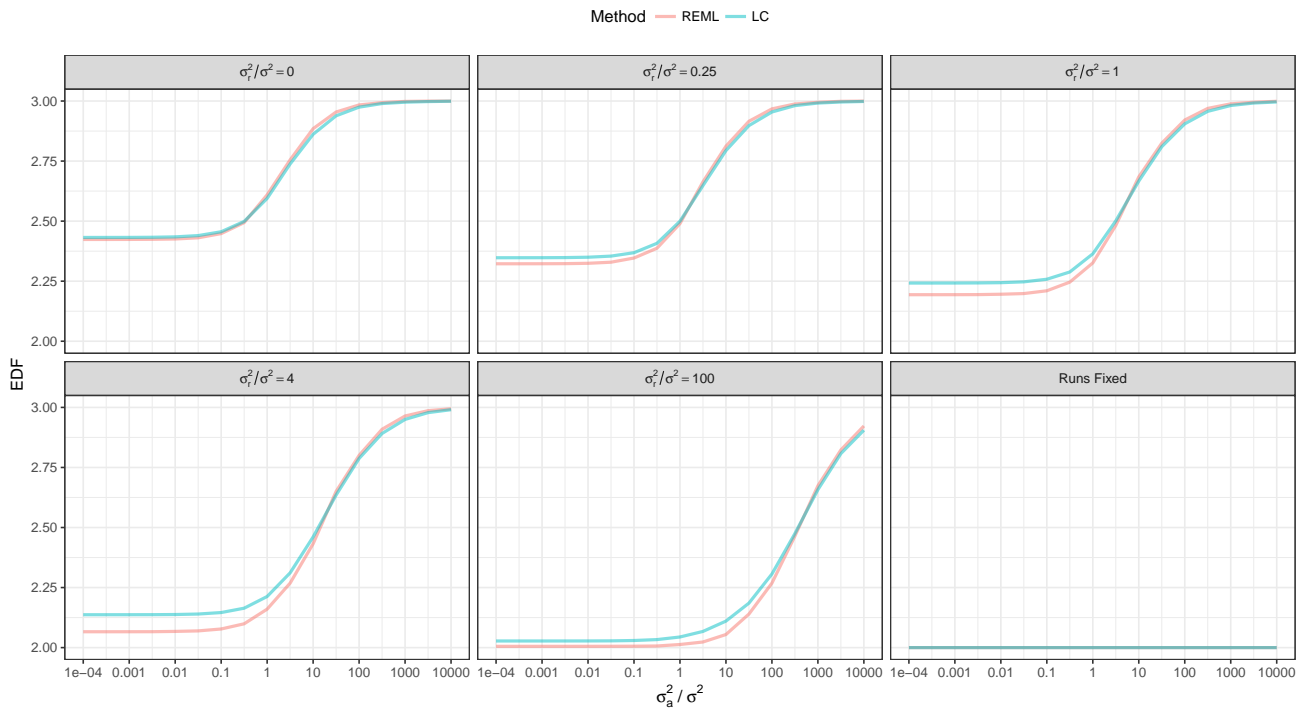


Figure 5.1: EDF for the variance of the treatment effects for two-phase experiment, i.e. for the optimal design of the Phase 2 experiment which takes account of the design of the Phase 1 experiment involves $\nu = 2$ treatments assigned to $n_a = 6$ animals. Each sample is further split into $n_s = 2$ sub-samples labelled by $n_\gamma = 4$ tags and measured in $n_r = 3$ runs, based on REML estimates of the variance components and on a linear combination of the mean squares.

5.5 EDF when Phase 1 experiment is arrange in a CRD

This section compares the EDF obtained from optimal designs of the Phase 2 experiment found when the Phase 1 experiment is arranged in a CRD. The main consideration is on comparing four-plex and eight-plex experiments using an identical Phase 1 experiment. Three different cases are presented, showing that different sets of design parameters can work better depending on whether the four-plex or eight-plex experiment is used.

5.5.1 Example 1: A CRD with 2 treatments and 12 animals

The first example experiment to be considered is the Phase 1 experiment with $\nu = 2$ treatments assigned to $n_a = 12$ animals. Based on the methods presented in Chapter 3, two optimal designs are found for the Phase 2 proteomics experiment: one assuming the four-plex iTRAQTM system is used and the other assuming that the eight-plex system is used, which are presented in Tables 5.3a and 5.3b, respectively.

Table 5.3: Optimal (a) four- and (b) eight-plex designs of Phase 2 proteomics experiment when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to each of $n_a = 12$ animals, with $n_s = 2$ sub-samples taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment. Animal IDs are denoted by upper case letters, while the lower case letters a and b denote the two treatments.

(a) Four-plex system.					(b) Eight-plex system.								
Run	Tag				Run	Tag							
	114	115	116	117		113	114	115	116	117	118	119	121
1	<i>Jb</i>	<i>Aa</i>	<i>Lb</i>	<i>Ca</i>	1	<i>Ia</i>	<i>Ea</i>	<i>Ga</i>	<i>Db</i>	<i>Hb</i>	<i>Ca</i>	<i>Bb</i>	<i>Jb</i>
2	<i>Aa</i>	<i>Jb</i>	<i>Ca</i>	<i>Lb</i>	2	<i>Ea</i>	<i>Ia</i>	<i>Db</i>	<i>Ga</i>	<i>Ca</i>	<i>Hb</i>	<i>Jb</i>	<i>Bb</i>
3	<i>Ia</i>	<i>Fb</i>	<i>Hb</i>	<i>Ka</i>	3	<i>Fb</i>	<i>Fb</i>	<i>Lb</i>	<i>Lb</i>	<i>Ka</i>	<i>Ka</i>	<i>Aa</i>	<i>Aa</i>
4	<i>Fb</i>	<i>Ia</i>	<i>Ka</i>	<i>Hb</i>									
5	<i>Bb</i>	<i>Ea</i>	<i>Db</i>	<i>Ga</i>									
6	<i>Ea</i>	<i>Bb</i>	<i>Ga</i>	<i>Db</i>									

The theoretical ANOVA tables for the four- and eight-plex optimal designs are presented in Tables 5.4 and 5.5, respectively. Based solely on these two theoretical ANOVA tables, the four-plex design is shown to be the better design, because it has higher Residual DF for estimating the Residual MS and therefore for testing treatment effects (7 DF compared to with 6 DF for the eight-plex design), and Treatment effects are fully estimated in the desired stratum, namely Between Animals within Runs. In comparison, the average efficiency factor for treatment effects in the eight-plex design is 0.889 due to the 1 DF of treatment contrast being partially confounded with the contrast of Tag 113, 114, 117, 118 versus Tag 115, 116, 119, 121. In addition, the theoretical ANOVA table of the four-plex experiment shows there are 2 DF associated with the residual variance in the Between Animals Between Runs stratum, potentially enabling the recovery of up to 2 additional DF, i.e. yielding up to 9 EDF. The eight-plex experiment has 1 DF associated with the Between Animals Between Runs stratum, which can be recovered giving EDF as high as 7 DF.

The EDF plots, in Figure 5.2, show that the EDF from the four-plex experiment is always higher than that from eight-plex experiment. The EDF of the four-plex experiment can be as low as 7 DF when there is large run-to-run variation, but the EDF approach 9 when the between-animal variation dominates. As for the eight-plex experiment, the EDF can be as low as 6 DF with large Run effects, but the EDF approach 7 with high between animal variation. This suggests that the optimal design using the four-plex experiment is preferred over the eight-plex experiment in this case. In addition, the EDF appears to be very similar between the REML and LC methods.

Table 5.4: Theoretical ANOVA table for the optimal design of the Phase 2 experiment in Table 5.3a.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	2	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Within Animals	3	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 6\theta_\gamma$	1	
Treatment	1	$\sigma^2 + 2\sigma_a^2 + 12\theta_\tau$		1
Residual	7	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 6\theta_\gamma$	1	
Residual	7	σ^2		

Table 5.5: Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.3b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 8\sigma_r^2$		
Within Animals	1	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 2\sigma_a^2 + 3\theta_\gamma + 1.33\theta_\tau$	1	0.1111
Treatment	1	$\sigma^2 + 2\sigma_a^2 + 10.67\theta_\tau$		0.8889
Residual	6	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	4	$\sigma^2 + 3\theta_\gamma$	1	
Residual	7	σ^2		

5.5.2 Example 2: A CRD with 8 treatments and 16 animals

The second example experiment to be considered is a Phase 1 experiment with $\nu = 8$ treatments each assigned to $n_a = 16$ animals. Based on the methods presented in Chapter 3, two optimal designs are found for the Phase 2 proteomics experiment: one assuming the four-plex iTRAQTM system is used and the other assuming that the eight-plex system is used. These designs are presented in Tables 5.6a and 5.6b, respectively.

The theoretical ANOVA tables for the four- and eight-plex optimal designs in Tables 5.6a and 5.6b are presented in Tables 5.7 and 5.8, respectively. Based solely on these two theoretical ANOVA tables, both designs have the 4 Residual DF for estimating the Residual MS and therefore for testing Treatment effects. Furthermore, in both designs the Treatment effects are estimated, in the desired stratum, namely Between Animals within Runs, and have average efficiency factor $E_\tau = 0.8077$.

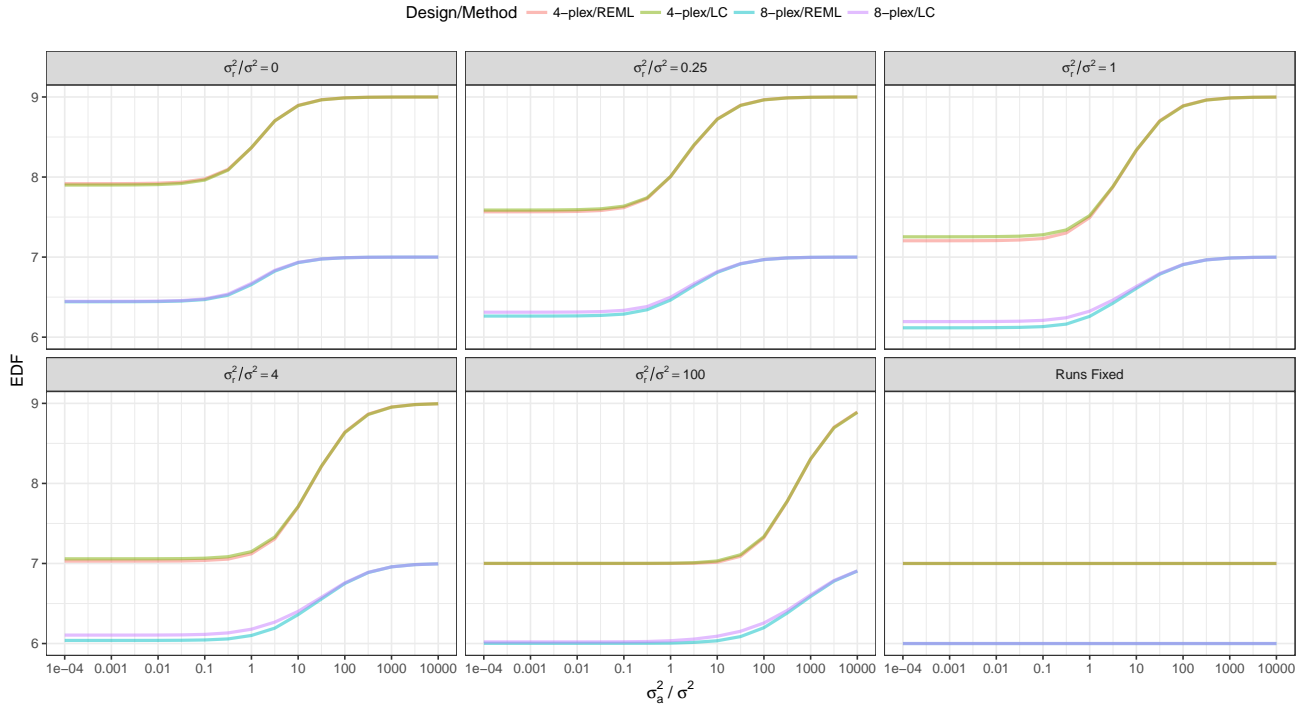


Figure 5.2: EDF plots for optimal designs shown in Tables 5.3a and 5.3b, where EDF is calculated using VCs estimated by both the REML and LC methods.

Table 5.6: Optimal (a) four- and (b) eight-plex designs of Phase 2 proteomics experiment when the Phase 1 experiment consists of $\nu = 8$ treatments each assigned to $n_a = 16$ animals, with $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

(a) Four-plex system.					(b) Eight-plex system.								
Run	Tag				Run	Tag							
	114	115	116	117		113	114	115	116	117	118	119	121
1	<i>Jb</i>	<i>Dd</i>	<i>Ph</i>	<i>Kc</i>	1	<i>Ff</i>	<i>Dd</i>	<i>Kc</i>	<i>Hh</i>	<i>Me</i>	<i>Aa</i>	<i>Og</i>	<i>Bb</i>
2	<i>Dd</i>	<i>Jb</i>	<i>Kc</i>	<i>Ph</i>	2	<i>Dd</i>	<i>Ff</i>	<i>Hh</i>	<i>Kc</i>	<i>Aa</i>	<i>Me</i>	<i>Bb</i>	<i>Og</i>
3	<i>Hh</i>	<i>Ee</i>	<i>Nf</i>	<i>Bb</i>	3	<i>Cc</i>	<i>Ia</i>	<i>Nf</i>	<i>Jb</i>	<i>Gg</i>	<i>Ph</i>	<i>Ld</i>	<i>Ee</i>
4	<i>Ee</i>	<i>Hh</i>	<i>Bb</i>	<i>Nf</i>	4	<i>Ia</i>	<i>Cc</i>	<i>Jb</i>	<i>Nf</i>	<i>Ph</i>	<i>Gg</i>	<i>Ee</i>	<i>Ld</i>
5	<i>Aa</i>	<i>Og</i>	<i>Me</i>	<i>Ld</i>									
6	<i>Og</i>	<i>Aa</i>	<i>Ld</i>	<i>Me</i>									
7	<i>Ff</i>	<i>Cc</i>	<i>Ia</i>	<i>Gg</i>									
8	<i>Cc</i>	<i>Ff</i>	<i>Gg</i>	<i>Ia</i>									

For the four-plex experiment, the treatment average efficiency factor of 0.8077 is due to the 3 DF associated with Treatment effects being confounded with Run effects. Thus, the Between Animals Between Runs EMS, which in previous cases provided a measure of pure error, now includes a contribution from the Treatment fixed effect. The 3 DF associated with this effect are not available for recovery of information, i.e. the existing 4 Residual DF for estimating the variance of the treatment effects cannot be improved upon. In comparison, although the

treatment average efficiency factor, in the eight-plex design, is also 0.8077, 3 of 7 DF associated with Treatment effects are now partially confounded with Tag effects. As a result, the 1 DF associated with the Between Animals MS in the Between Runs stratum remains pure error and can be recovered in the estimation of the VCs, resulting in the EDF to being as high as 5 DF.

Table 5.7: Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.6a.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals				
Treatment	3	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2 + 1.2\theta_\tau$		0.3
Within Animals	4	$\sigma^2 + 4\sigma_r^2$		
Within Run				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 8\theta_\gamma$	1	
Treatment	7	$\sigma^2 + 2\sigma_a^2 + 3.23\theta_\tau$		0.8077
Residual	4	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 8\theta_\gamma$	1	
Residual	10	σ^2		

Table 5.8: Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.6b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 8\sigma_r^2$		
Within Animals	2	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma + 1.2\theta_\tau$	1	0.3
Treatment	7	$\sigma^2 + 2\sigma_a^2 + 3.23\theta_\tau$		0.8077
Residual	4	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	4	$\sigma^2 + 4\theta_\gamma$	1	
Residual	10	σ^2		

The EDF plots, in Figure 5.3, show that the EDF for the design using the eight-plex system is always 4 irrespective of the values of the ratios σ_a^2/σ^2 and σ_r^2/σ^2 . For the four-plex design, the EDF can be as low as 4 when the run-to-run variation is much larger than the between animal variation, and as high as 5 DF when the between-animal variation dominates. The EDF are the same (4 DF) between four- and eight-plex experiments when $\sigma_r^2/\sigma^2 = 100$ and σ_a^2/σ^2 is between 1×10^{-4} to 10, i.e. when the run-to-run variation is 10 to 1×10^6 times of run-to-run variation over animal-to-animal variation. The EDF are also the same (4 DF) between four- and eight-plex experiments when Run effects are assumed to be fixed. In addition, the EDF are again shown to be very similar between the REML and LC methods.

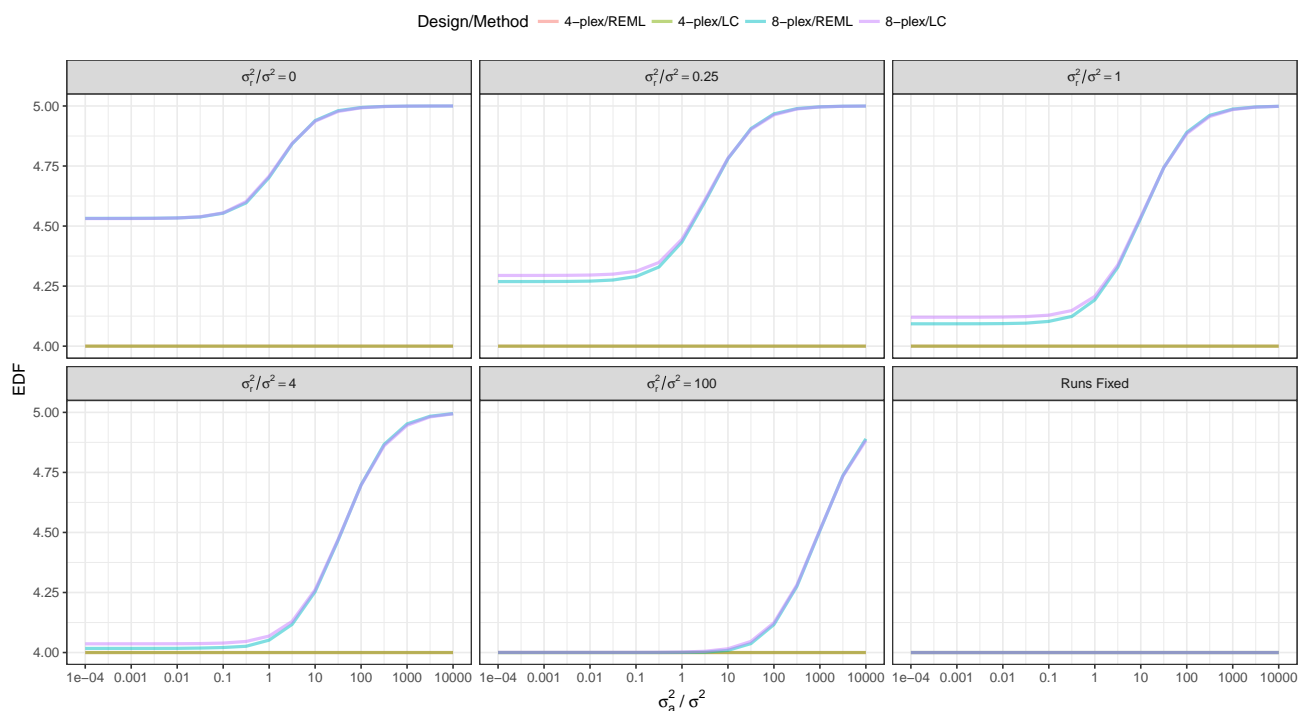


Figure 5.3: EDF plots for optimal designs shown in Tables 5.6a and 5.6b, where EDF is calculated using VCs estimated by both the REML and LC methods.

5.5.3 Example 3: A CRD with 4 treatments and 24 animals

In the third example, the Phase 1 experiment involves $\nu = 4$ treatments assigned to $n_a = 24$ animals. Based on the methods presented in Chapter 3, two optimal designs are found for the Phase 2 proteomics experiment: one assuming the four-plex iTRAQTM system is used and the other assuming that the eight-plex system is used. These designs are presented in Tables 5.9a and 5.9b, respectively.

The theoretical ANOVA tables for the four- and eight-plex optimal designs in Tables 5.9a and 5.9b are presented in Tables 5.10 and 5.11, respectively. Based solely on these two theoretical ANOVA tables, the eight-plex design has higher Residual DF for estimating the Residual MS and for testing treatment effects (15 DF compared 14 DF for the four-plex design). Comparing the treatment average efficiency factor from the two designs: $E_\tau = 0.9623$ for the eight-plex design which is slightly lower than $E_\tau = 1$ for the four-plex design. In addition, the theoretical ANOVA table of the four-plex design has 5 DF associated with the Between Animals Between Runs stratum, which can be recovered giving EDF as high as 19, whereas the eight-plex experiment has 2 DF associated with the Between Animals Between Runs stratum, which can be recovered giving EDF as high as 17.

The EDF plots, in Figure 5.4, show that, for the design using the four-plex experiment, the EDF can be as low as 14 when the run-to-run variation is much larger than the between-animal

Table 5.9: Optimal (a) four- and (b) eight-plex designs of Phase 2 proteomics experiment, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to each of $n_a = 24$ animals, $n_s = 2$ sub-samples are then taken from each animal and analysed in the Phase 2 MudPIT-iTRAQTM experiment. Upper case letters denote animal IDs, while the lower case letters denote the treatments.

(a) Four-plex system.					(b) Eight-plex system.								
Run	Tag				Run	Tag							
	114	115	116	117		113	114	115	116	117	118	119	121
1	<i>Aa</i>	<i>Rb</i>	<i>Cc</i>	<i>Dd</i>	1	<i>Vb</i>	<i>Ma</i>	<i>Qa</i>	<i>Pd</i>	<i>Dd</i>	<i>Wc</i>	<i>Bb</i>	<i>Oc</i>
2 2	<i>Rb</i>	<i>Aa</i>	<i>Dd</i>	<i>Cc</i>	2	<i>Ma</i>	<i>Vb</i>	<i>Pd</i>	<i>Qa</i>	<i>Wc</i>	<i>Dd</i>	<i>Oc</i>	<i>Bb</i>
3 3	<i>Ea</i>	<i>Fb</i>	<i>Gc</i>	<i>Hd</i>	3	<i>Kc</i>	<i>Rb</i>	<i>Td</i>	<i>Cc</i>	<i>Jb</i>	<i>Ea</i>	<i>Aa</i>	<i>Hd</i>
4 4	<i>Fb</i>	<i>Ea</i>	<i>Hd</i>	<i>Gc</i>	4	<i>Rb</i>	<i>Kc</i>	<i>Cc</i>	<i>Td</i>	<i>Ea</i>	<i>Jb</i>	<i>Hd</i>	<i>Aa</i>
5 5	<i>Ia</i>	<i>Jb</i>	<i>Kc</i>	<i>Ld</i>	5	<i>Sc</i>	<i>Ld</i>	<i>Fb</i>	<i>Nb</i>	<i>Xd</i>	<i>Ua</i>	<i>Gc</i>	<i>Ia</i>
6 6	<i>Jb</i>	<i>Ia</i>	<i>Ld</i>	<i>Kc</i>	6	<i>Ld</i>	<i>Sc</i>	<i>Nb</i>	<i>Fb</i>	<i>Ua</i>	<i>Xd</i>	<i>Ia</i>	<i>Gc</i>
7 7	<i>Oc</i>	<i>Pd</i>	<i>Ma</i>	<i>Nb</i>									
8 8	<i>Pd</i>	<i>Oc</i>	<i>Nb</i>	<i>Ma</i>									
9 9	<i>Sc</i>	<i>Td</i>	<i>Qa</i>	<i>Bb</i>									
10 10	<i>Td</i>	<i>Sc</i>	<i>Bb</i>	<i>Qa</i>									
11 11	<i>Wc</i>	<i>Xd</i>	<i>Ua</i>	<i>Vb</i>									
12 12	<i>Xd</i>	<i>Wc</i>	<i>Vb</i>	<i>Ua</i>									

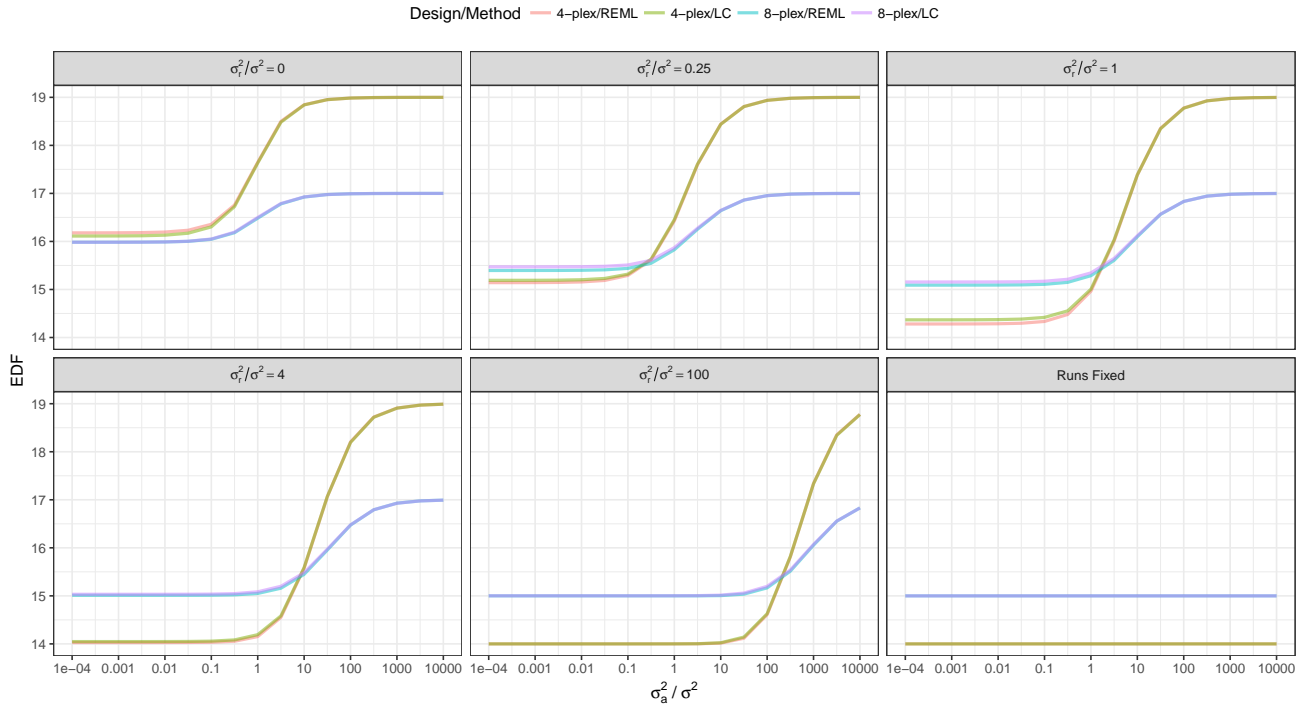


Figure 5.4: EDF plots for optimal designs shown in Tables 5.9a and 5.9b, where EDF is calculated using VCs estimated by both the REML and LC methods.

variation, and as high as 19 when between-animal variation dominates. For the the design using eight-plex experiment, the EDF can be as low as 15 DF when the run-to-run variation is much larger than the between-animal variation, and as high as 17 DF when the between-animal

Table 5.10: Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.9a.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	5	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Within Animals	6	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 12\theta_\gamma$	1	
Treatment	3	$\sigma^2 + 2\sigma_a^2 + 12\theta_\tau$		1
Residual	14	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 12\theta_\gamma$	1	
Residual	16	σ^2		

Table 5.11: Theoretical ANOVA table for the optimal design of Phase 2 experiment in Table 5.9b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	2	$\sigma^2 + 2\sigma_a^2 + 8\sigma_r^2$		
Within Animals	3	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 2\sigma_a^2 + 6\theta_\gamma + 0.67\theta_\tau$	1	0.0556
Treatment	3	$\sigma^2 + 2\sigma_a^2 + 11.55\theta_\tau$		0.9623
Residual	15	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	4	$\sigma^2 + 6\theta_\gamma$	1	
Residual	17	σ^2		

variation dominates. Furthermore, recall from the theoretical ANOVA tables (in Tables 5.10 and 5.11) when the eight-plex design has 15 Residual DF compared with 14 Residual DF of the four-plex design. From the EDF plots, we can see the EDF for the four-plex design exceed those of the eight-plex design as σ_a^2/σ^2 increases. For example, when $\sigma_r^2/\sigma^2 = 0.25$, the EDF become higher for the four-plex experiment when $\sigma_a^2/\sigma^2 = 1 \times 10^{-0.5}$, that is about 0.79 times of the run-to-run variation over animal-to-animal variation. Table 5.12 lists the σ_r^2/σ^2 and σ_a^2/σ^2 combinations when the EDF become higher for the four-plex experiment than the eight-plex experiment, and the magnitudes of the run-to-run variation over animal-to-animal variation based on Figure 5.4.

Table 5.12: Magnitudes of the run-to-run variation over animal-to-animal variation when the EDF become higher for four-plex experiment than eight-plex experiments based on Figure 5.4.

σ_r^2/σ^2	0.25	1	4	100
σ_a^2/σ^2	$10^{-0.5}$	$10^{0.25}$	10	$10^{2.25}$
σ_r^2/σ_a^2	0.79	0.56	0.4	0.32

5.6 Comparing the EDF when Phase 1 experiment is arranged in a RCBD

This section compares the EDF of optimal designs between Phase 2 four-plex and eight-plex experiments when the Phase 1 experiment is arranged in a RCBD. There are four types of designs that can be compared, given the same Phase 1 experiment is arranged in a RCBD:

1. Phase 1 Block effects are intentionally confounded with Run effects using the four-plex system,
2. Phase 1 Block effects are intentionally confounded with Tag effects using the four-plex system,
3. Phase 1 Block effects are intentionally confounded with Run effects using the eight-plex system and
4. Phase 1 Block effects are intentionally confounded with Tag effects using the eight-plex system,

The same method for estimating the VCs and approximating the EDF can also be applied to this example, because we can still construct an ANOVA table with Residual MS which assumed to have a chi-square distribution. In addition, the EDF are still approximated based on the Residual MS of the Between Experimental units (Plants) Within Blocks (Trays) Within Runs stratum, i.e. $\sigma^2 + \sigma_p^2$.

This section first compares designs from two different confounding schemes for each four- and eight-plex experiment, followed by an overall comparison between the four- and eight-plex systems. The Phase 2 experiment uses plants as an example when the Phase 1 experiment involves $\nu = 4$ treatments assigned to $n_p = 16$ plants in $n_b = 4$ trays. Then $n_s = 2$ sub-samples are obtained from each plant, i.e. a total of 32 sub-samples, for the Phase 2 proteomics experiment.

The simulation study was done on the basis that MS has at chi-square distribution, with the ratio of Between Plants VCs to measurement error, denoted by σ_p^2/σ^2 , set with 17 values ranging

from 10^{-4} to 10^4 , the ratio of Between Trays VCs to measurement error and Between Runs VCs to measurement error, denoted by σ_b^2/σ^2 and σ_r^2/σ^2 , respectively, set to 0, 0.25, 1, 5, 100, as well as having effects of tray and run fixed are also considered (effectively when $\sigma_b^2 = \infty$ and $\sigma_r^2 = \infty$).

5.6.1 Four-plex system

Given the Phase 1 experiment with $\nu = 4$ treatments assigned to $n_a = 16$ plants in $n_b = 4$ trays, based on the methods presented in Chapter 4, two optimal designs are found for the Phase 2 four-plex proteomics experiment: one assume that Tray effects are intentionally confounded with Run effects, and the other that Tray effects are intentionally confounded with Tag effects, which are presented in Tables 5.9a and 5.9b, respectively.

Table 5.13: Optimal design of Phase 2 proteomics experiment showing allocation of sub-samples from trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 16$ plants in $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment using $n_\gamma = 4$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

(a) Tray effects are intentionally confounded with Run effects.					(b) Tray effects are intentionally confounded with Tag effects.				
Run	Tag				Run	Tag			
	114	115	116	117		114	115	116	117
1	1Cc	1Dd	1Bb	1Aa	1	1Bb	1Dd	3Kc	3Ia
2	1Dd	1Cc	1Aa	1Bb	2	1Dd	1Bb	3Ia	3Kc
3	2Hd	2Fb	2Ea	2Gc	3	1Cc	1Aa	3Ld	3Jb
4	2Fb	2Hd	2Gc	2Ea	4	1Aa	1Cc	3Jb	3Ld
5	3Ia	3Jb	3Ld	3Kc	5	2Gc	2Fb	4Ma	4Pd
6	3Jb	3Ia	3Kc	3Ld	6	2Fb	2Gc	4Pd	4Ma
7	4Ma	4Oc	4Pd	4Nb	7	2Ea	2Hd	4Nb	4Oc
8	4Oc	4Ma	4Nb	4Pd	8	2Hd	2Ea	4Oc	4Nb

The theoretical ANOVA tables for the optimal designs of the Phase 2 experiment using the four-plex system are presented in Tables 5.14 and 5.15, respectively. Based solely on these two theoretical ANOVA tables, the design when Tray effects are intentionally confounded with Run effects is shown to be the better design, because it has higher Residual DF for estimating the Residual MS and therefore for testing Treatment effects (8 DF compared to 7 DF for the four-plex design) in the Between Plants Within Trays stratum. Both designs can estimate Treatment effects with full efficiency in the Between Plants Within Trays stratum, because the treatment average efficiency factors for both designs are 100%.

From the theoretical ANOVA of both designs, there is extra information on the Between Plants VC σ_p^2 in the Between Trays Between Runs MS, however, we may not be able to recover

this information, because we cannot equate Residual MS in Between Trays Between Runs stratum based on the EMS to estimate σ_p^2 . For the design when the Tray effects are intentionally confounded Run effects (see Table 5.15), there are 2 DF associated with the Between Plants Within Blocks Between Runs stratum which can be recovered giving EDF as high as 9.

Table 5.14: Theoretical ANOVA table from the Phase 2 experiment in Table 5.13a.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	3	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\sigma_r^2$		
Within Plants Within Trays	4	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Plants Within Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 8\theta_\gamma$	1	
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$		1
Residual	8	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 8\theta_\gamma$	1	
Residual	10	σ^2		

Table 5.15: Theoretical ANOVA table from the Phase 2 experiment in Table 5.13b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\sigma_r^2$		
Between Plants Within Trays	2	$\sigma^2 + 2\sigma_p^2 + 4\sigma_r^2$		
Within Plants Within Trays	4	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 8\theta_\gamma$	1	
Residual	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2$		
Between Plants Within Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$		1
Residual	7	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	2	$\sigma^2 + 8\theta_\gamma$	1	
Residual	10	σ^2		

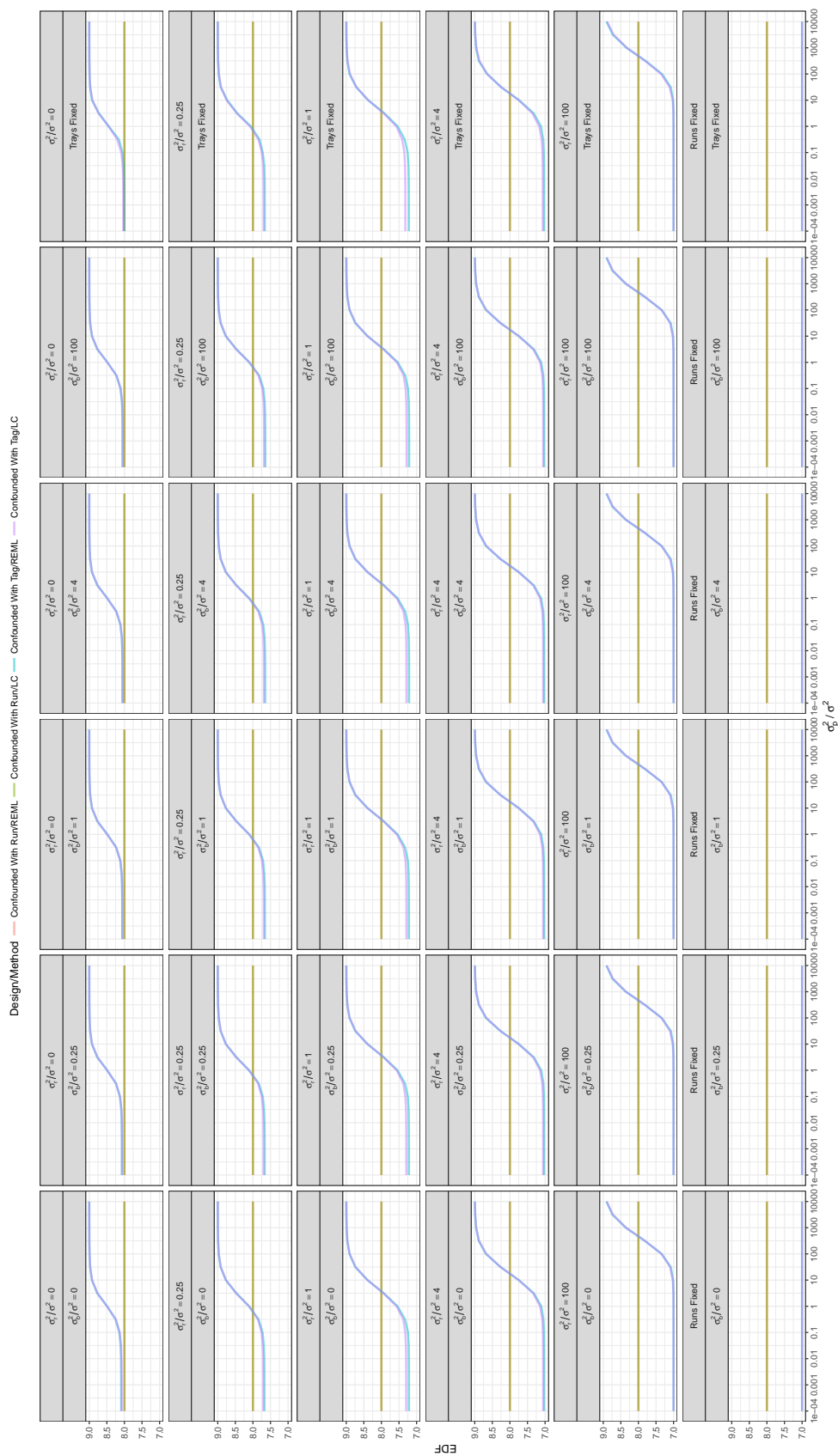


Figure 5.5: EDF plots for optimal designs shown in Tables 5.13a and 5.13b, where EDF is calculated using VCs estimated by both the REML and LC methods.

The EDF plots in Figure 5.5 are presented as a 6-by-6 panels having different combination of ranges of ratios for Between Plants VCs, Between Trays VCs and Between Runs VCs to measurement error, denoted by σ_p^2/σ^2 , σ_b^2/σ^2 and σ_r^2/σ^2 , respectively. We can first observe that different values of the ratio σ_b^2/σ^2 do not change the EDF. For the design when Tray effects are intentionally confounded with Run effects, the EDF are always 8 DF. For the design when Tray effects are intentionally confounded with Tag effects, the EDF can be as low as 7 DF when run-to-run variation is much larger than the between-plants variation, but the EDF approach 9 DF when the between-plants variation dominates. From the EDF plot, we can see that the EDF of the design when Tray effects are intentionally confounded with Tag effects become better than the design when Tray effects are intentionally confounded with Run effects as σ_p^2/σ^2 increases. For example, when $\sigma_r^2/\sigma^2 = 0.25$, the EDF become higher when $\sigma_p^2/\sigma^2 = 1$, that is about 0.25 times of the run-to-run variation over plant-to-plant variation. Table 5.12 lists the σ_r^2/σ^2 and σ_p^2/σ^2 combinations when the EDF become higher for the design when Tray effects are intentionally confounded with Tag effects than the design when Tray effects are intentionally confounded with Run effects, and the magnitudes of the run-to-run variation over plant-to-plant variation are based on Figure 5.4. In addition, the EDF are again shown to be very similar between the REML and LC methods.

Table 5.16: Magnitudes of the run-to-run variation over plant-to-plant variation when the EDF become higher for design when Tray effects are intentionally confounded with Tag effects than the design when Tray effects are intentionally confounded with Run effect based on Figure 5.5.

σ_r^2/σ^2	0.25	1	4	100
σ_p^2/σ^2	1	$10^{0.5}$	$10^{1.25}$	$10^{2.75}$
σ_r^2/σ_p^2	0.25	0.32	0.22	0.18

5.6.2 Eight-plex system

Using the same Phase 1 experiment with $\nu = 4$ treatments assigned to $n_a = 16$ plants in $n_b = 4$ trays, there are two optimal design that are found in the Phase 2 proteomics experiment using the eight-plex system. The first design assume the Tray effects are intentionally confounded with Run effects (see Tables 5.9a). The second design assume Tray effects are intentionally confounded with Tag effects (see Tables 5.9a).

The theoretical ANOVA tables for the optimal designs of the Phase 2 experiment using the eight-plex system are presented in Tables 5.14 and 5.15, respectively. Based solely on these two theoretical ANOVA tables, the design when Tray effects are intentionally confounded with Tag effects is shown to be the better design, because it has higher Residual DF for estimating the

Table 5.17: Optimal design of Phase 2 proteomics experiment showing allocation of sub-samples from trays, plants and treatments to runs and tags, when the Phase 1 experiment consists of $\nu = 2$ treatments assigned to $n_a = 16$ plants in $n_b = 4$ trays, $n_s = 2$ sub-samples are then taken from each plant and analysed in the Phase 2 MudPIT-iTRAQTM experiment using $n_\gamma = 8$ tags. Numbers denote trays, upper case letters denote plant IDs, while the lower case letters denote the treatments.

(a) Tray effects are intentionally confounded with Run effects.

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Bb	1Dd	1Aa	1Cc	2Ea	2Fb	2Gc	2Hd
2	1Dd	1Bb	1Cc	1Aa	2Fb	2Ea	2Hd	2Gc
3	3Ia	3Kc	3Ld	3Jb	4Oc	4Pd	4Ma	4Nb
4	3Kc	3Ia	3Jb	3Ld	4Pd	4Oc	4Nb	4Ma

(b) Tray effects are intentionally confounded with Run effects.

