

# Extending the Loop Design for Microarray Experiments

Naomi S. Altman  
Department of Statistics  
Pennsylvania State University

## Abstract

The loop design (Kerr and Churchill, 2001a,b) is a clever application of incomplete blocks of size 2 to 2-channel microarray experiments. In this paper, I discuss the use of optimal design principles to extending the loop design to include more replication, multi-factor experiments, and blocking. Loop and extended loop designs are shown to be more efficient than the reference design for any given number of arrays. Adding new treatments to the design and the effects of missing data are briefly discussed.

Keywords: optimal design; mixed effects; incomplete block design; reference design;

## 1.0 Introduction

Microarray data are both expensive to collect and noisy compared to the signal. This is a perfect situation to apply principles of optimal design to improve efficiency.

A popular design for 2-channel microarray studies is the reference design, in each one channel on each array is used for a reference sample. The same reference is used for all arrays, and the difference or ratio of expression between the treatment sample and the reference sample is usually used as the unit of analysis.

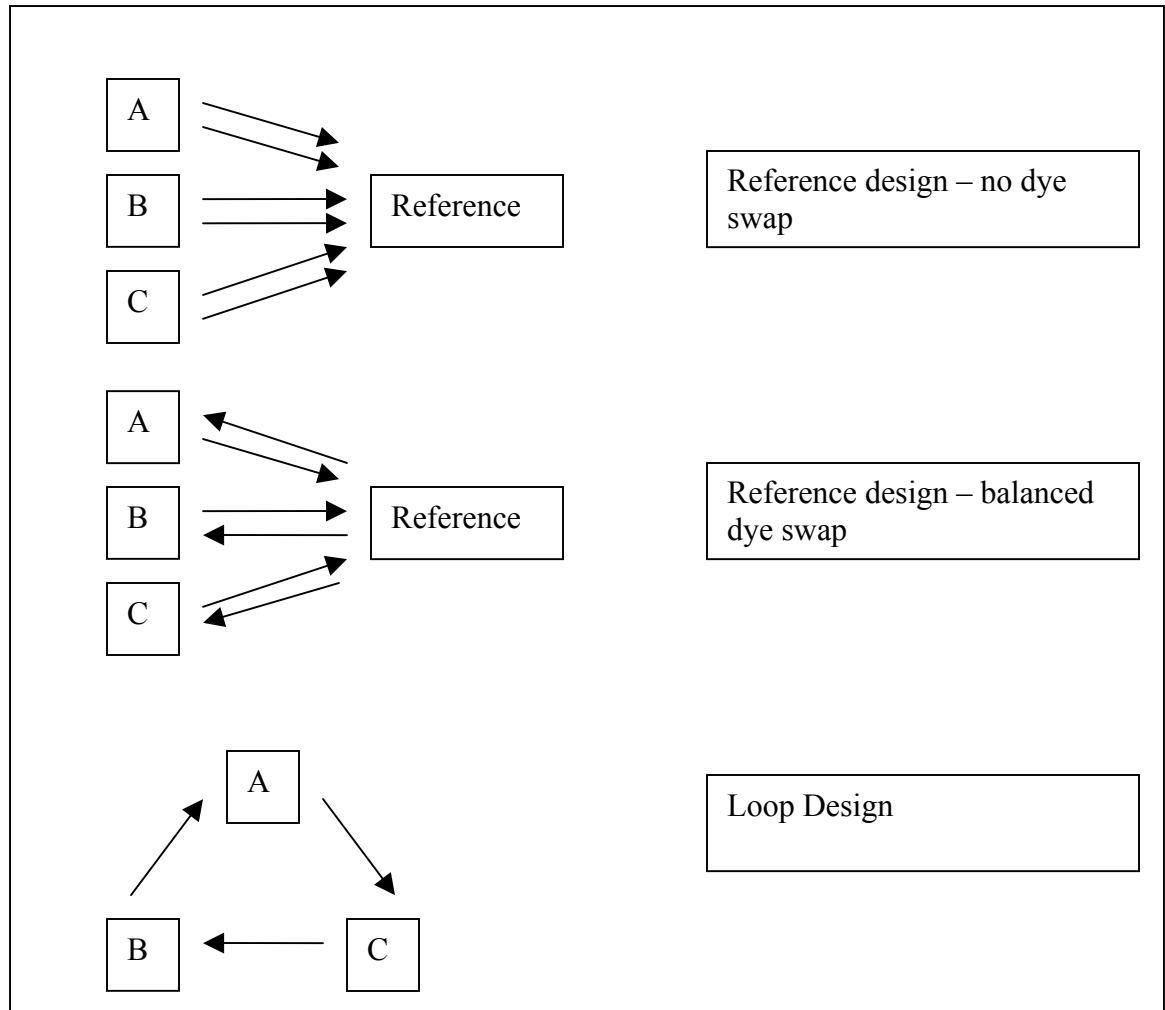
Kerr and Churchill, (2001a,b), introduced the loop design, which is an application of incomplete block designs to 2-channel microarrays. The loop design does not use reference samples, as a result of which only half the arrays are required to attain the same sample size as the reference design. Each loop includes 2 replicates of each treatment. Kerr and Churchill did not discuss how to include further replication in this design.