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Cc	2Gc	2Fb	3Ld	3Ia	4Pd	4Nb
2	1Cc	1Aa	2Fb	2Gc	3Ia	3Ld	4Nb	4Pd
3	1Bb	1Dd	2Hd	2Ea	3Kc	3Jb	4Ma	4Oc
4	1Dd	1Bb	2Ea	2Hd	3Jb	3Kc	4Oc	4Ma

Residual MS and therefore for testing Treatment effects (8 DF compared to 7 DF for the four-plex design) in the Between Plants Within Trays stratum. Both design can estimate Treatment effects with full efficiency in the Between Plants Within Trays stratum, because the treatment average efficiency factors E_τ for both designs are 100%.

For the design when the Tray effects are intentionally confounded with Run effects (see Table 5.18), there is no extra information on the Between Plants VC σ_p^2 that can be recovered, because these σ_p^2 are all in Between Trays Between Runs MS and we cannot equate residual MS in the Between Trays Between Runs strata based on the EMS to estimate σ_p^2 for this design.

For the design when the Tray effects are intentionally confounded with Tag effects (see Table 5.19), there are 2 DF associated with the Between Plants Within Blocks Between Runs stratum which can be recovered giving EDF as high as 10 DF. In addition, since the 3 DF associated with Tray effects are confounded with Tag effects, Tray effects can only be considered to be fixed.

The EDF plots, in Figure 5.2, show that the EDF for the design when the Tray effects are intentionally confounded with Run effects is always 7 DF under different ranges of values of the ratios σ_p^2/σ^2 , σ_b^2/σ^2 and σ_r^2/σ^2 .

For the design when Tray effects are intentionally confounded with Tag effects, the change of

Table 5.18: Theoretical ANOVA table of the Phase 2 experiment in Table 5.17a.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Trays	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 8\sigma_r^2$		
Within Plants Within Trays	2	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Trays				
Tag	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\theta_\gamma$	1	
Residual	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2$		
Between Plants Within Trays				
Tag	2	$\sigma^2 + 2\sigma_p^2 + 4\theta_\gamma$	1	
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$		1
Residual	7	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	4	$\sigma^2 + 4\theta_\gamma$	1	
Residual	10	σ^2		

Table 5.19: Theoretical ANOVA table of the Phase 2 experiment in Table 5.17b.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Plants Within Trays	1	$\sigma^2 + 2\sigma_p^2 + 8\sigma_r^2$		
Within Plants Within Trays	2	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Trays				
Tag	3	$\sigma^2 + 2\sigma_p^2 + 8\sigma_b^2 + 4\theta_\gamma$	1	
Between Plants Within Trays				
Treatment	3	$\sigma^2 + 2\sigma_p^2 + 8\theta_\tau$		1
Residual	8	$\sigma^2 + 2\sigma_p^2$		
Within Plants Within Trays				
Tag	4	$\sigma^2 + 4\theta_\gamma$	1	
Residual	10	σ^2		

EDF can only be observed when the Tray effects are assumed to be fixed. Further, the EDF can be as low as 8 DF when run-to-run variation is much larger than the between-plants variation, but the EDF approach 9 DF when the between plants variation dominates. Therefore, the design when Tray effects are intentionally confounded with Tag effects is always better than the design when the Tray effects are intentionally confounded with Run effects using the eight-plex system. The EDF are again shown to be very similar between the REML and LC methods.

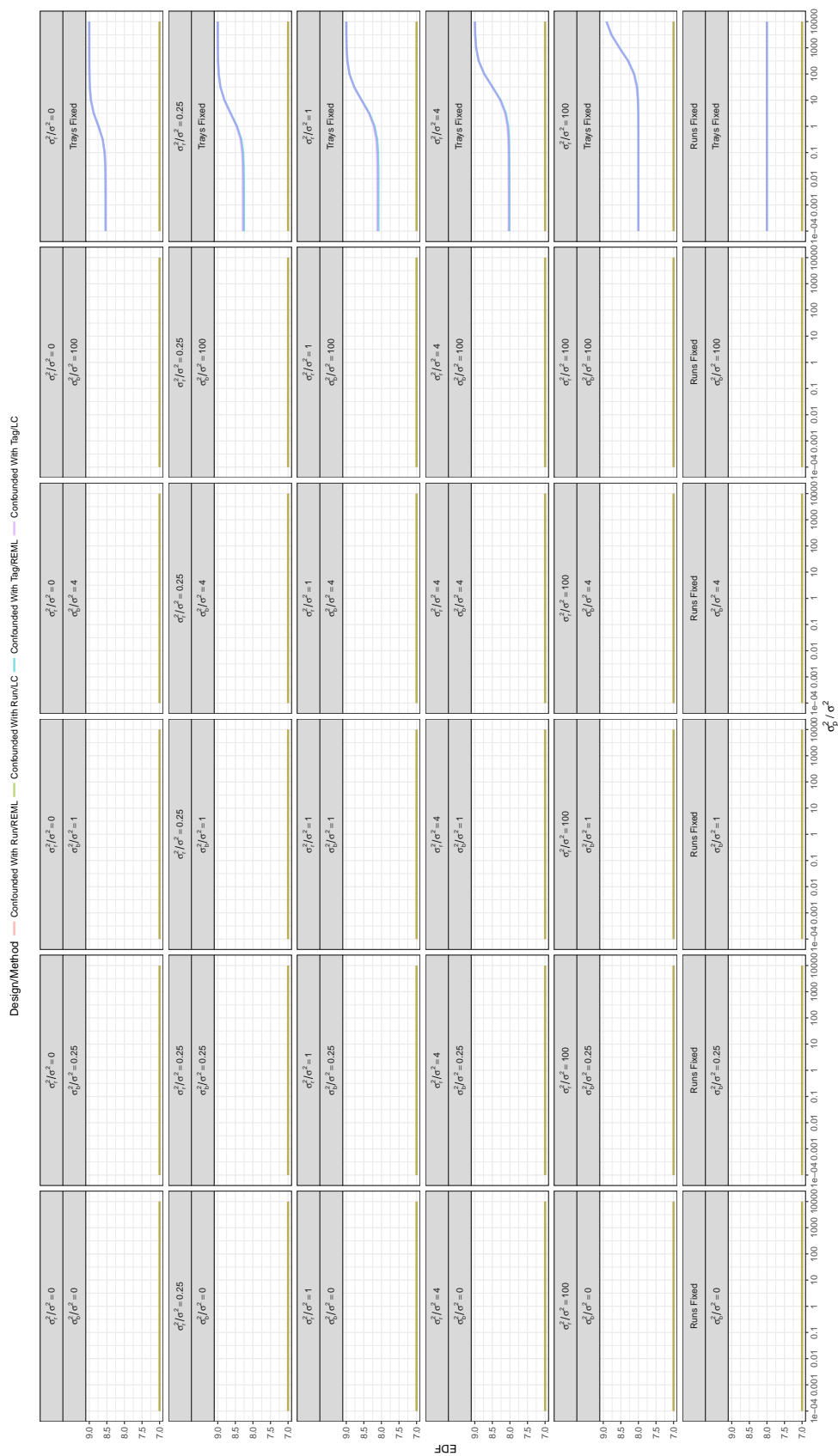


Figure 5.6: EDF plots for optimal designs shown in Tables 5.17a and 5.17b, where EDF is calculated using VCs estimated by both the REML and LC methods.

5.6.3 Four-plex versus Eight-plex system

From Figures 5.5 and 5.6, different ranges of values of the Between Tray VCs to measurement error ratio, denoted by σ_b^2/σ^2 , did not change the EDF, because no extra information on the Between Plants VCs can be recovered from the Residual MS of the Between Tray Within Runs stratum and the Between Tray Within Runs stratum. Thus, Figure 5.7 only presents the EDF when Tray effects are assumed to be fixed, i.e. $\sigma_b^2 = \infty$, to make an overall comparison between the four- and eight-plex systems.

Figure 5.7 shows the design when Tray effects are intentionally confounded with Tag effects with the eight-plex system being superior over the other three Phase 2 design options. However, there are three occasions when the EDF are the same between designs when Tray effects are intentionally confounded with Tag effects using the eight-plex system, and the design when Tray effects are intentionally confounded with Run effects using the four-plex system. The first occasion is when $\sigma_r^2/\sigma^2 = 4$ and σ_p^2/σ^2 is less than 0.1, thus, run-to-run variation is 40 times that of plant-to-plant variation. The second occasion is when $\sigma_r^2/\sigma^2 = 100$ and σ_p^2/σ^2 is less than 10, thus, run-to-run variation is 10 times that of plant-to-plant variation. The last occasion is when the Run effects are assumed to be fixed. Hence, the four-plex system will have the same precision as the eight-plex system when the run-to-run variation is 10 times higher than the plant-to-plant variation.

5.7 Summary

The Chapter described methods in estimating the VCs and approximating the EDF of Phase 2 experiments. EDF indicate how well we estimate the variances of Treatment effects, i.e. the Residual MS of the Between Experimental units stratum. Thus, the higher the EDF the better the estimates of the variance of Treatment effects and the valid F-test of the Treatment effects. We use EDF as another property with which to compare different optimal designs of the Phase 2 experiment found in Chapters 3 and 4.

From all the EDF plots, we have shown that the REML method described here did not improve the approximation of the EDF from the optimal designs found in Chapters 3 and 4. This is due to these optimal designs having the property when the Phase 1 experimental units to the Phase 2 Blocks are always balanced, which ensures that we always have a valid F-test for testing for Treatment effects. Thus, these optimal designs are robust to the VC estimation procedure.

Three different cases were described when each case consists of the same Phase 1 design

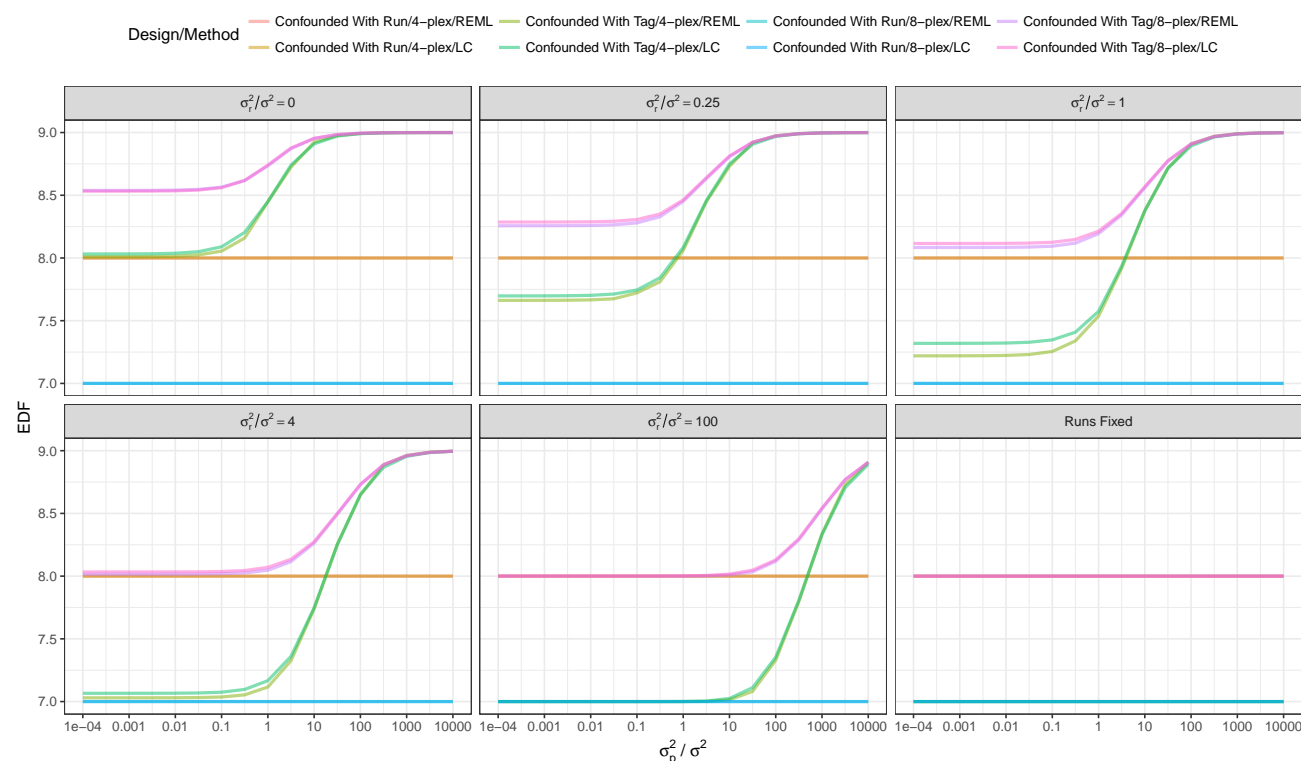


Figure 5.7: EDF for optimal designs shown in Tables 5.13a, 5.13b, 5.17a and 5.17b, with Tray effects are assumed to be fixed, where EDF is calculated using VCs estimated by both the REML and LC methods.

arranged in a CRD, comparing between using four-and eight-plex systems. The first case showed the four-plex system always generates higher EDF than the eight-plex system under different ranges of the VCs ratios. The second case showed when the EDF are always the same under different ranges of VCs ratios, because Treatment effects are completely confounded with base 2 Run effects, thus, the Run effects have to be assumed to be fixed. The last case showed an example when the four-plex system can have higher EDF than the eight-plex system, with run-to-run variation being higher than animal-to-animal variation, but the eight-plex system become better, with higher EDF, than the four-plex system when the animal-to-animal variation dominates.

The last part of the comparison was on the four different types of Phase 2 design with the same Phase 1 design arranged in a RCBD. These four different types of Phase 2 design comprised four-and eight-plex systems with two different confounding schemes when the Phase 1 Block (Tray) effects are intentionally confounded with the Phase 2 Tag and Run effects. We first showed that different ratios of Between Trays VCs to measurement error have no effects on the EDF. Further, the design when Phase 1 Block (Tray) effects are intentionally confounded with the Phase 2 Tag effects using the eight-plex system is preferable, as it generates the highest

EDF among the four designs under all different combinations of the VCs.

Chapter 6

Discussion

6.1 Summary

The primary purpose of this study was to develop a method for the computer generation of optimal designs of two-phase multiplex proteomics experiments. This method for generating optimal designs uses a combination of theory to define objective functions and computing, to improve the simulated annealing (SA) algorithm. Since designs are computer generated, there is no restriction on the design parameters (of the Phase 1 experiment), and end-users do not need to be expert in designing experiments to use this tool to generate their design.

The first part of this thesis applied the method of information decomposition to the design of any single- and two-phase experiment, and automated the construction of theoretical ANOVA tables. For single-phase experiments, the decomposition method was straightforward, as once the strata were defined based on the block structure, the treatment structure was then decomposed within each stratum. In two-phase experiments, however, decomposition began with the strata corresponding to the block structure in the Phase 2 experiment, followed by decomposition of the treatment structure into the strata corresponding to the Phase 1 experiment block structure. The procedure for the Phase 1 block-information decomposition was undertaken by regarding the Phase 1 block factors just as we would treatment factors.

The method for applying information decomposition to designs of any single- and two-phase experiments is implemented in a newly developed R package called **infoDecompuTE** which is available on the Comprehensive R Archive Network (CRAN). This R Package allows the user to automate the construction of theoretical ANOVA tables to enable fast assessment of the attributes of designs. These attributes are the degrees of freedom (DF), expected mean squares (EMS), along with the variance components, fixed effects components, and the treatment average efficiency factor for every source of variation.

For researchers who have no R experience, a Shiny application for the **infoDecompuTE** package is hosted at: https://kcha193.shinyapps.io/infoDecompuTE_Shiny/. There are three type of outputs that can be generated from this Shiny application: 1) output from the R console as a text file, 2) latex code as a text file, and 3) latex compiled portable document format file.

The second part of this thesis described a computational approach for finding optimal designs for Phase 2 proteomics experiments using MudPIT-iTRAQTM technologies. Chapter 3 presented the Phase 1 experiment arranged in a completely randomised design (CRD). The objective function was constructed to minimise confounding between Phase 1 Experiment units and Treatment effects with Phase 2 Run and Tag effects. The information matrix was constructed with an orthogonal projection matrix which projects \mathbf{y} onto the Within Runs and Tags vector subspace, by assuming that Tag effects are random.

A three-criterion objective function was derived for generating the optimal design with three properties:

1. information of the Phase 1 Experimental Units is maximised in the Within Runs stratum, based on A-optimal criteria,
2. treatment information is maximised in the Between Experimental Units Within Runs stratum, based on A-optimal criteria, and
3. DF of the Treatment effects must still be intact in the Between Experimental Units Within Runs stratum.

The modified nested SA algorithm presented consisted of two further improvements. The first improvement was applying the swapping method to only two of the experimental units of the Phase 1 experiment instead of the observational units. The second improvement was the three-stage swapping procedure, which divides a single large search space into three smaller search spaces, swapping the experiment units: 1) within the same runs, 2) within the same tags, and 3) not within the same runs and tags. These improvements were aimed at speeding up the process by optimising the objective function and then obtaining the optimal design.

Chapter 4 extended the concept to finding the optimal design when the Phase 1 experiment is arranged in blocks, more specifically, a randomised complete block design (RCBD), or a balanced incomplete block design (BIBD). Having this additional Block factor from the Phase 1 experiment required us to adjust the objective function to have another criterion in maximising the Residual DF in the Between Plots Within Blocks Within Runs stratum. In addition, instead

of having a single equation combining these four criteria with some weights, we optimise this new four-criterion objective function with three incremental steps:

1. The first step is to locate designs when the Phase 1 Plots average efficiency factor in the Within Blocks Within Runs and Tags vector subspace equals 1, and the DF associated with Treatment effects in the Between Plots Within Blocks Within Runs stratum are intact.
2. Then from among the designs located in the first step, the second step uses the modified nested SA algorithm to find optimal designs when the Residual DF in the Between Plots Within Blocks Within Runs stratum are maximised.
3. From among the designs found in the second step, the third step is to find the optimal design when the treatment average efficiency factor in the Between Plots Within Blocks Within Runs and Tags vector subspace is maximised.

Furthermore, two different types of confounding schemes were investigated, when Phase 1 Block effects are intentionally confounded with Tag effects, and when Phase 1 Block effects were intentionally confounded with Run effects. In general, designs when Phase 1 Block effects were intentionally confounded with Tag effects were shown to have higher Residual DF in the Between Plots Within Blocks Within Runs stratum, because some DF associated with Tag effects were then estimated in the Between Block stratum.

From the optimal designs generated, it was found that, if the Phase 1 experiment was arranged in a CRD with fewer than 16 animals (experimental units), it was better to use the four-plex system instead of the eight-plex system, due to the two extra DF available in the Between Animals Within Runs stratum. However, when more Phase 1 animals (experimental units) were used, the degrees of confounding between Animal effects and Run effects increased in the Phase 2 experiment; thus, it became preferable to use the eight-plex system over the four-plex system. In general, if the Phase 1 experiment is arranged in Blocks, it is recommended that the four-plex system should still be used when there are fewer than 16 animals (experimental units). However, no clear cut-off number of experimental units was identified, at which the eight-plex system become better than the four-plex system. This is because having the additional Block component could generate designs with higher Residual DF when Blocks effects that confounded with Tag effects.

The main purpose of Chapters 3 and 4 was to describe the development of an automated process for finding the optimal design for a wide range of two-phase multiplexing experiments. Even though the main examples were comprise of four- and eight-plex experiments, the methods presented were more general and could be applied to all two-phase designs. This allows researchers

using these technologies to design their experiments without requiring expert knowledge in experimental design. In addition, having this tool available also allows the consulting statisticians to present a quick solution to their client. (A set of resulting optimal designs were presented in Appendices C, E and G, and their properties were presented as tables in Appendices D, F and H.)

The last part of the thesis showed how to estimate the variance components (VCs) using a restricted maximum likelihood (REML) when the Phase 2 Run effects are assumed to be random. We then showed how to approximate the effective degrees of freedom (EDF), which indicated how well we estimate the variances of Treatment effects, i.e. the residual MS of the stratum associated with the experimental unit. A design with higher EDF provided a better estimated of the variance of Treatment effects. However, the REML method described here did not improve the approximation of the EDF from the optimal designs found in Chapters 3 and 4. This was due to these optimal designs having the characteristic of balanced arrangement for Phase 1 experimental units to the Phase 2 Blocks, which ensured that we always had a valid F-test for testing Treatment effects. Thus, these optimal designs were robust to the VCs estimation procedure.

6.2 Future lines of research

6.2.1 Shiny application for generating optimal designs of Phase 2 experiments

Scientists are very adaptive at using these technologies, and they also have a good intuitive sense of needing to design their experiments to protect against unwanted systematic sources of variation. The introduction of labelling technologies in multiplexing for the “omics” experiments is evidence of this.

In Appendices C, E and G, we provided a set of design that were generated from the work in Chapters 3 and 4. Researchers can used these to match their design paramter and select a design for their two-phase experiment.

Some additional R functions for the optimisation algorithm have been written, which will be published as a publicly available package on CRAN. Furthermore, this R package will also be turned into a Shiny application, so that it is are easily accessible to end-users from a wide range of scientific disciplines. Thus, even scientists who are unfamiliar with R will feel comfortable using it and our design methods will become publicly available to all researchers with a user-friendly

interface.

6.2.2 Effective degrees of freedom versus average treatment efficiency factor

Chapter 5 included an example of a Phase 1 experiment involving $\nu = 8$ treatments assigned to $n_a = 16$ animals. We compared the theoretical ANOVA from the designs of the Phase 2 experiment using four-plex and eight-plex in Tables 6.1 and 6.2.

In Table 6.1, when the Phase 2 experiment used the four-plex system, there were 3 DF associated with the Treatment effects estimated in the Between Runs stratum, with a treatment efficiency factors of 0.3. Thus, the Run effects were assumed to be fixed, because we could not recover the extra information on Between Animals VC, σ_a^2 , from the MS in the Between Animals Between Run stratum for estimating the variance of the Treatment effects. Hence, the EDF of the Between Animal Within Run stratum in this case were always 4 DF. As in Table 6.1 when the Phase 2 experiment used the eight-plex system, confounding occurred between Treatment and Tag effects, with Tag effects containing 0.3 of the treatment information. Since the Run effects were assumed as random, we could recover the extra information from the Between Animals Between Runs stratum for estimating the variance of Treatment effects, thus the EDF could be as high as 5 DF.

Table 6.1: Theoretical ANOVA table from the Phase 1 experiment arranged in CRD with $\nu = 8$ and $r_b = 2$, and from the Phase 2 experiment using the four-plex system.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals				
Treatment	3	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2 + 1.2\theta_\tau$		0.3
Within Animals	4	$\sigma^2 + 4\sigma_r^2$		
Within Run				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 8\theta_\gamma$	1	
Treatment	7	$\sigma^2 + 2\sigma_a^2 + 3.23\theta_\tau$		0.8077
Residual	4	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 8\theta_\gamma$	1	
Residual	10	σ^2		

Additional work can be done in comparing between recovering the treatment information across runs, and recovering the extra DF in EDF to get a better estimate of the variance. To achieve this, it would mean performing more extensive simulation studies to understand which

Table 6.2: Theoretical ANOVA table from the Phase 1 experiment arranged in CRD with $\nu = 8$ and $r_b = 2$ and the Phase 2 experiment using the eight-plex system.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 8\sigma_r^2$		
Within Animals	2	$\sigma^2 + 8\sigma_r^2$		
Within Runs				
Between Animals				
Tag	3	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma + 1.2\theta_\tau$	1	0.3
Treatment	7	$\sigma^2 + 2\sigma_a^2 + 3.23\theta_\tau$		0.8077
Residual	4	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	4	$\sigma^2 + 4\theta_\gamma$	1	
Residual	10	σ^2		

of these two designs would be preferable and under which circumstances. These circumstances are not just different ranges of values of VCs, but also different ranges of values in the fixed effects for the simulation study.

6.2.3 Missing values

One of the issues that arises with high-throughput multiplexing experiments is that of missing data. For a single protein, there are various ways in which the missing values can arise in a MudPIT-iTRAQTM proteomics experiment. One form of missing data which we are most interested, is when a unique peptide, which only belongs to a specific protein, is simply not found in one run of the experiment, but can be found on the other runs of the experiment. Thus, during database searching, bioinformatics software cannot re-construct this specific protein; and this protein would then be considered as missing for one entire run of the Phase 2 experiment. This can be problematic in the analysis stage, as the design is likely to become unbalanced due to unequal replication of the treatment group or the experimental units from the Phase 1 experiment.

For example, consider the Phase 2 experiment with the Phase 1 experiment consisting of $\nu = 4$ treatments assigned to $n_a = 12$ animals. Each animal is then further split into $n_s = 2$ sub-samples and measured in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags. An optimal design of Phase 2 experiment or this scenario is presented in Table 6.3.

The theoretical ANOVA of the full design in Table 6.3 is presented in Table 6.4. The total of 23 DF were partitioned to 5 DF for Between Runs stratum and 18 DF for Within Runs

Table 6.3: Optimal design for Phase 2 experiment showing the allocation of sub-samples from treatments assigned to animals, when the Phase 1 experiment consists of $\nu = 4$ treatments assigned to $n_a = 12$ animals, $n_s = 2$ sub-samples are then taken from each animals and measured in the Phase 2 MudPIT-iTRAQTM experiment comprising $n_r = 6$ runs and $n_\gamma = 4$ tags.

Run	Tag			
	114	115	116	117
1	<i>Jb</i>	<i>Ld</i>	<i>Ea</i>	<i>Cc</i>
2	<i>Ld</i>	<i>Jb</i>	<i>Cc</i>	<i>Ea</i>
3	<i>Aa</i>	<i>Gc</i>	<i>Fb</i>	<i>Dd</i>
4	<i>Gc</i>	<i>Aa</i>	<i>Dd</i>	<i>Fb</i>
5	<i>Hd</i>	<i>Ia</i>	<i>Kc</i>	<i>Bb</i>
6	<i>Ia</i>	<i>Hd</i>	<i>Bb</i>	<i>Kc</i>

stratum. In the Between Animals Within Runs stratum, Treatment effects could be estimated with 0.96 amount of the treatment information with 5 Residual DF for estimating the variance of Treatment effects. In addition, there was a valid F-test for comparing between treatments, because the coefficients of VCs were the same for the Treatment and Residual EMS in the Between Animals Within Runs stratum.

Table 6.4: Theoretical ANOVA table of design in Table 6.3.

Source of Variation	DF	EMS	E_γ	E_τ
Between Runs				
Between Animals	2	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Within Animals	3	$\sigma^2 + 4\sigma_r^2$		
Within Runs				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 6\theta_\gamma + 0.67\theta_\tau$	1	0.1111
Treatment	3	$\sigma^2 + 2\sigma_a^2 + 5.76\theta_\tau$		0.96
Residual	5	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 6\theta_\gamma$	1	
Residual	7	σ^2		

If a given protein was not detected in Run 6, then there were four observation missing for the Phase 2 experiment. The theoretical ANOVA is presented Table 6.5, which showed the total DF were reduce to 19 DF. The Residual DF in the Between Animals Within Runs stratum were also reduced to 3 DF, which is 2 DF less than the full design. Furthermore, there was no a direct valid F-test for this design, as coefficients of the VCs from the Treatment and Residual EMS were different in the Between Animals Within Runs stratum. Finally, the amount of the treatment information was also reduced from 0.96 to 0.8.

If a given protein is not detected in Runs 5 and 6, we then are left with 16 observations for the Phase 2 experiment. The theoretical ANOVA of the new design is presented in Table 6.6.

Table 6.5: Theoretical ANOVA table of design in Table 6.3 with Run 6 missing.

Source of Variation	DF	EMS	E_γ	E_τ
Between Run				
Between Animals	2	$\sigma^2 + 1.6\sigma_a^2 + 4\sigma_r^2$		
Within Animals	2	$\sigma^2 + 4\sigma_r^2$		
Within Run				
Between Animals				
Tag	3	$\sigma^2 + 1.27\sigma_a^2 + 1.36\theta_\gamma + 0.43\theta_\tau$	0.2727	0.0857
Treatment	3	$\sigma^2 + 1.96\sigma_a^2 + 4.23\theta_\tau$		0.8471
Residual	3	$\sigma^2 + 1.78\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 4\theta_\gamma$	0.8	
Residual	4	σ^2		

The Residual DF in the Between Animals Within Runs stratum were reduced to 2 DF, which is 3 DF less than the full design. However, there was a valid F-test for Treatment effects, with Treatment effects being fully estimable in the Between Animals Within Runs stratum. This was due to how we structured our initial designs with a 2-run-by-2-tag array system. Hence, if the last two runs of the experiment were to be missing, we basically lose one biological replicate, i.e. there would now be 8 animals from the Phase 1 experiments, so that the allocation of subsamples of animals and treatments, to be labelled with tags and analysed with runs still would have a balanced arrangement. Hence, the optimal design presented in Table 6.3 is shown to be robust in dealing with certain pattern of missing data, i.e. when Runs 1 and 2, or Runs 3 and 4, or Runs 5 and 6 are missing. Other different patterns of missing data will result in designs that have no valid F-test for treatment effects, or make it difficult to estimate the VCs from the theoretical ANOVA.

Table 6.6: Theoretical ANOVA table of design in Table 6.3 with Runs 5 and 6 missing.

Source of Variation	DF	EMS	E_γ	E_τ
Between Run				
Between Animals	1	$\sigma^2 + 2\sigma_a^2 + 4\sigma_r^2$		
Within Animals	2	$\sigma^2 + 4\sigma_r^2$		
Within Run				
Between Animals				
Tag	1	$\sigma^2 + 2\sigma_a^2 + 4\theta_\gamma$	1	
Treatment	3	$\sigma^2 + 2\sigma_a^2 + 4\theta_\tau$		1
Residual	2	$\sigma^2 + 2\sigma_a^2$		
Within Animals				
Tag	2	$\sigma^2 + 4\theta_\gamma$	1	
Residual	4	σ^2		

Further simulation studies can be done to explore what happens to the properties of the designs considered in Chapters 3 and 4 with different patterns of missingness. We can investigate how the design can start to breakdown as observed in Table 6.5, when there is one run of the experiment that is missing. We can further examine any alternative designs which have more desirable properties in terms of their robustness for downstream statistical analyses when we have missing values.

An alternative approach would be to construct an imputation model under a Bayesian multivariate and multilevel inference framework (Zeng et al., 2017). This model would use the information from experimental factors, such as the physical properties of the peptides, the effects from iTRAQTM tags and MudPIT runs, along with the clinical factors of each patient to construct a likelihood model. Each parameter in the likelihood model would be estimated by using Empirical Bayesian Hamiltonian MC algorithm, which integrates prior information for missing data and the distribution of missing values. The resultant posterior distribution of these parameters, including parameters of interest, would therefore be estimated utilising both the pattern of missing data and information for missing values. We could incorporate this framework into how to better design the Phase 2 experiment, which would enable us to impute reliable values for the final analysis.

6.3 More general future research directions

Another multiplexing technology, which started becoming popular only a few years ago is *Next-Generation Sequencing* (NGS). This multi-plexing technology can be carried out by attaching unique index sequences, namely *barcodes*, on the end of each DNA or RNA fragments (Smith et al., 2010). Therefore, different barcodes are attached to different biological samples, allowing NGS instrument to sequence multiple samples simultaneously. The abundance levels of sequences are then measured based on the number of barcodes present in each sample. These barcodes are very similar to the iTRAQTM tags, where measuring protein abundances. Note that MudPIT runs of the proteomics experiments are referred to as *lanes* of the NGS experiments. Thus, the methods of optimal designs described in this thesis also apply to this technology.

We can currently obtain a kit with 96 and 384 barcodes, meaning that we can quantify up to 96 or 384 unique samples at the same time (Smith et al., 2010; Shapland et al., 2015). However, using more barcodes is not always ideal, because as more barcodes are used the number of DNA or RNA sequences for each barcode decreases (Campbell et al., 2015). Hence, deciding on the number of barcodes is more practical than theoretical.

Let us consider a Phase 1 experiment arranged in a CRD with $\nu = 8$ treatments assigned to $n_a = 48$ animals, and the sample from each animal split into $n_s = 2$ sub-samples, which gives us a total of $n = 96$ sub-samples to be measured using the NGS technology. If a researcher decides to use the kit with 96 barcodes for just one lane of the experiment, then the Treatment effects are completely confounded with Tag effects.

Using the objective function and SA algorithm derived in this thesis, we can quickly generate four optimal designs with multiple lanes, where all have a valid F-test for Treatment effects, with different numbers of barcodes used in the Phase 2 experiment. The Residual DF and the treatment average efficiency factors of these four designs are presented in Table 6.7. This showed the best option is to use 8 lanes of the experiment with 12 barcodes, which generated the highest Residual DF (32 DF) and the treatment average efficiency factors (0.9837) in the Between Animals Within Runs stratum. However, given that each lane of experiment costs about five thousand dollars, it may be ideal to advise the researcher to use 4 lanes with 24 barcodes, because there would not a lot of improvement compared to using 8 lanes of the experiment with 12 barcodes. Therefore, more work can be done in examining the efficiency of using different numbers of barcodes for generating a better optimal design of the Phase 2 experiment.

Table 6.7: Residual DF and treatment average efficiency factors from the optimal design with different number of lanes and barcodes for Next-Generation Sequencing technology.

Number of lanes	Number of barcodes	Residual DF	E_τ
2	48	17	0.56
4	24	28	0.8532
8	12	32	0.9837
16	6	31	0.9510

Finally, the NGS experiment returns counts as the response. The method in this thesis assumes that the response, once log transformed, is normally distributed; thus, all of the designs we have generated assume unit-treatment additivity. Having a count as the response violates this assumption, and so further research could be undertaken on how to obtain optimal designs of the two-phase experiment where the response exhibits a non-normal distribution.

Further research outlined here will help to maximise the benefits of new technologies, such as NGS, while at the same time extending the capabilities of our method, for generating optimal designs, to a wider range of settings in two-phase multiplex proteomics experiments.

Appendices

Appendix A

Reference manual of infoDecompuTE package

Package ‘infoDecompuTE’

March 6, 2017

Title Information Decomposition of Two-Phase Experiments

Version 0.6.0

Date 2017-03-05

Description The main purpose of this package is to generate the structure of the analysis of variance (ANOVA) table of the two-phase experiments. The user only need to input the design and the relationships of the random and fixed factors using the Wilkinson-Rogers' syntax, this package can then quickly generate the structure of the ANOVA table with the coefficients of the variance components for the expected mean squares. Thus, the balanced incomplete block design and provides the efficiency factors of the fixed effects can also be studied and compared much easily.

Depends R (>= 3.0.0)

Imports MASS

License GPL (>= 3)

Encoding UTF-8

LazyLoad yes

URL <https://github.com/kcha193/infoDecompuTE>

BugReports <https://github.com/kcha193/infoDecompuTE/issues>

RoxygenNote 6.0.1

NeedsCompilation no

Author Kevin Chang [aut, cre]

Maintainer Kevin Chang <k.chang@auckland.ac.nz>

Repository CRAN

Date/Publication 2017-03-06 15:54:30

R topics documented:

infoDecompuTE-package	2
adjustEffectNames	3
adjustMissingLevels	4
getCoefVC	5

getEffFactor	6
getFixedEF	8
getTrtCoef	10
getTrtRep	11
getVMat.onePhase	12
identityMat	14
infoDecompMat	15
invInfMat	16
J	17
K	18
makeBlkDesMat	18
makeContrMat	19
makeOrthProjectors	21
makeOverDesMat	22
projMat	23
summaryAovOnePhase	24
summaryAovTwoPhase	26
toLatexTable	29
tr	31
unity	32

Index 33

infoDecompuTE-package *Information Decomposition of Two-phase Experiments*

Description

InfoDecompuTE is capable of generating the structure of the analysis of variance (ANOVA) table of the two-phase experiments. By inputting the design and the relationships of the random and fixed factors using the Wilkinson-Rogers' syntax, infoDecompuTE can generate the structure of the ANOVA table with the coefficients of the variance components for the expected mean squares. This package can also study the balanced incomplete block design and provides the efficiency factors of the fixed effects.

Details

Package:	infoDecompuTE
Type:	Package
Version:	0.5.1
Date:	2013-01-04
License:	GPL (>= 3)
LazyLoad:	yes

Author(s)

Kevin Chang and Katya Ruggiero

Maintainer: Kevin Chang <kcha193@aucklanduni.ac.nz>

adjustEffectNames	<i>Adjust the Effects' Names</i>
-------------------	----------------------------------

Description

Adjust for appropriate syntax describing the effects matching the structural formula.

Usage

```
adjustEffectNames(effectsMatrix, effectNames)
```

Arguments

effectNames	a vector of character containing the labels of the treatment or block terms in the model generated by the terms .
effectsMatrix	a matrix of variables by terms showing which variables appear in which terms generated by the terms .

Value

A vector of character containing the labels of the terms in the model with appropriate syntax describing the effects.

Author(s)

Kevin Chang

Examples

```
str.for = "A*(B/E/C)*D"
effectsMatrix= attr(terms(as.formula(paste("~", str.for)), keep.order = TRUE) , "factors")
effectNames = attr(terms(as.formula(paste("~", str.for)), keep.order = TRUE) , "term.labels")

adjustEffectNames(effectsMatrix, effectNames)
```

adjustMissingLevels	<i>Adjust the Missing Levels</i>
---------------------	----------------------------------

Description

Adjust for appropriate syntax describing the effects matching the structural formula.

Usage

```
adjustMissingLevels(design.df, str.for)
```

Arguments

design.df	a data frame containing the experimental design. Requires every column be a factor .
str.for	a single string of characters containing the structural formula using the Wilkinson-Rogers' syntax.

Value

A list containing a data frame with the experimental design and a single string of characters containing the structural formula.

Author(s)

Kevin Chang

Examples

```
design.df = data.frame( Blk = factor(1:16),
                       Ani = factor(c( 1,1,2,2,
                                       1,1,2,2,
                                       1,1,2,2,
                                       1,1,2,2)),
                       Trt = factor(c( 1,2,3,4,
                                       1,2,3,4,
                                       1,2,3,4,
                                       1,2,3,4)))

adjustMissingLevels(design.df, str.for = "Ani/Trt")
```

getCoefVC	<i>Get Variance Components' Coefficients and Mean Squares for Single-Phase or Two-Phase Experiments</i>
-----------	---

Description

Compute the variance components' coefficients and corresponding to random effects in the expected mean squares of ANOVA table in single-phase or two-phase experiments. These coefficients are then inserted to a matrix where the rows correspond to each source of variation and column correspond to DF and every variance component. The mean squares is calculated if the response argument is used.

Usage

```
getCoefVC.onePhase(Pb, design.df, v.mat, response, table.legend, decimal, digits)
getCoefVC.twoPhase(Pb, design.df, v.mat, response, table.legend, decimal, digits)
```

Arguments

Pb	a list of matrices generated by infoDecompMat function.
design.df	a data frame containing the experimental design. Requires every column be a factor .
v.mat	a list of matrix generated by getVMat.onePhase or getVMat.twoPhase .
response	a numeric vector contains the responses from the experiment.
table.legend	a logical allows users to generate a legend for the variance components of the ANOVA table for large designs. Default is FALSE, resulting in the use of original block factor names.
decimal	a logical allows users to display the coefficients as the decimals. Default is FALSE, resulting in the use of fractions.
digits	a integer indicating the number of decimal places. Default is 2, resulting in 2 decimal places.

Details

The main purpose of this function is to combine the matrices presenting every source of variation of the ANOVA table and the variance matrix to compute the coefficients of the variance components. The complication arise in giving the row names of the matrix for the source of variation in the ANOVA table.

Value

A matrix containing the characters.

Author(s)

Kevin Chang

Examples

```

design1 <- local({
  Ani <- as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt <- as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str <- "Ani"

rT <- terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)
blkTerm <- attr(rT, "term.labels")

Z <- makeBlkDesMat(design1, blkTerm)

trt.str <- "Trt"
fT <- terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms

trtTerm <- attr(fT, "term.labels")
effectsMatrix <- attr(fT, "factor")

T <- makeContrMat(design1, trtTerm, effectsMatrix, contr.vec = NA)

N = makeOverDesMat(design1, trtTerm)

Replist = getTrtRep(design1, trtTerm)

Rep <- Replist$Rep
trt.Sca <- Replist$Sca

effFactors = lapply(makeOrthProjectors(Z), function(z)
  getEffFactor(z, T, N, Rep, trt.Sca))

#Now construct variance matrices
Pb <- effFactors[sort(1:length(effFactors), decreasing=TRUE)]

v.mat <- getVMat.onePhase(Z.Phase1 = Z, design.df = design.df, var.comp = NA)

getCoefVC.onePhase(Pb = Pb, design.df = design1, v.mat = v.mat, response = NA,
  table.legend = FALSE, decimal = FALSE, digit = 2)

```

getEffFactor

Construct the Matrix from Information Decomposition and Compute the Efficiency Factors of Treatment effects

Description

Perform the information decomposition for either the block or treatment effects within a single stratum and Compute the Efficiency Factors for every treatment effect within a single stratum.

Usage

```
getEffFactor(z, T, N, Rep, trt.Sca)
```

Arguments

<code>z</code>	a matrix containing the orthogonal projector of a stratum generated by makeOrthProjectors .
<code>T</code>	a list of contrast matrices generated by makeContrMat .
<code>N</code>	a matrix containing the design matrix generated by makeOverDesMat .
<code>Rep</code>	a matrix containing the treatment replication number and is generated by getTrtRep .
<code>trt.Sca</code>	a numeric vector for computing a coefficients of the fixed effect parameter in EMS and is generated by getTrtRep .

Details

The main purpose of this function is to construct a list of resultant matrices associated with each source of variation after the information decomposition and to compute the canonical or average efficiency factors for each treatment effects in each stratum of ANOVA table.

The canonical efficiency factors are generated when the user input the treatment contrasts, otherwise the average efficiency factors, which is the harmonic mean of the canonical efficiency factors, are generated.

Value

A list of matrices and numeric vectors containing the efficiency factors of every treatment effect.