One issue of concern is gene-specific dye effects (gene by dye interaction) which can bias comparisons between treatments. In reference designs, dye effects can be eliminated either by using the same channel for the reference sample on each array (so that the treated samples are always labeled with the same dye) or by using balanced dye-swap pairs – i.e. for each treatment, the same number of arrays have the reference labeled red and green. In loop designs, a balanced dye-swap is built into the design. Each treatment has 2 samples, one labeled with each dye. Figure 1 is a graphical depiction of the two reference designs and the loop design for 3 treatments. At each node, the arrow tail is the sample labeled with the green dye and the arrow head is the sample labeled with the red dye.

In this paper, I assume that correlation among expression of genes on the same array is induced by a random array effect. I then compute optimal weights for analysis of the reference design and basic loop design. I then consider various ways of replicating loop designs – by replicating loops, or rearranging the treatments into a different loop.

Two commonly cited “problems” with loop designs are the loss of information for some spots on some arrays, leading to unbalanced designs, and the problem of adding treatments to the experiment. I suggest a simple method of “patching” a new treatment into an existing loop. I also note that the main problem with the loss of information from an array is due to the imbalance in the dye effect, a problem that also arises with reference designs.

Figure 1: Some basic designs for 2-channel microarrays



One source of confusion in the microarray experiment design and analysis literature is the handling of technical replicates and biological replicates in the same experiment. In this paper, we assume that we have only biological replicates. All the computations in this paper are readily extended to the cases in which there are multiple spots for the same gene on an array, or multiple arrays from the same sample.

## 2.0 To Difference or Not to Difference?

The objective of 2-channel microarray experiments is comparative analysis. As a result, the two channels on the array are often differenced (after converting to logarithms base 2), and the difference is treated as the observation of interest. That is, denoting the  $\log_2$  expression as  $Y_{ica}$  for gene  $i$  in channel  $c$  (red or green) on array  $a$ , analysis is generally done on  $M_{ia} = Y_{iRa} - Y_{iGa}$ . This is particularly convenient, because normalization is often done by normalizing  $M_{ia}$  against the average of the two channels,  $A_{ia} = (Y_{iRa} + Y_{iGa})/2$

In this paper, we use the individual channels as the observations of interest. Thus, each array provides 2 observations for each gene. As needed, we convert the normalized differences,  $\tilde{M}_{ia}$  back to channel level data via the transformation:

$$\begin{aligned}\tilde{Y}_{iRa} &= (2 A_{ia} + \tilde{M}_{ia}) / 2 \\ \tilde{Y}_{iGa} &= (2 A_{ia} - \tilde{M}_{ia}) / 2\end{aligned}$$

To see the efficacy of this approach, we look at some data from a drosophila array (B. McIver, personal communication). In this experiment, 2 biological samples were labeled and split. Both samples were then hybridized to the same array, as well as to other arrays with different samples. The normalized data are displayed in Figure 2. The left two panels show the normalized  $\log_2(\text{expression})$  values for the two samples on the same array. The right two panels show the values for the different arrays. While the right 2 panels are much noisier ( $\text{Var}(M)=0.453$  as opposed to  $\text{Var}(M)=0.126$  in the left 2 panels) there is still considerable information in the individual channels.

In the next section we show that a partial difference of the form  $\tilde{Y}_{iRa} - w \tilde{Y}_{iGa}$  is the optimal unit of analysis for a reference design. The optimal value of  $w$  depends on the relative magnitudes of the array effect (correlation between samples hybridized to the same spot) and the error variance. These quantities are unknown, but can be estimated from a mixed effects ANOVA. The same computations can be used to determine the relative efficiency of different experimental designs and optimal designs for microarray experiments. While the optimal weights vary from gene to gene, the ranking of the relative efficiency of different designs does not depend on the value of the weights, and hence the optimal design is the same for all genes.

We will use a simplified mixed model for the normalized expression data by channel:

$$\tilde{Y}_{ijca} = \mu_i + \tau_{ij} + \gamma_{ic} + \alpha_{ia} + \varepsilon_{ica}$$

where

- $\mu_i$  is the mean normalized expression for gene  $i$  over the entire experiment
- $\tau_{ij}$  is the mean effect of treatment  $j$  on gene  $i$
- $\gamma_{ic}$  is the mean effect of dye  $c$  on gene  $i$  (dye by gene interaction)
- $\alpha_{ia}$  is the random array (or spot) effect for gene  $i$  with variance  $\sigma_{ia}^2$
- $\varepsilon_{ica}$  is the random error with variance  $\sigma_{i\varepsilon}^2$

This paper focuses on the variance of a comparative analysis. As well, the focus is on a gene by gene analysis. Hence we do not make distribution assumptions except for the independence of data for the same gene on different arrays, and a mean of 0 for the random effects.