Author(s)

Kevin Chang

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str = "Ani"

rT = terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)
blkTerm = attr(rT, "term.labels")

Z = makeBlkDesMat(design1, blkTerm)

trt.str = "Trt"
fT <- terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms
```

```

trtTerm <- attr(fT, "term.labels")
effectsMatrix <- attr(fT, "factor")

T <- makeContrMat(design1, trtTerm, effectsMatrix, contr.vec = NA)

N = makeOverDesMat(design1, trtTerm)

Replist = getTrtRep(design1, trtTerm)

Rep <- Replist$Rep
trt.Sca <- Replist$Sca

effFactors = lapply(makeOrthProjectors(Z), function(z)
  getEffFactor(z, T, N, Rep, trt.Sca))

```

getFixedEF	<i>Get the Fixed Components' coefficients and Efficiency Factors of Single-Phase or Two-Phase Experiments.</i>
------------	--

Description

Calculate coefficients of fixed effects components of EMS and Treatment Efficiency Factors within each stratum in Single-Phase or two-phase experiment.

Constructs a matrix containing the coefficients of the coefficients of fixed effects components of EMS within each stratum. Also calculates and the average efficiency factors of each treatment effect across all strata

Construct a matrix contain the coefficients of the fixed Components and the average efficiency factors of single-phase experiments.

Usage

```

getFixedEF.onePhase(effFactors, trt.Sca, T, Rep, table.legend, decimal, digits, list.sep)
getFixedEF.twoPhase(effFactors, trt.Sca, T, Rep, table.legend, decimal, digits, list.sep)

```

Arguments

effFactors	a list of numeric vector generated by getEffFactor function.
trt.Sca	a numeric vector generated by getTrtRep function.
T	a list of matrices generated by makeContrMat function.
Rep	a numeric matrix generated by getTrtRep function.
table.legend	a logical allows users to generate a legend for the variance components of the ANOVA table for large designs. Default is FALSE, resulting in the use of original treatment factor names.
decimal	a logical allows users to display the coefficients as the decimals. Default is FALSE, resulting in the use of fractions.

digits	a integer indicating the number of decimal places. Default is 2, resulting in 2 decimal places.
list.sep	a logical allows users to present the efficiency factors and coefficients of the fixed effects a list of separate matrices.

Details

The function uses the efficiency factors generated by `getEffFactor` to calculate the coefficients of fixed Effects components of EMS and insert the treatment efficiency factor within each stratum.

The complication arises in giving the row names of the matrix for the source of variation in the ANOVA table.

Value

A matrix.

Author(s)

Kevin Chang

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str <- "Ani"

rT <- terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)
blkTerm = attr(rT, "term.labels")

Z <- makeBlkDesMat(design1, blkTerm)

trt.str <- "Trt"
fT <- terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms

trtTerm <- attr(fT, "term.labels")
effectsMatrix <- attr(fT, "factor")

T <- makeContrMat(design1, trtTerm, effectsMatrix, contr.vec = NA)

N <- makeOverDesMat(design1, trtTerm)

Replist = getTrtRep(design1, trtTerm)

Rep <- Replist$Rep
trt.Sca <- Replist$Sca
```



```

effFactors = lapply(makeOrthProjectors(Z), function(z) getEffFactor(z, T, N, Rep, trt.Sca))

effFactors <- effFactors[sort(1:length(effFactors), decreasing=TRUE)]

getFixedEF.onePhase(effFactors = effFactors, trt.Sca = trt.Sca, T = T, Rep = Rep,
table.legend = FALSE, decimal = FALSE, digits = 2, list.sep = TRUE)

```

getTrtCoef

Get the Treatment Coefficients

Description

Compute the overall coefficients every treatment term including the interaction.

Usage

```
getTrtCoef(design.df, trtTerm)
```

Arguments

design.df	a data frame containing the experimental design. Requires every column be a factor .
trtTerm	a vector of character containing the labels of the treatment terms in the model generated by terms .

Value

The numeric vector.

Author(s)

Kevin Chang

Examples

```

design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

trt.str = "Trt"

fT = terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms

```

```
trtTerm = attr(fT,"term.labels")
effectsMatrix = attr(fT,"factor")

trt.Coeff = getTrtCoef(design1, trtTerm)
```

getTrtRep

Calculate the Treatment Replication number

Description

Calculate the replication number of every treatment term including the interaction. This is used to compute the treatment efficiency factors.

Usage

```
getTrtRep(design.df, trtTerm)
```

Arguments

design.df	a data frame containing the experimental design. Requires every column be a factor .
trtTerm	a vector of character containing the labels of the treatment terms in the model generated by the terms .

Value

A list containing two objects. The first object is a matrix called Rep which contains the replication numbers, where the rows correspond to each treatment combination and the columns correspond to the treatment factors, i.e. the replication number with respect to each treatment factor based on the treatment combination. The second object called Sca which is a numeric vector for computing a coefficients of the fixed effect parameter in EMS.

Author(s)

Kevin Chang

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

trt.str = "Trt"

fT = terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms

trtTerm = attr(fT,"term.labels")
effectsMatrix = attr(fT,"factor")

getTrtRep(design1, trtTerm)
```

getVMat.onePhase	<i>Get the Variance Matrices for Sngle-Phase or Two-Phase experiment</i>
------------------	--

Description

Construct the matrix for each variance components for the single-phase or two-phase experiment.

Usage

```
getVMat.onePhase(Z.Phase1, design.df, var.comp = NA)
getVMat.twoPhase(Z.Phase1, Z.Phase2, design.df, var.comp = NA)
```

Arguments

Z.Phase1	a list of block design matrix from makeBlkDesMat function from Phase 1 block structure.
Z.Phase2	a list of block design matrix from makeBlkDesMat function from Phase 2 block structure.
design.df	a data frame containing the experimental design. Requires every column be a factor .
var.comp	a vector of characters containing the variance components of interest this allows the user to specify the variance components to be shown on the ANOVA table. This also allows the user to specify artificial stratum to facilitate decomposition. Default is NA, which uses every random factor as the variance components with the first phase's variance components in appear before the second phase's variance components.

Value

A list of matrices.

Author(s)

Kevin Chang

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str = "Ani"

rT = terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)

blkTerm = attr(rT,"term.labels")
Z = makeBlkDesMat(design1, rev(attr(rT,"term.labels"))))

V = getVMat.onePhase(Z, design1)

design2 <- local({
  Run = as.factor(rep(1:4, each = 4))
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8,
                           3,4,1,2,
                           7,8,5,6)])
  Tag = as.factor(c(114,115,116,117)[rep(1:4, 4)])
  Trt = as.factor(letters[c(1,2,1,2,
                           2,1,2,1,
                           1,2,1,2,
                           2,1,2,1)])
  data.frame(Run, Ani, Tag, Trt)
})

blk.str1 = "Ani"
blk.str2 = "Run"

rT1 = terms(as.formula(paste("~", blk.str1, sep = "")), keep.order = TRUE)
#random terms phase 1
rT2 = terms(as.formula(paste("~", blk.str2, sep = "")), keep.order = TRUE)
#random terms phase 2

blkTerm1 = attr(rT1,"term.labels")
blkTerm2 = attr(rT2,"term.labels")

Z1 = makeBlkDesMat(design2, rev(blkTerm1))
```

```
Z2 = makeBlkDesMat(design2, rev(blkTerm2))  
  
V = getVMat.twoPhase(Z1, Z2, design2, var.comp = NA)
```

identityMat*Identity Matrix*

Description

Construct an identity matrix.

Usage

```
identityMat(n)
```

Arguments

n a numeric describes the dimension of the identity matrix.

Value

This function returns a matrix with the diagonal elements equal to one and the off-diagonal elements equal to zero.

Author(s)

Kevin

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

See Also

[diag](#)

Examples

```
identityMat(10)
```

infoDecompMat	<i>Construct the Matrix from Information Decomposition</i>
---------------	--

Description

Perform the information decomposition for either the block or treatment effects within a single stratum.

Usage

```
infoDecompMat(z, T, N)
```

Arguments

z	a matrix containing the orthogonal projector for a single stratum generated by makeOrthProjectors .
T	a list of contrast matrices generated by makeContrMat .
N	a matrix containing the design matrix generated by makeOverDesMat .

Details

The main purpose of this function is to construct a list of resultant matrices associated with each source of variation after the information decomposition.

This list of matrices are then used to compute the coefficient of the variance components in the expected mean squares.

Value

A list of matrices.

Author(s)

Kevin Chang

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str = "Ani"

rT = terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)
blkTerm = attr(rT,"term.labels")
```

```

Z = makeBlkDesMat(design1, blkTerm)
Pb = makeOrthProjectors(Z)

trt.str = "Trt"
fT <- terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms

trtTerm <- attr(fT, "term.labels")
effectsMatrix <- attr(fT, "factor")

T <- makeContrMat(design1, trtTerm, effectsMatrix, contr.vec = NA)

N = makeOverDesMat(design1, trtTerm)

infoDecompMat(Pb[[1]], T, N)

```

invInfMat

Invert the Information Matrix

Description

Using the eigenvalue decomposition method to invert the information matrix.

Usage

```
invInfMat(C, N, T)
```

Arguments

C	a matrix of block projector for a single stratum.
N	a matrix representation the smallest unit of block or treatment effects generated by makeOverDesMat .
T	a list of contrast matrices from makeContrMat .

Value

This function returns a matrix.

Author(s)

Kevin Chang

References

Nelder JA (1965b). "The Analysis of Randomized Experiments with Orthogonal Block Structure. II. Treatment Structure and the General Analysis of Variance." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283(1393), 163-178.

Examples

```
m <- matrix(rnorm(10), 10, 10)

invInfMat(m, identityMat(10), identityMat(10))
```

J*Identity Matrix Minus Averging Matrix*

Description

Construct a square matrix which the identity matrix minus the averging matrix.

Usage

`J(n)`

Arguments

`n` a numeric describes the dimension of the square matrix.

Value

This function return a square matrix which the identity matrix minus the averging matrix.

Author(s)

Kevin Chang

Examples

```
J(10)
```

K	<i>Averaging Matrix</i>
---	-------------------------

Description

Construct a n-by-n averaging matrix.

Usage

K(n)

Arguments

n a numeric describes the dimension of the averaging matrix.

Value

This function returns a $n \times n$ square matrix with all elements equal 1/n.

Author(s)

Kevin Chang

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

Examples

K(10)

makeBlkDesMat	<i>Construct Block Design Matrix</i>
---------------	--------------------------------------

Description

Construct a binary matrix representing the block design. The rows are corresponding to the observations and the columns are corresponding to the blocks.

Usage

makeBlkDesMat(design.df, blkTerm)

Arguments

- design.df a data frame containing the experimental design. Requires every column be a [factor](#).
- blkTerm a vector of character containing the labels of the block terms in the model generated by the [terms](#).

Value

A list of the binary matrices.

Author(s)

Kevin Chang

See Also

[terms](#)

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str = "Ani*Trt"

rT = terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)

blkTerm = attr(rT,"term.labels")
Z = makeBlkDesMat(design1, blkTerm)
```

makeContrMat

Make Contrast Matrix

Description

Construct a list of contrast matrices for block for treatment effects.

Usage

```
makeContrMat(design.df, effectNames, effectsMatrix, contr.vec)
```



```

trt.str = "Trt"
fT <- terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE) #fixed terms

trtTerm <- attr(fT, "term.labels")
effectsMatrix <- attr(fT, "factor")

T <- makeContrMat(design1, trtTerm, effectsMatrix, contr.vec = NA)

#Fit each treatment contrasts as a vector seperately
Trt1 <- rep(c(1,-1), each = 4)
Trt2 <- rep(c(1,-1), time = 4)
Trt3 <- Trt1*Trt2

T <- makeContrMat(design1, trtTerm, effectsMatrix,
  contr.vec =list(Trt = list(Trt1 = Trt1, Trt2 = Trt2, Trt3 = Trt3)))

```

makeOrthProjectors	<i>Construct Orthogonal Projector Matrices</i>
--------------------	--

Description

Construct a list of orthogonal projector matrices corresponding to all strata of the experiment.

Usage

```
makeOrthProjectors(BlkDesList)
```

Arguments

BlkDesList a list of block design matrices generated by [makeBlkDesMat](#).

Details

The strata decomposition is performed within this function. The first step is to convert the list of block design matrices generated by [makeBlkDesMat](#) to projection matrices using [projMat](#). The second step is to use these projection matrices to project the raw data vector from one stratum to next stratum of the experiment; the resulting matrix corresponds to each stratum is the orthogonal projector matrix of the given stratum.

Value

A list containing matrices.

Author(s)

Kevin Chang

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str = "Ani"

rT = terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)
blkTerm = attr(rT,"term.labels")

Z = makeBlkDesMat(design1, blkTerm)
Pb = makeOrthProjectors(Z)
```

makeOverDesMat	<i>Construct the Overall Treatment or Block design Matrix</i>
----------------	---

Description

Construct the treatment or block matrix of the smallest unit based from the experimental design.

Usage

```
makeOverDesMat(design.df, effectNames)
```

Arguments

design.df	a data frame containing the experimental design. Requires every column be a factor .
effectNames	a vector of character containing the labels of the treatment or block terms in the model generated by the terms .

Details

The main purpose this matrix is used in information decomposition. For the factorial experiment, this matrix is typically the treatment design matrix associated with the interaction effects, because the interaction effects are the smallest unit for the treatment effects.

For the two-phase experiments, the same method of information decomposition is used for the block effects of Phase 1 experiment in the stratum defined from the block structure of the Phase 2 experiment. Hence, the block design matrix of the smallest unit for the block effects of Phase 1 experiment can also be constructed using this function.

Value

A matrix where the rows correspond to the observation and columns correspond to the overall combination of the treatment factors or the block factors of the Phase 1 experiment.

Author(s)

Kevin Chang

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

trt.str = "Trt"

fT = terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE)

trtTerm = attr(fT, "term.labels")
effectsMatrix = attr(fT, "factor")

makeOverDesMat(design1, trtTerm)
```

projMat

Construct a Projection Matrix

Description

Compute the projection matrix from a square matrix.

Usage

```
projMat(X)
```

Arguments

X a square matrix.

Value

A square matrix.

Author(s)

Kevin Chang

Examples

```
m = matrix(1, nrow = 10, ncol = 3)
projMat(m)
```

summaryAovOnePhase	<i>Summarize an Theoretical Analysis of Variance Model of Single-Phase Experiments</i>
--------------------	--

Description

Computes the coefficients of the variance components for the expected mean squares for single-phase experiments. The function accepts a data frame of the experimental design with the structural formulae of the block and treatment factors. Two tables containing the variance components of the random effects and fixed effects are returned.

Usage

```
summaryAovOnePhase(design.df, blk.str, trt.str, var.comp = NA,
                    trt.contr = NA, table.legend = FALSE,
                    response = NA, latex = FALSE, fixed.names = NA,
                    decimal = FALSE, digits = 2, list.sep = TRUE)
```

Arguments

- design.df a data frame containing the experimental design. Requires every column be a **factor**. Any punctuation or symbol such as dots or parentheses should be avoid for the c
- blk.str a single string of characters containing the structural formula for the block fac-
tors using the Wilkinson-Rogers' syntax.
- trt.str a single string of characters containing the structural formula for the treatment
factors using the Wilkinson-Rogers' syntax.
- var.comp a vector of characters containing the variance components of interest this allows
the user to specify the variance components to be shown on the ANOVA table.
This also allows the user to specify artificial stratum to facilitate decomposition.
Default is NA, which uses every random factor as the variance components from
random.terms.
- trt.contr a list of treatment contrast vectors, this allows the user to specify the contrasts for
each treatment factor. Note that if this argument is used, it is necessary to specify
the contrasts for every treatment factor with the same order as fixed.terms.
Default is NA, which uses the C matrix described by John and Williams (1987).

table.legend	a logical allows the users to use the legend for the variance components of the ANOVA table for a large design. Default is FALSE, which uses the original names.
response	a numeric vector contains the responses from the experiment.
latex	a logical allows the users to output the Latex script to Latex table. Once the Latex script is generated, it requires the user to install and load two Latex packages: booktabs and bm to compile the Latex script.
fixed.names	a vector of character allows the users to modify symbols for the fixed effects for the Latex outputs.
decimal	a logical allows users to display the coefficients as the decimals. Default is FALSE, resulting in the use of function fractions.
digits	a integer indicating the number of decimal places. Default is 2, resulting in 2 decimal places.
list.sep	a logical allows users to present the efficiency factors and coefficients of the fixed effects a list of separate matrices. Default is TRUE.

Value

The values returned depends on the value of the `table.legend` argument. If `table.legend = FALSE`, this function will return a list of two data frames. The first data frame contains the random effects and the second data frame contains the fixed effects. If the `table.legend` argument is TRUE, then it will return a list containing two lists. The first list consists of a data frame of random effects and a character string for the legend. The second list consists of a data frame of fixed effects and a character string for the legend. If `response` argument is used, the random effect table will have one extra column with of mean squares computed from the responses from the experiment.

Author(s)

Kevin Chang

References

- John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.
- Nelder JA (1965b). "The Analysis of Randomized Experiments with Orthogonal Block Structure. II. Treatment Structure and the General Analysis of Variance." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283(1393), 163-178.
- Wilkinson GN, Rogers CE (1973). "Symbolic Description of Factorial Models for Analysis of Variance." *Applied Statistics*, 22(3), 392-399.

See Also

[terms](#) for more information on the structural formula.

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

summaryAovOnePhase(design1, blk.str = "Ani", trt.str = "Trt")

summaryAovOnePhase(design1, blk.str = "Ani", trt.str = "Trt",
  latex = TRUE, fixed.names = c("\\tau"))
```

summaryAovTwoPhase	<i>Summarize an Theoretical Analysis of Variance Model of Two-Phase Experiments</i>
--------------------	---

Description

Computes the coefficients of the variance components for the expected mean squares for two-phase experiments. The function accepts a data frame of the experimental design with the structural formulae of the block and treatment factors. Two tables containing the variance components of the random effects and fixed effects are returned.

Usage

```
summaryAovTwoPhase(design.df, blk.str1, blk.str2, trt.str, var.comp = NA,
  blk.contr = NA, trt.contr = NA, table.legend = FALSE,
  response = NA, latex = FALSE, fixed.names = NA,
  decimal = FALSE, digits = 2, list.sep = TRUE)
```

Arguments

design.df	a data frame containing the experimental design. Requires every column be a factor . Any punctuation or symbol such as dots or parentheses should be avoid for the c
blk.str1	a single string of characters containing the structural formula for the block factors of the first-phase experiment using the Wilkinson-Rogers' syntax.
blk.str2	a single string of characters containing the structural formula for the block factors of the second-phase experiment using the Wilkinson-Rogers' syntax.
trt.str	a single string of characters containing the structural formula for the treatment factors using the Wilkinson-Rogers' syntax.
var.comp	a vector of characters containing the variance components of interest this allows the user to specify the variance components to be shown on the ANOVA table. This also allows the user to specify artificial stratum to facilitate decomposition. Default is NA, which uses every random factor as the variance components

	with the first phase's variance components in <code>random.terms1</code> appear before the second phase's variance components in <code>random.terms2</code> .
<code>blk.contr</code>	a list of first-phase block contrast vectors, this allows the user to specify the contrasts for each block factor in the first phase experiment. Note that if this argument is used, it is necessary to specify the contrasts for every treatment factor with the same order as <code>fixed.terms</code> . Default is NA, which uses the C matrix described by John and Williams (1987).
<code>trt.contr</code>	a list of treatment contrast vectors, this allows the user to specify the contrasts for each treatment factor. Note that if this argument is used, it is necessary to specify the contrasts for every treatment factor with the same order as <code>fixed.terms</code> . Default is NA, which uses the C matrix described by John and Williams (1987).
<code>table.legend</code>	a logical allows the users to use the legend for the variance components of the ANOVA table for a large design. Default is FALSE, which uses the original names.
<code>response</code>	a numeric vector contains the responses from the experiment.
<code>latex</code>	a logical allows the users to output the Latex script to Latex table. Once the Latex script is generated, it requires the user to install and load two Latex packages: <code>booktabs</code> and <code>bm</code> to compile the Latex script.
<code>fixed.names</code>	a vector of character allows the users to modify symbols for the fixed effects for the Latex outputs.
<code>decimal</code>	a logical allows users to display the coefficients as the decimals. Default is FALSE, resulting in the use of function <code>fractions</code> .
<code>digits</code>	a integer indicating the number of decimal places. Default is 2, resulting in 2 decimal places.
<code>list.sep</code>	a logical allows users to present the efficiency factors and coefficients of the fixed effects a list of separate matrices. Default is TRUE.

Value

The values returned depends on the value of the `table.legend` argument. If `table.legend = FALSE`, this function will return a list of two data frames. The first data frame contains the random effects and the second data frame contains the fixed effects. If the `table.legend` argument is TRUE, then it will return a list containing two lists. The first list consists of a data frame of random effects and a character string for the legend. The second list consists of a data frame of fixed effects and a character string for the legend. If `response` argument is used, the random effect table will have one extra column with of mean squares computed from the responses from the experiment.

Author(s)

Kevin Chang

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

Nelder JA (1965b). "The Analysis of Randomized Experiments with Orthogonal Block Structure. II. Treatment Structure and the General Analysis of Variance." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283(1393), 163-178.

Wilkinson GN, Rogers CE (1973). "Symbolic Description of Factorial Models for Analysis of Variance." *Applied Statistics*, 22(3), 392-399.

See Also

[terms](#) for more information on the structural formula.

Examples

```
#Phase 2 experiment
design2 <- local({
  Run = as.factor(rep(1:4, each = 4))
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8,
                           3,4,1,2,
                           7,8,5,6)])
  Sam = as.factor(as.numeric(duplicated(Ani)) + 1)
  Tag = as.factor(c(114,115,116,117)[rep(1:4, 4)])
  Trt = as.factor(c("healthy", "diseased")[c(1,2,1,2,
                                             2,1,2,1,
                                             1,2,1,2,
                                             2,1,2,1)])
  data.frame(Run, Ani, Sam, Tag, Trt)
})
design2

summaryAovTwoPhase(design2, blk.str1 = "Ani", blk.str2 = "Run",
trt.str = "Tag + Trt")

#Add the sample into the Phase 1 block structure
summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
trt.str = "Tag + Trt")

#Assuming there is crossing between the animals and samples
summaryAovTwoPhase(design2, blk.str1 = "Ani*Sam", blk.str2 = "Run",
trt.str = "Tag + Trt")

#Set Artificial stratum
design2$AniSet = as.factor(c(2, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1))
design2

summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "AniSet/Run",
trt.str = "Tag + Trt", var.comp = c("Ani:Sam", "Ani", "Run"))

#Define treatment contrasts
TagA = rep(c(1,1,-1,-1),time = 4)
```

```

TagB = rep(c(1,-1,1,-1),time = 4)
TagC = TagA * TagB
Tag = list(TagA = TagA, TagB = TagB, TagC = TagC)
Tag

Trt = as.numeric(design2$Trt)-1.5
Trt

summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
trt.str = "Tag + Trt",
trt.contr = list(Tag = list(TagA = TagA, TagB = TagB, TagC = TagC), Trt = Trt),
table.legend = TRUE)

#Compute MS
set.seed(527)
summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
trt.str = "Tag + Trt", response = rnorm(16))$ANOVA

#Generate Latex scripts
summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
trt.str = "Tag + Trt", latex = TRUE, fixed.names = c("\\gamma", "\\tau"))

#Generate Latex scripts with MS
set.seed(527)
summaryAovTwoPhase(design2, blk.str1 = "Ani/Sam", blk.str2 = "Run",
trt.str = "Tag + Trt", response = rnorm(16), latex = TRUE,
fixed.names = c("\\gamma", "\\tau") )

```

toLatexTable

*Convert the R output to Latex Table***Description**

Print the Latex scripts on the screen for the user to output the table from the Latex output.

Usage

```
toLatexTable(ANOVA, EF, fixed.names)
```

Arguments

ANOVA	a matrix containing the coefficients of the variance components in EMS of ANOVA table generated by getCoefVC.onePhase or getCoefVC.twoPhase .
EF	a matrix containing the coefficient of the fixed effects components and the treatment average efficiency factors generated by getFixedEF.onePhase or getFixedEF.onePhase function.
fixed.names	a vector of character allows the users to modify symbols for the fixed effects.

Details

Once the Latex script is generated, it requires the user to install and load two Latex packages: booktabs and bm to compile the Latex script.

Author(s)

Kevin Chang

Examples

```
design1 <- local({
  Ani = as.factor(LETTERS[c(1,2,3,4,
                           5,6,7,8)])
  Trt = as.factor(letters[c(1,1,1,1,
                           2,2,2,2)])
  data.frame(Ani, Trt)
})

blk.str <- "Ani"

rT <- terms(as.formula(paste("~", blk.str, sep = "")), keep.order = TRUE)
blkTerm <- attr(rT, "term.labels")

Z <- makeBlkDesMat(design1, blkTerm)

trt.str = "Trt"
fT <- terms(as.formula(paste("~", trt.str, sep = "")), keep.order = TRUE)

trtTerm <- attr(fT, "term.labels")
effectsMatrix <- attr(fT, "factor")

T <- makeContrMat(design1, trtTerm, effectsMatrix, contr.vec = NA)

N <- makeOverDesMat(design1, trtTerm)

Replist = getTrtRep(design1, trtTerm)

Rep <- Replist$Rep
trt.Sca <- Replist$Sca

effFactors = lapply(makeOrthProjectors(Z), function(z)
  getEffFactor(z, T, N, Rep, trt.Sca))

effFactors <- effFactors[sort(1:length(effFactors), decreasing=TRUE)]

v.mat <- getVMat.onePhase(Z.Phase1 = Z, design.df = design.df, var.comp = NA)

ANOVA <- getCoefVC.onePhase(Pb = effFactors, design.df = design1, v.mat = v.mat,
  response = NA, table.legend = FALSE, decimal = FALSE, digits = 2)

EF <- getFixedEF.onePhase(effFactors = effFactors, trt.Sca = trt.Sca, T = T,
```

```
Rep = Rep,  
table.legend = FALSE, decimal = FALSE, digits = 2, list.sep = FALSE)  
  
toLatexTable(ANOVA = ANOVA, EF = EF, fixed.names = c("\\tau"))
```

tr	<i>Trace of the Matrix</i>
----	----------------------------

Description

Compute the trace of the square matrix.

Usage

```
tr(X)
```

Arguments

X a square matrix.

Value

A numeric value.

Author(s)

Kevin

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

See Also

[diag](#)

Examples

```
m = matrix(1, nrow = 10, ncol = 10)  
tr(m)
```

`unity`*Construct a unity vector*

Description

Construct a vector with all elements unity.

Usage

```
unity(n)
```

Arguments

`n` a numeric describe the length of vector.

Value

This function returns a n
*times*1 matrix will all elements unity.

Author(s)

Kevin Chang

References

John J, Williams E (1987). *Cyclic and computer generated Designs*. Second edition. Chapman & Hall.

Examples