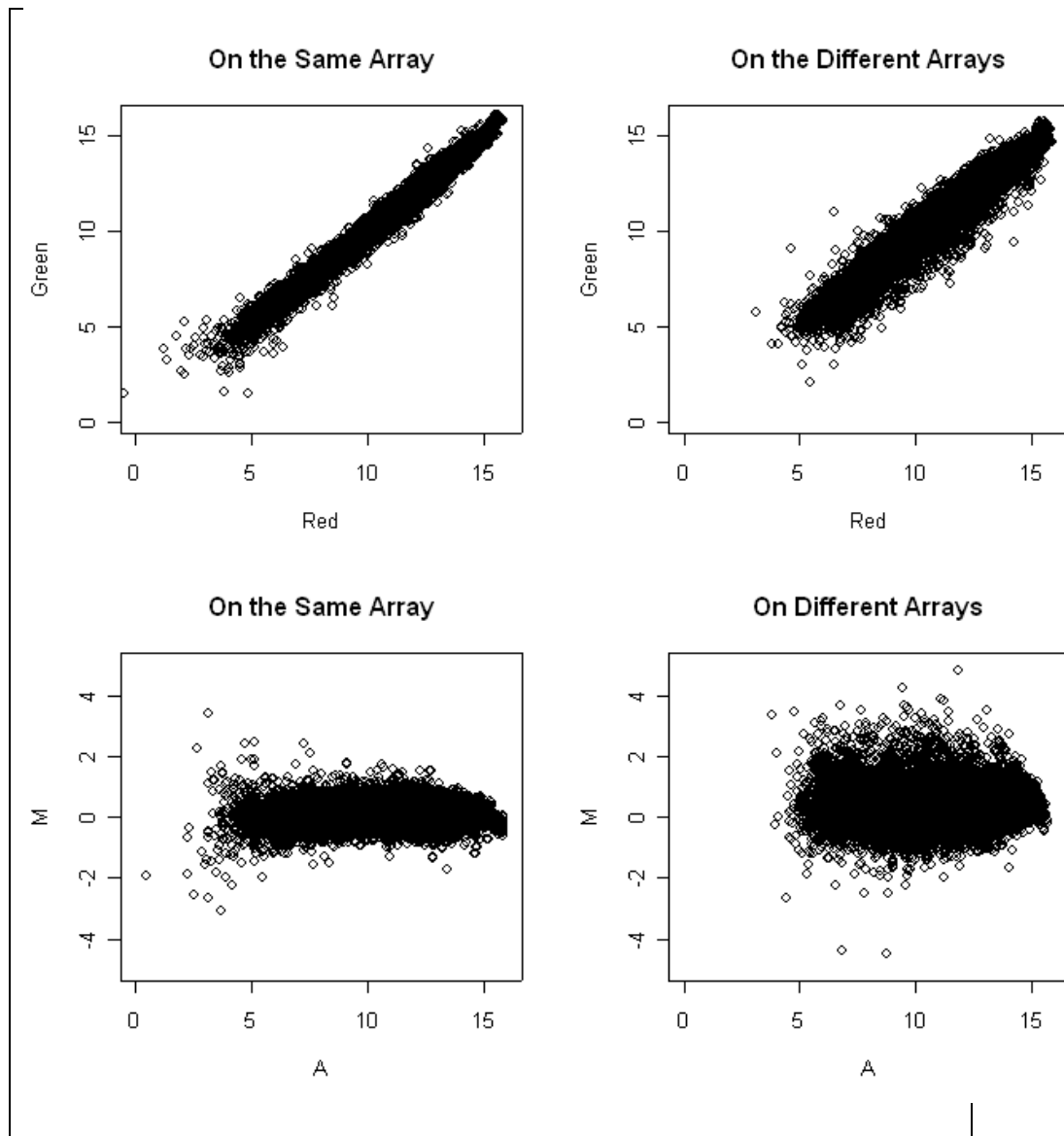


Figure 2:  $\text{Log}_2(\text{Expression})$  from a drosophila experiment. The Green and Red samples are different biological samples that were split after labeling into 2 aliquots. In the 2 left panels, the samples were hybridized to one array. In the 2 right panels, the samples were hybridized to different arrays.

### 3.0 Reference Design

In a reference design, one channel of each array is a reference sample, which is the same biological material on every array. With  $T$  treatments and  $k$  replicates per treatment, we use  $kT$  arrays. If there are technical dye-swaps, these are averaged to form 1 replicate.

If all comparisons are between treatments, there is no need to dye-swap. If there are dye-swaps, these should be balanced by treatment.

Usually the analysis is done on the difference between the treatment and reference samples on each array. For convenience of exposition (i.e. so we do not have to keep track of +/- signs) we assume that the treatment sample is in the red channel on all arrays, so we can denote the difference as,  $\tilde{M}_{ia}$ .

Then if treatment A is on array 1 and treatment B is on array 2, the contrast A-B is estimated by

$$\tilde{M}_{i1} - \tilde{M}_{i2} \text{ with variance } 4\sigma_{i\epsilon}^2.$$

With k replicates, the estimated contrast would have variance  $4\sigma_{i\epsilon}^2/k$ .

We now consider using the partial difference  $\tilde{M}_{ia}^w = \tilde{Y}_{iRa} - w\tilde{Y}_{iGa}$  and estimate the contrast with  $\tilde{M}_{i1}^w - \tilde{M}_{i2}^w$ . We find that the optimal weight is  $w = \sigma_{ia}^2 / (\sigma_{ia}^2 + \sigma_{i\epsilon}^2)$  with resulting variance  $4\sigma_{i\epsilon}^2 - 2\sigma_{ia}^4 / (\sigma_{ia}^2 + \sigma_{i\epsilon}^2) = 2\sigma_{i\epsilon}^2 + 2\sigma_{ia}^2\sigma_{i\epsilon}^2 / (\sigma_{ia}^2 + \sigma_{i\epsilon}^2)$  and with k replicates the variance is  $4\sigma_{i\epsilon}^2/k - 2\sigma_{ia}^4 / [k(\sigma_{ia}^2 + \sigma_{i\epsilon}^2)] = 2\sigma_{i\epsilon}^2/k + 2\sigma_{ia}^2\sigma_{i\epsilon}^2 / [k(\sigma_{ia}^2 + \sigma_{i\epsilon}^2)]$ .