```
unity(10)
```

Index

*Topic **design**

summaryAovOnePhase, [24](#)

summaryAovTwoPhase, [26](#)

*Topic **package**

infoDecompuTE-package, [2](#)

adjustEffectNames, [3](#)

adjustMissingLevels, [4](#)

diag, [14](#), [31](#)

factor, [4](#), [5](#), [10–12](#), [19](#), [20](#), [22](#), [24](#), [26](#)

getCoefVC, [5](#)

getCoefVC.onePhase, [29](#)

getCoefVC.twoPhase, [29](#)

getEffFactor, [6](#), [8](#)

getFixedEF, [8](#)

getFixedEF.onePhase, [29](#)

getTrtCoef, [10](#)

getTrtRep, [7](#), [8](#), [11](#)

getVMat.onePhase, [5](#), [12](#)

getVMat.twoPhase, [5](#)

getVMat.twoPhase (getVMat.onePhase), [12](#)

identityMat, [14](#)

infoDecompMat, [5](#), [15](#)

infoDecompuTE (infoDecompuTE-package), [2](#)

infoDecompuTE-package, [2](#)

invInfMat, [16](#)

J, [17](#)

K, [18](#)

makeBlkDesMat, [18](#), [21](#)

makeContrMat, [7](#), [8](#), [15](#), [16](#), [19](#)

makeOrthProjectors, [7](#), [15](#), [21](#)

makeOverDesMat, [7](#), [15](#), [16](#), [22](#)

projMat, [21](#), [23](#)

summaryAovOnePhase, [24](#)

summaryAovTwoPhase, [26](#)

terms, [3](#), [10](#), [11](#), [19](#), [20](#), [22](#), [25](#), [28](#)

toLatexTable, [29](#)

tr, [31](#)

unity, [32](#)

Appendix B

Information matrix in more detail

This Chapter is to show information matrix $\mathbf{Z}'(\mathbf{I} - \mathbf{P}_r)(\mathbf{I} - \mathbf{P}_\gamma)\mathbf{Z}$ used for the objective function is symmetric.

Let \mathbf{P}_r and \mathbf{P}_γ be orthogonal projection matrices, then

$$\mathbf{P}_r = \mathbf{P}_r^2 = \mathbf{P}_r\mathbf{P}_r \quad \text{and} \quad \mathbf{P}_r = \mathbf{P}_r'$$

So

$$\begin{aligned} & \mathbf{Z}'(\mathbf{I} - \mathbf{P}_r)(\mathbf{I} - \mathbf{P}_\gamma)\mathbf{Z} \\ = & \mathbf{Z}'(\mathbf{I} - \mathbf{P}_r - \mathbf{P}_\gamma + \mathbf{P}_r\mathbf{P}_\gamma)\mathbf{Z} \\ = & \mathbf{Z}'\mathbf{Z} - \mathbf{Z}'\mathbf{P}_r\mathbf{Z} - \mathbf{Z}'\mathbf{P}_\gamma\mathbf{Z} + 2\mathbf{Z}'\mathbf{P}_r\mathbf{P}_\gamma\mathbf{Z} \\ = & \mathbf{Z}'\mathbf{Z} - (\mathbf{P}_r\mathbf{Z})'\mathbf{Z}\mathbf{P}_r - (\mathbf{P}_\gamma\mathbf{Z})'\mathbf{P}_\gamma\mathbf{Z} + 2\mathbf{Z}'\mathbf{P}_r\mathbf{P}_\gamma\mathbf{Z} \end{aligned} \tag{B.1}$$

$$\tag{B.2}$$

where $\mathbf{Z}'\mathbf{Z}$, $(\mathbf{P}_r\mathbf{Z})'\mathbf{Z}\mathbf{P}_r$ and $(\mathbf{P}_\gamma\mathbf{Z})'\mathbf{P}_\gamma\mathbf{Z}$ are symmetric.

The projection matrix which projects \mathbf{y} onto the Between Runs vector subspace is

$$\mathbf{P}_r = \mathbf{W}_r(\mathbf{W}_r'\mathbf{W}_r)^{-1}\mathbf{W}_r'$$

where \mathbf{W}_r is design matrix for Run factor of the Phase 2 experiment and can be expressed as

$$\mathbf{W}_r = \mathbf{1}_{n_\gamma} \otimes \mathbf{I}_{n_r}.$$

Thus, projection matrix for Between Runs can be re-written as

$$\mathbf{P}_r = \frac{1}{n_\gamma} \mathbf{1}_{n_\gamma} \mathbf{1}_{n_\gamma}' \otimes \mathbf{I}_{n_r}.$$

The projection matrix which projects \mathbf{y} onto the Between Tags vector subspace is

$$\mathbf{P}_\gamma = \mathbf{X}_\gamma (\mathbf{X}'_\gamma \mathbf{X}_\gamma)^{-1} \mathbf{X}_\gamma$$

where \mathbf{X}_γ is design matrix for Tag factor of the Phase 2 experiment and can be expressed as

$$\mathbf{X}_\gamma = \mathbf{I}_{n_\gamma} \otimes \mathbf{1}_{n_r}$$

Thus, projection matrix for Between Tags can be re-written as

$$\mathbf{P}_\gamma = \frac{1}{n_r} \mathbf{I}_{n_\gamma} \otimes \mathbf{1}_{n_r} \mathbf{1}'_{n_r}.$$

This follow by $\mathbf{P}_r \mathbf{P}_\gamma$ in B.1 becomes

$$\begin{aligned} \mathbf{P}_r \mathbf{P}_\gamma &= \frac{1}{n_r n_\gamma} \mathbf{1}_{n_\gamma} \mathbf{1}'_{n_\gamma} \otimes \mathbf{1}_{n_r} \mathbf{1}'_{n_r} \\ &= \frac{1}{n} \mathbf{1}_{n_\gamma n_r} \mathbf{1}'_{n_\gamma n_r} \\ &= \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n \\ &= \mathbf{K}_n. \end{aligned}$$

(B.3)

Thus, $\mathbf{P}_r \mathbf{P}_\gamma = \mathbf{K}$, then

$$\mathbf{Z}' \mathbf{P}_r \mathbf{P}_\gamma \mathbf{Z} = \frac{1}{n} \mathbf{Z}' \mathbf{1} \mathbf{1}' \mathbf{Z} = \frac{1}{n} (\mathbf{Z}' \mathbf{1}) (\mathbf{Z}' \mathbf{1})'.$$

Let

$$\mathbf{Z} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

then

$$\mathbf{Z}' \mathbf{1} = \begin{pmatrix} a_{.1} \\ a_{.2} \\ a_{.3} \end{pmatrix}.$$

So,

$$\begin{aligned}
 (\mathbf{Z}'\mathbf{1})(\mathbf{Z}'\mathbf{1})' &= \begin{pmatrix} a_{.1} \\ a_{.2} \\ a_{.3} \end{pmatrix} \begin{pmatrix} a_{.1} & a_{.2} & a_{.3} \end{pmatrix} \\
 &= \begin{pmatrix} a_{.1}^2 & a_{.1}a_{.2} & a_{.1}a_{.3} \\ a_{.2}a_{.1} & a_{.2}^2 & a_{.2}a_{.3} \\ a_{.3}a_{.1} & a_{.3}a_{.2} & a_{.3}^2 \end{pmatrix}.
 \end{aligned}$$

Thus, $\mathbf{Z}'\mathbf{P}_r\mathbf{P}_\gamma\mathbf{Z}$ is also symmetric.

Appendix C

Tables of optimal designs when Phase 1 is CRD

Two Treatments

Four-plex system

4 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Aa</i>	<i>Bb</i>	<i>Ca</i>	<i>Db</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Db</i>	<i>Ca</i>

8 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Ea</i>	<i>Db</i>	<i>Bb</i>	<i>Ca</i>
2	<i>Db</i>	<i>Ea</i>	<i>Ca</i>	<i>Bb</i>
3	<i>Hb</i>	<i>Ga</i>	<i>Fb</i>	<i>Aa</i>
4	<i>Ga</i>	<i>Hb</i>	<i>Aa</i>	<i>Fb</i>

10 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Hb</i>	<i>Fb</i>	<i>Ga</i>	<i>Ea</i>
2	<i>Fb</i>	<i>Hb</i>	<i>Ea</i>	<i>Ga</i>
3	<i>Bb</i>	<i>Aa</i>	<i>Ca</i>	<i>Db</i>
4	<i>Aa</i>	<i>Bb</i>	<i>Db</i>	<i>Ca</i>
5	<i>Ia</i>	<i>Ia</i>	<i>Jb</i>	<i>Jb</i>

12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Aa</i>	<i>Bb</i>	<i>Ca</i>	<i>Db</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Db</i>	<i>Ca</i>
3	<i>Ea</i>	<i>Fb</i>	<i>Ga</i>	<i>Hb</i>
4	<i>Fb</i>	<i>Ea</i>	<i>Hb</i>	<i>Ga</i>
5	<i>Ia</i>	<i>Jb</i>	<i>Ka</i>	<i>Lb</i>
6	<i>Jb</i>	<i>Ia</i>	<i>Lb</i>	<i>Ka</i>

Eight-plex system

8 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Aa</i>	<i>Bb</i>	<i>Ca</i>	<i>Db</i>	<i>Ea</i>	<i>Fb</i>	<i>Ga</i>	<i>Hb</i>
2	<i>Bb</i>	<i>Aa</i>	<i>Db</i>	<i>Ca</i>	<i>Fb</i>	<i>Ea</i>	<i>Hb</i>	<i>Ga</i>

12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Ca</i>	<i>Ia</i>	<i>Fb</i>	<i>Hb</i>	<i>Jb</i>	<i>Aa</i>	<i>Ea</i>	<i>Bb</i>
2	<i>Ia</i>	<i>Ca</i>	<i>Hb</i>	<i>Fb</i>	<i>Aa</i>	<i>Jb</i>	<i>Bb</i>	<i>Ea</i>
3	<i>Db</i>	<i>Db</i>	<i>Ga</i>	<i>Ga</i>	<i>Ka</i>	<i>Ka</i>	<i>Lb</i>	<i>Lb</i>

16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Ga</i>	<i>Ia</i>	<i>Jb</i>	<i>Aa</i>	<i>Ka</i>	<i>Bb</i>	<i>Pb</i>	<i>Fb</i>
2	<i>Ia</i>	<i>Ga</i>	<i>Aa</i>	<i>Jb</i>	<i>Bb</i>	<i>Ka</i>	<i>Fb</i>	<i>Pb</i>
3	<i>Nb</i>	<i>Db</i>	<i>Lb</i>	<i>Ca</i>	<i>Ma</i>	<i>Hb</i>	<i>Ea</i>	<i>Oa</i>
4	<i>Db</i>	<i>Nb</i>	<i>Ca</i>	<i>Lb</i>	<i>Hb</i>	<i>Ma</i>	<i>Oa</i>	<i>Ea</i>

14 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Jb</i>	<i>Ga</i>	<i>Ia</i>	<i>Db</i>
2	<i>Ga</i>	<i>Jb</i>	<i>Db</i>	<i>Ia</i>
3	<i>Bb</i>	<i>Ea</i>	<i>Lb</i>	<i>Aa</i>
4	<i>Ea</i>	<i>Bb</i>	<i>Aa</i>	<i>Lb</i>
5	<i>Fb</i>	<i>Ca</i>	<i>Nb</i>	<i>Ka</i>
6	<i>Ca</i>	<i>Fb</i>	<i>Ka</i>	<i>Nb</i>
7	<i>Hb</i>	<i>Hb</i>	<i>Ma</i>	<i>Ma</i>

16 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Bb</i>	<i>Aa</i>	<i>Lb</i>	<i>Ka</i>
2	<i>Aa</i>	<i>Bb</i>	<i>Ka</i>	<i>Lb</i>
3	<i>Ma</i>	<i>Hb</i>	<i>Fb</i>	<i>Ca</i>
4	<i>Hb</i>	<i>Ma</i>	<i>Ca</i>	<i>Fb</i>
5	<i>Ia</i>	<i>Ga</i>	<i>Pb</i>	<i>Nb</i>
6	<i>Ga</i>	<i>Ia</i>	<i>Nb</i>	<i>Pb</i>
7	<i>Jb</i>	<i>Db</i>	<i>Oa</i>	<i>Ea</i>
8	<i>Db</i>	<i>Jb</i>	<i>Ea</i>	<i>Oa</i>

Three Treatments

Four-plex system

6 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Bb</i>	<i>Da</i>	<i>Aa</i>	<i>Cc</i>
2	<i>Da</i>	<i>Bb</i>	<i>Cc</i>	<i>Aa</i>
3	<i>Fc</i>	<i>Fc</i>	<i>Eb</i>	<i>Eb</i>

12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Da</i>	<i>Eb</i>	<i>Kb</i>	<i>Lc</i>
2	<i>Eb</i>	<i>Da</i>	<i>Lc</i>	<i>Kb</i>
3	<i>Hb</i>	<i>Ja</i>	<i>Ga</i>	<i>Ic</i>
4	<i>Ja</i>	<i>Hb</i>	<i>Ic</i>	<i>Ga</i>
5	<i>Fc</i>	<i>Cc</i>	<i>Aa</i>	<i>Bb</i>
6	<i>Cc</i>	<i>Fc</i>	<i>Bb</i>	<i>Aa</i>

18 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Ja</i>	<i>Ga</i>	<i>Hb</i>	<i>Lc</i>
2	<i>Ga</i>	<i>Ja</i>	<i>Lc</i>	<i>Hb</i>
3	<i>Kb</i>	<i>Ic</i>	<i>Da</i>	<i>Cc</i>
4	<i>Ic</i>	<i>Kb</i>	<i>Cc</i>	<i>Da</i>
5	<i>Eb</i>	<i>Aa</i>	<i>Bb</i>	<i>Fc</i>
6	<i>Aa</i>	<i>Eb</i>	<i>Fc</i>	<i>Bb</i>
7	<i>Nb</i>	<i>Oc</i>	<i>Pa</i>	<i>Ma</i>
8	<i>Oc</i>	<i>Nb</i>	<i>Ma</i>	<i>Pa</i>
9	<i>Rc</i>	<i>Rc</i>	<i>Qb</i>	<i>Qb</i>

Eight-plex system

12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Cc</i>	<i>Bb</i>	<i>Ja</i>	<i>Eb</i>	<i>Lc</i>	<i>Kb</i>	<i>Aa</i>	<i>Fc</i>
2	<i>Bb</i>	<i>Cc</i>	<i>Eb</i>	<i>Ja</i>	<i>Kb</i>	<i>Lc</i>	<i>Fc</i>	<i>Aa</i>
3	<i>Ga</i>	<i>Ga</i>	<i>Ic</i>	<i>Ic</i>	<i>Da</i>	<i>Da</i>	<i>Hb</i>	<i>Hb</i>

18 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Va</i>	<i>Rc</i>	<i>Kb</i>	<i>Fc</i>	<i>Hb</i>	<i>Aa</i>	<i>Ga</i>	<i>Nb</i>
2	<i>Rc</i>	<i>Va</i>	<i>Fc</i>	<i>Kb</i>	<i>Aa</i>	<i>Hb</i>	<i>Nb</i>	<i>Ga</i>
3	<i>Pa</i>	<i>Bb</i>	<i>Eb</i>	<i>Da</i>	<i>Oc</i>	<i>Wb</i>	<i>Xc</i>	<i>Cc</i>
4	<i>Bb</i>	<i>Pa</i>	<i>Da</i>	<i>Eb</i>	<i>Wb</i>	<i>Oc</i>	<i>Cc</i>	<i>Xc</i>
5	<i>Tb</i>	<i>Uc</i>	<i>Ic</i>	<i>Ja</i>	<i>Sa</i>	<i>Lc</i>	<i>Qb</i>	<i>Ma</i>
6	<i>Uc</i>	<i>Tb</i>	<i>Ja</i>	<i>Ic</i>	<i>Lc</i>	<i>Sa</i>	<i>Ma</i>	<i>Qb</i>

24 Phase 1 experimental units

Run	Tag			
	114	115	116	117
1	<i>Ga</i>	<i>Nb</i>	<i>Fc</i>	<i>Ja</i>
2	<i>Nb</i>	<i>Ga</i>	<i>Ja</i>	<i>Fc</i>
3	<i>Lc</i>	<i>Da</i>	<i>Ic</i>	<i>Qb</i>
4	<i>Da</i>	<i>Lc</i>	<i>Qb</i>	<i>Ic</i>
5	<i>Ma</i>	<i>Hb</i>	<i>Aa</i>	<i>Oc</i>
6	<i>Hb</i>	<i>Ma</i>	<i>Oc</i>	<i>Aa</i>
7	<i>Xc</i>	<i>Rc</i>	<i>Sa</i>	<i>Wb</i>
8	<i>Rc</i>	<i>Xc</i>	<i>Wb</i>	<i>Sa</i>
9	<i>Kb</i>	<i>Bb</i>	<i>Va</i>	<i>Uc</i>
10	<i>Bb</i>	<i>Kb</i>	<i>Uc</i>	<i>Va</i>
11	<i>Cc</i>	<i>Pa</i>	<i>Eb</i>	<i>Tb</i>
12	<i>Pa</i>	<i>Cc</i>	<i>Tb</i>	<i>Eb</i>

Four Treatments

Four-plex system

8 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Ea</i>	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>
2	<i>Bb</i>	<i>Ea</i>	<i>Dd</i>	<i>Cc</i>
3	<i>Gc</i>	<i>Hd</i>	<i>Aa</i>	<i>Fb</i>
4	<i>Hd</i>	<i>Gc</i>	<i>Fb</i>	<i>Aa</i>

12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Bb</i>	<i>Kc</i>	<i>la</i>	<i>Hd</i>
2	<i>Kc</i>	<i>Bb</i>	<i>Hd</i>	<i>la</i>
3	<i>Dd</i>	<i>Aa</i>	<i>Cc</i>	<i>Fb</i>
4	<i>Aa</i>	<i>Dd</i>	<i>Fb</i>	<i>Cc</i>
5	<i>Ea</i>	<i>Jb</i>	<i>Ld</i>	<i>Gc</i>
6	<i>Jb</i>	<i>Ea</i>	<i>Gc</i>	<i>Ld</i>

16 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Dd</i>	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>
2	<i>Aa</i>	<i>Dd</i>	<i>Cc</i>	<i>Bb</i>
3	<i>Gc</i>	<i>Fb</i>	<i>Hd</i>	<i>Ea</i>
4	<i>Fb</i>	<i>Gc</i>	<i>Ea</i>	<i>Hd</i>
5	<i>Kc</i>	<i>Jb</i>	<i>Ld</i>	<i>la</i>
6	<i>Jb</i>	<i>Kc</i>	<i>la</i>	<i>Ld</i>
7	<i>Ma</i>	<i>Pd</i>	<i>Oc</i>	<i>Nb</i>
8	<i>Pd</i>	<i>Ma</i>	<i>Nb</i>	<i>Oc</i>

20 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Kc</i>	<i>Rb</i>	<i>Aa</i>	<i>Ld</i>
2	<i>Rb</i>	<i>Kc</i>	<i>Ld</i>	<i>Aa</i>
3	<i>Jb</i>	<i>Td</i>	<i>Oc</i>	<i>Ea</i>
4	<i>Td</i>	<i>Jb</i>	<i>Ea</i>	<i>Oc</i>
5	<i>Hd</i>	<i>Gc</i>	<i>Fb</i>	<i>la</i>
6	<i>Gc</i>	<i>Hd</i>	<i>la</i>	<i>Fb</i>
7	<i>Qa</i>	<i>Pd</i>	<i>Sc</i>	<i>Bb</i>
8	<i>Pd</i>	<i>Qa</i>	<i>Bb</i>	<i>Sc</i>
9	<i>Ma</i>	<i>Nb</i>	<i>Cc</i>	<i>Dd</i>
10	<i>Nb</i>	<i>Ma</i>	<i>Dd</i>	<i>Cc</i>

24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Cc</i>	<i>Bb</i>	<i>Aa</i>	<i>Dd</i>
2	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>	<i>Aa</i>
3	<i>Hd</i>	<i>Ea</i>	<i>Fb</i>	<i>Gc</i>
4	<i>Ea</i>	<i>Hd</i>	<i>Gc</i>	<i>Fb</i>
5	<i>Jb</i>	<i>Kc</i>	<i>Ld</i>	<i>la</i>
6	<i>Kc</i>	<i>Jb</i>	<i>la</i>	<i>Ld</i>
7	<i>Ma</i>	<i>Pd</i>	<i>Oc</i>	<i>Nb</i>
8	<i>Pd</i>	<i>Ma</i>	<i>Nb</i>	<i>Oc</i>
9	<i>Sc</i>	<i>Qa</i>	<i>Td</i>	<i>Rb</i>
10	<i>Qa</i>	<i>Sc</i>	<i>Rb</i>	<i>Td</i>
11	<i>Vb</i>	<i>Xd</i>	<i>Ua</i>	<i>Wc</i>
12	<i>Xd</i>	<i>Vb</i>	<i>Wc</i>	<i>Ua</i>

28 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Ala</i>	<i>ALd</i>	<i>AOc</i>	<i>ARb</i>
2	<i>ALd</i>	<i>Ala</i>	<i>ARb</i>	<i>AOc</i>
3	<i>AHd</i>	<i>BAc</i>	<i>AEa</i>	<i>AZb</i>
4	<i>BAc</i>	<i>AHd</i>	<i>AZb</i>	<i>AEa</i>
5	<i>AGc</i>	<i>ATd</i>	<i>ANb</i>	<i>AYa</i>
6	<i>ATd</i>	<i>AGc</i>	<i>AYa</i>	<i>ANb</i>
7	<i>ACc</i>	<i>AVb</i>	<i>AUa</i>	<i>AXd</i>
8	<i>AVb</i>	<i>ACc</i>	<i>AXd</i>	<i>AUa</i>
9	<i>AJb</i>	<i>ADd</i>	<i>ASc</i>	<i>AAa</i>
10	<i>ADd</i>	<i>AJb</i>	<i>AAa</i>	<i>ASc</i>
11	<i>AQa</i>	<i>ABb</i>	<i>APd</i>	<i>AWc</i>
12	<i>ABb</i>	<i>AQa</i>	<i>AWc</i>	<i>APd</i>
13	<i>AKc</i>	<i>AMa</i>	<i>AFb</i>	<i>BBd</i>
14	<i>AMa</i>	<i>AKc</i>	<i>BBd</i>	<i>AFb</i>

32 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>ACc</i>	<i>ADd</i>	<i>AAa</i>	<i>ABb</i>
2	<i>ADd</i>	<i>ACc</i>	<i>ABb</i>	<i>AAa</i>
3	<i>AGc</i>	<i>AEa</i>	<i>AHd</i>	<i>AFb</i>
4	<i>AEa</i>	<i>AGc</i>	<i>AFb</i>	<i>AHd</i>
5	<i>ALd</i>	<i>AJb</i>	<i>AKc</i>	<i>Ala</i>
6	<i>AJb</i>	<i>ALd</i>	<i>Ala</i>	<i>AKc</i>
7	<i>APd</i>	<i>AMa</i>	<i>AOc</i>	<i>ANb</i>
8	<i>AMa</i>	<i>APd</i>	<i>ANb</i>	<i>AOc</i>
9	<i>AQa</i>	<i>ASc</i>	<i>ARb</i>	<i>ATd</i>
10	<i>ASc</i>	<i>AQa</i>	<i>ATd</i>	<i>ARb</i>
11	<i>AVb</i>	<i>AUa</i>	<i>AWc</i>	<i>AXd</i>
12	<i>AUa</i>	<i>AVb</i>	<i>AXd</i>	<i>AWc</i>
13	<i>BBd</i>	<i>AZb</i>	<i>BAc</i>	<i>AYa</i>
14	<i>AZb</i>	<i>BBd</i>	<i>AYa</i>	<i>BAc</i>
15	<i>BDb</i>	<i>BEc</i>	<i>BCa</i>	<i>BFd</i>
16	<i>BEc</i>	<i>BDb</i>	<i>BFd</i>	<i>BCa</i>

Eight-plex system

8 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Aa</i>	<i>Gc</i>	<i>Cc</i>	<i>Fb</i>	<i>Hd</i>	<i>Bb</i>	<i>Ea</i>	<i>Dd</i>
2	<i>Gc</i>	<i>Aa</i>	<i>Fb</i>	<i>Cc</i>	<i>Bb</i>	<i>Hd</i>	<i>Dd</i>	<i>Ea</i>

12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Bb</i>	<i>Ld</i>	<i>Ea</i>	<i>Jb</i>	<i>Aa</i>	<i>Cc</i>	<i>Hd</i>	<i>Kc</i>
2	<i>Ld</i>	<i>Bb</i>	<i>Jb</i>	<i>Ea</i>	<i>Cc</i>	<i>Aa</i>	<i>Kc</i>	<i>Hd</i>
3	<i>Gc</i>	<i>Gc</i>	<i>Dd</i>	<i>Dd</i>	<i>Fb</i>	<i>Fb</i>	<i>la</i>	<i>la</i>

16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Kc</i>	<i>Ld</i>	<i>Jb</i>	<i>Dd</i>	<i>Fb</i>	<i>Ea</i>	<i>Gc</i>	<i>Ma</i>
2	<i>Ld</i>	<i>Kc</i>	<i>Dd</i>	<i>Jb</i>	<i>Ea</i>	<i>Fb</i>	<i>Ma</i>	<i>Gc</i>
3	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>	<i>la</i>	<i>Oc</i>	<i>Pd</i>	<i>Nb</i>	<i>Hd</i>
4	<i>Bb</i>	<i>Aa</i>	<i>la</i>	<i>Cc</i>	<i>Pd</i>	<i>Oc</i>	<i>Hd</i>	<i>Nb</i>

20 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Cc</i>	<i>Qa</i>	<i>la</i>	<i>Fb</i>	<i>Sc</i>	<i>Td</i>	<i>Rb</i>	<i>Ld</i>
2	<i>Qa</i>	<i>Cc</i>	<i>Fb</i>	<i>la</i>	<i>Td</i>	<i>Sc</i>	<i>Ld</i>	<i>Rb</i>
3	<i>Dd</i>	<i>Jb</i>	<i>Kc</i>	<i>Ma</i>	<i>Bb</i>	<i>Aa</i>	<i>Pd</i>	<i>Gc</i>
4	<i>Jb</i>	<i>Dd</i>	<i>Ma</i>	<i>Kc</i>	<i>Aa</i>	<i>Bb</i>	<i>Gc</i>	<i>Pd</i>
5	<i>Oc</i>	<i>Oc</i>	<i>Hd</i>	<i>Hd</i>	<i>Nb</i>	<i>Nb</i>	<i>Ea</i>	<i>Ea</i>

24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Qa</i>	<i>Sc</i>	<i>Rb</i>	<i>Ld</i>	<i>Vb</i>	<i>Kc</i>	<i>Ma</i>	<i>Pd</i>
2	<i>Sc</i>	<i>Qa</i>	<i>Ld</i>	<i>Rb</i>	<i>Kc</i>	<i>Vb</i>	<i>Pd</i>	<i>Ma</i>
3	<i>Hd</i>	<i>Ua</i>	<i>Wc</i>	<i>Dd</i>	<i>Cc</i>	<i>Nb</i>	<i>Jb</i>	<i>Aa</i>
4	<i>Ua</i>	<i>Hd</i>	<i>Dd</i>	<i>Wc</i>	<i>Nb</i>	<i>Cc</i>	<i>Aa</i>	<i>Jb</i>
5	<i>Bb</i>	<i>Fb</i>	<i>Oc</i>	<i>Ea</i>	<i>Td</i>	<i>la</i>	<i>Xd</i>	<i>Gc</i>
6	<i>Fb</i>	<i>Bb</i>	<i>Ea</i>	<i>Oc</i>	<i>la</i>	<i>Td</i>	<i>Gc</i>	<i>Xd</i>

28 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>AHd</i>	<i>AVb</i>	<i>AZb</i>	<i>AGc</i>	<i>APd</i>	<i>AWc</i>	<i>AYa</i>	<i>AEa</i>
2	<i>AVb</i>	<i>AHd</i>	<i>AGc</i>	<i>AZb</i>	<i>AWc</i>	<i>APd</i>	<i>AEa</i>	<i>AYa</i>
3	<i>AJb</i>	<i>AUa</i>	<i>BAc</i>	<i>Ala</i>	<i>AFb</i>	<i>ADd</i>	<i>ASc</i>	<i>AXd</i>
4	<i>AUa</i>	<i>AJb</i>	<i>Ala</i>	<i>BAc</i>	<i>ADd</i>	<i>AFb</i>	<i>AXd</i>	<i>ASc</i>
5	<i>ALd</i>	<i>AAa</i>	<i>ARb</i>	<i>BBd</i>	<i>AKc</i>	<i>AMa</i>	<i>ABb</i>	<i>AOc</i>
6	<i>AAa</i>	<i>ALd</i>	<i>BBd</i>	<i>ARb</i>	<i>AMa</i>	<i>AKc</i>	<i>AOc</i>	<i>ABb</i>
7	<i>ACc</i>	<i>ACc</i>	<i>AQa</i>	<i>AQa</i>	<i>ANb</i>	<i>ANb</i>	<i>ATd</i>	<i>ATd</i>

32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	AAa	ABb	ACc	ADd	AEa	AFb	AGc	AHd
2	ABb	AAa	ADd	ACc	AFb	AEa	AHd	AGc
3	Ala	AJb	AKc	ALd	AMa	ANb	AOc	APd
4	AJb	Ala	ALd	AKc	ANb	AMa	APd	AOc
5	ASc	ATd	AQa	ARb	AWc	AXd	AUa	AVb
6	ATd	ASc	ARb	AQa	AXd	AWc	AVb	AUa
7	BAC	BBd	AYa	AZb	BEc	BFd	BCa	BDb
8	BBd	BAC	AZb	AYa	BFd	BEc	BDb	BCa

Five Treatments

Four-plex system

10 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	Gb	Cc	Aa	Dd
2	Cc	Gb	Dd	Aa
3	Je	Fa	Bb	Hc
4	Fa	Je	Hc	Bb
5	Id	Id	Ee	Ee

20 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	Te	Rc	Dd	Bb
2	Rc	Te	Bb	Dd
3	Gb	Sd	Mc	Fa
4	Sd	Gb	Fa	Mc
5	Cc	Ka	Lb	Oe
6	Ka	Cc	Oe	Lb
7	Qb	Nd	Aa	Je
8	Nd	Qb	Je	Aa
9	Pa	Ee	Id	Hc
10	Ee	Pa	Hc	Id

30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	BBc	AEe	AZa	AGb
2	AEe	BBc	AGb	AZa
3	AKa	ASd	AMc	ATe
4	ASd	AKa	ATe	AMc
5	AYe	ACc	AFa	ABb
6	ACc	AYe	ABb	AFa
7	BDe	APa	AXd	BAb
8	APa	BDe	BAb	AXd
9	AUa	AVb	AJe	ADd
10	AVb	AUa	ADd	AJe
11	AQb	ARc	BCd	AOe
12	ARc	AQb	AOe	BCd
13	Ala	ALb	AHc	AAa
14	ALb	Ala	AAa	AHc
15	ANd	ANd	AWc	AWc

40 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	Ala	AFa	ABb	ACc
2	AFa	Ala	ACc	ABb
3	BLc	BJa	BAb	AOe
4	BJa	BLc	AOe	BAb
5	BDe	BBc	BCd	AUa
6	BBc	BDe	AUa	BCd
7	ADd	AWc	AEe	BFb
8	AWc	ADd	BFb	AEe
9	Ble	BKb	BEa	AMc
10	BKb	Ble	AMc	BEa
11	BHd	AQb	BNe	AHc
12	AQb	BHd	AHc	BNe
13	AGb	ASd	BGc	APa
14	ASd	AGb	APa	BGc
15	AAa	ARc	AXd	AJe
16	ARc	AAa	AJe	AXd
17	ATe	ALb	AKa	BMd
18	ALb	ATe	BMd	AKa
19	AZa	AYe	ANd	AVb
20	AYe	AZa	AVb	ANd

Eight-plex system

20 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Ka</i>	<i>Cc</i>	<i>Ee</i>	<i>Sd</i>	<i>Je</i>	<i>Mc</i>	<i>Fa</i>	<i>Qb</i>
2	<i>Cc</i>	<i>Ka</i>	<i>Sd</i>	<i>Ee</i>	<i>Mc</i>	<i>Je</i>	<i>Qb</i>	<i>Fa</i>
3	<i>Dd</i>	<i>Lb</i>	<i>Bb</i>	<i>Rc</i>	<i>Id</i>	<i>Pa</i>	<i>Te</i>	<i>Hc</i>
4	<i>Lb</i>	<i>Dd</i>	<i>Rc</i>	<i>Bb</i>	<i>Pa</i>	<i>Id</i>	<i>Hc</i>	<i>Te</i>
5	<i>Oe</i>	<i>Oe</i>	<i>Aa</i>	<i>Aa</i>	<i>Gb</i>	<i>Gb</i>	<i>Nd</i>	<i>Nd</i>

40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>BLc</i>	<i>BNe</i>	<i>AMc</i>	<i>ADd</i>	<i>BKb</i>	<i>AAa</i>	<i>AOe</i>	<i>AUa</i>
2	<i>BNe</i>	<i>BLc</i>	<i>ADd</i>	<i>AMc</i>	<i>AAa</i>	<i>BKb</i>	<i>AUa</i>	<i>AOe</i>
3	<i>ASd</i>	<i>AVb</i>	<i>AYe</i>	<i>Ble</i>	<i>BBc</i>	<i>AQb</i>	<i>ANd</i>	<i>BJa</i>
4	<i>AVb</i>	<i>ASd</i>	<i>Ble</i>	<i>AYe</i>	<i>AQb</i>	<i>BBc</i>	<i>BJa</i>	<i>ANd</i>
5	<i>AWc</i>	<i>AKa</i>	<i>BMd</i>	<i>AGb</i>	<i>AXd</i>	<i>BDe</i>	<i>ALb</i>	<i>AHc</i>
6	<i>AKa</i>	<i>AWc</i>	<i>AGb</i>	<i>BMd</i>	<i>BDe</i>	<i>AXd</i>	<i>AHc</i>	<i>ALb</i>
7	<i>AFa</i>	<i>BFb</i>	<i>APa</i>	<i>BGc</i>	<i>BHd</i>	<i>AJe</i>	<i>BAb</i>	<i>ACc</i>
8	<i>BFb</i>	<i>AFa</i>	<i>BGc</i>	<i>APa</i>	<i>AJe</i>	<i>BHd</i>	<i>ACc</i>	<i>BAb</i>
9	<i>ATe</i>	<i>BCd</i>	<i>ABb</i>	<i>BEa</i>	<i>AZa</i>	<i>ARc</i>	<i>AEe</i>	<i>Ald</i>
10	<i>BCd</i>	<i>ATe</i>	<i>BEa</i>	<i>ABb</i>	<i>ARc</i>	<i>AZa</i>	<i>Ald</i>	<i>AEe</i>

Six Treatments

Four-plex iTRAQ system

12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Hb</i>	<i>Cc</i>	<i>Aa</i>	<i>Lf</i>
2	<i>Cc</i>	<i>Hb</i>	<i>Lf</i>	<i>Aa</i>
3	<i>Ff</i>	<i>Jd</i>	<i>Ke</i>	<i>Bb</i>
4	<i>Jd</i>	<i>Ff</i>	<i>Bb</i>	<i>Ke</i>
5	<i>Ga</i>	<i>Ee</i>	<i>Ic</i>	<i>Dd</i>
6	<i>Ee</i>	<i>Ga</i>	<i>Dd</i>	<i>Ic</i>

18 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Jd</i>	<i>Oc</i>	<i>Ke</i>	<i>Lf</i>
2	<i>Oc</i>	<i>Jd</i>	<i>Lf</i>	<i>Ke</i>
3	<i>Aa</i>	<i>Bb</i>	<i>Cc</i>	<i>Rf</i>
4	<i>Bb</i>	<i>Aa</i>	<i>Rf</i>	<i>Cc</i>
5	<i>Ee</i>	<i>Ff</i>	<i>Ma</i>	<i>Nb</i>
6	<i>Ff</i>	<i>Ee</i>	<i>Nb</i>	<i>Ma</i>
7	<i>Qe</i>	<i>Ic</i>	<i>Ga</i>	<i>Pd</i>
8	<i>Ic</i>	<i>Qe</i>	<i>Pd</i>	<i>Ga</i>
9	<i>Hb</i>	<i>Hb</i>	<i>Dd</i>	<i>Dd</i>

24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Jd</i>	<i>Uc</i>	<i>Ga</i>	<i>Nb</i>
2	<i>Uc</i>	<i>Jd</i>	<i>Nb</i>	<i>Ga</i>
3	<i>Bb</i>	<i>Dd</i>	<i>Cc</i>	<i>Xf</i>
4	<i>Dd</i>	<i>Bb</i>	<i>Xf</i>	<i>Cc</i>
5	<i>Ee</i>	<i>Lf</i>	<i>Hb</i>	<i>Pd</i>
6	<i>Lf</i>	<i>Ee</i>	<i>Pd</i>	<i>Hb</i>
7	<i>Ke</i>	<i>Aa</i>	<i>Vd</i>	<i>Rf</i>
8	<i>Aa</i>	<i>Ke</i>	<i>Rf</i>	<i>Vd</i>
9	<i>Oc</i>	<i>Ff</i>	<i>We</i>	<i>Sa</i>
10	<i>Ff</i>	<i>Oc</i>	<i>Sa</i>	<i>We</i>
11	<i>Tb</i>	<i>Ma</i>	<i>Qe</i>	<i>Ic</i>
12	<i>Ma</i>	<i>Tb</i>	<i>Ic</i>	<i>Qe</i>

30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>AWe</i>	<i>BDf</i>	<i>Alc</i>	<i>ANb</i>
2	<i>BDf</i>	<i>AWe</i>	<i>ANb</i>	<i>Alc</i>
3	<i>AMa</i>	<i>ACc</i>	<i>BBd</i>	<i>AXf</i>
4	<i>ACc</i>	<i>AMa</i>	<i>AXf</i>	<i>BBd</i>
5	<i>AVd</i>	<i>ARf</i>	<i>ASa</i>	<i>AKe</i>
6	<i>ARf</i>	<i>AVd</i>	<i>AKe</i>	<i>ASa</i>
7	<i>ATb</i>	<i>AYa</i>	<i>AEe</i>	<i>ADd</i>
8	<i>AYa</i>	<i>ATb</i>	<i>ADd</i>	<i>AEe</i>
9	<i>AGa</i>	<i>ABb</i>	<i>APd</i>	<i>AUc</i>
10	<i>ABb</i>	<i>AGa</i>	<i>AUc</i>	<i>APd</i>
11	<i>AQe</i>	<i>AJd</i>	<i>AHb</i>	<i>BAC</i>
12	<i>AJd</i>	<i>AQe</i>	<i>BAC</i>	<i>AHb</i>
13	<i>BCe</i>	<i>AOc</i>	<i>AAa</i>	<i>ALf</i>
14	<i>AOc</i>	<i>BCe</i>	<i>ALf</i>	<i>AAa</i>
15	<i>AZb</i>	<i>AZb</i>	<i>AFf</i>	<i>AFf</i>

36 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	BHd	AMa	AUc	AHb
2	AMa	BHd	AHb	AUc
3	BGc	AZb	AKe	AGa
4	AZb	BGc	AGa	AKe
5	ASa	Alc	AXf	Ble
6	Alc	ASa	Ble	AXf
7	AYa	ACc	BDf	APd
8	ACc	AYa	APd	BDf
9	ARf	ATb	BBd	BCe
10	ATb	ARf	BCe	BBd
11	AWe	ALf	AAa	AVd
12	ALf	AWe	AVd	AAa
13	AFf	AQe	BAc	BFb
14	AQe	AFf	BFb	BAc
15	ADd	AEe	ABb	BEa
16	AEe	ADd	BEa	ABb
17	AJd	ANb	BJf	AOc
18	ANb	AJd	AOc	BJf

42 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	BFb	AQe	AMa	ACc
2	AQe	BFb	ACc	AMa
3	AEe	BLb	AJd	ASa
4	BLb	AEe	ASa	AJd
5	AXf	BBd	AUc	AZb
6	BBd	AXf	AZb	AUc
7	AOc	AHb	AKe	ARf
8	AHb	AOc	ARf	AKe
9	AVd	ANb	AFf	BOe
10	ANb	AVd	BOe	AFf
11	ADd	AGa	BMc	Ble
12	AGa	ADd	Ble	BMc
13	AYa	BAc	ALf	ATb
14	BAc	AYa	ATb	ALf
15	AAa	BCe	Alc	BJf
16	BCe	AAa	BJf	Alc
17	APd	BPf	ABb	BKa
18	BPf	APd	BKa	ABb
19	BEa	BDf	BNd	AWe
20	BDf	BEa	AWe	BNd
21	BGc	BGc	BHd	BHd

48 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	APd	BAc	BRb	BUe
2	BAc	APd	BUe	BRb
3	ARf	AQe	BMc	BFb
4	AQe	ARf	BFb	BMc
5	BTd	BKa	BVf	ACc
6	BKa	BTd	ACc	BVf
7	BSc	AVd	BQa	AKe
8	AVd	BSc	AKe	BQa
9	BEa	AUc	BOe	AFf
10	AUc	BEa	AFf	BOe
11	BDf	AHb	ADd	BGc
12	AHb	BDf	BGc	ADd
13	AOc	AYa	BHd	AZb
14	AYa	AOc	AZb	BHd
15	AMa	BPf	ABb	BNd
16	BPf	AMa	BNd	ABb
17	BLb	AEe	Alc	AAa
18	AEe	BLb	AAa	Alc
19	AWe	ANb	AJd	ALf
20	ANb	AWe	ALf	AJd
21	BCe	BBd	AGa	BJf
22	BBd	BCe	BJf	AGa
23	AXf	ATb	ASa	Ble
24	ATb	AXf	Ble	ASa

Eight-plex system

24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Qe</i>	<i>Ic</i>	<i>Lf</i>	<i>Ke</i>	<i>Vd</i>	<i>Ma</i>	<i>Oc</i>	<i>Hb</i>
2	<i>Ic</i>	<i>Qe</i>	<i>Ke</i>	<i>Lf</i>	<i>Ma</i>	<i>Vd</i>	<i>Hb</i>	<i>Oc</i>
3	<i>Rf</i>	<i>Bb</i>	<i>Cc</i>	<i>Dd</i>	<i>Nb</i>	<i>Ee</i>	<i>Xf</i>	<i>Ga</i>
4	<i>Bb</i>	<i>Rf</i>	<i>Dd</i>	<i>Cc</i>	<i>Ee</i>	<i>Nb</i>	<i>Ga</i>	<i>Xf</i>
5	<i>Jd</i>	<i>Aa</i>	<i>Sa</i>	<i>Tb</i>	<i>Uc</i>	<i>Ff</i>	<i>We</i>	<i>Pd</i>
6	<i>Aa</i>	<i>Jd</i>	<i>Tb</i>	<i>Sa</i>	<i>Ff</i>	<i>Uc</i>	<i>Pd</i>	<i>We</i>

36 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>ARf</i>	<i>BEa</i>	<i>AJd</i>	<i>Alc</i>	<i>AQe</i>	<i>AGa</i>	<i>ALf</i>	<i>AZb</i>
2	<i>BEa</i>	<i>ARf</i>	<i>Alc</i>	<i>AJd</i>	<i>AGa</i>	<i>AQe</i>	<i>AZb</i>	<i>ALf</i>
3	<i>AVd</i>	<i>AKe</i>	<i>AYa</i>	<i>BDf</i>	<i>ABb</i>	<i>BAC</i>	<i>AUc</i>	<i>AWe</i>
4	<i>AKe</i>	<i>AVd</i>	<i>BDf</i>	<i>AYa</i>	<i>BAC</i>	<i>ABb</i>	<i>AWe</i>	<i>AUc</i>
5	<i>BCe</i>	<i>BJf</i>	<i>ACc</i>	<i>BFb</i>	<i>AOc</i>	<i>ADd</i>	<i>AAa</i>	<i>BHd</i>
6	<i>BJf</i>	<i>BCe</i>	<i>BFb</i>	<i>ACc</i>	<i>ADd</i>	<i>AOc</i>	<i>BHd</i>	<i>AAa</i>
7	<i>AMa</i>	<i>BGc</i>	<i>ANb</i>	<i>AEe</i>	<i>AFf</i>	<i>AHb</i>	<i>ASa</i>	<i>BBd</i>
8	<i>BGc</i>	<i>AMa</i>	<i>AEe</i>	<i>ANb</i>	<i>AHb</i>	<i>AFf</i>	<i>BBd</i>	<i>ASa</i>
9	<i>ATb</i>	<i>ATb</i>	<i>APd</i>	<i>APd</i>	<i>AXf</i>	<i>AXf</i>	<i>Ble</i>	<i>Ble</i>

48 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>APd</i>	<i>AMa</i>	<i>BCe</i>	<i>AHb</i>	<i>BOe</i>	<i>Alc</i>	<i>BNd</i>	<i>ALf</i>
2	<i>AMa</i>	<i>APd</i>	<i>AHb</i>	<i>BCe</i>	<i>Alc</i>	<i>BOe</i>	<i>ALf</i>	<i>BNd</i>
3	<i>BVf</i>	<i>BFb</i>	<i>BGc</i>	<i>AVd</i>	<i>AFf</i>	<i>BAC</i>	<i>AAa</i>	<i>AQe</i>
4	<i>BFb</i>	<i>BVf</i>	<i>AVd</i>	<i>BGc</i>	<i>BAC</i>	<i>AFf</i>	<i>AQe</i>	<i>AAa</i>
5	<i>BPf</i>	<i>AJd</i>	<i>ARf</i>	<i>AZb</i>	<i>ANb</i>	<i>AEe</i>	<i>AUc</i>	<i>ASa</i>
6	<i>AJd</i>	<i>BPf</i>	<i>AZb</i>	<i>ARf</i>	<i>AEe</i>	<i>ANb</i>	<i>ASa</i>	<i>AUc</i>
7	<i>BSc</i>	<i>Ble</i>	<i>ACc</i>	<i>AGa</i>	<i>AXf</i>	<i>BKa</i>	<i>BLb</i>	<i>BBd</i>
8	<i>Ble</i>	<i>BSc</i>	<i>AGa</i>	<i>ACc</i>	<i>BKa</i>	<i>AXf</i>	<i>BBd</i>	<i>BLb</i>
9	<i>BRb</i>	<i>BEa</i>	<i>BUe</i>	<i>ADd</i>	<i>BHd</i>	<i>ABb</i>	<i>AOc</i>	<i>BJf</i>
10	<i>BEa</i>	<i>BRb</i>	<i>ADd</i>	<i>BUe</i>	<i>ABb</i>	<i>BHd</i>	<i>BJf</i>	<i>AOc</i>
11	<i>AKe</i>	<i>BMc</i>	<i>BDf</i>	<i>BQa</i>	<i>AYa</i>	<i>BTd</i>	<i>AWe</i>	<i>ATb</i>
12	<i>BMc</i>	<i>AKe</i>	<i>BQa</i>	<i>BDf</i>	<i>BTd</i>	<i>AYa</i>	<i>ATb</i>	<i>AWe</i>

Seven Treatments

Four-plex system

14 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Aa</i>	<i>Dd</i>	<i>Jc</i>	<i>Ib</i>
2	<i>Dd</i>	<i>Aa</i>	<i>Ib</i>	<i>Jc</i>
3	<i>Ff</i>	<i>Ng</i>	<i>Ee</i>	<i>Kd</i>
4	<i>Ng</i>	<i>Ff</i>	<i>Kd</i>	<i>Ee</i>
5	<i>Le</i>	<i>Cc</i>	<i>Ha</i>	<i>Mf</i>
6	<i>Cc</i>	<i>Le</i>	<i>Mf</i>	<i>Ha</i>
7	<i>Bb</i>	<i>Bb</i>	<i>Gg</i>	<i>Gg</i>

28 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>AAa</i>	<i>AJc</i>	<i>ALe</i>	<i>ANg</i>
2	<i>AJc</i>	<i>AAa</i>	<i>ANg</i>	<i>ALe</i>
3	<i>AUg</i>	<i>AKd</i>	<i>BAf</i>	<i>AVa</i>
4	<i>AKd</i>	<i>AUg</i>	<i>AVa</i>	<i>BAf</i>
5	<i>AHa</i>	<i>AWb</i>	<i>AZe</i>	<i>AMf</i>
6	<i>AWb</i>	<i>AHa</i>	<i>AMf</i>	<i>AZe</i>
7	<i>AQc</i>	<i>ADd</i>	<i>AOa</i>	<i>ABb</i>
8	<i>ADd</i>	<i>AQc</i>	<i>ABb</i>	<i>AOa</i>
9	<i>Alb</i>	<i>ASe</i>	<i>AGg</i>	<i>ARd</i>
10	<i>ASe</i>	<i>Alb</i>	<i>ARd</i>	<i>AGg</i>
11	<i>BBg</i>	<i>AFf</i>	<i>ACc</i>	<i>APb</i>
12	<i>AFf</i>	<i>BBg</i>	<i>APb</i>	<i>ACc</i>
13	<i>AEe</i>	<i>ATf</i>	<i>AYd</i>	<i>AXc</i>
14	<i>ATf</i>	<i>AEe</i>	<i>AXc</i>	<i>AYd</i>

42 Phase 1 experimental units

Run	Tag			
	114	115	116	117
1	APb	AHa	AMf	AXc
2	AHa	APb	AXc	AMf
3	BLc	AKd	AOa	AGg
4	AKd	BLc	AGg	AOa
5	ATf	BNe	ACc	BCa
6	BNe	ATf	BCa	ACc
7	AZe	BBg	BMd	AVa
8	BBg	AZe	AVa	BMd
9	AAa	BFd	AWb	BHf
10	BFd	AAa	BHf	AWb
11	BDb	ANg	AQc	ARd
12	ANg	BDb	ARd	AQc
13	BGe	BJa	Alb	AUg
14	BJa	BGe	AUg	Alb
15	Blg	BEc	AEe	BKb
16	BEc	Blg	BKb	AEe
17	AYd	ABb	ALe	AFf
18	ABb	AYd	AFf	ALe
19	BAf	AJc	ADd	ASe
20	AJc	BAf	ADd	ASe
21	BOf	BOf	BPg	BPg

56 Phase 1 experimental units

Run	Tag			
	114	115	116	117
1	AXc	BGe	BBg	Alb
2	BGe	AXc	Alb	BBg
3	BFd	BVf	BNe	BDb
4	BVf	BFd	BDb	BNe
5	AKd	CCf	Blg	BUe
6	CCf	AKd	BUe	Blg
7	AUg	AOa	CBe	BOf
8	AOa	AUg	BOf	CBe
9	AZe	APb	BJa	BMd
10	APb	AZe	BMd	BJa
11	BCa	ANg	BKb	BTd
12	ANg	BCa	BTd	BKb
13	CAd	BRb	AJc	AMf
14	BRb	CAd	AMf	AJc
15	AFf	ABb	AAa	BPg
16	ABb	AFf	BPg	AAa
17	BYb	AEe	BZc	BWg
18	AEe	BYb	BWg	BZc
19	CDg	ACc	BHf	AYd
20	ACc	CDg	AYd	BHf
21	BSc	BAf	AHa	ASe
22	BAf	BSc	ASe	AHa
23	BXa	AGg	ADd	BLc
24	AGg	BXa	BLc	ADd
25	AQc	BQa	ATf	AWb
26	BQa	AQc	AWb	ATf
27	ALe	ARd	AVa	BEc
28	ARd	ALe	BEc	AVa

Eight-plex system

28 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	AGg	ABb	ALe	AJc	AOa	BAf	AYd	AVa
2	ABb	AGg	AJc	ALe	BAf	AOa	AVa	AYd
3	AQc	AEe	ATf	AHa	AKd	AZe	ANg	AWb
4	AEe	AQc	AHa	ATf	AZe	AKd	AWb	ANg
5	AMf	AAa	Alb	ARd	AXc	AUg	AFf	ASe
6	AAa	AMf	ARd	Alb	AUg	AXc	ASe	AFf
7	ADd	ADd	BBg	BBg	APb	APb	ACc	ACc

56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	ACc	AHa	AFf	AGg	ABb	ADd	AEe	AAa
2	AHa	ACc	AGg	AFf	ADd	ABb	AAa	AEe
3	ANg	AKd	APb	AMf	AOa	ALe	AJc	Alb
4	AKd	ANg	AMf	APb	ALe	AOa	Alb	AJc
5	ARd	AWb	AXc	AVa	ATf	ASe	AQc	AUg
6	AWb	ARd	AVa	AXc	ASe	ATf	AUg	AQc
7	BDb	AZe	AYd	BCa	BEc	BBg	BFd	BAf
8	AZe	BDb	BCa	AYd	BBg	BEc	BAf	BFd
9	BGe	BHf	Blg	BMd	BLc	BJa	BNe	BKb
10	BHf	BGe	BMd	Blg	BJa	BLc	BKb	BNe
11	BOf	BSc	BYb	CBe	BTd	CDg	BPg	BQa
12	BSc	BOf	CBe	BYb	CDg	BTd	BQa	BPg
13	BXa	BWg	BZc	BUe	CCf	BRb	CAd	BVf
14	BWg	BXa	BUe	BZc	BRb	CCf	BVf	CAd

Eight Treatments

Four-plex system

16 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	Og	la	Ee	Hh
2	la	Og	Hh	Ee
3	Jb	Me	Dd	Ff
4	Me	Jb	Ff	Dd
5	Ph	Nf	Kc	Gg
6	Nf	Ph	Gg	Kc
7	Ld	Cc	Aa	Bb
8	Cc	Ld	Bb	Aa

24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	Bb	Vf	Ue	Wg
2	Vf	Bb	Wg	Ue
3	Ld	Me	Sc	Gg
4	Me	Ld	Gg	Sc
5	Qa	Cc	Jb	Td
6	Cc	Qa	Td	Jb
7	Rb	Ph	Kc	Nf
8	Ph	Rb	Nf	Kc
9	Ee	Og	la	Hh
10	Og	Ee	Hh	la
11	Xh	Dd	Ff	Aa
12	Dd	Xh	Aa	Ff

32 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	APh	ANf	ASc	AJb
2	ANf	APh	AJb	ASc
3	AVf	AXh	AMe	ALd
4	AXh	AVf	ALd	AMe
5	ADd	BAC	BCe	AYa
6	BAC	ADd	AYa	BCe
7	ARb	AQa	BFh	ATd
8	AQa	ARb	ATd	BFh
9	BEg	ABb	AFf	AAa
10	ABb	BEg	AAa	AFf
11	AOg	BBd	ACc	BDf
12	BBd	AOg	BDf	ACc
13	Ala	AEe	AHh	AWg
14	AEe	Ala	AWg	AHh
15	AUe	AKc	AGg	AZb
16	AKc	AUe	AZb	AGg

40 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>BFh</i>	<i>AVf</i>	<i>AKc</i>	<i>AGg</i>
2	<i>AVf</i>	<i>BFh</i>	<i>AGg</i>	<i>AKc</i>
3	<i>ARb</i>	<i>BBd</i>	<i>BCe</i>	<i>ACc</i>
4	<i>BBd</i>	<i>ARb</i>	<i>ACc</i>	<i>BCe</i>
5	<i>BAc</i>	<i>AMe</i>	<i>AHh</i>	<i>AYa</i>
6	<i>AMe</i>	<i>BAc</i>	<i>AYa</i>	<i>AHh</i>
7	<i>ANf</i>	<i>AUe</i>	<i>AXh</i>	<i>ALd</i>
8	<i>AUe</i>	<i>ANf</i>	<i>ALd</i>	<i>AXh</i>
9	<i>AJb</i>	<i>BJd</i>	<i>AOg</i>	<i>APh</i>
10	<i>BJd</i>	<i>AJb</i>	<i>APh</i>	<i>AOg</i>
11	<i>Ala</i>	<i>BMg</i>	<i>BKe</i>	<i>ATd</i>
12	<i>BMg</i>	<i>Ala</i>	<i>ATd</i>	<i>BKe</i>
13	<i>AQa</i>	<i>BEg</i>	<i>BHb</i>	<i>ASc</i>
14	<i>BEg</i>	<i>AQa</i>	<i>ASc</i>	<i>BHb</i>
15	<i>AFf</i>	<i>BNh</i>	<i>AZb</i>	<i>BGa</i>
16	<i>BNh</i>	<i>AFf</i>	<i>BGa</i>	<i>AZb</i>
17	<i>AEe</i>	<i>ABb</i>	<i>BDf</i>	<i>AWg</i>
18	<i>ABb</i>	<i>AEe</i>	<i>AWg</i>	<i>BDf</i>
19	<i>Blc</i>	<i>AAa</i>	<i>BLf</i>	<i>ADd</i>
20	<i>AAa</i>	<i>Blc</i>	<i>ADd</i>	<i>BLf</i>

48 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>Blc</i>	<i>BBd</i>	<i>BSe</i>	<i>BDf</i>
2	<i>BBd</i>	<i>Blc</i>	<i>BDf</i>	<i>BSe</i>
3	<i>BRd</i>	<i>AQa</i>	<i>BKe</i>	<i>BPb</i>
4	<i>AQa</i>	<i>BRd</i>	<i>BPb</i>	<i>BKe</i>
5	<i>BQc</i>	<i>BNh</i>	<i>BGa</i>	<i>AOg</i>
6	<i>BNh</i>	<i>BQc</i>	<i>AOg</i>	<i>BGa</i>
7	<i>BJd</i>	<i>BUg</i>	<i>ASc</i>	<i>AZb</i>
8	<i>BUg</i>	<i>BJd</i>	<i>AZb</i>	<i>ASc</i>
9	<i>AAa</i>	<i>AVf</i>	<i>BCe</i>	<i>AXh</i>
10	<i>AVf</i>	<i>AAa</i>	<i>AXh</i>	<i>BCe</i>
11	<i>BAc</i>	<i>BVh</i>	<i>ATd</i>	<i>BOa</i>
12	<i>BVh</i>	<i>BAc</i>	<i>BOa</i>	<i>ATd</i>
13	<i>BMg</i>	<i>AEe</i>	<i>BTf</i>	<i>ACc</i>
14	<i>AEe</i>	<i>BMg</i>	<i>ACc</i>	<i>BTf</i>
15	<i>ARb</i>	<i>AUe</i>	<i>AWg</i>	<i>AYa</i>
16	<i>AUe</i>	<i>ARb</i>	<i>AYa</i>	<i>AWg</i>
17	<i>ABb</i>	<i>AMe</i>	<i>AKc</i>	<i>AHh</i>
18	<i>AMe</i>	<i>ABb</i>	<i>AHh</i>	<i>AKc</i>
19	<i>BEg</i>	<i>BFh</i>	<i>BLf</i>	<i>BHb</i>
20	<i>BFh</i>	<i>BEg</i>	<i>BHb</i>	<i>BLf</i>
21	<i>Ala</i>	<i>AFf</i>	<i>ADd</i>	<i>AGg</i>
22	<i>AFf</i>	<i>Ala</i>	<i>AGg</i>	<i>ADd</i>
23	<i>ANf</i>	<i>AJb</i>	<i>APh</i>	<i>ALd</i>
24	<i>AJb</i>	<i>ANf</i>	<i>ALd</i>	<i>APh</i>

56 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>BLf</i>	<i>AJb</i>	<i>BNh</i>	<i>BYc</i>
2	<i>AJb</i>	<i>BLf</i>	<i>BYc</i>	<i>BNh</i>
3	<i>BAc</i>	<i>ALd</i>	<i>ANf</i>	<i>BCe</i>
4	<i>ALd</i>	<i>BAc</i>	<i>BCe</i>	<i>ANf</i>
5	<i>AAa</i>	<i>BVh</i>	<i>BBd</i>	<i>AEe</i>
6	<i>BVh</i>	<i>AAa</i>	<i>AEe</i>	<i>BBd</i>
7	<i>AQa</i>	<i>ASc</i>	<i>AVf</i>	<i>ADd</i>
8	<i>ASc</i>	<i>AQa</i>	<i>ADd</i>	<i>AVf</i>
9	<i>AHh</i>	<i>BJd</i>	<i>AGg</i>	<i>CBf</i>
10	<i>BJd</i>	<i>AHh</i>	<i>CBf</i>	<i>AGg</i>
11	<i>BRd</i>	<i>BSe</i>	<i>AWg</i>	<i>ARb</i>
12	<i>BSe</i>	<i>BRd</i>	<i>ARb</i>	<i>AWg</i>
13	<i>BPb</i>	<i>BGa</i>	<i>BZd</i>	<i>BFh</i>
14	<i>BGa</i>	<i>BPb</i>	<i>BFh</i>	<i>BZd</i>
15	<i>BDf</i>	<i>BKe</i>	<i>BXb</i>	<i>AXh</i>
16	<i>BKe</i>	<i>BDf</i>	<i>AXh</i>	<i>BXb</i>
17	<i>CCg</i>	<i>BTf</i>	<i>AMe</i>	<i>BOa</i>
18	<i>BTf</i>	<i>CCg</i>	<i>BOa</i>	<i>AMe</i>
19	<i>CAe</i>	<i>Blc</i>	<i>AYa</i>	<i>AZb</i>
20	<i>Blc</i>	<i>CAe</i>	<i>AZb</i>	<i>AYa</i>
21	<i>APh</i>	<i>AUe</i>	<i>BQc</i>	<i>AOg</i>
22	<i>AUe</i>	<i>APh</i>	<i>AOg</i>	<i>BQc</i>
23	<i>Ala</i>	<i>CDh</i>	<i>BMg</i>	<i>AKc</i>
24	<i>CDh</i>	<i>Ala</i>	<i>AKc</i>	<i>BMg</i>
25	<i>BEg</i>	<i>AFf</i>	<i>BWa</i>	<i>BHb</i>
26	<i>AFf</i>	<i>BEg</i>	<i>BHb</i>	<i>BWa</i>
27	<i>ABb</i>	<i>BUg</i>	<i>ATd</i>	<i>ACc</i>
28	<i>BUg</i>	<i>ABb</i>	<i>ACc</i>	<i>ATd</i>

64 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	<i>BWa</i>	<i>AXh</i>	<i>AUe</i>	<i>ALd</i>
2	<i>AXh</i>	<i>BWa</i>	<i>ALd</i>	<i>AUe</i>
3	<i>Ala</i>	<i>BEg</i>	<i>ATd</i>	<i>ACc</i>
4	<i>BEg</i>	<i>Ala</i>	<i>ACc</i>	<i>ATd</i>
5	<i>BYc</i>	<i>BPb</i>	<i>APh</i>	<i>CKg</i>
6	<i>BPb</i>	<i>BYc</i>	<i>CKg</i>	<i>APh</i>
7	<i>CHd</i>	<i>BXb</i>	<i>BKe</i>	<i>Blc</i>
8	<i>BXb</i>	<i>CHd</i>	<i>Blc</i>	<i>BKe</i>
9	<i>AHh</i>	<i>BCe</i>	<i>BUg</i>	<i>BQc</i>
10	<i>BCe</i>	<i>AHh</i>	<i>BQc</i>	<i>BUg</i>
11	<i>BFh</i>	<i>BJd</i>	<i>BTf</i>	<i>CEa</i>
12	<i>BJd</i>	<i>BFh</i>	<i>CEa</i>	<i>BTf</i>
13	<i>AEe</i>	<i>BLf</i>	<i>AGg</i>	<i>CDh</i>
14	<i>BLf</i>	<i>AEe</i>	<i>CDh</i>	<i>AGg</i>
15	<i>BHb</i>	<i>CAe</i>	<i>AVf</i>	<i>AYa</i>
16	<i>CAe</i>	<i>BHb</i>	<i>AYa</i>	<i>AVf</i>
17	<i>ASc</i>	<i>AQa</i>	<i>CFb</i>	<i>CLh</i>
18	<i>AQa</i>	<i>ASc</i>	<i>CLh</i>	<i>CFb</i>
19	<i>ANf</i>	<i>BAc</i>	<i>BGa</i>	<i>AMe</i>
20	<i>BAc</i>	<i>ANf</i>	<i>AMe</i>	<i>BGa</i>
21	<i>BSe</i>	<i>CGc</i>	<i>CBf</i>	<i>ABb</i>
22	<i>CGc</i>	<i>BSe</i>	<i>ABb</i>	<i>CBf</i>
23	<i>AJb</i>	<i>ADd</i>	<i>BDf</i>	<i>CCg</i>
24	<i>ADd</i>	<i>AJb</i>	<i>CCg</i>	<i>BDf</i>
25	<i>BMg</i>	<i>BBd</i>	<i>Cle</i>	<i>AAa</i>
26	<i>BBd</i>	<i>BMg</i>	<i>AAa</i>	<i>Cle</i>
27	<i>BVh</i>	<i>AFf</i>	<i>BRd</i>	<i>AZb</i>
28	<i>AFf</i>	<i>BVh</i>	<i>AZb</i>	<i>BRd</i>
29	<i>AWg</i>	<i>BOa</i>	<i>ARb</i>	<i>BNh</i>
30	<i>BOa</i>	<i>AWg</i>	<i>BNh</i>	<i>ARb</i>
31	<i>AOg</i>	<i>CJf</i>	<i>BZd</i>	<i>AKc</i>
32	<i>CJf</i>	<i>AOg</i>	<i>AKc</i>	<i>BZd</i>

Eight-plex system

16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Hh</i>	<i>Ia</i>	<i>Cc</i>	<i>Ld</i>	<i>Bb</i>	<i>Ee</i>	<i>Gg</i>	<i>Ff</i>
2	<i>Ia</i>	<i>Hh</i>	<i>Ld</i>	<i>Cc</i>	<i>Ee</i>	<i>Bb</i>	<i>Ff</i>	<i>Gg</i>
3	<i>Nf</i>	<i>Jb</i>	<i>Aa</i>	<i>Og</i>	<i>Kc</i>	<i>Dd</i>	<i>Ph</i>	<i>Me</i>
4	<i>Jb</i>	<i>Nf</i>	<i>Og</i>	<i>Aa</i>	<i>Dd</i>	<i>Kc</i>	<i>Me</i>	<i>Ph</i>

24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	<i>Ph</i>	<i>Ff</i>	<i>Ee</i>	<i>Qa</i>	<i>Rb</i>	<i>Og</i>	<i>Ld</i>	<i>Sc</i>
2	<i>Ff</i>	<i>Ph</i>	<i>Qa</i>	<i>Ee</i>	<i>Og</i>	<i>Rb</i>	<i>Sc</i>	<i>Ld</i>
3	<i>Bb</i>	<i>Ue</i>	<i>Xh</i>	<i>Dd</i>	<i>Kc</i>	<i>Aa</i>	<i>Vf</i>	<i>Gg</i>
4	<i>Ue</i>	<i>Bb</i>	<i>Dd</i>	<i>Xh</i>	<i>Aa</i>	<i>Kc</i>	<i>Gg</i>	<i>Vf</i>
5	<i>Wg</i>	<i>Td</i>	<i>Jb</i>	<i>Cc</i>	<i>Hh</i>	<i>Nf</i>	<i>Me</i>	<i>Ia</i>
6	<i>Td</i>	<i>Wg</i>	<i>Cc</i>	<i>Jb</i>	<i>Nf</i>	<i>Hh</i>	<i>Ia</i>	<i>Me</i>

32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	AAa	ABb	ACc	ADd	AEe	AFf	AGg	AHh
2	ABb	AAa	ADd	ACc	AFf	AEe	AHh	AGg
3	AKc	ALd	Ala	AJb	AOg	APh	AMe	ANf
4	ALd	AKc	AJb	Ala	APh	AOg	ANf	AMe
5	AUe	AVf	AWg	AXh	AQa	ARb	ASc	ATd
6	AVf	AUe	AXh	AWg	ARb	AQa	ATd	ASc
7	BEg	BFh	BCe	BDf	BAc	BBd	AYa	AZb
8	BFh	BEg	BDf	BCe	BBd	BAc	AZb	AYa

40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	AAa	AFf	ACc	AEe	ADd	ABb	AHh	AGg
2	AFf	AAa	AEe	ACc	ABb	ADd	AGg	AHh
3	AOg	APh	ALd	Ala	AMe	ANf	AJb	AKc
4	APh	AOg	Ala	ALd	ANf	AMe	AKc	AJb
5	ARb	ATd	AXh	AQa	AUe	AWg	ASc	AVf
6	ATd	ARb	AQa	AXh	AWg	AUe	AVf	ASc
7	AZb	BBd	BEg	BFh	BAc	AYa	BCe	BDf
8	BBd	AZb	BFh	BEg	AYa	BAc	BDf	BCe
9	BKe	Blc	BHb	BLf	BMg	BNh	BJd	BGa
10	Blc	BKe	BLf	BHb	BNh	BMg	BGa	BJd

48 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	AGg	ABb	AFf	AAa	ADd	ACc	AHh	AEe
2	ABb	AGg	AAa	AFf	ACc	ADd	AEe	AHh
3	APh	AOg	AKc	ALd	Ala	AMe	ANf	AJb
4	AOg	APh	ALd	AKc	AMe	Ala	AJb	ANf
5	AUe	AVf	ASc	AWg	AXh	ARb	AQa	ATd
6	AVf	AUe	AWg	ASc	ARb	AXh	ATd	AQa
7	BDf	AYa	BFh	BEg	BAc	AZb	BBd	BCe
8	AYa	BDf	BEg	BFh	AZb	BAc	BCe	BBd
9	Blc	BJd	BHb	BKe	BLf	BMg	BGa	BNh
10	BJd	Blc	BKe	BHb	BMg	BLf	BNh	BGa
11	BVh	BOa	BRd	BPb	BSe	BTf	BUg	BQc
12	BOa	BVh	BPb	BRd	BTf	BSe	BQc	BUg

56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	ADd	AGg	AEe	AHh	AFf	AAa	ACc	ABb
2	AGg	ADd	AHh	AEe	AAa	AFf	ABb	ACc
3	ANf	Ala	ALd	AJb	AOg	APh	AMe	AKc
4	Ala	ANf	AJb	ALd	APh	AOg	AKc	AMe
5	ARb	AUe	AWg	AVf	ATd	ASc	AQa	AXh
6	AUe	ARb	AVf	AWg	ASc	ATd	AXh	AQa
7	AZb	BDf	BAc	AYa	BFh	BCe	BBd	BEg
8	BDf	AZb	AYa	BAc	BCe	BFh	BEg	BBd
9	Blc	BVh	BGa	BRd	BTf	BPb	BSe	BUg
10	BVh	Blc	BRd	BGa	BPb	BTf	BUg	BSe
11	BQc	BMg	BNh	BHb	BKe	BJd	BLf	BOa
12	BMg	BQc	BHb	BNh	BJd	BKe	BOa	BLf
13	CDh	BWa	BYc	CAe	CCg	BXb	BZd	CBf
14	BWa	CDh	CAe	BYc	BXb	CCg	CBf	BZd

64 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	AAa	ABb	ACc	ADd	AEe	AFf	AGg	AHh
2	ABb	AAa	ADd	ACc	AFf	AEe	AHh	AGg
3	Ala	AJb	AKc	ALd	AMe	ANf	AOg	APh
4	AJb	Ala	ALd	AKc	ANf	AMe	APh	AOg
5	ASc	ATd	AQa	ARb	AWg	AXh	AUe	AVf
6	ATd	ASc	ARb	AQa	AXh	AWg	AVf	AUe
7	BAC	BBd	AYa	AZb	BEg	BFh	BCe	BDf
8	BBd	BAC	AZb	AYa	BFh	BEg	BDf	BCe
9	BKe	BLf	BMg	BNh	BGa	BHb	Blc	BJd
10	BLf	BKe	BNh	BMg	BHb	BGa	BJd	Blc
11	BSe	BTf	BUg	BVh	BOa	BPb	BQc	BRd
12	BTf	BSe	BVh	BUg	BPb	BOa	BRd	BQc
13	CCg	CDh	CAe	CBf	BYc	BZd	BWa	BXb
14	CDh	CCg	CBf	CAe	BZd	BYc	BXb	BWa
15	CKg	CLh	Cle	CJf	CGc	CHd	CEa	CFb
16	CLh	CKg	CJf	Cle	CHd	CGc	CFb	CEa

Appendix D

Tables of properties of optimal designs when Phase 1 is CRD

Phase 1 Experiment			Phase 2 Experiment						
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	<i>Between Runs stratum</i>	<i>Between Animals within Runs stratum</i>			
					Animal DF	Tag DF	Residual DF	Tag \perp Trt (E)	<i>Treatment efficiency factors</i>
									E
2	4	2	2	4	0	1	1	Yes	1
	6	4	3		1	1	2	No (1/9)	0.8889
	8	6	4		1	1	4	Yes	1
	10	8	5		2	1	5	No (1/25)	0.96
	12	10	6		2	1	7	Yes	1
	14	12	7		3	1	8	No (1/49)	0.9796
	16	14	8		3	1	10	Yes	1
	8	6	2	8	0	3	3	Yes	1
	12	10	3		1	3	6	No (1/9)	0.8889
	16	14	4		1	3	10	Yes	1

Phase 1 Design				Phase 2 Design							
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	<i>Between Runs stratum</i>		<i>Between Animals within Runs stratum</i>				
					Animal DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	<i>Treatment efficiency factors</i>	
										e_i	E
3	6	3	3	4	1	1	1	1	Yes	1, 3/4	0.8571
	12	9	6		2	2	1	6	Yes	15/16(2)	0.9375
	18	15	9		4	2	1	10	Yes	23/24, 7/8	0.9148
	24	21	12		5	2	1	15	Yes	15/16 (2)	0.9375
	12	9	3	8	1	1	3	5	Yes	1, 15/16	0.9677
	24	21	6		2	2	3	16	Yes	63/64 (2)	0.9844

Phase 1 Design				Phase 2 Design							
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	Between Runs stratum		Between Animals within Runs stratum				
					Animal DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt (E)	Treatment efficiency factors	
										e_i	E
4	8	4	4	4	1	0	1	2	Yes	1	1
	12	8	6		2	0	1	5	No (1/9)	1(2), 8/9	0.96
	16	12	8		3	0	1	8	Yes	1	1
	20	16	10		4	0	1	11	No (1/25)	1(2), 24/25	0.9863
	24	20	12		5	0	1	14	Yes	1	1
	28	24	14		6	0	1	17	No (1/49)	1(2), 48/49	0.9931
	32	28	16		7	0	1	20	Yes	1	1
	8	4	2	8	0	0	3	2	No (1/2)	1,1/2(2)	0.6
	12	8	3		1	0	3	4	No (1/9)	8/9 (3)	0.8889
	16	12	4		1	0	3	8	Yes	1	1
	20	16	5		2	0	3	11	No (1/25)	24/25(3)	0.96
	24	20	6		2	0	3	15	No (1/18)	1, 17/18(2)	0.9623
	28	24	7		3	0	3	18	No (1/49)	48/49(3)	0.9796
	32	28	8		3	0	3	22	Yes	1	1

Phase 1 Design				Phase 2 Design							
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	<i>Between Runs stratum</i>		<i>Between Animals within Runs stratum</i>				
					Animal DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	<i>Treatment efficiency factors</i>	
										e_i	E
5	10	5	5	4	2	2	1	2	Yes	1(2), 7/8, 5/8	0.8434
	20	15	10		4	4	1	10	Yes	15/16(4)	0.9375
	30	25	15		7	4	1	17	Yes	23/24(2), 11/12 5/6	0.9137
	40	35	20		9	4	1	25	Yes	15/16(4)	0.9375
	50	45	25		12	4	1	32	Yes	19/20(2), 37/40, 7/8	0.9240
	20	15	5	8	2	2	3	10	Yes	1(2), 15/16(2)	0.9677
	40	35	10		4	4	3	28	Yes	0.994 (2), 0.959(2)	0.9763

Phase 1 Design				Phase 2 Design							
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	Between Runs stratum		Between Animals within Runs stratum				
					Animal DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt (E)	Treatment efficiency factors	
										e_i	E
6	12	6	6	4	2	2	1	3	Yes	1(3), 3/4(2)	0.8824
	18	12	9		4	4	1	7	No (1/9)	11/12(2), 8/9, 3/4(2)	0.8370
	24	18	12		5	4	1	12	Yes	1, 15/16(2), 13/16(2)	0.8937
	30	24	15		7	5	1	16	No (1/25)	0.953, 9/10, 0.8836, 0.8235, 4/5	0.8686
	36	30	18		8	4	1	21	Yes	1, 7/8 (4)	0.8974
	42	36	21		10	5	1	25	No (1/49)	13/14, 0.9164, 6/7(2), 0.8489	0.8804
	48	42	24		11	5	1	30	Yes	15/16 (2), 7/8 (3)	0.8990
	12	6	3	8	1	1	3	2	No (1/3)	1, 3/4, 2/3(3)	0.7317
	24	18	6		2	2	3	13	Yes	1(3), 15/16(2)	0.9740
	36	30	9		4	4	3	23	No (4/81)	0.9792, 0.9601, 0.9421 0.9375 0.9033	0.9438
	48	42	12		5	4	3	34	Yes	1, 63/64(2), 61/64(2)	0.9746

Phase 1 Design				Phase 2 Design							
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	<i>Between Runs stratum</i>		<i>Between Animals within Runs stratum</i>				
					Animal DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	<i>Treatment efficiency factors</i>	
										e_i	E
7	14	7	7	4	3	3	1	3	Yes	1(3), 7/8, 5/8, 1/2	0.7749
	28	21	14		6	6	1	14	Yes	7/8 (6)	0.875
	42	35	21		10	6	1	24	Yes	7/8(5), 19/24	0.8599
	56	49	28		13	6	1	35	Yes	7/8 (6)	0.875
	28	21	7	8	3	3	3	15	Yes	1(3),31/32(2), 7/8	0.9666
	56	49	14		6	6	3	40	Yes	63/64 (6)	0.9844

Phase 1 Design				Phase 2 Design							
v	n_a	Between Animals Residual DF	n_{Runs}	n_{Tags}	Between Runs stratum		Between Animals within Runs stratum				
					Animal DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt (E)	Treatment efficiency factors	
										e_i	E
8	16	8	8	4	3	3	1	4	Yes	1(4), 3/4(2), 1/2	0.8077
	24	16	12		5	5	1	10	No (1/9)	1, 11/12(2), 8/9, 3/4(2), 2/3	0.8261
	32	24	16		7	7	1	16	Yes	0.963 (2), 0.875 (2), 0.7866 (2), 0.75	0.8498
	40	32	20		9	7	1	22	No (1/25)	9/10(3), 43/50, 4/5(3)	0.8489
	48	40	24		11	7	1	28	Yes	1, 5/6(6)	0.8537
	56	48	28		13	7	1	34	No (1/49)	6/7(6), 41/49	0.8542
	64	56	32		15	7	1	40	Yes	0.9192(2), 0.875, 0.8308(2), 0.8125(2)	0.8550
	16	8	4	8	1	0	3	4	No (3/10)	1(4), 3/4(2), 1/2	0.8077
	24	16	6		2	0	3	11	No (1/9)	1(4), 8/9(3)	0.9492
	32	24	8		3	0	3	18	Yes	1(7)	1
	40	32	10		4	0	3	25	No (1/25)	1(4), 24/25(3)	0.9825
	48	40	12		5	0	3	32	No (1/30)	1(4), 35/36(2), 17/18	0.9837
	56	48	14		6	0	3	39	No (1/49)	1(4), 48/49(3)	0.9912
	64	56	16		7	0	3	46	Yes	1(7)	1

Appendix E

Tables of optimal designs when Phase 1 is RCBD

Two Treatments

Four-plex system

2 Blocks and 4 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Aa	2Ca	2Db
2	1Aa	1Bb	2Db	2Ca

3 Blocks and 6 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	3Ea	1Bb	3Fb	1Aa
2	1Bb	3Ea	1Aa	3Fb
3	2Ca	2Ca	2Db	2Db

2 Blocks and 8 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Db	1Ca	2Ca	2Db
2	1Ca	1Db	2Db	2Ca
3	1Bb	1Aa	2Bb	2Aa
4	1Aa	1Bb	2Aa	2Bb

4 Blocks and 8 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Ca	2Db	4Ga	4Hb
2	2Db	2Ca	4Hb	4Ga
3	1Aa	1Bb	3Ea	3Fb
4	1Bb	1Aa	3Fb	3Ea

5 Blocks and 10 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Db	1Aa	1Bb	2Ca
2	1Aa	2Db	2Ca	1Bb
3	4Ga	3Fb	4Hb	3Ea
4	3Fb	4Ga	3Ea	4Hb
5	5Ia	5Ia	5Jb	5Jb

2 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Jb	2Lb	1Ca	1Ea
2	2Lb	2Jb	1Ea	1Ca
3	2Ka	2Ia	1Db	1Bb
4	2Ia	2Ka	1Bb	1Db
5	2Hb	2Ga	1Fb	1Aa
6	2Ga	2Hb	1Aa	1Fb

3 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	1Ca	1Db
2	1Bb	1Aa	1Db	1Ca
3	2Fb	2Ea	2Ga	2Hb
4	2Ea	2Fb	2Hb	2Ga
5	3Jb	3Ia	3Ka	3Lb
6	3Ia	3Jb	3Lb	3Ka

6 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Aa	6Lb	6Ka
2	1Aa	1Bb	6Ka	6Lb
3	3Fb	3Ea	4Ga	4Hb
4	3Ea	3Fb	4Hb	4Ga
5	2Db	2Ca	5Ia	5Jb
6	2Ca	2Db	5Jb	5Ia

7 Blocks and 14 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Aa	2Db	1Bb	2Ca
2	2Db	1Aa	2Ca	1Bb
3	3Ea	4Hb	3Fb	4Ga
4	4Hb	3Ea	4Ga	3Fb
5	6Ka	5Jb	5Ia	6Lb
6	5Jb	6Ka	6Lb	5Ia
7	7Nb	7Nb	7Ma	7Ma

2 Blocks and 16 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Db	1Ga	2Oa	2Nb
2	1Ga	1Db	2Nb	2Oa
3	1Aa	1Bb	2Ia	2Pb
4	1Bb	1Aa	2Pb	2Ia
5	1Ca	1Hb	2Lb	2Ma
6	1Hb	1Ca	2Ma	2Lb
7	1Fb	1Ea	2Jb	2Ka
8	1Ea	1Fb	2Ka	2Jb

4 Blocks and 16 Phase 1 Experimental units
(Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1Bb	1Ca	1Aa	1Db
2	1Ca	1Bb	1Db	1Aa
3	2Ea	2Fb	2Ga	2Hb
4	2Fb	2Ea	2Hb	2Ga
5	3Lb	3Ka	3Ia	3Jb
6	3Ka	3Lb	3Jb	3Ia
7	4Nb	4Oa	4Pb	4Ma
8	4Oa	4Nb	4Ma	4Pb

4 Blocks and 16 Phase 1 Experimental units
(Higher EDF)

Run	Tag			
	114	115	116	117
1	1Ca	1Bb	3Lb	3Ka
2	1Bb	1Ca	3Ka	3Lb
3	1Aa	1Db	3Jb	3Ia
4	1Db	1Aa	3Ia	3Jb
5	2Ga	2Hb	4Oa	4Pb
6	2Hb	2Ga	4Pb	4Oa
7	2Fb	2Ea	4Nb	4Ma
8	2Ea	2Fb	4Ma	4Nb

8 Blocks and 16 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Aa	6Lb	6Ka
2	1Aa	1Bb	6Ka	6Lb
3	3Ea	3Fb	2Db	2Ca
4	3Fb	3Ea	2Ca	2Db
5	5Jb	5Ia	8Oa	8Pb
6	5Ia	5Jb	8Pb	8Oa
7	7Ma	7Nb	4Ga	4Hb
8	7Nb	7Ma	4Hb	4Ga

Eight-plex system

2 Blocks and 8 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Bb	1Ca	1Db	2Fb	2Ea	2Hb	2Ga
2	1Bb	1Aa	1Db	1Ca	2Ea	2Fb	2Ga	2Hb

4 Blocks and 8 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Bb	2Ca	2Db	3Ea	3Fb	4Hb	4Ga
2	1Bb	1Aa	2Db	2Ca	3Fb	3Ea	4Ga	4Hb

2 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Bb	1Ea	1Aa	1Db	2Ka	2Hb	2Jb	2Ga
2	1Ea	1Bb	1Db	1Aa	2Hb	2Ka	2Ga	2Jb
3	1Fb	1Fb	1Ca	1Ca	2Lb	2Lb	2Ia	2Ia

3 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2Ga	2Fb	2Ea	2Hb	1Bb	1Aa	1Ca	1Db
2	2Fb	2Ga	2Hb	2Ea	1Aa	1Bb	1Db	1Ca
3	3Ka	3Ka	3Jb	3Jb	3Lb	3Lb	3Ia	3Ia

6 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	6Ka	1Bb	6Lb	1Aa	4Hb	2Ca	2Db	4Ga
2	1Bb	6Ka	1Aa	6Lb	2Ca	4Hb	4Ga	2Db
3	3Fb	3Fb	3Ea	3Ea	5Ia	5Ia	5Jb	5Jb

2 Blocks and 16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Fb	1Ea	1Bb	2Lb	2Oa	2Pb	2Ka
2	1Fb	1Aa	1Bb	1Ea	2Oa	2Lb	2Ka	2Pb
3	1Ca	1Hb	1Db	1Ga	2Ia	2Nb	2Jb	2Ma
4	1Hb	1Ca	1Ga	1Db	2Nb	2Ia	2Ma	2Jb

4 Blocks and 16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Db	2Fb	2Ga	3Lb	3Ka	4Oa	4Nb
2	1Db	1Aa	2Ga	2Fb	3Ka	3Lb	4Nb	4Oa
3	1Bb	1Ca	2Hb	2Ea	3Ia	3Jb	4Pb	4Ma
4	1Ca	1Bb	2Ea	2Hb	3Jb	3Ia	4Ma	4Pb

8 Blocks and 16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Bb	3Ea	3Fb	5Jb	5Ia	7Ma	7Nb
2	1Bb	1Aa	3Fb	3Ea	5Ia	5Jb	7Nb	7Ma
3	2Ca	2Db	4Hb	4Ga	6Lb	6Ka	8Oa	8Pb
4	2Db	2Ca	4Ga	4Hb	6Ka	6Lb	8Pb	8Oa

Three Treatments

Four-plex system

2 Blocks and 6 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	2Fc	2Da
2	1Bb	1Aa	2Da	2Fc
3	1Cc	1Cc	2Eb	2Eb

2 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	2Lc	2Hb
2	1Bb	1Aa	2Hb	2Lc
3	1Cc	1Da	2Ga	2Kb
4	1Da	1Cc	2Kb	2Ga
5	1Eb	1Fc	2Ic	2Ja
6	1Fc	1Eb	2Ja	2Ic

4 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	2Eb	3Hb	4Ja
2	2Eb	1Cc	4Ja	3Hb
3	1Aa	2Fc	3Ga	4Kb
4	2Fc	1Aa	4Kb	3Ga
5	1Bb	2Da	3Ic	4Lc
6	2Da	1Bb	4Lc	3Ic

2 Blocks and 18 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Hb	1Cc	2Ma	2Nb
2	1Cc	1Hb	2Nb	2Ma
3	1Aa	1Fc	2Qb	2Pa
4	1Fc	1Aa	2Pa	2Qb
5	1Da	1Eb	2Oc	2Rc
6	1Eb	1Da	2Rc	2Oc
7	1Ic	1Bb	2Kb	2Ja
8	1Bb	1Ic	2Ja	2Kb
9	1Ga	1Ga	2Lc	2Lc

3 Blocks and 18 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Ic	1Bb	2Hb	1Da
2	1Bb	2Ic	1Da	2Hb
3	1Cc	2Ja	2Ga	1Eb
4	2Ja	1Cc	1Eb	2Ga
5	2Kb	1Aa	1Fc	2Lc
6	1Aa	2Kb	2Lc	1Fc
7	3Ma	3Nb	3Pa	3Oc
8	3Nb	3Ma	3Oc	3Pa
9	3Rc	3Rc	3Qb	3Qb

6 Blocks and 18 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Aa	2Eb	4Lc	5Ma
2	2Eb	1Aa	5Ma	4Lc
3	1Cc	2Da	4Kb	5Nb
4	2Da	1Cc	5Nb	4Kb
5	2Fc	1Bb	5Oc	4Ja
6	1Bb	2Fc	4Ja	5Oc
7	3Ga	3Hb	6Rc	6Qb
8	3Hb	3Ga	6Qb	6Rc
9	3Ic	3Ic	6Pa	6Pa

2 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Ic	1Da	2Nb	2Xc
2	1Da	1Ic	2Xc	2Nb
3	1Ja	1Fc	2Sa	2Qb
4	1Fc	1Ja	2Qb	2Sa
5	1Cc	1Kb	2Tb	2Pa
6	1Kb	1Cc	2Pa	2Tb
7	1Bb	1Lc	2Va	2Wb
8	1Lc	1Bb	2Wb	2Va
9	1Ga	1Hb	2Oc	2Uc
10	1Hb	1Ga	2Uc	2Oc
11	1Aa	1Eb	2Rc	2Ma
12	1Eb	1Aa	2Ma	2Rc

4 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Eb	1Fc	3Pa	3Nb
2	1Fc	1Eb	3Nb	3Pa
3	1Bb	1Cc	3Rc	3Ma
4	1Cc	1Bb	3Ma	3Rc
5	1Da	1Aa	3Oc	3Qb
6	1Aa	1Da	3Qb	3Oc
7	2Kb	2Ja	4Wb	4Uc
8	2Ja	2Kb	4Uc	4Wb
9	2Hb	2Ic	4Sa	4Xc
10	2Ic	2Hb	4Xc	4Sa
11	2Lc	2Ga	4Tb	4Va
12	2Ga	2Lc	4Va	4Tb

8 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Eb	1Aa	6Pa	5Oc
2	1Aa	2Eb	5Oc	6Pa
3	1Bb	2Da	6Rc	5Nb
4	2Da	1Bb	5Nb	6Rc
5	1Cc	2Fc	6Qb	5Ma
6	2Fc	1Cc	5Ma	6Qb
7	4Lc	3Hb	7Sa	8Xc
8	3Hb	4Lc	8Xc	7Sa
9	4Ja	3Ga	8Wb	7Uc
10	3Ga	4Ja	7Uc	8Wb