While we do not know the optimal weights, if we use mixed model ANOVA such as those available in SAS, Splus or R, the weights are approximated from the data – leading to more efficient computations.

#### 4. Loop Designs

A loop is balanced for dye effects and has two replicates at each node. For T treatments using Tk arrays we have 2k replicates as compared to a reference design for which the same number of arrays yields only k replicates.

The computation of variance of a contrast in a loop design depends on the design and the number of treatments. For expository purposes we demonstrate with 4 treatments, but the conclusions are similar regardless of the number of treatments.

The 4 treatment loop design is depicted in Figure 3.



Figure 3: 4 treatment loop design

Using the optimal weighting, we find that the variance of the contrast between adjacent treatments is

$\sigma_{i\epsilon}^2 + \sigma_{ia}^2 \sigma_{i\epsilon}^2 / 2(\sigma_{ia}^2 + \sigma_{i\epsilon}^2)$  while the variance of the contrast between diagonally opposite treatments is  $\sigma_{i\epsilon}^2 + \sigma_{ia}^2 \sigma_{i\epsilon}^2 / (\sigma_{ia}^2 + \sigma_{i\epsilon}^2)$ . Both of these are smaller than the variance of the contrast from a reference design with the same number of treatments and arrays, primarily because there are two replicates per sample, rather than one. As well, the loop design has sufficient replication for statistical analysis, while the reference design has no replicates of the treatments unless the number of arrays is increased.

We can also consider replicating our design. With 12 arrays, the variance of any treatment difference in the reference design is  $2/3[\sigma_{i\epsilon}^2 + \sigma_{ia}^2 \sigma_{i\epsilon}^2 / (\sigma_{ia}^2 + \sigma_{i\epsilon}^2)]$ . If we use all of the 3 possible loops as our replicates, the variance of any treatment difference for the loop design is  $\sigma_{i\epsilon}^2 / 3 + 2\sigma_{ia}^2 \sigma_{i\epsilon}^2 / (12\sigma_{ia}^2 + 9\sigma_{i\epsilon}^2)$ , which is considerably smaller.

Table 1 shows the variance of a treatment difference for 3 ratios of  $\sigma_{ia}^2 / \sigma_{i\epsilon}^2$  with 12 arrays. We see that the reference design is always the least efficient (by a considerable amount) and that the design using all 3 loops is an excellent choice for pairwise comparisons.

Design	$\sigma_{ia}^2 / \sigma_{i\epsilon}^2 = 1/3$	$\sigma_{ia}^2 / \sigma_{i\epsilon}^2 = 1$	$\sigma_{ia}^2 / \sigma_{i\epsilon}^2 = 3$
Reference Design	1.17	1.0	2.50
3 x Loop 1 (adjacent)	0.46	0.42	1.13
3 x Loop 1 (diagonal)	0.58	0.50	1.25
All Loops	0.47	0.43	1.15

Table 1: Variance of a pair-wise comparison of treatments (as a multiple of  $\sigma_{i\epsilon}^2$ ) for a design with 4 treatments and 12 arrays.

### 5.0 More Complicated Loop Designs

Factorial treatment designs can readily be incorporated into loop designs. For example, a 2-factor design with 2 levels per factor (Aa and Bb) and main effects of primary interest could be laid out as in Figure 4:

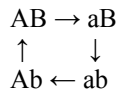


Figure 4: A 2 factor experiment with two levels of factor 1 (A and a) and 2 levels of treatment 2 (B and b).