11	4Kb	3Ic	7Tb	8Va
12	3Ic	4Kb	8Va	7Tb

Eight-plex system

2 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Cc	1Da	1Eb	1Aa	2Lc	2Hb	2Kb	2Ga
2	1Da	1Cc	1Aa	1Eb	2Hb	2Lc	2Ga	2Kb
3	1Bb	1Bb	1Fc	1Fc	2Ja	2Ja	2Ic	2Ic

4 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Cc	2Fc	2Eb	3Ga	3Hb	4Kb	4Ja
2	1Cc	1Aa	2Eb	2Fc	3Hb	3Ga	4Ja	4Kb
3	1Bb	1Bb	2Da	2Da	3Ic	3Ic	4Lc	4Lc

2 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Ga	1Hb	1Eb	1Fc	2Rc	2Uc	2Sa	2Ma
2	1Hb	1Ga	1Fc	1Eb	2Uc	2Rc	2Ma	2Sa
3	1Kb	1Lc	1Cc	1Da	2Tb	2Va	2Nb	2Oc
4	1Lc	1Kb	1Da	1Cc	2Va	2Tb	2Oc	2Nb
5	1Ja	1Ic	1Aa	1Bb	2Pa	2Qb	2Wb	2Xc
6	1Ic	1Ja	1Bb	1Aa	2Qb	2Pa	2Xc	2Wb

4 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Fc	2Kb	2Ga	3Ma	3Qb	4Wb	4Xc
2	1Fc	1Aa	2Ga	2Kb	3Qb	3Ma	4Xc	4Wb
3	1Eb	1Cc	2Ja	2Lc	3Nb	3Oc	4Va	4Sa
4	1Cc	1Eb	2Lc	2Ja	3Oc	3Nb	4Sa	4Va
5	1Bb	1Da	2Ic	2Hb	3Rc	3Pa	4Uc	4Tb
6	1Da	1Bb	2Hb	2Ic	3Pa	3Rc	4Tb	4Uc

8 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2Eb	1Aa	4Ja	3Hb	6Rc	5Oc	8Wb	7Uc
2	1Aa	2Eb	3Hb	4Ja	5Oc	6Rc	7Uc	8Wb
3	1Cc	2Da	4Lc	3Ga	6Qb	5Nb	8Xc	7Sa
4	2Da	1Cc	3Ga	4Lc	5Nb	6Qb	7Sa	8Xc
5	1Bb	2Fc	4Kb	3Ic	6Pa	5Ma	7Tb	8Va
6	2Fc	1Bb	3Ic	4Kb	5Ma	6Pa	8Va	7Tb

Four Treatments

Four-plex system

2 Blocks and 8 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Aa	2Gc	2Hd
2	1Aa	1Bb	2Hd	2Gc
3	1Dd	1Cc	2Fb	2Ea
4	1Cc	1Dd	2Ea	2Fb

3 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Dd	1Cc	1Bb	1Aa
2	1Cc	1Dd	1Aa	1Bb
3	2Fb	2Hd	2Ea	2Gc
4	2Hd	2Fb	2Gc	2Ea
5	3Ia	3Jb	3Kc	3Ld
6	3Jb	3Ia	3Ld	3Kc

2 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Hd	1Cc	2Jb	2Ia
2	1Cc	1Hd	2Ia	2Jb
3	1Gc	1Dd	2Ma	2Nb
4	1Dd	1Gc	2Nb	2Ma
5	1Ea	1Bb	2Ld	2Oc
6	1Bb	1Ea	2Oc	2Ld
7	1Aa	1Fb	2Pd	2Kc
8	1Fb	1Aa	2Kc	2Pd

4 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	1Dd	1Bb	1Aa
2	1Dd	1Cc	1Aa	1Bb
3	2Hd	2Fb	2Ea	2Gc
4	2Fb	2Hd	2Gc	2Ea
5	3Ia	3Jb	3Kc	3Ld
6	3Jb	3Ia	3Ld	3Kc
7	4Oc	4Ma	4Pd	4Nb
8	4Ma	4Oc	4Nb	4Pd

4 Blocks and 12 Phase 1 Experimental units
(Higher EDF)

Run	Tag			
	114	115	116	117
1	1Bb	1Aa	3Kc	3Ld
2	1Aa	1Bb	3Ld	3Kc
3	2Hd	2Gc	4Nb	4Ma
4	2Gc	2Hd	4Ma	4Nb
5	2Ea	2Fb	4Oc	4Pd
6	2Fb	2Ea	4Pd	4Oc
7	1Cc	1Dd	3Ia	3Jb
8	1Dd	1Cc	3Jb	3Ia

5 Blocks and 20 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	1Aa	1Bb	1Dd
2	1Aa	1Cc	1Dd	1Bb
3	2Fb	2Hd	2Ea	2Gc
4	2Hd	2Fb	2Gc	2Ea
5	3Kc	3Ia	3Jb	3Ld
6	3Ia	3Kc	3Ld	3Jb
7	4Ma	4Pd	4Oc	4Nb
8	4Pd	4Ma	4Nb	4Oc
9	5Sc	5Rb	5Qa	5Td
10	5Rb	5Sc	5Td	5Qa

2 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	1Aa	2Td	2Rb
2	1Aa	1Cc	2Rb	2Td
3	1Dd	1Ia	2Oc	2Nb
4	1Ia	1Dd	2Nb	2Oc
5	1Ea	1Bb	2Pd	2Wc
6	1Bb	1Ea	2Wc	2Pd
7	1Jb	1Kc	2Xd	2Ua
8	1Kc	1Jb	2Ua	2Xd
9	1Hd	1Gc	2Vb	2Ma
10	1Gc	1Hd	2Ma	2Vb
11	1Ld	1Fb	2Qa	2Sc
12	1Fb	1Ld	2Sc	2Qa

3 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Dd	1Bb	1Cc	1Aa
2	1Bb	1Dd	1Aa	1Cc
3	1Hd	1Fb	1Ea	1Gc
4	1Fb	1Hd	1Gc	1Ea
5	2Ia	2Kc	2Ld	2Jb
6	2Kc	2Ia	2Jb	2Ld
7	2Nb	2Ma	2Pd	2Oc
8	2Ma	2Nb	2Oc	2Pd
9	3Sc	3Td	3Qa	3Rb
10	3Td	3Sc	3Rb	3Qa
11	3Wc	3Ua	3Vb	3Xd
12	3Ua	3Wc	3Xd	3Vb

6 Blocks and 24 Phase 1 Experimental units
(Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1Dd	1Cc	1Bb	1Aa
2	1Cc	1Dd	1Aa	1Bb
3	2Ea	2Gc	2Fb	2Hd
4	2Gc	2Ea	2Hd	2Fb
5	3Ld	3Ia	3Jb	3Kc
6	3Ia	3Ld	3Kc	3Jb
7	4Ma	4Nb	4Oc	4Pd
8	4Nb	4Ma	4Pd	4Oc
9	5Rb	5Sc	5Qa	5Td
10	5Sc	5Rb	5Td	5Qa
11	6Xd	6Vb	6Ua	6Wc
12	6Vb	6Xd	6Wc	6Ua

6 Blocks and 24 Phase 1 Experimental units
(Higher EDF)

Run	Tag			
	114	115	116	117
1	1Aa	1Bb	4Pd	4Oc
2	1Bb	1Aa	4Oc	4Pd
3	1Cc	1Dd	4Nb	4Ma
4	1Dd	1Cc	4Ma	4Nb
5	2Gc	2Fb	5Td	5Qa
6	2Fb	2Gc	5Qa	5Td
7	2Ea	2Hd	5Sc	5Rb
8	2Hd	2Ea	5Rb	5Sc
9	3Ia	3Kc	6Vb	6Xd
10	3Kc	3Ia	6Xd	6Vb
11	3Jb	3Ld	6Wc	6Ua
12	3Ld	3Jb	6Ua	6Wc

7 Blocks and 28 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AAa	1ADd	1ABb	1ACc
2	1ADd	1AAa	1ACc	1ABb
3	2AGc	2AFb	2AEa	2AHd
4	2AFb	2AGc	2AHd	2AEa
5	3AKc	3AJb	3Ala	3ALd
6	3AJb	3AKc	3ALd	3Ala
7	4AOc	4ANb	4AMa	4APd
8	4ANb	4AOc	4APd	4AMa
9	5ASc	5AQa	5ATd	5ARb
10	5AQa	5ASc	5ARb	5ATd
11	6AVb	6AXd	6AWc	6AUa
12	6AXd	6AVb	6AUa	6AWc
13	7AYa	7BBd	7AZb	7BAc
14	7BBd	7AYa	7BAc	7AZb

2 Blocks and 32 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Ala	1APd	2BEc	2AVb
2	1APd	1Ala	2AVb	2BEc
3	1AJb	1AMa	2BBd	2ASc
4	1AMa	1AJb	2ASc	2BBd
5	1AFb	1AKc	2AXd	2AUa
6	1AKc	1AFb	2AUa	2AXd
7	1ALd	1AEa	2AWc	2BDb
8	1AEa	1ALd	2BDb	2AWc
9	1AHd	1AAa	2ARb	2BAc
10	1AAa	1AHd	2BAc	2ARb
11	1ANb	1AOc	2BCa	2BFd
12	1AOc	1ANb	2BFd	2BCa
13	1AGc	1ADd	2AYa	2AZb
14	1ADd	1AGc	2AZb	2AYa
15	1ABb	1ACc	2AQa	2ATd
16	1ACc	1ABb	2ATd	2AQa

4 Blocks and 2 Phase 1 Experimental units
(Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1AGc	1AEa	1AFb	1AHd
2	1AEa	1AGc	1AHd	1AFb
3	1ADd	1AAa	1ACc	1ABb
4	1AAa	1ADd	1ABb	1ACc
5	2ALd	2AKc	2Ala	2AJb
6	2AKc	2ALd	2AJb	2Ala
7	2AMa	2ANb	2APd	2AOc
8	2ANb	2AMa	2AOc	2APd
9	3ATd	3ARb	3AQa	3ASc
10	3ARb	3ATd	3ASc	3AQa
11	3AUa	3AVb	3AXd	3AWc
12	3AVb	3AUa	3AWc	3AXd
13	4AZb	4BAc	4AYa	4BBd
14	4BAc	4AZb	4BBd	4AYa
15	4BEc	4BFd	4BDb	4BCa
16	4BFd	4BEc	4BCa	4BDb

4 Blocks and 32 Phase 1 Experimental units
(Higher EDF)

Run	Tag			
	114	115	116	117
1	1ADd	1AEa	3ASc	3AVb
2	1AEa	1ADd	3AVb	3ASc
3	1ACc	1AFb	3AQa	3AXd
4	1AFb	1ACc	3AXd	3AQa
5	1AGc	1ABb	3ATd	3AUa
6	1ABb	1AGc	3AUa	3ATd
7	1AAa	1AHd	3AWc	3ARb
8	1AHd	1AAa	3ARb	3AWc
9	2Ala	2AJb	4BFd	4BEc
10	2AJb	2Ala	4BEc	4BFd
11	2AOc	2ALd	4AZb	4AYa
12	2ALd	2AOc	4AYa	4AZb
13	2ANb	2AKc	4BCa	4BBd
14	2AKc	2ANb	4BBd	4BCa
15	2APd	2AMa	4BDb	4BAc
16	2AMa	2APd	4BAc	4BDb

8 Blocks and 32 Phase 1 Experimental units
(Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1ABb	1ADd	1AAa	1ACc
2	1ADd	1ABb	1ACc	1AAa
3	2AEa	2AFb	2AHd	2AGc
4	2AFb	2AEa	2AGc	2AHd
5	3ALd	3Ala	3AKc	3AJb
6	3Ala	3ALd	3AJb	3AKc
7	4AMa	4AOc	4APd	4ANb
8	4AOc	4AMa	4ANb	4APd
9	5ARb	5ASc	5AQa	5ATd
10	5ASc	5ARb	5ATd	5AQa
11	6AVb	6AXd	6AWc	6AUa
12	6AXd	6AVb	6AUa	6AWc
13	7AYa	7BAc	7AZb	7BBd
14	7BAc	7AYa	7BBd	7AZb
15	8BFd	8BEc	8BDb	8BCa
16	8BEc	8BFd	8BCa	8BDb

8 Blocks and 32 Phase 1 Experimental units
(Higher EDF)

Run	Tag			
	114	115	116	117
1	1ADd	1ABb	5AQa	5ASc
2	1ABb	1ADd	5ASc	5AQa
3	1AAa	1ACc	5ATd	5ARb
4	1ACc	1AAa	5ARb	5ATd
5	2AEa	2AGc	6AVb	6AXd
6	2AGc	2AEa	6AXd	6AVb
7	2AHd	2AFb	6AWc	6AUa
8	2AFb	2AHd	6AUa	6AWc
9	3AJb	3AKc	7AYa	7BBd
10	3AKc	3AJb	7BBd	7AYa
11	3ALd	3Ala	7AZb	7BAc
12	3Ala	3ALd	7BAc	7AZb
13	4ANb	4AOc	8BFd	8BCa
14	4AOc	4ANb	8BCa	8BFd
15	4AMa	4APd	8BEc	8BDb
16	4APd	4AMa	8BDb	8BEc

Eight-plex system

2 Blocks and 8 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Dd	1Aa	1Bb	1Cc	2Ea	2Fb	2Gc	2Hd
2	1Aa	1Dd	1Cc	1Bb	2Fb	2Ea	2Hd	2Gc

3 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2Ea	1Dd	2Hd	1Bb	1Cc	2Fb	2Gc	1Aa
2	1Dd	2Ea	1Bb	2Hd	2Fb	1Cc	1Aa	2Gc
3	3Kc	3Kc	3la	3la	3Ld	3Ld	3Jb	3Jb

2 Blocks and 32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Hd	1Cc	1Dd	1Gc	2la	2Nb	2Jb	2Ma
2	1Cc	1Hd	1Gc	1Dd	2Nb	2la	2Ma	2Jb
3	1Fb	1Aa	1Ea	1Bb	2Ld	2Oc	2Kc	2Pd
4	1Aa	1Fb	1Bb	1Ea	2Oc	2Ld	2Pd	2Kc

4 Blocks and 32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Cc	2Gc	2Fb	3Ld	3la	4Nb	4Pd
2	1Cc	1Aa	2Fb	2Gc	3la	3Ld	4Pd	4Nb
3	1Bb	1Dd	2Hd	2Ea	3Jb	3Kc	4Oc	4Ma
4	1Dd	1Bb	2Ea	2Hd	3Kc	3Jb	4Ma	4Oc

5 Blocks and 40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Bb	2Ea	1Aa	2Fb	2Gc	1Dd	2Hd	1Cc
2	2Ea	1Bb	2Fb	1Aa	1Dd	2Gc	1Cc	2Hd
3	3Ld	4Pd	4Oc	3Kc	4Nb	3la	4Ma	3Jb
4	4Pd	3Ld	3Kc	4Oc	3la	4Nb	3Jb	4Ma
5	5Sc	5Sc	5Td	5Td	5Qa	5Qa	5Rb	5Rb

2 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Fb	1Ea	1Gc	1Dd	2Xd	2Nb	2Wc	2Qa
2	1Ea	1Fb	1Dd	1Gc	2Nb	2Xd	2Qa	2Wc
3	1Cc	1Ia	1Aa	1Bb	2Sc	2Vb	2Pd	2Td
4	1Ia	1Cc	1Bb	1Aa	2Vb	2Sc	2Td	2Pd
5	1Kc	1Ld	1Jb	1Hd	2Oc	2Ma	2Rb	2Ua
6	1Ld	1Kc	1Hd	1Jb	2Ma	2Oc	2Ua	2Rb

3 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Dd	1Gc	1Fb	1Bb	1Aa	1Ea	1Hd	1Cc
2	1Gc	1Dd	1Bb	1Fb	1Ea	1Aa	1Cc	1Hd
3	2Ma	2Pd	2Oc	2Kc	2Jb	2Nb	2Ld	2Ia
4	2Pd	2Ma	2Kc	2Oc	2Nb	2Jb	2Ia	2Ld
5	3Vb	3Sc	3Td	3Ua	3Xd	3Wc	3Rb	3Qa
6	3Sc	3Vb	3Ua	3Td	3Wc	3Xd	3Qa	3Rb

6 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	3Kc	2Fb	2Ea	3Ld	5Td	4Ma	4Nb	5Sc
2	2Fb	3Kc	3Ld	2Ea	4Ma	5Td	5Sc	4Nb
3	3Ia	1Dd	1Cc	3Jb	4Oc	6Ua	4Pd	6Vb
4	1Dd	3Ia	3Jb	1Cc	6Ua	4Oc	6Vb	4Pd
5	1Bb	2Gc	2Hd	1Aa	5Rb	6Wc	6Xd	5Qa
6	2Gc	1Bb	1Aa	2Hd	6Wc	5Rb	5Qa	6Xd

7 Blocks and 28 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2AEa	1AAa	1ABb	2AHd	1ACc	2AFb	1ADd	2AGc
2	1AAa	2AEa	2AHd	1ABb	2AFb	1ACc	2AGc	1ADd
3	3AKc	4ANb	3AJb	4APd	4AMa	3ALd	4AOc	3Ala
4	4ANb	3AKc	4APd	3AJb	3ALd	4AMa	3Ala	4AOc
5	5ASc	6AXd	6AUa	5AQa	5ARb	6AWc	6AVb	5ATd
6	6AXd	5ASc	5AQa	6AUa	6AWc	5ARb	5ATd	6AVb
7	7AZb	7AZb	7BAc	7BAc	7BBd	7BBd	7AYa	7AYa

2 Blocks and 32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFb	1AAa	1AKc	1AMa	2ATd	2BFd	2AVb	2AWc
2	1AAa	1AFb	1AMa	1AKc	2BFd	2ATd	2AWc	2AVb
3	1ACc	1AHd	1AJb	1APd	2BCa	2BDb	2AQa	2ASc
4	1AHd	1ACc	1APd	1AJb	2BDb	2BCa	2ASc	2AQa
5	1AGc	1ABb	1ALd	1Ala	2BAc	2AZb	2AUa	2AXd
6	1ABb	1AGc	1Ala	1ALd	2AZb	2BAc	2AXd	2AUa
7	1ADd	1AEa	1ANb	1AOc	2AYa	2BEc	2ARb	2BBd
8	1AEa	1ADd	1AOc	1ANb	2BEc	2AYa	2BBd	2ARb

4 Blocks and 32 Phase 1 Experimental units (Higher Residual EDF)

Run	Tag							
	113	114	115	116	117	118	119	121
1	1ABb	1AAa	2A1a	2AOc	3ATd	3ARb	4BEc	4BFd
2	1AAa	1ABb	2AOc	2A1a	3ARb	3ATd	4BFd	4BEc
3	1ACc	1AGc	2ALd	2ANb	3AQa	3AXd	4AYa	4BDb
4	1AGc	1ACc	2ANb	2ALd	3AXd	3AQa	4BDb	4AYa
5	1AFb	1AHd	2AJb	2APd	3AUa	3ASc	4BAc	4BCa
6	1AHd	1AFb	2APd	2AJb	3ASc	3AUa	4BCa	4BAc
7	1AEa	1ADd	2AKc	2AMa	3AWc	3AVb	4BBd	4AZb
8	1ADd	1AEa	2AMa	2AKc	3AVb	3AWc	4AZb	4BBd

8 Blocks and 32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AAa	1ABb	3ALd	3AJb	5ATd	5ASc	7AYa	7BAc
2	1ABb	1AAa	3AJb	3ALd	5ASc	5ATd	7BAc	7AYa
3	1ADd	1ACc	3A1a	3AKc	5AQa	5ARb	7AZb	7BBd
4	1ACc	1ADd	3AKc	3A1a	5ARb	5AQa	7BBd	7AZb
5	2AHd	2AFb	4AMa	4AOc	6AVb	6AWc	8BFd	8BCa
6	2AFb	2AHd	4AOc	4AMa	6AWc	6AVb	8BCa	8BFd
7	2AGc	2AEa	4APd	4ANb	6AXd	6AUa	8BDb	8BEc
8	2AEa	2AGc	4ANb	4APd	6AUa	6AXd	8BEc	8BDb

Five Treatments

Four-plex system

2 Blocks and 10 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	1Bb	2Fa	2Id
2	1Bb	1Cc	2Id	2Fa
3	1Aa	1Ee	2Gb	2Hc
4	1Ee	1Aa	2Hc	2Gb
5	1Dd	1Dd	2Je	2Je

2 Blocks and 20 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Gb	1Cc	2Nd	2Oe
2	1Cc	1Gb	2Oe	2Nd
3	1Bb	1Ee	2Pa	2Sd
4	1Ee	1Bb	2Sd	2Pa
5	1Id	1Aa	2Mc	2Qb
6	1Aa	1Id	2Qb	2Mc
7	1Je	1Hc	2Lb	2Ka
8	1Hc	1Je	2Ka	2Lb
9	1Fa	1Dd	2Te	2Rc
10	1Dd	1Fa	2Rc	2Te

4 Blocks and 20 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	2Je	4Qb	3Ka
2	2Je	1Cc	3Ka	4Qb
3	1Bb	2Fa	3Nd	4Rc
4	2Fa	1Bb	4Rc	3Nd
5	1Dd	2Gb	4Te	3Mc
6	2Gb	1Dd	3Mc	4Te
7	2Hc	1Aa	3Oe	4Sd
8	1Aa	2Hc	4Sd	3Oe
9	2Id	1Ee	4Pa	3Lb
10	1Ee	2Id	3Lb	4Pa

2 Blocks and 30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AGb	1AId	2ATe	2AZa
2	1AId	1AGb	2AZa	2ATe
3	1AMc	1ADd	2BAb	2AYe
4	1ADd	1AMc	2AYe	2BAb
5	1ANd	1AKa	2BDe	2AQb
6	1AKa	1ANd	2AQb	2BDe
7	1AHc	1AOe	2BCd	2AVb
8	1AOe	1AHc	2AVb	2BCd
9	1AFa	1ALb	2AWc	2ASd
10	1ALb	1AFa	2ASd	2AWc
11	1AJe	1ACc	2AUa	2AXd
12	1ACc	1AJe	2AXd	2AUa
13	1AEe	1ABb	2APa	2ARc
14	1ABb	1AEe	2ARc	2APa
15	1AAa	1AAa	2BBc	2BBc

3 Blocks and 30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2ATe	1AAa	1ACc	2AQb
2	1AAa	2ATe	2AQb	1ACc
3	2ASd	1AHc	2ALb	1AEe
4	1AHc	2ASd	1AEe	2ALb
5	1AGb	2AKa	1AId	2AOe
6	2AKa	1AGb	2AOe	1AId
7	2ARc	1ABb	2APa	1ADd
8	1ABb	2ARc	1ADd	2APa
9	1AJe	1AFa	2ANd	2AMc
10	1AFa	1AJe	2AMc	2ANd
11	3AYe	3BBc	3AZa	3ABb
12	3BBc	3AYe	3ABb	3AZa
13	3AXd	3AVb	3AUa	3AWc
14	3AVb	3AXd	3AWc	3AUa
15	3BCd	3BCd	3BDe	3BDe

Six Blocks and 30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AAa	2AJe	4AQb	5AWc
2	2AJe	1AAa	5AWc	4AQb
3	2AHc	1ABb	5AXd	4APa
4	1ABb	2AHc	4APa	5AXd
5	2AId	1ACc	5AVb	4ATe
6	1ACc	2AId	4ATe	5AVb
7	2AGb	1AEe	4ASd	5AUa
8	1AEe	2AGb	5AUa	4ASd
9	2AFa	1ADd	4ARc	5AYe
10	1ADd	2AFa	5AYe	4ARc
11	3AOe	3AMc	6AZa	6BCd
12	3AMc	3AOe	6BCd	6AZa
13	3ANd	3AKa	6BDe	6BAb
14	3AKa	3ANd	6BAb	6BDe
15	3ALb	3ALb	6BBc	6BBc

2 Blocks and 40 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AFa	1AEe	2BAb	2BLc
2	1AEe	1AFa	2BLc	2BAb
3	1AKa	1ALb	2Ble	2BGc
4	1ALb	1AKa	2BGc	2Ble
5	1AId	1APa	2BBc	2BDe
6	1APa	1AId	2BDe	2BBc
7	1ANd	1AHc	2AVb	2BJa
8	1AHc	1ANd	2BJa	2AVb
9	1ADd	1AMc	2AYe	2BEa
10	1AMc	1ADd	2BEa	2AYe
11	1AAa	1ABb	2AXd	2AWc
12	1ABb	1AAa	2AWc	2AXd
13	1ARc	1ATe	2BKb	2BCd
14	1ATe	1ARc	2BCd	2BKb
15	1ASd	1AGb	2AUa	2BNe
16	1AGb	1ASd	2BNe	2AUa
17	1AQb	1AJe	2AZa	2BHd
18	1AJe	1AQb	2BHd	2AZa
19	1AOe	1ACc	2BFb	2BMD
20	1ACc	1AOe	2BMD	2BFb

4 Blocks and 40 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AGb	1AAa	3BCd	3AYe
2	1AAa	1AGb	3AYe	3BCd
3	1AHc	1ADd	3BAb	3BDe
4	1ADd	1AHc	3BDe	3BAb
5	1ABb	1AEe	3AUa	3BBc
6	1AEe	1ABb	3BBc	3AUa
7	1AId	1AJe	3AZa	3AWc
8	1AJe	1AId	3AWc	3AZa
9	1AFa	1ACc	3AVb	3AXd
10	1ACc	1AFa	3AXd	3AVb
11	2AQb	2AKa	4BMd	4BGc
12	2AKa	2AQb	4BGc	4BMd
13	2AOe	2ALb	4BJa	4BHd
14	2ALb	2AOe	4BHd	4BJa
15	2AMc	2ATe	4BEa	4BFb
16	2ATe	2AMc	4BFb	4BEa
17	2AND	2ARc	4BKb	4BNe
18	2ARc	2AND	4BNe	4BKb
19	2APa	2ASd	4Blc	4BLc
20	2ASd	2APa	4BLc	4Blc

8 Blocks and 40 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2AHc	1AAa	6BAb	5AYe
2	1AAa	2AHc	5AYe	6BAb
3	2AGb	1AEe	5AXd	6AZa
4	1AEe	2AGb	6AZa	5AXd
5	2AId	1ACc	5AUa	6BDe
6	1ACc	2AId	6BDe	5AUa
7	2AFa	1ADd	6BBc	5AVb
8	1ADd	2AFa	5AVb	6BBc
9	2AJe	1ABb	6BCd	5AWc
10	1ABb	2AJe	5AWc	6BCd
11	3AND	4ARc	8BNe	7BFb
12	4ARc	3AND	7BFb	8BNe
13	4ATe	3ALb	8BMd	7BEa
14	3ALb	4ATe	7BEa	8BMd
15	3AOe	4APa	7BGc	8BKb
16	4APa	3AOe	8BKb	7BGc
17	3AMc	4AQb	7BHd	8BJa
18	4AQb	3AMc	8BJa	7BHd
19	4ASd	3AKa	8BLc	7Ble
20	3AKa	4ASd	7Ble	8BLc

Eight-plex iTRAQ system

2 Blocks and 20 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Aa	1Cc	1Gb	1Id	2Qb	2Pa	2Nd	2Oe
2	1Cc	1Aa	1Id	1Gb	2Pa	2Qb	2Oe	2Nd
3	1Dd	1Bb	1Je	1Fa	2Te	2Rc	2Ka	2Mc
4	1Bb	1Dd	1Fa	1Je	2Rc	2Te	2Mc	2Ka
5	1Ee	1Ee	1Hc	1Hc	2Sd	2Sd	2Lb	2Lb

4 Blocks and 20 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Bb	1Cc	2Id	2Je	3Nd	3Oe	4Pa	4Qb
2	1Cc	1Bb	2Je	2Id	3Oe	3Nd	4Qb	4Pa
3	1Dd	1Aa	2Fa	2Hc	3Lb	3Mc	4Te	4Sd
4	1Aa	1Dd	2Hc	2Fa	3Mc	3Lb	4Sd	4Te
5	1Ee	1Ee	2Gb	2Gb	3Ka	3Ka	4Rc	4Rc

2 Blocks and 40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AGb	1AQb	1AOe	1ADd	2BNe	2BJa	2BLc	2AWc
2	1AQb	1AGb	1ADd	1AOe	2BJa	2BNe	2AWc	2BLc
3	1ACc	1APa	1ATe	1ABb	2BBc	2Ble	2BCd	2BMd
4	1APa	1ACc	1ABb	1ATe	2Ble	2BBc	2BMd	2BCd
5	1ANd	1AJe	1ALb	1AAa	2BGc	2AXd	2AUa	2BAb
6	1AJe	1ANd	1AAa	1ALb	2AXd	2BGc	2BAb	2AUa
7	1AHc	1AFa	1AId	1ARc	2BHd	2AVb	2BDe	2AZa
8	1AFa	1AHc	1ARc	1AId	2AVb	2BHd	2AZa	2BDe
9	1ASd	1AEe	1AMc	1AKa	2BEa	2BKb	2AYe	2BFb
10	1AEe	1ASd	1AKa	1AMc	2BKb	2BEa	2BFb	2AYe

4 Blocks and 40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AAa	1AHc	2APa	2ALb	3AYe	3AWc	4Ble	4BMd
2	1AHc	1AAa	2ALb	2APa	3AWc	3AYe	4BMd	4Ble
3	1AJe	1AGb	2ATe	2AMc	3AZa	3BAb	4BHd	4BLc
4	1AGb	1AJe	2AMc	2ATe	3BAb	3AZa	4BLc	4BHd
5	1ADd	1ABb	2AQb	2ASd	3BBc	3BDe	4BEa	4BNe
6	1ABb	1ADd	2ASd	2AQb	3BDe	3BBc	4BNe	4BEa
7	1AId	1ACc	2AKa	2AOe	3AUa	3AXd	4BFb	4BKb
8	1ACc	1AId	2AOe	2AKa	3AXd	3AUa	4BKb	4BFb
9	1AFa	1AEe	2ARc	2AND	3AVb	3BCd	4BJa	4BGc
10	1AEe	1AFa	2AND	2ARc	3BCd	3AVb	4BGc	4BJa

8 Blocks and 40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2AJe	1ACc	3ALb	4APa	5AUa	6BCd	8BMd	7BFb
2	1ACc	2AJe	4APa	3ALb	6BCd	5AUa	7BFb	8BMd
3	2AFa	1ABb	3AND	4ATe	5AWc	6BAb	7BEa	8BLc
4	1ABb	2AFa	4ATe	3AND	6BAb	5AWc	8BLc	7BEa
5	1AAa	2AHc	4ASd	3AOe	6AZa	5AYe	7BHd	8BKb
6	2AHc	1AAa	3AOe	4ASd	5AYe	6AZa	8BKb	7BHd
7	2AGb	1ADd	4ARc	3AKa	5AXd	6BBc	8BNe	7Ble
8	1ADd	2AGb	3AKa	4ARc	6BBc	5AXd	7Ble	8BNe
9	2AId	1AEe	4AQb	3AMc	5AVb	6BDe	8BJa	7BGc
10	1AEe	2AId	3AMc	4AQb	6BDe	5AVb	7BGc	8BJa

Six Treatments

Four-plex system

2 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Ee	2Jd	2Ga
2	1Ee	1Bb	2Ga	2Jd
3	1Aa	1Ff	2Ic	2Ke
4	1Ff	1Aa	2Ke	2Ic
5	1Cc	1Dd	2Lf	2Hb
6	1Dd	1Cc	2Hb	2Lf

2 Blocks and 18 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Jd	1Bb	1Ee	2Ic
2	1Bb	2Jd	2Ic	1Ee
3	1Cc	2Ke	2Lf	1Aa
4	2Ke	1Cc	1Aa	2Lf
5	1Ff	2Ga	2Hb	1Dd
6	2Ga	1Ff	1Dd	2Hb
7	3Pd	3Ma	3Oc	3Nb
8	3Ma	3Pd	3Nb	3Oc
9	3Qe	3Qe	3Rf	3Rf

2 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Cc	1Jd	2Tb	2We
2	1Jd	1Cc	2We	2Tb
3	1Bb	1Ee	2Ma	2Oc
4	1Ee	1Bb	2Oc	2Ma
5	1Ic	1Aa	2Xf	2Nb
6	1Aa	1Ic	2Nb	2Xf
7	1Ga	1Ke	2Rf	2Vd
8	1Ke	1Ga	2Vd	2Rf
9	1Lf	1Hb	2Qe	2Pd
10	1Hb	1Lf	2Pd	2Qe
11	1Dd	1Ff	2Uc	2Sa
12	1Ff	1Dd	2Sa	2Uc

4 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Ee	1Dd	3Nb	3Ma
2	1Dd	1Ee	3Ma	3Nb
3	1Cc	1Aa	3Rf	3Qe
4	1Aa	1Cc	3Qe	3Rf
5	1Ff	1Bb	3Oc	3Pd
6	1Bb	1Ff	3Pd	3Oc
7	2Ga	2Jd	4Uc	4Xf
8	2Jd	2Ga	4Xf	4Uc
9	2Ic	2Ke	4Vd	4Tb
10	2Ke	2Ic	4Tb	4Vd
11	2Hb	2Lf	4We	4Sa
12	2Lf	2Hb	4Sa	4We

5 Blocks and 30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2AJd	2AHb	1AEe	1AFf
2	2AHb	2AJd	1AFf	1AEe
3	2ALf	1ACc	2AGa	1ABb
4	1ACc	2ALf	1ABb	2AGa
5	1ADd	1AAa	2AKe	2AIlc
6	1AAa	1ADd	2AIlc	2AKe
7	3APd	4AWe	3AOc	4AXf
8	4AWe	3APd	4AXf	3AOc
9	4ASa	3ARf	4ATb	3AQe
10	3ARf	4ASa	3AQe	4ATb
11	3ANb	4AUc	3AMa	4AVd
12	4AUc	3ANb	4AVd	3AMa
13	5AYa	5BCe	5AZb	5BBd
14	5BCe	5AYa	5BBd	5AZb
15	5BDf	5BDf	5BAC	5BAC

Two Blocks and 36 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1APd	1AEe	2BEa	2BGc
2	1AEe	1APd	2BGc	2BEa
3	1AQe	1AOc	2BJf	2AZb
4	1AOc	1AQe	2AZb	2BJf
5	1Alc	1AKe	2ATb	2BBd
6	1AKe	1Alc	2BBd	2ATb
7	1ANb	1AAa	2AVd	2BDf
8	1AAa	1ANb	2BDf	2AVd
9	1AGa	1ALf	2BFb	2BAc
10	1ALf	1AGa	2BAc	2BFb
11	1ABb	1AJd	2AUc	2AYa
12	1AJd	1ABb	2AYa	2AUc
13	1AHb	1ADd	2AXf	2Ble
14	1ADd	1AHb	2Ble	2AXf
15	1ARf	1ACc	2ASa	2AWe
16	1ACc	1ARf	2AWe	2ASa
17	1AMa	1AFf	2BHd	2BCe
18	1AFf	1AMa	2BCe	2BHd

Three Blocks and 36 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AFf	1AKe	1Alc	1AAa
2	1AKe	1AFf	1AAa	1Alc
3	1ADd	1AEe	1ACc	1ABb
4	1AEe	1ADd	1ABb	1ACc
5	1AHb	1AGa	1AJd	1ALf
6	1AGa	1AHb	1ALf	1AJd
7	2AOc	2APd	2ARf	2ANb
8	2APd	2AOc	2ANb	2ARf
9	2AUc	2AMa	2AQe	2AVd
10	2AMa	2AUc	2AVd	2AQe
11	2ASa	2ATb	2AWe	2AXf
12	2ATb	2ASa	2AXf	2AWe
13	3BGc	3BDf	3BEa	3BBd
14	3BDf	3BGc	3BBd	3BEa
15	3BJf	3AZb	3BCe	3BAC
16	3AZb	3BJf	3BAC	3BCe
17	3BHd	3Ble	3AYa	3BFb
18	3Ble	3BHd	3BFb	3AYa

Six Blocks and 36 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AAa	1ACc	4AXf	4ATb
2	1ACc	1AAa	4ATb	4AXf
3	1ADd	1AFf	4ASa	4AWe
4	1AFf	1ADd	4AWe	4ASa
5	1ABb	1AEe	4AUc	4AVd
6	1AEe	1ABb	4AVd	4AUc
7	2AHb	2AGa	5BCe	5BAC
8	2AGa	2AHb	5BAC	5BCe
9	2AIlc	2ALf	5AYa	5BBd
10	2ALf	2AIlc	5BBd	5AYa
11	2AKe	2AJd	5AZb	5BDf
12	2AJd	2AKe	5BDf	5AZb
13	3AOc	3AQe	6BHd	6BEa
14	3AQe	3AOc	6BEa	6BHd
15	3APd	3ANb	6BJf	6BGc
16	3ANb	3APd	6BGc	6BJf
17	3AMa	3ARf	6Ble	6BFb
18	3ARf	3AMa	6BFb	6Ble

7 Blocks and 42 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ADd	2ALf	1ABb	2AKe
2	2ALf	1ADd	2AKe	1ABb
3	1AFf	2Alc	1AAa	2AHb
4	2Alc	1AFf	2AHb	1AAa
5	1ACc	2AGa	2AJd	1AEe
6	2AGa	1ACc	1AEe	2AJd
7	4AWe	3APd	3ARf	4AUc
8	3APd	4AWe	4AUc	3ARf
9	4ASa	3AQe	3ANb	4AXf
10	3AQe	4ASa	4AXf	3ANb
11	4ATb	3AOc	3AMa	4AVd
12	3AOc	4ATb	4AVd	3AMa
13	6BFb	6BHd	5BAc	5BDf
14	6BHd	6BFb	5BDf	5BAc
15	5BCe	6BJf	6BEa	5BBd
16	6BJf	5BCe	5BBd	6BEa
17	5AZb	5AYa	6Ble	6BGc
18	5AYa	5AZb	6BGc	6Ble
19	7BLb	7BOe	7BMc	7BPf
20	7BOe	7BLb	7BPf	7BMc
21	7BNd	7BNd	7BKa	7BKa

2 Blocks and 48 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ABb	1AAa	2BUe	2BBd
2	1AAa	1ABb	2BBd	2BUe
3	1AOc	1AWe	2AZb	2BDf
4	1AWe	1AOc	2BDf	2AZb
5	1AJd	1AGa	2Ble	2BAc
6	1AGa	1AJd	2BAc	2Ble
7	1ARf	1Alc	2BOe	2BHd
8	1Alc	1ARf	2BHd	2BOe
9	1APd	1ASa	2BRb	2BVf
10	1ASa	1APd	2BVf	2BRb
11	1AQe	1ACc	2BLb	2BKa
12	1ACc	1AQe	2BKa	2BLb
13	1ALf	1ADd	2BQa	2BSc
14	1ADd	1ALf	2BSc	2BQa
15	1ANb	1AMa	2BJf	2BGc
16	1AMa	1ANb	2BGc	2BJf
17	1AVd	1ATb	2BCe	2BMc
18	1ATb	1AVd	2BMc	2BCe
19	1AEe	1AXf	2BEa	2BFb
20	1AXf	1AEe	2BFb	2BEa
21	1AHb	1AUc	2BNd	2BPf
22	1AUc	1AHb	2BPf	2BNd
23	1AFf	1AKe	2BTd	2AYa
24	1AKe	1AFf	2AYa	2BTd

4 Blocks and 48 Phase 1 Experimental units
(Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1AGa	1AEe	1ABb	1AFf
2	1AEe	1AGa	1AFf	1ABb
3	1ALf	1ADd	1ACc	1AAa
4	1ADd	1ALf	1AAa	1ACc
5	1AHb	1AJd	1Alc	1AKe
6	1AJd	1AHb	1AKe	1Alc
7	2ARf	2ANb	2AUc	2APd
8	2ANb	2ARf	2APd	2AUc
9	2AVd	2AQe	2ASa	2ATb
10	2AQe	2AVd	2ATb	2ASa
11	2AOc	2AXf	2AMa	2AWe
12	2AXf	2AOc	2AWe	2AMa
13	3Ble	3BAc	3BDf	3BBd
14	3BAc	3Ble	3BBd	3BDf
15	3BEa	3BGc	3BCe	3AZb
16	3BGc	3BEa	3AZb	3BCe
17	3BFb	3BJf	3BHd	3AYa
18	3BJf	3BFb	3AYa	3BHd
19	4BUe	4BMc	4BRb	4BVf
20	4BMc	4BUe	4BVf	4BRb
21	4BTd	4BKa	4BPf	4BOe
22	4BKa	4BTd	4BOe	4BPf
23	4BQa	4BLb	4BNd	4BSc
24	4BLb	4BQa	4BSc	4BNd

4 Blocks and 48 Phase 1 Experimental units
(Higher Residual EDF)

Run	Tag			
	114	115	116	117
1	1ALf	1AAa	3BCe	3BBd
2	1AAa	1ALf	3BBd	3BCe
3	1Alc	1ADd	3BJf	3BEa
4	1ADd	1Alc	3BEa	3BJf
5	1AFf	1AHb	3BHd	3BAc
6	1AHb	1AFf	3BAc	3BHd
7	1AGa	1ABb	3Ble	3BGc
8	1ABb	1AGa	3BGc	3Ble
9	1AJd	1AEe	3BFb	3AYa
10	1AEe	1AJd	3AYa	3BFb
11	1ACc	1AKe	3AZb	3BDf
12	1AKe	1ACc	3BDf	3AZb
13	2ATb	2ARf	4BSc	4BKa
14	2ARf	2ATb	4BKa	4BSc
15	2APd	2AMa	4BVf	4BRb
16	2AMa	2APd	4BRb	4BVf
17	2ANb	2ASa	4BPf	4BUe
18	2ASa	2ANb	4BUe	4BPf
19	2AXf	2AUc	4BTd	4BOe
20	2AUc	2AXf	4BOe	4BTd
21	2AVd	2AWe	4BLb	4BMc
22	2AWe	2AVd	4BMc	4BLb
23	2AOc	2AQe	4BNd	4BQa
24	2AQe	2AOc	4BQa	4BNd

8 Blocks and 48 Phase 1 Experimental units (Higher EDF)

Run	Tag			
	114	115	116	117
1	1AAa	1ADd	5BDf	5AZb
2	1ADd	1AAa	5AZb	5BDf
3	1AEe	1AFf	5BBd	5BAc
4	1AFf	1AEe	5BAc	5BBd
5	1ACc	1ABb	5AYa	5BCe
6	1ABb	1ACc	5BCe	5AYa
7	2Alc	2AJd	6BJf	6BFb
8	2AJd	2Alc	6BFb	6BJf
9	2AGa	2AKe	6BGc	6BHd
10	2AKe	2AGa	6BHd	6BGc
11	2AHb	2ALf	6Ble	6BEa
12	2ALf	2AHb	6BEa	6Ble
13	3AMa	3ARf	7BMc	7BNd
14	3ARf	3AMa	7BNd	7BMc
15	3AQe	3APd	7BKa	7BLb
16	3APd	3AQe	7BLb	7BKa
17	3ANb	3AOc	7BOe	7BPf
18	3AOc	3ANb	7BPf	7BOe
19	4ASa	4AWe	8BVf	8BSc
20	4AWe	4ASa	8BSc	8BVf
21	4ATb	4AXf	8BUe	8BTd
22	4AXf	4ATb	8BTd	8BUe
23	4AVd	4AUc	8BRb	8BQa
24	4AUc	4AVd	8BQa	8BRb

Eight-plex system

2 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Bb	1Cc	1Ee	1Aa	2Lf	2Ga	2Jd	2Hb
2	1Cc	1Bb	1Aa	1Ee	2Ga	2Lf	2Hb	2Jd
3	1Ff	1Ff	1Dd	1Dd	2lc	2lc	2Ke	2Ke

2 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Hb	1lc	1Cc	1Dd	2Rf	2Qe	2Vd	2Sa
2	1lc	1Hb	1Dd	1Cc	2Qe	2Rf	2Sa	2Vd
3	1Ke	1Jd	1Ff	1Aa	2Oc	2Nb	2Tb	2We
4	1Jd	1Ke	1Aa	1Ff	2Nb	2Oc	2We	2Tb
5	1Lf	1Ga	1Ee	1Bb	2Pd	2Ma	2Uc	2Xf
6	1Ga	1Lf	1Bb	1Ee	2Ma	2Pd	2Xf	2Uc

4 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Bb	1Cc	2lc	2Ke	3Pd	3Rf	4Vd	4Sa
2	1Cc	1Bb	2Ke	2lc	3Rf	3Pd	4Sa	4Vd
3	1Aa	1Dd	2Lf	2Hb	3Oc	3Qe	4Xf	4Tb
4	1Dd	1Aa	2Hb	2Lf	3Qe	3Oc	4Tb	4Xf
5	1Ee	1Ff	2Ga	2Jd	3Ma	3Nb	4Uc	4We
6	1Ff	1Ee	2Jd	2Ga	3Nb	3Ma	4We	4Uc

2 Blocks and 36 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1A1c	1AGa	1AHb	1AJd	2BCe	2BEa	2AXf	2AUc
2	1AGa	1A1c	1AJd	1AHb	2BEa	2BCe	2AUc	2AXf
3	1ALf	1AQe	1ABb	1AKe	2BGc	2ASa	2BHd	2AZb
4	1AQe	1ALf	1AKe	1ABb	2ASa	2BGc	2AZb	2BHd
5	1ADd	1ARf	1AOc	1AMa	2AVd	2AWe	2ATb	2BAc
6	1ARf	1ADd	1AMa	1AOc	2AWe	2AVd	2BAc	2ATb
7	1APd	1AEe	1AAa	1ACc	2BFb	2BJf	2BDf	2B1e
8	1AEe	1APd	1ACc	1AAa	2BJf	2BFb	2B1e	2BDf
9	1ANb	1ANb	1AFf	1AFf	2BBd	2BBd	2AYa	2AYa

3 Blocks and 36 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2AXf	1AEe	2AOc	1ACc	1AFf	2AVd	2AMa	1ABb
2	1AEe	2AXf	1ACc	2AOc	2AVd	1AFf	1ABb	2AMa
3	2ANb	1A1c	1ADd	1AGa	2AWe	1AKe	2ASa	2ARf
4	1A1c	2ANb	1AGa	1ADd	1AKe	2AWe	2ARf	2ASa
5	1AJd	1AAa	1AHb	2ATb	2AUc	2APd	1ALf	2AQe
6	1AAa	1AJd	2ATb	1AHb	2APd	2AUc	2AQe	1ALf
7	3BJf	3BHd	3AYa	3B1e	3BEa	3AZb	3BBd	3BAc
8	3BHd	3BJf	3B1e	3AYa	3AZb	3BEa	3BAc	3BBd
9	3BCe	3BCe	3BDf	3BDf	3BGc	3BGc	3BFb	3BFb

6 Blocks and 36 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFf	2A1c	2AJd	3AQe	4AWe	5AYa	5BBd	6BFb
2	2A1c	1AFf	3AQe	2AJd	5AYa	4AWe	6BFb	5BBd
3	2AKe	3APd	1ABb	2ALf	6BEa	5BDf	4AUc	5BAc
4	3APd	2AKe	2ALf	1ABb	5BDf	6BEa	5BAc	4AUc
5	2AHb	1AAa	2AGa	3AOc	5AZb	4AXf	5BCe	6BHd
6	1AAa	2AHb	3AOc	2AGa	4AXf	5AZb	6BHd	5BCe
7	3ANb	1ADd	3AMa	1AEe	6BGc	4AVd	6BJf	4ASa
8	1ADd	3ANb	1AEe	3AMa	4AVd	6BGc	4ASa	6BJf
9	3ARf	3ARf	1ACc	1ACc	6B1e	6B1e	4ATb	4ATb

2 Blocks and 48 Phase 1 Experimental units

Run	Tag	V2	V3	V4	V5	V6	V7	V8
	113	114	115	116	117	118	119	121
1	1AFf	1AWe	1ATb	1AOc	2BQa	2BGc	2BVf	2BHd
2	1AWe	1AFf	1AOc	1ATb	2BGc	2BQa	2BHd	2BVf
3	1AMa	1ARf	1AKe	1AXf	2BTd	2BLb	2AZb	2BMc
4	1ARf	1AMa	1AXf	1AKe	2BLb	2BTd	2BMc	2AZb
5	1A1c	1AHb	1AJd	1AGa	2BNd	2BPf	2B1e	2BEa
6	1AHb	1A1c	1AGa	1AJd	2BPf	2BNd	2BEa	2B1e
7	1ADd	1ACc	1ASa	1AVd	2BOe	2BDf	2BRb	2BCe
8	1ACc	1ADd	1AVd	1ASa	2BDf	2BOe	2BCe	2BRb
9	1APd	1AQe	1ALf	1AUc	2AYa	2BFb	2BAc	2BKa
10	1AQe	1APd	1AUc	1ALf	2BFb	2AYa	2BKa	2BAc
11	1ANb	1AAa	1ABb	1AEe	2BSc	2BUe	2BJf	2BBd
12	1AAa	1ANb	1AEe	1ABb	2BUe	2BSc	2BBd	2BJf

4 Blocks and 48 Phase 1 Experimental units (Higher Residual EDF)

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFf	1ADd	2AOc	2ATb	3BEa	3BCe	4BUe	4BTd
2	1ADd	1AFf	2ATb	2AOc	3BCe	3BEa	4BTd	4BUe
3	1AHb	1Afc	2ASa	2AVd	3BGc	3BDf	4BOe	4BPf
4	1Afc	1AHb	2AVd	2ASa	3BDf	3BGc	4BPf	4BOe
5	1ALf	1ABb	2AXf	2AQe	3AYa	3BFb	4BMc	4BNd
6	1ABb	1ALf	2AQe	2AXf	3BFb	3AYa	4BNd	4BMc
7	1AJd	1AAa	2ARf	2AUc	3BBd	3Ble	4BQa	4BLb
8	1AAa	1AJd	2AUc	2ARf	3Ble	3BBd	4BLb	4BQa
9	1AEe	1ACc	2APd	2AWe	3BAc	3BJf	4BRb	4BKa
10	1ACc	1AEe	2AWe	2APd	3BJf	3BAc	4BRb	4BKa
11	1AKe	1AGa	2AMa	2ANb	3AZb	3BHd	4BVf	4BSc
12	1AGa	1AKe	2ANb	2AMa	3BHd	3AZb	4BSc	4BVf

8 Blocks and 48 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AEe	1ABb	3ARf	3AQe	5AYa	5BBd	7BPf	7BMc
2	1ABb	1AEe	3AQe	3ARf	5BBd	5AYa	7BMc	7BPf
3	1ADd	1ACc	3AMa	3AOc	5BDf	5BCe	7BLb	7BKa
4	1ACc	1ADd	3AOc	3AMa	5BCe	5BDf	7BKa	7BLb
5	1AAa	1AFf	3ANb	3APd	5BAc	5AZb	7BOe	7BNd
6	1AFf	1AAa	3APd	3ANb	5AZb	5BAc	7BNd	7BOe
7	2ALf	2Afc	4ATb	4ASa	6Ble	6BHd	8BRb	8BSc
8	2Afc	2ALf	4ASa	4ATb	6BHd	6Ble	8BSc	8BRb
9	2AGa	2AJd	4AXf	4AUc	6BJf	6BFb	8BTd	8BUe
10	2AJd	2AGa	4AUc	4AXf	6BFb	6BJf	8BUe	8BTd
11	2AKe	2AHb	4AWe	4AVd	6BEa	6BGc	8BQa	8BVf
12	2AHb	2AKe	4AVd	4AWe	6BGc	6BEa	8BVf	8BQa

Seven Treatments

Four-plex system

2 Blocks and 14 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Dd	1Aa	2Le	2Mf
2	1Aa	1Dd	2Mf	2Le
3	1Ff	1Gg	2lb	2Ha
4	1Gg	1Ff	2Ha	2lb
5	1Bb	1Ee	2Jc	2Ng
6	1Ee	1Bb	2Ng	2Jc
7	1Cc	1Cc	2Kd	2Kd

2 Blocks and 28 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AGg	1AMf	2ARd	2AWb
2	1AMf	1AGg	2AWb	2ARd
3	1ACc	1Alb	2ASe	2BBg
4	1Alb	1ACc	2BBg	2ASe
5	1ABb	1AJc	2AOa	2AYd
6	1AJc	1ABb	2AYd	2AOa
7	1ALe	1AHa	2APb	2ATf
8	1AHa	1ALe	2ATf	2APb
9	1ANg	1AKd	2AVa	2AZe
10	1AKd	1ANg	2AZe	2AVa
11	1AEe	1ADd	2AXc	2BAf
12	1ADd	1AEe	2BAf	2AXc
13	1AFf	1AAa	2AQc	2AUg
14	1AAa	1AFf	2AUg	2AQc

4 Blocks and 28 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AFf	2ANg	4AYd	3AQc
2	2ANg	1AFf	3AQc	4AYd
3	2ALe	1AGg	3ATf	4AWb
4	1AGg	2ALe	4AWb	3ATf
5	2AHa	1ACc	3APb	4BAf
6	1ACc	2AHa	4BAf	3APb
7	2AJc	1AEe	3AOa	4BBg
8	1AEe	2AJc	4BBg	3AOa
9	1ADd	2Alb	4AXc	3ASe
10	2Alb	1ADd	3ASe	4AXc
11	2AMf	1AAa	4AZe	3ARd
12	1AAa	2AMf	3ARd	4AZe
13	2AKd	1ABb	3AUg	4AVa
14	1ABb	2AKd	4AVa	3AUg

2 Blocks and 42 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ARd	1AGg	2BHf	2AXc
2	1AGg	1ARd	2AXc	2BHf
3	1AEe	1AOa	2BAf	2BLc
4	1AOa	1AEe	2BLc	2BAf
5	1ALe	1ABb	2BJa	2BPg
6	1ABb	1ALe	2BPg	2BJa
7	1ACc	1AHa	2AYd	2BDb
8	1AHa	1ACc	2BDb	2AYd
9	1ADd	1ATf	2Blg	2BKb
10	1ATf	1ADd	2BKb	2Blg
11	1AAa	1AFf	2AZe	2BBg
12	1AFf	1AAa	2BBg	2AZe
13	1ASe	1AJc	2BOf	2AWb
14	1AJc	1ASe	2AWb	2BOf
15	1ANg	1Alb	2BNe	2BEc
16	1Alb	1ANg	2BEc	2BNe
17	1APb	1AMf	2AVa	2BFd
18	1AMf	1APb	2BFd	2AVa
19	1AUg	1AQc	2BCa	2Bmd
20	1AQc	1AUg	2Bmd	2BCa
21	1AKd	1AKd	2BGe	2BGe

6 Blocks and 42 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2AKd	1ACc	4BBg	5BHf
2	1ACc	2AKd	5BHf	4BBg
3	2ALe	1ADd	4AWb	5BCa
4	1ADd	2ALe	5BCa	4AWb
5	1AFf	2AJc	5BDb	4AZe
6	2AJc	1AFf	4AZe	5BDb
7	2AMf	1ABb	4AVa	5BEc
8	1ABb	2AMf	5BEc	4AVa
9	1AAa	2ANg	5BGe	4BAf
10	2ANg	1AAa	4BAf	5BGe
11	1AGg	2AHa	5BFd	4AXc
12	2AHa	1AGg	4AXc	5BFd
13	1AEe	2Alb	4AYd	5Blg
14	2Alb	1AEe	5Blg	4AYd
15	3AQc	3AUg	6BKb	6BJa
16	3AUg	3AQc	6BJa	6BKb
17	3AOa	3ARd	6BOf	6BNe
18	3ARd	3AOa	6BNe	6BOf
19	3APb	3ATf	6BPg	6Bmd
20	3ATf	3APb	6Bmd	6BPg
21	3ASe	3ASe	6BLc	6BLc

3 Blocks and 42 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AEe	2BBg	2AVa	1AMf
2	2BBg	1AEe	1AMf	2AVa
3	2AXc	2ASe	1Alb	1AAa
4	2ASe	2AXc	1AAa	1Alb
5	1AJc	1ABb	2AUg	2AZe
6	1ABb	1AJc	2AZe	2AUg
7	1ALe	1ANg	2ATf	2AYd
8	1ANg	1ALe	2AYd	2ATf
9	2AOa	1ADd	1AFf	2APb
10	1ADd	2AOa	2APb	1AFf
11	1AHa	2ARd	1AGg	2AQc
12	2ARd	1AHa	2AQc	1AGg
13	2AWb	2BAf	1AKd	1ACc
14	2BAf	2AWb	1ACc	1AKd
15	3BDb	3BPg	3Bmd	3BJa
16	3BPg	3BDb	3BJa	3Bmd
17	3BCa	3BHf	3BEc	3BNe
18	3BHf	3BCa	3BNe	3BEc
19	3BLc	3BOf	3BKb	3Blg
20	3BOf	3BLc	3Blg	3BKb
21	3BFd	3BFd	3BGe	3BGe

2 Blocks and 56 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AHa	1AQc	2BYb	2BHf
2	1AQc	1AHa	2BHf	2BYb
3	1AFf	1AKd	2Blg	2BCa
4	1AKd	1AFf	2BCa	2Blg
5	1ALe	1Alb	2BSc	2CAd
6	1Alb	1ALe	2CAd	2BSc
7	1ASe	1ATf	2BRb	2BFd
8	1ATf	1ASe	2BFd	2BRb
9	1ADd	1AJc	2BOf	2BPg
10	1AJc	1ADd	2BPg	2BOf
11	1ARd	1ACc	2BQa	2BGe
12	1ACc	1ARd	2BGe	2BQa
13	1AVa	1AYd	2CBe	2CCf
14	1AYd	1AVa	2CCf	2CBe
15	1AEe	1BBg	2BVf	2BZc
16	1BBg	1AEe	2BZc	2BVf
17	1AWb	1AMf	2BUe	2BWg
18	1AMf	1AWb	2BWg	2BUe
19	1AZe	1AXc	2BJa	2CDg
20	1AXc	1AZe	2CDg	2BJa
21	1ABb	1AUg	2BXa	2BNe
22	1AUg	1ABb	2BNe	2BXa
23	1BAf	1AOa	2BEc	2BKb
24	1AOa	1BAf	2BKb	2BEc
25	1ANg	1AAa	2BDb	2BTd
26	1AAa	1ANg	2BTd	2BDb
27	1AGg	1APb	2BLc	2Bmd
28	1APb	1AGg	2Bmd	2BLc

4 Blocks and 56 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AHa	1AEe	3BMd	3BEc
2	1AEe	1AHa	3BEc	3BMd
3	1ALe	1AAa	3BHf	3BLc
4	1AAa	1ALe	3BLc	3BHf
5	1AKd	1AFf	3BGe	3BKb
6	1AFf	1AKd	3BKb	3BGe
7	1ACc	1ADd	3Blg	3BOf
8	1ADd	1ACc	3BOf	3Blg
9	1AJc	1AMf	3BPg	3BDb
10	1AMf	1AJc	3BDb	3BPg
11	1ANg	1ABb	3BJa	3BFd
12	1ABb	1ANg	3BFd	3BJa
13	1AGg	1Alb	3BCa	3BNe
14	1Alb	1AGg	3BNe	3BCa
15	2AYd	2AUg	4BXa	4CCf
16	2AUg	2AYd	4CCf	4BXa
17	2AZe	2APb	4BWg	4BZc
18	2APb	2AZe	4BZc	4BWg
19	2BBg	2ASe	4BTd	4BSc
20	2ASe	2BBg	4BSc	4BTd
21	2AXc	2AVa	4BYb	4BVf
22	2AVa	2AXc	4BVf	4BYb
23	2AWb	2AQc	4BQa	4CAd
24	2AQc	2AWb	4CAd	4BQa
25	2ARd	2ATf	4BRb	4CBe
26	2ATf	2ARd	4CBe	4BRb
27	2AOa	2BAf	4BUe	4CDg
28	2BAf	2AOa	4CDg	4BUe

8 Blocks and 56 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ABb	2AMf	6BPg	5BEc
2	2AMf	1ABb	5BEc	6BPg
3	1AGg	2ALe	6BMd	5BCa
4	2ALe	1AGg	5BCa	6BMd
5	2AJc	1AFf	6BNe	5Blg
6	1AFf	2AJc	5Blg	6BNe
7	1ACc	2AKd	6BKb	5BGe
8	2AKd	1ACc	5BGe	6BKb
9	2ANg	1ADd	5BHf	6BJa
10	1ADd	2ANg	6BJa	5BHf
11	2AHa	1AEe	6BOf	5BDb
12	1AEe	2AHa	5BDb	6BOf
13	1AAa	2Alb	6BLc	5BFd
14	2Alb	1AAa	5BFd	6BLc
15	4BAf	3ARd	7BRb	8CBe
16	3ARd	4BAf	8CBe	7BRb
17	4AZe	3AQc	7BVf	8BXa
18	3AQc	4AZe	8BXa	7BVf
19	4AWb	3ATf	8CDg	7BTd
20	3ATf	4AWb	7BTd	8CDg
21	4AYd	3AOa	8CCf	7BSc
22	3AOa	4AYd	7BSc	8CCf
23	4BBg	3APb	7BQa	8BZc
24	3APb	4BBg	8BZc	7BQa
25	3ASe	4AVa	8BYb	7BWg
26	4AVa	3ASe	7BWg	8BYb
27	4AXc	3AUg	7BUe	8CAd
28	3AUg	4AXc	8CAd	7BUe

Eight-plex system

2 Blocks and 28 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1ADd	1AGg	1AFf	1AAa	2ASe	2AQc	2APb	2AVa
2	1AGg	1ADd	1AAa	1AFf	2AQc	2ASe	2AVa	2APb
3	1ACc	1AMf	1AKd	1ABb	2AUg	2AOa	2AXc	2AZe
4	1AMf	1ACc	1ABb	1AKd	2AOa	2AUg	2AZe	2AXc
5	1AEe	1AHa	1AJc	1ANg	2ATf	2AWb	2BBg	2ARd
6	1AHa	1AEe	1ANg	1AJc	2AWb	2ATf	2ARd	2BBg
7	1Alb	1Alb	1ALe	1ALe	2AYd	2AYd	2BAf	2BAf

4 Blocks and 28 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFf	1AAa	2ANg	2AJc	3ASe	3ARd	4BBg	4AWb
2	1AAa	1AFf	2AJc	2ANg	3ARd	3ASe	4AWb	4BBg
3	1AGg	1AEe	2Alb	2AMf	3AQc	3ATf	4AVa	4AYd
4	1AEe	1AGg	2AMf	2Alb	3ATf	3AQc	4AYd	4AVa
5	1ACc	1ADd	2ALe	2AHa	3AUg	3APb	4BAf	4AZe
6	1ADd	1ACc	2AHa	2ALe	3APb	3AUg	4AZe	4BAf
7	1ABb	1ABb	2AKd	2AKd	3AOa	3AOa	4AXc	4AXc

2 Blocks and 56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AZe	1Alb	1ANg	1AMf	2BQa	2BSc	2BMd	2BZc
2	1Alb	1AZe	1AMf	1ANg	2BSc	2BQa	2BZc	2BMd
3	1AGg	1ADd	1AWb	1AJc	2BCa	2CCf	2BNe	2BXa
4	1ADd	1AGg	1AJc	1AWb	2CCf	2BCa	2BXa	2BNe
5	1ACc	1ATf	1AVa	1AKd	2BRb	2CAd	2CDg	2CBe
6	1ATf	1ACc	1AKd	1AVa	2CAd	2BRb	2CBe	2CDg
7	1AAa	1AQc	1ALe	1APb	2BDb	2BFd	2BPg	2BHf
8	1AQc	1AAa	1APb	1ALe	2BFd	2BDb	2BHf	2BPg
9	1AHa	1AFf	1AYd	1AEe	2Blg	2BGe	2BEc	2BYb
10	1AFf	1AHa	1AEe	1AYd	2BGe	2Blg	2BYb	2BEc
11	1ASe	1ARd	1BAf	1BBg	2BOf	2BLc	2BJa	2BKb
12	1ARd	1ASe	1BBg	1BAf	2BLc	2BOf	2BKb	2BJa
13	1ABb	1AUg	1AOa	1AXc	2BUe	2BWg	2BTd	2BVf
14	1AUg	1ABb	1AXc	1AOa	2BWg	2BUe	2BVf	2BTd

4 Blocks and 56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AEe	1ANg	2AYd	2AOa	3BEc	3BDb	4CDg	4CCf
2	1ANg	1AEe	2AOa	2AYd	3BDb	3BEc	4CCf	4CDg
3	1AFf	1AJc	2AWb	2AVa	3BPg	3BMd	4BQa	4BUe
4	1AJc	1AFf	2AVa	2AWb	3BMd	3BPg	4BUe	4BQa
5	1Alb	1ACc	2ASe	2APb	3BJa	3BOf	4BWg	4BTd
6	1ACc	1Alb	2APb	2ASe	3BOf	3BJa	4BTd	4BWg
7	1AKd	1AGg	2BAf	2ATf	3BNe	3BKb	4BXa	4BZc
8	1AGg	1AKd	2ATf	2BAf	3BKb	3BNe	4BZc	4BXa
9	1ALe	1ADd	2AXc	2AUg	3BCa	3BHf	4CAd	4BRb
10	1ADd	1ALe	2AUg	2AXc	3BHf	3BCa	4BRb	4CAd
11	1AHa	1AMf	2AZe	2AQc	3BFd	3Blg	4CBe	4BYb
12	1AMf	1AHa	2AQc	2AZe	3Blg	3BFd	4BYb	4CBe
13	1ABb	1AAa	2ARd	2BBg	3BLc	3BGe	4BSc	4BVf
14	1AAa	1ABb	2BBg	2ARd	3BGe	3BLc	4BVf	4BSc

8 Blocks and 56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2AMf	1AEe	3AQc	4BBg	6BMd	5BDb	7BQa	8CBe
2	1AEe	2AMf	4BBg	3AQc	5BDb	6BMd	8CBe	7BQa
3	1AAa	2Alb	3ASe	4BAf	5BCa	6BPg	7BSc	8CAd
4	2Alb	1AAa	4BAf	3ASe	6BPg	5BCa	8CAd	7BSc
5	2AHa	1AFf	3AUg	4AYd	6BLc	5BGe	7BWg	8BYb
6	1AFf	2AHa	4AYd	3AUg	5BGe	6BLc	8BYb	7BWg
7	1ADd	2AJc	4AZe	3APb	6BKb	5Blg	8BXa	7BVf
8	2AJc	1ADd	3APb	4AZe	5Blg	6BKb	7BVf	8BXa
9	1AGg	2ALe	3ARd	4AVa	5BFd	6BOf	8BZc	7BRb
10	2ALe	1AGg	4AVa	3ARd	6BOf	5BFd	7BRb	8BZc
11	2ANg	1ACc	4AWb	3AOa	5BEc	6BNe	8CCf	7BTd
12	1ACc	2ANg	3AOa	4AWb	6BNe	5BEc	7BTd	8CCf
13	2AKd	1ABb	3ATf	4AXc	5BHf	6BJa	7BUe	8CDg
14	1ABb	2AKd	4AXc	3ATf	6BJa	5BHf	8CDg	7BUe

Eight Treatments

Four-plex system

2 Blocks and 16 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Cc	2Ld	2Og
2	1Cc	1Bb	2Og	2Ld
3	1Aa	1Dd	2Jb	2Me
4	1Dd	1Aa	2Me	2Jb
5	1Ff	1Hh	2Ia	2Kc
6	1Hh	1Ff	2Kc	2Ia
7	1Ee	1Gg	2Ph	2Nf
8	1Gg	1Ee	2Nf	2Ph

3 Blocks and 24 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1Bb	1Ff	1Gg	1Ee
2	1Ff	1Bb	1Ee	1Gg
3	1Cc	1Hh	1Dd	1Aa
4	1Hh	1Cc	1Aa	1Dd
5	2Ld	2Og	2Ia	2Me
6	2Og	2Ld	2Me	2Ia
7	2Kc	2Nf	2Jb	2Ph
8	2Nf	2Kc	2Ph	2Jb
9	3Td	3Ue	3Vf	3Xh
10	3Ue	3Td	3Xh	3Vf
11	3Qa	3Wg	3Rb	3Sc
12	3Wg	3Qa	3Sc	3Rb

2 Blocks and 32 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AJb	1AHh	2BBd	2BCe
2	1AHh	1AJb	2BCe	2BBd
3	1ADd	1AGg	2BFh	2AVf
4	1AGg	1ADd	2AVf	2BFh
5	1ANf	1Ala	2BAc	2ATd
6	1Ala	1ANf	2ATd	2BAc
7	1AMe	1ALd	2ASc	2BEg
8	1ALd	1AMe	2BEg	2ASc
9	1AKc	1AFf	2AWg	2ARb
10	1AFf	1AKc	2ARb	2AWg
11	1ACc	1APh	2AYa	2AZb
12	1APh	1ACc	2AZb	2AYa
13	1AAa	1AEe	2AXh	2BDf
14	1AEe	1AAa	2BDf	2AXh
15	1AOg	1ABb	2AQa	2AUe
16	1ABb	1AOg	2AUe	2AQa

4 Blocks and 32 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ABb	1AAa	1ADd	1ACc
2	1AAa	1ABb	1ACc	1ADd
3	1AFf	1AHh	1AEe	1AGg
4	1AHh	1AFf	1AGg	1AEe
5	2ALd	2AOg	2AJb	2APh
6	2AOg	2ALd	2APh	2AJb
7	2AKc	2ANf	2AMe	2Ala
8	2ANf	2AKc	2Ala	2AMe
9	3AXh	3AWg	3AQa	3ASc
10	3AWg	3AXh	3ASc	3AQa
11	3ARb	3AUe	3ATd	3AVf
12	3AUe	3ARb	3AVf	3ATd
13	4BCe	4AYa	4AZb	4BEg
14	4AYa	4BCe	4BEg	4AZb
15	4BBd	4BAc	4BFh	4BDf
16	4BAc	4BBd	4BDf	4BFh

4 Blocks and 32 Phase 1 Experimental units
(Higher Residual EDF)

Run	Tag			
	114	115	116	117
1	1ABb	1AHh	3ATd	3AQa
2	1AHh	1ABb	3AQa	3ATd
3	1AAa	1AFf	3AUe	3AWg
4	1AFf	1AAa	3AWg	3AUe
5	1AGg	1ACc	3ARb	3AVf
6	1ACc	1AGg	3AVf	3ARb
7	1AEe	1ADd	3ASc	3AXh
8	1ADd	1AEe	3AXh	3ASc
9	2AKc	2Ala	4BBd	4BEg
10	2Ala	2AKc	4BEg	4BBd
11	2AMe	2ALd	4BDf	4AZb
12	2ALd	2AMe	4AZb	4BDf
13	2AOg	2AJb	4BCe	4BFh
14	2AJb	2AOg	4BFh	4BCe
15	2ANf	2APh	4AYa	4BAc
16	2APh	2ANf	4BAc	4AYa

5 Blocks and 40 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ACc	1AGg	1AAa	1AEe
2	1AGg	1ACc	1AEe	1AAa
3	1AHh	1ADd	1ABb	1AFf
4	1ADd	1AHh	1AFf	1ABb
5	2AOg	2AKc	2APh	2AJb
6	2AKc	2AOg	2AJb	2APh
7	2ANf	2AMe	2Ala	2ALd
8	2AMe	2ANf	2ALd	2Ala
9	3AQa	3ARb	3ASc	3ATd
10	3ARb	3AQa	3ATd	3ASc
11	3AUe	3AWg	3AXh	3AVf
12	3AWg	3AUe	3AVf	3AXh
13	4BBd	4BDf	4BEg	4BAc
14	4BDf	4BBd	4BAc	4BEg
15	4AZb	4AYa	4BFh	4BCe
16	4AYa	4AZb	4BCe	4BFh
17	5BLf	5BHb	5Blc	5BKe
18	5BHb	5BLf	5BKe	5Blc
19	5BNh	5BGa	5BMg	5BJd
20	5BGa	5BNh	5BJd	5BMg

2 Blocks and 48 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1ADd	1ARb	2BNh	2BUg
2	1ARb	1ADd	2BUg	2BNh
3	1ANf	1ALd	2BSe	2BFh
4	1ALd	1ANf	2BFh	2BSe
5	1ATd	1AGg	2BAc	2BKe
6	1AGg	1ATd	2BKe	2BAc
7	1ASc	1Ala	2BTf	2BCe
8	1Ala	1ASc	2BCe	2BTf
9	1AHh	1AAa	2BBd	2BDf
10	1AAa	1AHh	2BDf	2BBd
11	1AKc	1AWg	2BJd	2BGa
12	1AWg	1AKc	2BGa	2BJd
13	1AXh	1AUe	2AYa	2BEg
14	1AUe	1AXh	2BEg	2AYa
15	1AVf	1AOg	2BOa	2BHb
16	1AOg	1AVf	2BHb	2BOa
17	1ABb	1AQa	2Blc	2BVh
18	1AQa	1ABb	2BVh	2Blc
19	1ACc	1AJb	2BRd	2BLf
20	1AJb	1ACc	2BLf	2BRd
21	1AEe	1APh	2AZb	2BQc
22	1APh	1AEe	2BQc	2AZb
23	1AMe	1AFf	2BPb	2BMg
24	1AFf	1AMe	2BMg	2BPb

3 Blocks and 48 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1APh	1AJb	1AKc	1AGg
2	1AJb	1APh	1AGg	1AKc
3	1AMe	1ACc	1AHh	1ANf
4	1ACc	1AMe	1ANf	1AHh
5	1AEe	1AFf	1Ala	1ALd
6	1AFf	1AEe	1ALd	1Ala
7	1ABb	1AOg	1AAa	1ADd
8	1AOg	1ABb	1ADd	1AAa
9	2ATd	2BCe	2AWg	2BFh
10	2BCe	2ATd	2BFh	2AWg
11	2BBd	2BEg	2BDf	2BAc
12	2BEg	2BBd	2BAc	2BDf
13	2AXh	2ARb	2AQa	2AUe
14	2ARb	2AXh	2AUe	2AQa
15	2ASc	2AYa	2AVf	2AZb
16	2AYa	2ASc	2AZb	2AVf
17	3BTf	3BJd	3BPb	3BNh
18	3BJd	3BTf	3BNh	3BPb
19	3BLf	3BUg	3BKe	3BHb
20	3BUg	3BLf	3BHb	3BKe
21	3BQc	3BOa	3BSe	3BMg
22	3BOa	3BQc	3BMg	3BSe
23	3BVh	3BGa	3Blc	3BRd
24	3BGa	3BVh	3BRd	3Blc

6 Blocks and 48 Phase 1 Experimental units (Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1AHh	1ACc	1ADd	1AEe
2	1ACc	1AHh	1AEe	1ADd
3	1AGg	1AAa	1ABb	1AFf
4	1AAa	1AGg	1AFf	1ABb
5	2AKc	2ALd	2AOg	2Ala
6	2ALd	2AKc	2Ala	2AOg
7	2AJb	2APh	2ANf	2AMe
8	2APh	2AJb	2AMe	2ANf
9	3ARb	3AWg	3ASc	3AXh
10	3AWg	3ARb	3AXh	3ASc
11	3AVf	3AQa	3ATd	3AUe
12	3AQa	3AVf	3AUe	3ATd
13	4BCe	4BAc	4AYa	4AZb
14	4BAc	4BCe	4AZb	4AYa
15	4BBd	4BDf	4BFh	4BEg
16	4BDf	4BBd	4BEg	4BFh
17	5BKe	5BMg	5BHb	5BJd
18	5BMg	5BKe	5BJd	5BHb
19	5BGa	5BNh	5BLf	5Blc
20	5BNh	5BGa	5Blc	5BLf
21	6BRd	6BPb	6BVh	6BOa
22	6BPb	6BRd	6BOa	6BVh
23	6BTf	6BSe	6BUg	6BQc
24	6BSe	6BTf	6BQc	6BUg

6 Blocks and 48 Phase 1 Experimental units
(Higher Residual EDF)

Run	Tag			
	114	115	116	117
1	1AGg	1AAa	4BCe	4BDf
2	1AAa	1AGg	4BDf	4BCe
3	1AEe	1ACc	4AZb	4BFh
4	1ACc	1AEe	4BFh	4AZb
5	1AHh	1Aff	4BBd	4AYa
6	1Aff	1AHh	4AYa	4BBd
7	1ABb	1Add	4BAc	4BEg
8	1Add	1ABb	4BEg	4BAc
9	2AOg	2ALd	5BNh	5BKe
10	2ALd	2AOg	5BKe	5BNh
11	2APh	2AJb	5BMg	5BGa
12	2AJb	2APh	5BGa	5BMg
13	2Ala	2AKc	5BLf	5BHb
14	2AKc	2Ala	5BHb	5BLf
15	2ANf	2AMe	5BJd	5Blc
16	2AMe	2ANf	5Blc	5BJd
17	3ASc	3AXh	6BOa	6BSe
18	3AXh	3ASc	6BSe	6BOa
19	3AQa	3ATd	6BUg	6BQc
20	3ATd	3AQa	6BQc	6BUg
21	3AWg	3AUe	6BTf	6BPb
22	3AUe	3AWg	6BPb	6BTf
23	3ARb	3AVf	6BRd	6BVh
24	3AVf	3ARb	6BVh	6BRd

7 Blocks and 56 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AGg	1Add	1AAa	1Aff
2	1Add	1AGg	1Aff	1AAa
3	1ACc	1ABb	1AHh	1AEe
4	1ABb	1ACc	1AEe	1AHh
5	2AMe	2APh	2ALd	2Ala
6	2APh	2AMe	2Ala	2ALd
7	2AOg	2ANf	2AJb	2AKc
8	2ANf	2AOg	2AKc	2AJb
9	3AQa	3AUe	3ASc	3AWg
10	3AUe	3AQa	3AWg	3ASc
11	3ARb	3AVf	3ATd	3AXh
12	3AVf	3ARb	3AXh	3ATd
13	4AYa	4BBd	4AZb	4BAc
14	4BBd	4AYa	4BAc	4AZb
15	4BFh	4BEg	4BCe	4BDf
16	4BEg	4BFh	4BDf	4BCe
17	5BHb	5BNh	5BGa	5BMg
18	5BNh	5BHb	5BMg	5BGa
19	5BLf	5BJd	5BKe	5Blc
20	5BJd	5BLf	5Blc	5BKe
21	6BQc	6BOa	6BTf	6BVh
22	6BOa	6BQc	6BVh	6BTf
23	6BUg	6BSe	6BPb	6BRd
24	6BSe	6BUg	6BRd	6BPb
25	7CBf	7BXb	7BWa	7CAe
26	7BXb	7CBf	7CAe	7BWa
27	7BZd	7BYc	7CDh	7CCg
28	7BYc	7BZd	7CCg	7CDh

2 Blocks and 64 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AGg	1AXh	2BXb	2CEa
2	1AXh	1AGg	2CEa	2BXb
3	1ABb	1ACc	2CHd	2CLh
4	1ACc	1ABb	2CLh	2CHd
5	1APh	1AWg	2Cle	2BQc
6	1AWg	1APh	2BQc	2Cle
7	1BAc	1AOg	2BZd	2BSe
8	1AOg	1BAc	2BSe	2BZd
9	1ALd	1Aff	2BUg	2BOa
10	1Aff	1ALd	2BOa	2BUg
11	1AKc	1BBd	2BVh	2BWa
12	1BBd	1AKc	2BWa	2BVh
13	1AHh	1AQa	2BLf	2CCg
14	1AQa	1AHh	2CCg	2BLf
15	1ATd	1ARb	2CDh	2CBf
16	1ARb	1ATd	2CBf	2CDh
17	1ANf	1Ala	2BPb	2BKe
18	1Ala	1ANf	2BKe	2BPb
19	1AZb	1BDf	2CGc	2CKg
20	1BDf	1AZb	2CKg	2CGc
21	1BCe	1AYa	2BRd	2BMg
22	1AYa	1BCe	2BMg	2BRd
23	1AAa	1AJb	2BJd	2Blc
24	1AJb	1AAa	2Blc	2BJd
25	1AEe	1Add	2CFb	2BTf
26	1Add	1AEe	2BTf	2CFb
27	1AUe	1BEg	2BNh	2BHb
28	1BEg	1AUe	2BHb	2BNh
29	1AVf	1AMe	2BGa	2BYc
30	1AMe	1AVf	2BYc	2BGa
31	1ASc	1BFh	2CJf	2CAe
32	1BFh	1ASc	2CAe	2CJf

4 Blocks and 64 Phase 1 Experimental units
(Higher Residual DF)

Run	Tag			
	114	115	116	117
1	1AKc	1ABb	1AAa	1APh
2	1ABb	1AKc	1APh	1AAa
3	1AJb	1AGg	1AMe	1ADd
4	1AGg	1AJb	1ADd	1AMe
5	1AEe	1AHh	1ANf	1Ala
6	1AHh	1AEe	1Ala	1ANf
7	1AFf	1ALd	1AOg	1ACc
8	1ALd	1AFf	1ACc	1AOg
9	2BBd	2AUe	2BDf	2AXh
10	2AUe	2BBd	2AXh	2BDf
11	2ATd	2BAc	2AZb	2BEg
12	2BAc	2ATd	2BEg	2AZb
13	2ARb	2BFh	2AQa	2ASc
14	2BFh	2ARb	2ASc	2AQa
15	2AWg	2AYa	2AVf	2BCe
16	2AYa	2AWg	2BCe	2AVf
17	3BMg	3Blc	3BSe	3BVh
18	3Blc	3BMg	3BVh	3BSe
19	3BJd	3BOa	3BHb	3BKe
20	3BOa	3BJd	3BKe	3BHb
21	3BTf	3BQc	3BGa	3BUg
22	3BQc	3BTf	3BUg	3BGa
23	3BPb	3BLf	3BNh	3BRd
24	3BLf	3BPb	3BRd	3BNh
25	4CAe	4BWa	4BZd	4CGc
26	4BWa	4CAe	4CGc	4BZd
27	4CCg	4CDh	4CFb	4CJf
28	4CDh	4CCg	4CJf	4CFb
29	4Cle	4CBf	4BYc	4BXb
30	4CBf	4Cle	4BXb	4BYc
31	4CLh	4CEa	4CHd	4CKg
32	4CEa	4CLh	4CKg	4CHd

4 Blocks and 64 Phase 1 Experimental units
(Higher Residual EDF)

Run	Tag			
	114	115	116	117
1	1AJb	1AOg	3BVh	3Blc
2	1AOg	1AJb	3Blc	3BVh
3	1APh	1Ala	3BQc	3BJd
4	1Ala	1APh	3BJd	3BQc
5	1AGg	1AHh	3BLf	3BKe
6	1AHh	1AGg	3BKe	3BLf
7	1ANf	1ALd	3BGa	3BSe
8	1ALd	1ANf	3BSe	3BGa
9	1AKc	1AFf	3BHb	3BOa
10	1AFf	1AKc	3BOa	3BHb
11	1ABb	1ADd	3BTf	3BUg
12	1ADd	1ABb	3BUg	3BTf
13	1AMe	1ACc	3BRd	3BMg
14	1ACc	1AMe	3BMg	3BRd
15	1AEe	1AAa	3BNh	3BPb
16	1AAa	1AEe	3BPb	3BNh
17	2BBd	2BEg	4CFb	4BWa
18	2BEg	2BBd	4BWa	4CFb
19	2AXh	2ATd	4Cle	4CGc
20	2ATd	2AXh	4CGc	4Cle
21	2BDf	2BCe	4BZd	4BXb
22	2BCe	2BDf	4BXb	4BZd
23	2AVf	2AQa	4CDh	4CKg
24	2AQa	2AVf	4CKg	4CDh
25	2BFh	2AYa	4CCg	4CHd
26	2AYa	2BFh	4CHd	4CCg
27	2ASc	2AUe	4CBf	4CEa
28	2AUe	2ASc	4CEa	4CBf
29	2AZb	2BAc	4CJf	4CLh
30	2BAc	2AZb	4CLh	4CJf
31	2ARb	2AWg	4CAe	4BYc
32	2AWg	2ARb	4BYc	4CAe

8 Blocks and 64 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AEe	1AAa	1ADd	1AFf
2	1AAa	1AEe	1AFf	1ADd
3	1AGg	1ACc	1AHh	1ABb
4	1ACc	1AGg	1ABb	1AHh
5	2APh	2AMe	2AKc	2Ala
6	2AMe	2APh	2Ala	2AKc
7	2AJb	2ANf	2AOg	2ALd
8	2ANf	2AJb	2ALd	2AOg
9	3AVf	3AQa	3ARb	3AXh
10	3AQa	3AVf	3AXh	3ARb
11	3ATd	3AUe	3ASc	3AWg
12	3AUe	3ATd	3AWg	3ASc
13	4BEg	4BDf	4AYa	4BAc
14	4BDf	4BEg	4BAc	4AYa
15	4BBd	4BFh	4BCe	4AZb
16	4BFh	4BBd	4AZb	4BCe
17	5BNh	5BJd	5BGa	5BMg
18	5BJd	5BNh	5BMg	5BGa
19	5BHb	5Blc	5BKe	5BLf
20	5Blc	5BHb	5BLf	5BKe
21	6BOa	6BPb	6BUg	6BSe
22	6BPb	6BOa	6BSe	6BUg
23	6BQc	6BVh	6BRd	6BTf
24	6BVh	6BQc	6BTf	6BRd
25	7BXb	7BWa	7BYc	7BZd
26	7BWa	7BXb	7BZd	7BYc
27	7CAe	7CCg	7CBf	7CDh
28	7CCg	7CAe	7CDh	7CBf
29	8CGc	8CHd	8CFb	8CEa
30	8CHd	8CGc	8CEa	8CFb
31	8CJf	8CKg	8CLh	8Cle
32	8CKg	8CJf	8Cle	8CLh

8 Blocks and 64 Phase 1 Experimental units (Higher Residual EDF)

Run	Tag			
	114	115	116	117
1	1AHh	4AYa	8CFb	5BJd
2	4AYa	1AHh	5BJd	8CFb
3	4BDf	4BFh	8CGc	5BGa
4	4BFh	4BDf	5BGa	8CGc
5	1ACc	1ABb	5BNh	8CKg
6	1ABb	1ACc	8CKg	5BNh
7	1AAa	1AEe	5BLf	8CHd
8	1AEe	1AAa	8CHd	5BLf
9	2AOg	2APh	6BTf	6BRd
10	2APh	2AOg	6BRd	6BTf
11	2AMe	2AKc	6BUg	6BOa
12	2AKc	2AMe	6BOa	6BUg
13	2Ala	2ALd	6BQc	6BPb
14	2ALd	2Ala	6BPb	6BQc
15	2AJb	2ANf	6BVh	6BSe
16	2ANf	2AJb	6BSe	6BVh
17	3ASc	3AUe	7BXb	7BWa
18	3AUe	3ASc	7BWa	7BXb
19	3AQa	3AXh	7CCg	7BZd
20	3AXh	3AQa	7BZd	7CCg
21	3ATd	3AWg	7CBf	7BYc
22	3AWg	3ATd	7BYc	7CBf
23	3AVf	3ARb	7CAe	7CDh
24	3ARb	3AVf	7CDh	7CAe
25	1AGg	4BCe	8CLh	5Blc
26	4BCe	1AGg	5Blc	8CLh
27	1AFf	4AZb	5BMg	8CEa
28	4AZb	1AFf	8CEa	5BMg
29	4BEg	4BBd	5BHb	8Cle
30	4BBd	4BEg	8Cle	5BHb
31	1ADd	4BAc	5BKe	8CJf
32	4BAc	1ADd	8CJf	5BKe

Eight-plex system

2 Blocks and 16 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Cc	1Hh	1Aa	1Gg	1Dd	1Ff	1Bb	1Ee
2	1Hh	1Cc	1Gg	1Aa	1Ff	1Dd	1Ee	1Bb
3	2Ld	2Jb	2Kc	2Ph	2Me	2Ia	2Og	2Nf
4	2Jb	2Ld	2Ph	2Kc	2Ia	2Me	2Nf	2Og

2 Blocks and 16 Phase 1 Experimental units (Higher Residual EDF)

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Gg	1Cc	1Ee	1Bb	2Ph	2Ia	2Nf	2Ld
2	1Cc	1Gg	1Bb	1Ee	2Ia	2Ph	2Ld	2Nf
3	1Hh	1Ff	1Dd	1Aa	2Og	2Jb	2Me	2Kc
4	1Ff	1Hh	1Aa	1Dd	2Jb	2Og	2Kc	2Me

3 Blocks and 24 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1Gg	1Bb	1Cc	1Dd	1Aa	1Ee	1Ff	1Hh
2	1Bb	1Gg	1Dd	1Cc	1Ee	1Aa	1Hh	1Ff
3	2Kc	2Me	2Ph	2Jb	2Ld	2Nf	2Ia	2Og
4	2Me	2Kc	2Jb	2Ph	2Nf	2Ld	2Og	2Ia
5	3Xh	3Vf	3Ue	3Qa	3Wg	3Rb	3Sc	3Td
6	3Vf	3Xh	3Qa	3Ue	3Rb	3Wg	3Td	3Sc

2 Blocks and 32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1ALd	1AMe	1APh	1ABb	2AWg	2AQa	2AVf	2BAc
2	1AMe	1ALd	1ABb	1APh	2AQa	2AWg	2BAc	2AVf
3	1AAa	1ANf	1AOg	1AEe	2ASc	2AXh	2BBd	2AZb
4	1ANf	1AAa	1AEe	1AOg	2AXh	2ASc	2AZb	2BBd
5	1AGg	1AJb	1Ala	1ACc	2BDf	2ATd	2BCe	2BFh
6	1AJb	1AGg	1ACc	1Ala	2ATd	2BDf	2BFh	2BCe
7	1AKc	1AHh	1AFf	1ADd	2ARb	2AUe	2AYa	2BEg
8	1AHh	1AKc	1ADd	1AFf	2AUe	2ARb	2BEg	2AYa

4 Blocks and 32 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AHh	1ABb	2ALd	2AOg	3AVf	3AUe	4AYa	4BAc
2	1ABb	1AHh	2AOg	2ALd	3AUe	3AVf	4BAc	4AYa
3	1AEe	1AFf	2APh	2AKc	3AQa	3AWg	4AZb	4BBd
4	1AFf	1AEe	2AKc	2APh	3AWg	3AQa	4BBd	4AZb
5	1ACc	1AAa	2AMe	2ANf	3ARb	3ATd	4BEg	4BFh
6	1AAa	1ACc	2ANf	2AMe	3ATd	3ARb	4BFh	4BEg
7	1ADd	1AGg	2Ala	2AJb	3ASc	3AXh	4BCe	4BDf
8	1AGg	1ADd	2AJb	2Ala	3AXh	3ASc	4BDf	4BCe

5 Blocks and 40 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFf	1ADd	1AAa	1AGg	1AHh	1ABb	1AEe	1ACc
2	1ADd	1AFf	1AGg	1AAa	1ABb	1AHh	1ACc	1AEe
3	2APh	2AMe	2AJb	2AOg	2ANf	2AKc	2ALd	2Ala
4	2AMe	2APh	2AOg	2AJb	2AKc	2ANf	2Ala	2ALd
5	3AWg	3ASc	3ARb	3AUe	3AVf	3ATd	3AXh	3AQa
6	3ASc	3AWg	3AUe	3ARb	3ATd	3AVf	3AQa	3AXh
7	4BCe	4AZb	4BDf	4BFh	4BBd	4AYa	4BEg	4BAc
8	4AZb	4BCe	4BFh	4BDf	4AYa	4BBd	4BAc	4BEg
9	5BNh	5BGa	5Blc	5BJd	5BMg	5BKe	5BHb	5BLf
10	5BGa	5BNh	5BJd	5Blc	5BKe	5BMg	5BLf	5BHb

2 Blocks and 48 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1ASc	1AVf	1ATd	1AEe	2BVh	2BPb	2BOa	2BEg
2	1AVf	1ASc	1AEe	1ATd	2BPb	2BVh	2BEg	2BOa
3	1AHh	1AAa	1ADd	1AUe	2BQc	2BMg	2BHb	2BLf
4	1AAa	1AHh	1AUe	1ADd	2BMg	2BQc	2BLf	2BHb
5	1ABb	1AOg	1Ala	1APh	2BRd	2BDf	2BKe	2Blc
6	1AOg	1ABb	1APh	1Ala	2BDf	2BRd	2Blc	2BKe
7	1AMe	1ACc	1AXh	1ARb	2BGa	2BUg	2BTf	2BBd
8	1ACc	1AMe	1ARb	1AXh	2BUg	2BGa	2BBd	2BTf
9	1AQa	1ALd	1ANf	1AWg	2BAc	2BSe	2AZb	2BNh
10	1ALd	1AQa	1AWg	1ANf	2BSe	2BAc	2BNh	2AZb
11	1AGg	1AJb	1AFf	1AKc	2BFh	2BJd	2AYa	2BCe
12	1AJb	1AGg	1AKc	1AFf	2BJd	2BFh	2BCe	2AYa

3 Blocks and 48 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFf	1AHh	1AGg	1AEe	1AAa	1AJb	1ACc	1ADd
2	1AHh	1AFf	1AEe	1AGg	1AJb	1AAa	1ADd	1ACc
3	1AOg	1AKc	1ALd	1ANf	1APh	1ABb	1AMe	1Ala
4	1AKc	1AOg	1ANf	1ALd	1ABb	1APh	1Ala	1AMe
5	2AUe	2AZb	2ATd	2AQa	2AVf	2ASc	2AWg	2AXh
6	2AZb	2AUe	2AQa	2ATd	2ASc	2AVf	2AXh	2AWg
7	2BFh	2BEg	2AYa	2ARb	2BAc	2BCe	2BDf	2BBd
8	2BEg	2BFh	2ARb	2AYa	2BCe	2BAc	2BBd	2BDf
9	3BGa	3BJd	3BMg	3Blc	3BKe	3BLf	3BNh	3BHb
10	3BJd	3BGa	3Blc	3BMg	3BLf	3BKe	3BHb	3BNh
11	3BPb	3BTf	3BQc	3BVh	3BRd	3BUg	3BOa	3BSe
12	3BTf	3BPb	3BVh	3BQc	3BUg	3BRd	3BSe	3BOa

6 Blocks and 48 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AHh	1AFf	1AGg	1ACc	1ABb	1ADd	1AAa	1AEe
2	1AFf	1AHh	1ACc	1AGg	1ADd	1ABb	1AEe	1AAa
3	2Ala	2APh	2ALd	2AMe	2AJb	2AKc	2ANf	2AOg
4	2APh	2Ala	2AMe	2ALd	2AKc	2AJb	2AOg	2ANf
5	3AUe	3AWg	3ARb	3AVf	3ATd	3AQa	3AXh	3ASc
6	3AWg	3AUe	3ARb	3AVf	3ATd	3AQa	3AXh	3ASc
7	4BAc	4AZb	4BFh	4AYa	4BCe	4BDf	4BEg	4BBd
8	4AZb	4BAc	4AYa	4BFh	4BDf	4BCe	4BBd	4BEg
9	5BJd	5Blc	5BHb	5BLf	5BGa	5BMg	5BNh	5BKe
10	5Blc	5BJd	5BLf	5BHb	5BMg	5BGa	5BKe	5BNh
11	6BRd	6BUg	6BQc	6BSe	6BVh	6BTf	6BPb	6BOa
12	6BUg	6BRd	6BSe	6BQc	6BTf	6BVh	6BOa	6BPb

7 Blocks and 56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1ADd	1ACc	1AGg	1AFf	1ABb	1AEe	1AHh	1AAa
2	1ACc	1ADd	1AFf	1AGg	1AEe	1ABb	1AAa	1AHh
3	2AId	2AOg	2ALd	2AKc	2APh	2AJb	2AMe	2ANf
4	2AOg	2AId	2AKc	2ALd	2AJb	2APh	2ANf	2AMe
5	3AVf	3AUe	3AXh	3ATd	3AWg	3ASc	3ARb	3AQa
6	3AUe	3AVf	3ATd	3AXh	3ASc	3AWg	3AQa	3ARb
7	4AZb	4BBd	4AYa	4BEg	4BCe	4BDf	4BAc	4BFh
8	4BBd	4AZb	4BEg	4AYa	4BDf	4BCe	4BFh	4BAc
9	5BNh	5BKe	5BHb	5Blc	5BGa	5BLf	5BJd	5BMg
10	5BKe	5BNh	5Blc	5BHb	5BLf	5BGa	5BMg	5BJd
11	6BVh	6BQc	6BSe	6BTf	6BOa	6BRd	6BUg	6BPb
12	6BQc	6BVh	6BTf	6BSe	6BRd	6BOa	6BPb	6BUg
13	7BXb	7BWA	7CDh	7CAe	7CCg	7BZd	7BYc	7CBf
14	7BWA	7BXb	7CAe	7CDh	7BZd	7CCg	7CBf	7BYc

2 Blocks and 64 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1ALd	1BCe	1AXh	1ABb	2BYc	2CEa	2CKg	2CJf
2	1BCe	1ALd	1ABb	1AXh	2CEa	2BYc	2CJf	2CKg
3	1AQa	1AGg	1AVf	1BFh	2BSe	2BQc	2BPb	2CHd
4	1AGg	1AQa	1BFh	1AVf	2BQc	2BSe	2CHd	2BPb
5	1ASc	1ATd	1BDf	1AOg	2BVh	2BHb	2CAe	2BOa
6	1ATd	1ASc	1AOg	1BDf	2BHb	2BVh	2BOa	2CAe
7	1ACc	1BEg	1AUe	1Ala	2BLf	2BJd	2BNh	2BXb
8	1BEg	1ACc	1Ala	1AUe	2BJd	2BLf	2BXb	2BNh
9	1AMe	1ANf	1AZb	1BAc	2BWA	2CCg	2CLh	2BRd
10	1ANf	1AMe	1BAc	1AZb	2CCg	2BWA	2BRd	2CLh
11	1AJb	1AFf	1AWg	1AKc	2CDh	2BZd	2BGa	2Cle
12	1AFf	1AJb	1AKc	1AWg	2BZd	2CDh	2Cle	2BGa
13	1ARb	1APh	1AAa	1BBd	2CBf	2BKe	2CGc	2BMg
14	1APh	1ARb	1BBd	1AAa	2BKe	2CBf	2BMg	2CGc
15	1AYa	1AHh	1ADd	1AEe	2CFb	2BUg	2Blc	2BTf
16	1AHh	1AYa	1AEe	1ADd	2BUg	2CFb	2BTf	2Blc

4 Blocks and 64 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AEe	1AGg	2BBd	2ASc	3BTf	3BVh	4CFb	4BWA
2	1AGg	1AEe	2ASc	2BBd	3BVh	3BTf	4BWA	4CFb
3	1AAa	1AFf	2AWg	2ARb	3BRd	3BSe	4BYc	4CLh
4	1AFf	1AAa	2ARb	2AWg	3BSe	3BRd	4CLh	4BYc
5	1ALd	1AOg	2BFh	2BAc	3BHb	3BOa	4Cle	4CJf
6	1AOg	1ALd	2BAc	2BFh	3BOa	3BHb	4CJf	4Cle
7	1AKc	1ANf	2AZb	2AXh	3BJd	3BKe	4CEa	4CCg
8	1ANf	1AKc	2AXh	2AZb	3BKe	3BJd	4CCg	4CEa
9	1AHh	1ACc	2AYa	2BCe	3BMg	3BPb	4CBf	4CHd
10	1ACc	1AHh	2BCe	2AYa	3BPb	3BMg	4CHd	4CBf
11	1ABb	1ADd	2BEg	2BDf	3BGa	3BNh	4CAe	4CGc
12	1ADd	1ABb	2BDf	2BEg	3BNh	3BGa	4CGc	4CAe
13	1Ala	1APh	2AUe	2AVf	3Blc	3BUg	4BZd	4BXb
14	1APh	1Ala	2AVf	2AUe	3BUg	3Blc	4BXb	4BZd
15	1AMe	1AJb	2AQa	2ATd	3BQc	3BLf	4CKg	4CDh