In this arrangement, the interaction effects will be estimated with more precision than the main effects, due to cancellation of the array effects..

As another example, the author has been working with a one-way design with 8 treatments. However, the investigator is able to handle a maximum of 5 arrays in a day.

We have incorporated “day of preparation” as a complete block in the loop as in Figure 5 below. The thick arrows are one block and the thin are a second block.

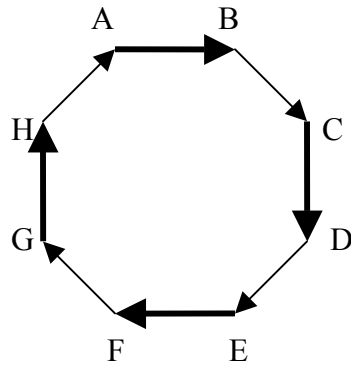


Figure 5: An 8 treatment loop design in two complete blocks. The bold lines are one block and the thin lines are the other.

We are also replicating this experiment. Using the same principles of optimal design, we find that the optimal replication is given (in 4 blocks) by Figure 6.

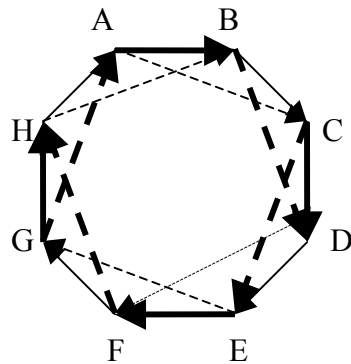


Figure 6: A replicated loop design for 8 treatments in 4 blocks. Each block is denoted by one of the 4 line types.

## 6.0 Adding Treatments and Missing Spots

In a reference design, it is quite obvious how to add a treatment to the design – just include new arrays with the new treatment. In a completed loop design, it is less obvious how to add a treatment without destroying the loop structure.

Figure 7 is one suggestion for adding a 5<sup>th</sup> treatment to a loop design with 4 treatments. Notice however, that this design is no longer balanced for the dye effect, since we now have an excess of red samples at A and green

samples at D. Hence, it is necessary to add an extra array AD for balance of the dye effect, or to ignore the data on array DA. The latter suggestion has the appeal that it enlarges the design to a larger loop design with the same structure as the original design. Adding an extra array means that the design will no longer be balanced for the number of replicates, which can make the analysis more complex.

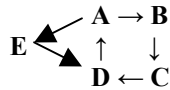


Figure 7: Adding treatment E to a 4 treatment loop design.

Loss of an array in a loop design is problematic primarily due to the loss of balance for the dye effect. Of course, if a whole array fails, the investigator is likely to take another sample and rehybridize. More commonly, some spots fail on some arrays, creating a loss of balance for some genes. The ANOVA computations automatically adjust for such imbalance, but the loss of information can severely affect statistical power. Each “lost” spot loses information about 2 treatments for the affected gene. However, the situation is not actually that much better for reference designs. In dye-swap designs, loss of balance due to spot failure can create imbalance. As well, since the initial power of a dye-swap design is about  $\frac{1}{2}$  the power of a loop design, each lost spot seriously depletes the power of the design.

## 7.0 Discussion

The main message of this paper is that generalized loop designs are much more efficient than reference designs. The emphasis is on designs that have two treatment samples on each array, (loop designs being an example) rather than loop designs per se. Such designs should be analyzed by treating the arrays as blocks of size 2 and analyzing the channels as individual observations.

In this set up, mixed model analysis and the principles of optimal design can usefully be employed to improve the efficiency of microarray experiments. It is likely that the use of “strength borrowing” methods such as empirical Bayes analysis add additional power over and above the gain by optimal design and gene by gene analysis.

## **Bibliography**

Kerr MK, Churchill GA. 2001a. Experimental design for gene expression microarrays *Biostatistics*, **2**, 183-201.

Kerr MK, Churchill GA. 2001b. Statistical design and the analysis of gene expression microarray data *Genet Res*, **77**, 123-128.