16	1AJb	1AMe	2ATd	2AQa	3BLf	3BQc	4CDh	4CKg

8 Blocks and 64 Phase 1 Experimental units (Higher Residual DF)

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AHh	1AAa	1ACc	1ABb	1AEe	1ADd	1AFf	1AGg
2	1AAa	1AHh	1ABb	1ACc	1ADd	1AEe	1AGg	1AFf
3	2AOg	2AMe	2Ala	2AKc	2ALd	2APh	2ANf	2AJb
4	2AMe	2AOg	2AKc	2Ala	2APh	2ALd	2AJb	2ANf
5	3AUe	3ATd	3ARb	3AXh	3AVf	3AWg	3ASc	3AQa
6	3ATd	3AUe	3AXh	3ARb	3AWg	3AVf	3AQa	3ASc
7	4BDf	4AYa	4BEg	4BFh	4AZb	4BCe	4BBd	4BAc
8	4AYa	4BDf	4BFh	4BEg	4BCe	4AZb	4BAc	4BBd
9	5Blc	5BHb	5BLf	5BJd	5BNh	5BGa	5BMg	5BKe
10	5BHb	5Blc	5BJd	5BLf	5BGa	5BNh	5BKe	5BMg
11	6BUg	6BQc	6BRd	6BTf	6BOa	6BPb	6BSe	6BVh
12	6BQc	6BUg	6BTf	6BRd	6BPb	6BOa	6BVh	6BSe
13	7CBf	7BXb	7CAe	7BWa	7BYc	7CCg	7CDh	7BZd
14	7BXb	7CBf	7BWa	7CAe	7CCg	7BYc	7BZd	7CDh
15	8CLh	8CHd	8Cle	8CKg	8CJf	8CGc	8CEa	8CFb
16	8CHd	8CLh	8CKg	8Cle	8CGc	8CJf	8CFb	8CEa

8 Blocks and 64 Phase 1 Experimental units (Higher Residual EDF)

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AFf	1ADd	3AXh	3ASc	5BMg	5BKe	7BWa	7BXb
2	1ADd	1AFf	3ASc	3AXh	5BKe	5BMg	7BXb	7BWa
3	1AEe	1ABb	3AWg	3AVf	5BGa	5BNh	7BYc	7BZd
4	1ABb	1AEe	3AVf	3AWg	5BNh	5BGa	7BZd	7BYc
5	1AAa	1AGg	3AUe	3ARb	5BJd	5Blc	7CDh	7CBf
6	1AGg	1AAa	3ARb	3AUe	5Blc	5BJd	7CBf	7CDh
7	1ACc	1AHh	3ATd	3AQa	5BLf	5BHb	7CAe	7CCg
8	1AHh	1ACc	3AQa	3ATd	5BHb	5BLf	7CCg	7CAe
9	2ANf	2APh	4AZb	4BEg	6BQc	6BRd	8CEa	8Cle
10	2APh	2ANf	4BEg	4AZb	6BRd	6BQc	8Cle	8CEa
11	2AOg	2ALd	4BAc	4BCe	6BVh	6BOa	8CJf	8CFb
12	2ALd	2AOg	4BCe	4BAc	6BOa	6BVh	8CFb	8CJf
13	2AKc	2Ala	4BBd	4BDf	6BPb	6BSe	8CKg	8CLh
14	2Ala	2AKc	4BDf	4BBd	6BSe	6BPb	8CLh	8CKg
15	2AJb	2AMe	4AYa	4BFh	6BTf	6BUg	8CHd	8CGc
16	2AMe	2AJb	4BFh	4AYa	6BUg	6BTf	8CGc	8CHd

Appendix F

Tables of properties of optimal designs when Phase 1 is RCBD

Phase 1 Design				Phase 2 Design								
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum			
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors
												E
2	2	4	1	2	4	0	0	0	0	1	Yes	1
	3	6	2	3		1	0	0	1	1	No(1/9)	0.8889
	2	8	5	4		0	1	0	0	4	Yes	1
	4		3			1	0	0	0	3	Yes	1
	5	10	4	5		2	0	0	1	3	No(1/25)	0.96
	2	12	9	6		0	2	0	0	7	Yes	1
	3		8			2	0	0	1	7	Yes	1
	6		5			2	0	0	0	5	Yes	1
	7	14	6	7		3	0	0	1	5	No(1/49)	0.9796
	2	16	13	8		0	3	0	0	10	Yes	1
	4		11			3	0	0	1	10	Yes	1
						1	2	0	0	9	Yes	1
	8		7			3	0	0	0	7	Yes	1
	2	8	5	2	8	0	0	0	2	3	Yes	1
	4		3			0	0	0	0	3	Yes	1
	2	12	9	3		0	1	0	2	6	No(1/9)	0.8889
	3		8			1	0	0	3	5	No(1/9)	0.8889
	6		5			1	0	0	2	3	No(1/9)	0.8889
	2	16	13	4		0	1	0	2	10	Yes	1
	4		11			0	1	0	0	10	Yes	1
	8		7			1	0	0	0	7	Yes	1

Phase 1 Design				Phase 2 Design									
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors	
												e _i	E
3	2	6	2	3	4	0	1	1	0	1	Yes	1, 3/4	0.8571
	2	12	8	6		0	2	2	0	6	Yes	15/16 (2)	0.9375
	4		6			0	2	2	0	4	Yes	15/16 (2)	0.9375
	2	18	14	9		0	4	2	0	10	Yes	23/24, 7/8	0.9148
	3		13			1	3	2	1	9	Yes	23/24, 7/8	0.9148
	6		10			1	3	2	0	7	Yes	23/24, 7/8	0.9148
	2	24	20	12		0	5	2	0	15	Yes	15/16 (2)	0.9375
	4		18			1	4	2	0	14	Yes	15/16 (2)	0.9375
	8		14			1	4	2	0	10	Yes	15/16 (2)	0.9375
	2	12	8	3	8	0	1	1	2	5	Yes	1, 15/16	0.9677
	4		6			0	1	1	0	5	Yes	1, 15/16	0.9677
	2	24	20	6		0	2	2	2	16	Yes	63/64(2)	0.9844
	4		18			0	2	2	0	16	Yes	63/64(2)	0.9844
	8		14			0	2	2	0	12	Yes	63/64(2)	0.9844

Phase 1 Design				Phase 2 Design									
v	n_b	n_p	Between Plots Residual DF	n_{Runs}	n_{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	Treatment efficiency factors	
												e_i	E
4	2	8	3	4	4	0	1	0	0	2	Yes	1(3)	1
	3	12	6	6		2	0	0	1	5	No(1/9)	1(2), 8/9	0.96
	2	16	11	8		0	3	0	0	8	Yes	1(3)	1
	4		9			3	0	0	1	8	Yes	1(3)	1
						1	2	0	0	7	Yes	1(3)	1
	5	20	12	10		4	0	0	1	11	No(1/25)	1(2), 24/25	0.9863
	2	24	19	12		0	5	0	0	14	Yes	1(3)	1
	3		18			2	3	0	1	14	Yes	1(3)	1
	6		15			5	0	0	1	14	Yes	1(3)	1
		28		14		2	3	0	0	12	Yes	1(3)	1
	7		18			6	0	0	1	17	No(1/49)	1(2), 48/49	0.9931
	2	32	27	16		0	7	0	0	20	Yes	1(3)	1
	4		25			3	4	0	1	20	Yes	1(3)	1
						1	6	0	0	19	Yes	1(3)	1
	8		21			7	0	0	1	20	Yes	1(3)	1
						3	4	0	0	17	Yes	1(3)	1

Phase 1 Design				Phase 2 Design									
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors	
												e _i	E
4	2	8	3	4	8	0	0	0	2	1	No (1/2)	1, 1/2(2)	0.6
	3	12	6	6		1	0	0	3	3	No (1/9)	8/9 (3)	0.8889
	2	16	11	8		0	1	0	2	8	Yes	1 (3)	1
	4		9			0	1	0	0	8	Yes	1 (3)	1
	5	20	12	10		2	0	0	3	9	No (1/25)	24/25 (3)	0.96
	2	24	19	12		0	2	0	2	15	No (1/18)	1, 17/18(2)	0.9623
	3		18			2	0	0	3	15	No (1/18)	1, 17/18(2)	0.9623
	6		15			2	0	0	2	13	No (1/18)	1, 17/18(2)	0.9623
	7	28	18	14		3	0	0	3	15	No (1/49)	48/49 (3)	0.9796
	2	32	27	16		0	3	0	2	22	Yes	1 (3)	1
	4		25			0	3	0	0	22	Yes	1 (3)	1
	8		21			1	2	0	0	19	Yes	1 (3)	1

Phase 1 Design				Phase 2 Design									
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors	
												e _i	E
5	2	10	4	5	4	0	2	2	0	2	Yes	1(2), 7/8, 5/8	0.8434
	2	20	14	10		0	4	4	0	10	Yes	15/16(4)	0.9375
	4		12			0	4	4	0	8	Yes	15/16(4)	0.9375
	2	30	24	15		0	7	4	0	17	Yes	23/24(2), 11/12, 5/6	0.9137
	3		23			1	6	4	1	16	Yes	23/24(2), 11/12, 5/6	0.9137
	6		20			1	6	4	0	14	Yes	23/24(2), 11/12, 5/6	0.9137
	2	40	34	20		0	9	4	0	25	Yes	15/16 (4),	0.9375
	4		32			1	8	4	0	24	Yes	15/16 (4),	0.9375
	8		28			1	8	4	0	20	Yes	15/16 (4),	0.9375
	2	20	14	5	8	0	2	2	2	10	Yes	1 (2), 15/16 (2)	0.9677
	4		12			0	2	2	0	10	Yes	1 (2), 15/16 (2)	0.9677
	2	40	34	10		0	4	4	2	28	Yes	0.994(2), 0.959(2)	0.9763
	4		32			0	4	4	0	28	Yes	0.994(2), 0.959(2)	0.9763
	8		28			0	4	4	0	24	Yes	0.994(2), 0.959(2)	0.9763

Phase 1 Design				Phase 2 Design									
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors	
												e _i	E
6	2	12	5	6	4	0	2	2	0	3	Yes	1(3), 3/4(2)	0.8824
	3	18	10	9		1	3	3	1	6	No(1/9)	1, 0.894, 8/9, 5/6, 0.606	0.8204
	2	24	17	12		0	5	4	0	12	Yes	1, 15/16(2), 13/16(2)	0.8937
	4		15	1		4	4	0	11	Yes	1, 15/16(2), 13/16(2)	0.8937	
	5	30	20	15		2	5	4	1	14	No(1/25)	0.96, 0.95, 0.85(2), 0.75	0.8650
	2	36	29	18		0	8	4	0	21	Yes	1, 7/8(4)	0.8974
	3		28			2	6	4	1	21	Yes	1, 7/8(4)	0.8974
	6		25			2	6	4	0	19	Yes	1, 7/8(4)	0.8974
	7	42	30	21		3	7	5	1	22	No(1/49)	0.947, 0.919, 25/28, 0.854, 0.795	0.8784
	2	48	41	24		0	11	5	0	30	Yes	15/16(2), 7/8 (3)	0.8990
	4		39			3	8	5	1	30	Yes	15/16(2), 7/8 (3)	0.8990
						1	10	5	0	29	Yes	15/16(2), 7/8 (3)	0.8990
	8		35			3	8	5	0	27	Yes	15/16(2), 7/8 (3)	0.8990

Phase 1 Design				Phase 2 Design									
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors	
												e _i	E
6	2	12	5	3	8	0	1	1	2	2	No(4/9)	1(2), 3/4, 2/3, 1/3	0.6383
	2	24	17	6		0	2	2	2	13	Yes	1(3), 5/16(2)	0.974
	4		15			0	2	2	0	13	Yes	1(3), 5/16(2)	0.974
	2	36	29	9		0	4	4	2	23	No(4/81)	0.979, 0.960, 0.942, 0.938, 0.903	0.9438
	3		28			1	3	3	3	22	No(4/81)	0.974, 0.963, 0.958, 0.926, 0.901	0.9437
	6		25			1	3	3	2	19	No(4/81)	0.974, 0.963, 0.958, 0.926 0.901	0.9437
	2	48	41	12		0	5	4	2	34	Yes	1, 63/64(2), 61/64(2)	0.9746
	4		39			0	5	4	0	34	Yes	1, 63/64(2), 61/64(2)	0.9746
	8		35			1	4	4	0	31	Yes	1, 63/64(2), 61/64(2)	0.9746

Phase 1 Design				Phase 2 Design									
v	n_b	n_p	Between Plots Residual DF	n_{Runs}	n_{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	Treatment efficiency factors	
												e_i	E
7	2	14	6	7	4	0	3	3	0	3	Yes	1(3), 7/8, 5/8, 1/2	0.7749
	2	28	20	14		0	6	6	0	14	Yes	7/8(6)	0.875
	4		18			0	6	6	0	12	Yes	7/8(6)	0.875
	2	42	34	21		0	10	6	0	24	Yes	7/8(5), 19/24	0.8599
	3		33			1	9	6	1	23	Yes	0.934, 7/8(3), 0.816, 19/24	0.8586
	6		30			1	9	6	0	21	Yes	0.934, 7/8(3), 0.816, 19/24	0.8586
	2	56	48	28		0	13	6	0	35	Yes	7/8(6)	0.875
	4		46			1	12	6	0	34	Yes	7/8(6)	0.875
	8		42			1	12	6	0	30	Yes	7/8(6)	0.875
	2	28	20	7	8	0	3	3	2	15	Yes	1(3), 31/32(2), 7/8	0.9666
	4		18			0	3	3	0	15	Yes	1(3), 31/32(2), 7/8	0.9666
	2	56	48	14		0	6	6	2	40	Yes	63/64 (6)	0.9844
	4		46			0	6	6	0	40	Yes	63/64 (6)	0.9844
	8		42			0	6	6	0	36	Yes	63/64 (6)	0.9844

Phase 1 Design				Phase 2 Design									
v	n_b	n_p	Between Plots Residual DF	n_{Runs}	n_{Tags}	<i>Between Runs stratum</i>			<i>Between Plots within Blocks within Runs stratum</i>				
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	<i>Treatment efficiency factors</i>	
												e_i	E
8	2	16	7	8	4	0	3	3	0	4	Yes	1(4), 3/4(2), 1/2	0.8077
	3	24	14	12		2	3	3	1	10	No (1/9)	1(3), 8/9 2/3(3)	0.8116
	2	32	23	16		0	7	7	0	16	Yes	0.963(2), 7/8(2), 0.787(2), 3/4	0.8498
	4		21			3	4	4	1	16	Yes	1(3), 3/4(4)	0.84
	1		6			6	0	15	Yes	1, 7/8(4), 3/4(2)	0.8497		
	5	40	28	20		4	5	5	1	22	No(1/25)	1, 24/25, 4/5(5)	0.8442
	2	48	39	24		0	11	6	0	28	Yes	1, 5/6(6)	0.8537
	3		38			2	9	6	1	28	Yes	1, 5/6(6)	0.8537
	6		35			5	6	6	1	28	Yes	1, 5/6(6)	0.8537
	2		9			6	0	26	Yes	1, 5/6(6)	0.8537		
	7	56	42	28		6	7	7	1	34	Yes	6/7(6), 41/49	0.8542
	2	64	55	32		0	15	7	0	40	Yes	0.919(2), 7/8, 0.831(2), 13/16 (2)	0.8550
	4		53			3	12	5	1	40	Yes	0.919(2), 7/8, 0.831(2), 13/16 (2)	0.8550
	1		14			7	0	39	Yes	0.919(2), 7/8, 0.831(2), 13/16 (2)	0.8550		
	7		8			7	1	40	Yes	7/8 (6), 3/4	0.8547		
	3		12			7	0	36	Yes	0.919(2), 7/8, 0.831(2), 13/16 (2)	0.8550		

Phase 1 Design				Phase 2 Design										
v	n _b	n _p	Between Plots Residual DF	n _{Runs}	n _{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum					
						Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag ⊥ Trt	Treatment efficiency factors		
												e _i	E	
8	2	16	7	4	8	1	0	0	3	4	No(3/10)	1(4), 3/4(2), 1/2	0.8077	
						0	1	0	2	4	No(1/2)	1(5), 1/2(2)	0.7778	
	3	24	14	6		2	0	0	3	11	No(1/9)	1(4), 8/9(3)	0.9492	
	2	32	23	8		0	3	0	2	18	Yes	1(7)	1	
	4					21	0	3	0	0	18	Yes	1(7)	1
	5	40	28	10		4	0	0	3	25	No(1/25)	1(4), 24/25(3)	0.9825	
	2	48	39	12		0	5	0	2	32	No(0.06)	1(5), 0.944(2)	0.9835	
	3					38	2	3	0	3	32	No(1/30)	1(4), 35/36(2), 17/18	0.9837
	6					35	5	0	0	3	32	No(1/30)	1(4), 35/36(2), 17/18	0.9837
	7	56	42	14		6	0	0	7	39	Yes	1(4), 48/49(3)	0.9912	
	2	64	55	16		0	7	0	0	46	Yes	1(7)	1	
	4					53	0	7	0	0	46	Yes	1(7)	1
	8					49	7	0	0	3	46	Yes	1(7)	1
							1	6	0	0	43	Yes	1(7)	1

Appendix G

Tables of optimal designs when Phase 1 is BIBD

Four Treatments

Four-plex system

4 Blocks and 12 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	3 lb	1 Cd	2 Fa	4 Lc
2	1 Cd	3 lb	4 Lc	2 Fa
3	3 Ha	1 Bc	4 Kb	2 Ed
4	1 Bc	3 Ha	2 Ed	4 Kb
5	3 Gd	1 Ab	2 Dc	4 Ja
6	1 Ab	3 Gd	4 Ja	2 Dc

Eight-plex system

4 Blocks and 12 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1 Cd	1 Bc	2 Ed	2 Dc	3 lb	3 Ha	4 Kb	4 Ja
2	1 Bc	1 Cd	2 Dc	2 Ed	3 Ha	3 lb	4 Ja	4 Kb
3	1 Ab	1 Ab	2 Fa	2 Fa	3 Gd	3 Gd	4 Lc	4 Lc

Five Treatments

Four-plex system

5 Blocks and 20 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1 Bc	1 Ab	1 De	1 Cd
2	1 Ab	1 Bc	1 Cd	1 De
3	2 Ge	2 Ha	2 Ec	2 Fd
4	2 Ha	2 Ge	2 Fd	2 Ec
5	3 Ka	3 ld	3 Lb	3 Je
6	3 ld	3 Ka	3 Je	3 Lb
7	4 Pc	4 Me	4 Ob	4 Na
8	4 Me	4 Pc	4 Na	4 Ob
9	5 Td	5 Rb	5 Sc	5 Qa
10	5 Rb	5 Td	5 Qa	5 Sc

Eight-plex system

5 Blocks and 20 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1 Cd	2 Ec	1 De	2 Ha	1 Bc	2 Ge	2 Fd	1 Ab
2	2 Ec	1 Cd	2 Ha	1 De	2 Ge	1 Bc	1 Ab	2 Fd
3	4 Na	3 Je	3 Lb	4 Pc	4 Ob	3 ld	3 Ka	4 Me
4	3 Je	4 Na	4 Pc	3 Lb	3 ld	4 Ob	4 Me	3 Ka
5	5 Rb	5 Rb	5 Td	5 Td	5 Qa	5 Qa	5 Sc	5 Sc

Six Treatments

Four-plex system

6 Blocks and 30 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2AJa	2Alf	5AXc	5AYd
2	2Alf	2AJa	5AYd	5AXc
3	1ADe	2AGd	4ASb	5AVa
4	2AGd	1ADe	5AVa	4ASb
5	1ABc	1ACd	5AWb	5AUf
6	1ACd	1ABc	5AUf	5AWb
7	2AFc	2AHe	4ARa	4AQf
8	2AHe	2AFc	4AQf	4ARa
9	1AAb	1AEf	4APe	4ATc
10	1AEf	1AAb	4ATc	4APe
11	3ANa	3AMf	6BDe	6BAb
12	3AMf	3ANa	6BAb	6BDe
13	3AOB	3AKd	6AZa	6BBc
14	3AKd	3AOB	6BBc	6AZa
15	3ALe	3ALe	6BCd	6BCd

Seven Treatments

Four-plex system

7 Blocks and 28 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	1AAAd	1ACf	1ADg	1ABe
2	1ACf	1AAAd	1ABe	1ADg
3	2AHg	2AEb	2AFc	2AGf
4	2AEb	2AHg	2AGf	2AFc
5	3AKd	3ALe	3Alb	3AJc
6	3ALe	3AKd	3AJc	3Alb
7	4APg	4ANc	4AOe	4AMa
8	4ANc	4APg	4AMa	4AOe
9	5ATf	5ARc	5ASd	5AQa
10	5ARc	5ATf	5AQa	5ASd
11	6AUa	6AWe	6AVb	6AXf
12	6AWe	6AUa	6AXf	6AVb
13	7AZb	7AYa	7BBg	7BAAd
14	7AYa	7AZb	7BAAd	7BBg

7 Blocks and 42 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2AJf	2AKg	1ABc	1ACd
2	2AKg	2AJf	1ACd	1ABc
3	1AEf	1AAb	1AFg	2Ale
4	1AAb	1AEf	2Ale	1AFg
5	2AGc	1ADe	2AHd	2ALa
6	1ADe	2AGc	2ALa	2AHd
7	4AVa	3ARb	3APg	4ASe
8	3ARb	4AVa	4ASe	3APg
9	3ANe	3AMd	3AOf	4AWb
10	3AMd	3ANe	4AWb	3AOf
11	3AQa	4AUg	4ATf	4AXc
12	4AUg	3AQa	4AXc	4ATf
13	5BDd	6BGb	6BHc	5AZg
14	6BGb	5BDd	5AZg	6BHc
15	6BEg	6BJe	6BId	5BAa
16	6BJe	6BEg	5BAa	6BId
17	6BFa	5BCc	5AYf	5BBb
18	5BCc	6BFa	5BBb	5AYf
19	7BNd	7BPf	7BLb	7BKa
20	7BPf	7BNd	7BKa	7BLb
21	7BMc	7BMc	7BOe	7BOe

Eight-plex system

7 Blocks and 28 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	2AGf	1ABe	1ADg	2AEb	1ACf	2AHg	2AFc	1AAAd
2	1ABe	2AGf	2AEb	1ADg	2AHg	1ACf	1AAAd	2AFc
3	4APg	3AJc	3AKd	4AOe	3ALe	4ANc	4AMa	3Alb
4	3AJc	4APg	4AOe	3AKd	4ANc	3ALe	3Alb	4AMa
5	5AQa	6AVb	5ARc	6AXf	5ASd	6AUa	5ATf	6AWe
6	6AVb	5AQa	6AXf	5ARc	6AUa	5ASd	6AWe	5ATf
7	7BAAd	7BAAd	7AYa	7AYa	7AZb	7AZb	7BBg	7BBg

Eight Treatments

Four-plex system

8 Blocks and 56 Phase 1 Experimental units

Run	Tag			
	114	115	116	117
1	2Alc	2AKe	6BJa	6BMd
2	2AKe	2Alc	6BMd	6BJa
3	2ANh	2ALf	5BCa	5BHg
4	2ALf	2ANh	5BHg	5BCa
5	1ADe	1AFg	6BKb	6BPh
6	1AFg	1ADe	6BPh	6BKb
7	2AJd	1AEf	6BOg	5BDb
8	1AEf	2AJd	5BDb	6BOg
9	2AMg	1AGh	5BFd	6BLc
10	1AGh	2AMg	6BLc	5BFd
11	1ABc	2AHa	5Blh	6BNe
12	2AHa	1ABc	6BNe	5Blh
13	1ACd	1AAb	5BEc	5BGf
14	1AAb	1ACd	5BGf	5BEc
15	3ASf	4BBh	8BXa	7BTd
16	4BBh	3ASf	7BTd	8BXa
17	4AZf	4AXc	7BUe	7BWh
18	4AXc	4AZf	7BWh	7BUe
19	3AOa	4AYe	7BVf	8BYb
20	4AYe	3AOa	8BYb	7BVf
21	3ATg	3ARe	8CAd	8CCf
22	3ARe	3ATg	8CCf	8CAd
23	4BAg	3APb	8CBe	7BSc
24	3APb	4BAg	7BSc	8CBe
25	3AUh	3AQd	7BRb	7BQa
26	3AQd	3AUh	7BQa	7BRb
27	4AWb	4AVa	8BZc	8CDg
28	4AVa	4AWb	8CDg	8BZc

Eight-plex system

8 Blocks and 56 Phase 1 Experimental units

Run	Tag							
	113	114	115	116	117	118	119	121
1	1AGh	2AJe	3ARg	4AZa	5BCf	6BNc	7BUd	8BYb
2	2AJe	1AGh	4AZa	3ARg	6BNc	5BCf	8BYb	7BUd
3	2AKf	1ACd	4BAb	3APe	6BJg	5BEh	7BTc	8BXa
4	1ACd	2AKf	3APe	4BAb	5BEh	6BJg	8BXa	7BTc
5	1AAb	2AMh	4BBc	3ATa	5Bld	6BPe	8CDg	7BWf
6	2AMh	1AAb	3ATa	4BBc	6BPe	5Bld	7BWf	8CDg
7	2ANa	1ABc	4AXg	3AQf	6BKd	5BGb	7BVe	8CAd
8	1ABc	2ANa	3AQf	4AXg	5BGb	6BKd	8CAd	7BVe
9	2Ald	1ADe	3AUb	4AWf	6BLa	5BDg	7BQh	8BZc
10	1ADe	2Ald	4AWf	3AUb	5BDg	6BLa	8BZc	7BQh
11	1AEf	2ALg	3AOd	4AYh	5BHc	6BMb	8CBe	7BRa
12	2ALg	1AEf	4AYh	3AOd	6BMb	5BHc	7BRa	8CBe
13	2AHc	1AFg	3ASh	4AVe	5BFa	6BOd	7BSb	8CCf
14	1AFg	2AHc	4AVe	3ASh	6BOd	5BFa	8CCf	7BSb

Appendix H

Tables of properties of optimal designs when Phase 1 is BIBD

Phase 1 Design					Phase 2 Design									
v	n_b	n_p	Between Plots Residual DF	E	n_{Runs}	n_{Tags}	Between Runs stratum			Between Plots within Blocks within Runs stratum				
							Phase 1 Block DF	Phase 1 Plot DF	Trt DF	Tag DF	Residual DF	Tag \perp Trt	Treatment efficiency factors	
													e_i	E
4	4	12	5	0.8889	6	4	0	2	0	0	3	Yes	0.8889 (3)	0.8889
					3	8	0	1	0	0	4		0.8889 (3)	0.8889
5	5	20	11	0.9375	10	4	4	0	4	1	10		0.9375 (4)	0.9375
					5	8	2	0	2	3	8		0.9375 (4)	0.9375
6	6	30	19	0.96	15	4	3	4	6	0	13		0.938, 0.9, 0.874, 0.822, 0.786	0.8606
7	7	28	15	0.875	14	4	6	0	6	1	14		0.875 (6)	0.875
					7	8	3	0	3	3	12		0.875 (6)	0.875
7	7	42	29	0.9722	21	4	6	4	10	1	21		0.921, 0.874, 0.863, 0.840, 0.814, 0.780	0.8462
8	8	56	41	0.9796	28	4	5	8	10	0	29		0.929(2), 0.857, 0.837(2), 0.786(2)	0.8478
					14	8	0	6	0	0	35		0.9796 (7)	0.9796

Bibliography

- Bailey, R. A. (2008). *Design of comparative experiment*. Cambridge University Press.
- Bailey, R. A., Brien, C., et al. (2016). Randomization-based models for multitiered experiments: I. a chain of randomizations. *The Annals of Statistics*, 44(3):1131–1164.
- Blow, D. M. (2002). *Outline of crystallography for biologists*. Oxford University Press.
- Boehm, A. M., Pütz, S., Altenhöfer, D., Sickmann, A., and Falk, M. (2007). Precise protein quantification based on peptide quantification using iTRAQ. *BMC Bioinformatics*, 8:214.
- Brien, C. (1992). Factorial linear model analysis.
- Brien, C. (2011). *dae: Functions useful in the design and ANOVA of experiments*. R package version 2.1-7.
- Brien, C. J. (1983). Analysis of Variance Tables Based on Experimental Structure. *Biometrics*, 39(1):53–59.
- Brien, C. J. and Bailey, R. A. (2006). Multiple randomizations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):571–609.
- Brien, C. J. and Bailey, R. A. (2009). Decomposition tables for experiments I. A chain of randomizations. *The Annals of Statistics*, 37(6B):4184–4213.
- Brien, C. J. and Bailey, R. A. (2010). Decomposition tables for experiments. II. Two-one randomizations. *The Annals of Statistics*, 38(5):3164–3190.
- Brien, C. J., Harch, B. D., Correll, R. L., and Bailey, R. a. (2011). Multiphase Experiments with at Least One Later Laboratory Phase. I. Orthogonal Designs. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(3):422–450.
- Brien, C. J. and Payne, R. W. (1999). Tiers, structure formulae and the analysis of complicated experiments. *Journal of the Royal Statistical Society Series D: The Statistician*, 48(1):41–52.

- Brien, C. J. and Payne, R. W. (2006). *GenStat Reference Manual Release 9, Part 3, Procedure Library PL 17*, chapter AMTIER Pro, pages 73–75. VSN International: Hemel Hempstead.
- Campbell, J. D., Liu, G., Luo, L., Xiao, J., Gerrein, J., Juan-Guardela, B., Tedrow, J., Alekseyev, Y. O., Yang, I. V., Correll, M., et al. (2015). Assessment of microrna differential expression and detection in multiplexed small rna sequencing data. *RNA*, 21(2):164–171.
- Chang, K. (2008). Protein expression data from MudPIT coupled with iTRAQTM: an experimental design for information decomposition and data analysis.
- Choe, L., D’Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., and Lee, K. H. (2007). 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer’s disease. *Proteomics*, 7(20):3651–3660.
- Colinge, J. and Bennett, K. L. (2007). Introduction to computational proteomics. *PLoS Comput Biol*, 3(7):1151–1160.
- Curnow, R. N. (1959). The analysis of a two phase experiment. *Biometrics*, 15(1):60–73.
- Eccleston, J. A. and Hedayat, A. (1974). On the theory of connected designs: Characterization and optimality. *The Annals of Statistics*, 2(6):pp. 1238–1255.
- Eidhammer, I., Flikka, K., Martens, L., and Mikalsen, S.-O. (2007). *Computational Methods for Mass Spectrometry Proteomics*. Wiley-Interscience.
- Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246.
- Fisher, R. (1935). *The Design of Experiments*. Oliver and Boyd.
- Goos, P. (2012). *The optimal design of blocked and split-plot experiments*, volume 164. Springer Science & Business Media.
- Hadi, A. S. (1996). *Matrix Algebra as a tool*. Duxbury Press at Wadsworth Publishing Company.
- James, A. T. and Wilkinson, G. N. (1971). Factorization of the Residual Operator and Canonical Decomposition of Nonorthogonal Factors in the Analysis of Variance. *Biometrika*, 58(2):279.
- Jarrett, R. G. and Ruggiero, K. (2008). Design and analysis of two-phase experiments for gene expression microarrays-part \textnormal{I}. *Biometrics*, 64(1):208–216.

- John, J., Ruggiero, K., and Williams, E. (2002). α_n -designs. *Australian & New Zealand Journal of Statistics*, 44(4):457–465.
- John, J. A. and Whitaker, D. (1993). Construction of resolvable row-column designs using simulated annealing. *Australian Journal of Statistics*, 35(August 1991):237–245.
- John, J. A. and Williams, E. R. (1995). *Cyclic and computer generated Designs*. Chapman & Hall, second edition.
- Kerr, M. K. and Churchill, G. A. (2001a). Experimental design for gene expression microarrays. *Biostatistics*, 2(2):183.
- Kerr, M. K. and Churchill, G. A. (2001b). Statistical design and the analysis of gene expression microarray data. *Genetics Research*, 77(2):123–128.
- Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science (New York, N.Y.)*, 220(4598):671–80.
- McIntyre, G. A. (1955). Design and Analysis of Two Phase Experiments. *Biometrics*, 11(3):324–334.
- Nelder, J. A. (1965a). The Analysis of Randomized Experiments with Orthogonal Block Structure. I. Block Structure and the Null Analysis of Variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283(1393):pp. 147–162.
- Nelder, J. A. (1965b). The Analysis of Randomized Experiments with Orthogonal Block Structure. II. Treatment Structure and the General Analysis of Variance. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 283(1393):163–178.
- Nesvizhskii, A. I. and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics*, 4(10):1419–1440.
- Nunn, B. L. and Timperman, A. T. (2007). Marine proteomics. *Marine Ecology Progress Series*, 332:281–289.
- Oberg, A. and Mahoney, D. (2012). Statistical methods for quantitative mass spectrometry proteomic experiments with labeling. *BMC Bioinformatics*, 13(Suppl 16):S7.
- Oberg, A. L. and Vitek, O. (2009). Statistical design of quantitative mass spectrometry-based proteomic experiments. *Journal of proteome research*, 8(5):2144–56.

- Patterson, H. D. and Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3):pp. 545–554.
- Payne, R. W. (2003). General balance, large data sets and extensions to unbalanced treatment structures. *Computational Statistics & Data Analysis*, 44(1-2):297–304.
- Payne, R. W. and Wilkinson, G. N. (1977). A general algorithm for analysis of variance. *Applied Statistics*.
- Pierce, A., Unwin, R. D., Evans, C. A., Griffiths, S., Carney, L., Zhang, L., Jaworska, E., Lee, C.-F., Blinco, D., Okoniewski, M. J., Miller, C. J., Bitton, D. A., Spooncer, E., and Whetton, A. D. (2008). Eight-channel iTRAQ Enables Comparison of the Activity of Six Leukemogenic Tyrosine Kinases. *Mol Cell Proteomics*, 7(5):853–863.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics*, 3(12):1154–1169.
- Sadygov, R. G., Cociorva, D., and Yates, J. R. (2004). Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods*, 1(3):195–202.
- Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, 6(5):309–316.
- Sciex (2004). *QSTAR[®] XL Hybrid LC/MS/MS System*. Applied Biosystems and MDS Sciex.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley-Interscience.
- Shah, K. (1960). Optimality criteria for incomplete block designs. *The Annals of Mathematical Statistics*, pages 791–794.
- Shapland, E. B., Holmes, V., Reeves, C. D., Sorokin, E., Durot, M., Platt, D., Allen, C., Dean, J., Serber, Z., Newman, J., and Chandran, S. (2015). Low-cost, high-throughput sequencing of dna assemblies using a highly multiplexed nextera process. *ACS Synthetic Biology*, 4(7):860–866. PMID: 25913499.

- Smith, A. M., Heisler, L. E., Onge, R. P. S., Farias-Hesson, E., Wallace, I. M., Bodeau, J., Harris, A. N., Perry, K. M., Giaever, G., Pourmand, N., et al. (2010). Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic acids research*, page gkq368.
- Thelen, J. J. and Peck, S. C. (2007). Quantitative proteomics in plants: choices in abundance. *Plant Cell*, 19(11):3339–3346.
- Washburn, M. P., Wolters, D., and Yates, J. R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol*, 19(3):242–247.
- Whitaker, D., Triggs, C. M., and John, J. A. (1990). Construction of Block Designs Using Mathematical Programming. *Journal of the Royal Statistical Society Series B (Methodological)*, 52(3):497–503.
- Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, 7(3):340–350.
- Wilkinson, G. N. (1970). A General Recursive Procedure for Analysis of Variance. *Biometrika*, 57(1):19.
- Wilkinson, G. N. and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, 22(3):392–399.
- Williams, E. R. and John, J. A. (1996). Row-column factorial designs for use in agricultural field trials. *Applied statistics*, 45(1):39–46.
- Wood, J. T., William, E. R., and Speed, T. P. (1988). Non-orthogonal block structure in two-phase designs. *Australian & New Zealand Journal of Statistics*, 30A:225–237.
- Yates, F. (1936). Incomplete Randomized Blocks. *Annals of Human Genetics*, 7(2):121–140.
- Zeng, I. S., Lumley, T., Ruggiero, K., and Middleditch, M. (2017). A bayesian approach to multivariate and multilevel modelling with non-random missingness for hierarchical clinical proteomics data.