Statistica Sinica Preprint No: SS-2017-0503	
Title	Efficient Experimental Plans for Second-Order Event-Related Functional Magnetic Resonance Imaging
Manuscript ID	SS-2017-0503
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0503
Complete List of Authors	Yuan-Lung Lin and Frederick Kin Hing Phoa
Corresponding Author	Frederick Kin Hing Phoa
E-mail	fredphoa@stat.sinica.edu.tw

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

Statistica Sinica

Efficient Experimental Plans for Second-Order Event-Related Functional Magnetic Resonance Imaging

Yuan-Lung Lin and Frederick Kin Hing Phoa

Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

Abstract: Functional magnetic resonance imaging (fMRI) experiments help to render correct statistical inferences on brain function. However, few theoretical constructions of efficient designs for these crucial experiments exist in the literature. Recent works on the construction of circulant orthogonal arrays using algebraic difference sets show promise. However, such arrays are limited by the assumption that any interactions between the effects of a hemodynamic response function (HRF) and its residual effects are negligible in magnitude estimations of BOLD signals collected from an fMRI. Therefore, we propose a theoretical construction of a circulant orthogonal array of high strength using an extension of the complete difference system. FMRI experiments conducted using our proposed design show that the main effects of the individual HRF on the signal are not biased by the main effects of other HRFs, or by the interaction with its residual effects. Several properties of this new class of designs are studied, and the statistical regression model associated with this class of designs is presented.

Keywords and phrases: Circulant Almost Orthogonal Arrays; Complete Difference System; Design efficiency; Hemodynamic response function.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

1. Introduction

Functional magnetic resonance imaging (fMRI) experiments provide guidelines for the prevention and treatment of brain disorders such as Alzheimer's disease. A highly efficient experimental design is an important step in a successful study of functional brain images. An event-related fMRI (ER-fMRI) yields a shape estimation of hemodynamic response functions (HRFs), which are associated with transient brain activations evoked by various mental stimuli. An ER-fMRI design is a sequence of stimuli administered by an experimenter, and is regarded as a circulant design (Kao (2013)). In an fMRI experiment, a design may contain tens to hundreds of stimuli from cerebral neuronal activities. These stimuli lead to a change, described in the form of HRFs, in the ratio of "oxy- to deoxy-blood," detected in the MRI scanner as a change in the strength of the magnetic field. After the onset of a stimulus, the HRF takes several seconds to return completely to the baseline. Researchers then make a statistical inference on the brain activity by using an MRI scanner that repeatedly scans the subject's brain to collect data. See (Lazar (2008)) for more details.

Buračas and Boynton (2002) proposed using an m-sequence in fMRI experiments. However, although this sequence performs well, as indicated in several studies (Liu and Frank (2004); Liu (2004); Jansma et al. (2013)),

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

its length is limited to $n = (Q+1)^r - 1$, where Q is the total number of stimulus types and r is a positive integer. To relax this constraint, Liu (2004) and Kao (2013) proposed using a truncated version and an extended version, respectively. However, the former suffers from an efficiency loss (Kao (2014)), and the latter is universally optimal only when a few effects were estimated. As a result, a new class of highly efficient fMRI designs with flexible run sizes is needed. Kao (2014) proposed an H-sequence, which exists when its length $n \equiv 3 \pmod 4$ is a prime, twin prime, or a power of two, for ER-fMRI experiments with one stimulus type. This sequence offers the advantage of run-size flexibility, but it fits for some specific n only. A matrix $(a_{i,j})_{n\times n}$ is *circulant* if $a_{i+1,j+1} = a_{i,j}$, where the subscripts i and j are reduced modulo n. Obtained from a computer search (Low et al. (2005)), Craigen et al. (2013) introduced the r-row-regular *circulant partial Hadamard matrix*, denoted as r-$H(k\times n)$, where H is a $k\times n$ circulant (± 1) matrix, such that $HH^T = nI_k$, and r is the row sum of H. A 0-$H(k\times n)$ is a two-symbol, n-run, k-factor *circulant orthogonal array* (COA) and is efficient for fMRI experiments (Kao (2015)) when $n \equiv 0 \pmod 4$. The optimal design properties of these classes are described comprehensively in (Buračas and Boynton (2002); Liu (2004); Kao (2013, 2015); Cheng and Kao (2015); **?**); Lin, Phoa, and Kao (2017b)).

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

Introduced in (Rao (1946)), an orthogonal array (OA) is a class of experimental designs widely used in areas, such as medicine, agriculture, manufacturing, and many others. See (Hedayat, Sloane, and Stufken (1999)) for more details. The main advantages of using an OA design are the orthogonality and projectivity of the effect estimates (Cheng (1980, 1995); Raktoe, Hedayat, and Federer (1981)). However, an OA is also constrained by its lack of flexibility in run sizes (i.e., they must be a multiple of s^t). Recently, Lin, Phoa, and Kao (2017b) proposed a unified method to obtain OAs with the circulant property, which can be used in fMRI experiments with any run sizes and two or more levels. This generalized structure of fMRI designs, called a *circulant (almost-)orthogonal array* (CAOA), guarantees that t-tuples appear almost equally often.

Definition 1. A circulant $k \times n$ array $\boldsymbol{A}$ with entries from Z_s is said to be a CAOA with s levels, strength t, and bandwidth b, if each ordered t-tuple α based on Z_s occurs $\lambda(\alpha)$ times as column vectors of any $t \times n$ submatrices of $\boldsymbol{A}$, such that $|\lambda(\alpha) - \lambda(\beta)| \leq b$, for any two t-tuples α and β. Such an array $\boldsymbol{A}$ is denoted as $CAOA(n, k, s, t, b)$.

Lin, Phoa, and Kao (2017a,b) studied CAOAs of strength two, and compiled a table of universally optimal $CAOA(n, K, 2, 2, 1)$ when $n \leq 600$. These CAOAs help researchers measure the magnitudes of HRFs at specific time points, under the assumption that no interactions are detected in the signal output from the MRI scanner between the direct effect and the residual effects. However, such an assumption is unlikely to be realistic. Nevertheless, these interaction effects are often ignored in practice for analytical simplicity, which can lead to a bias in the estimation of the direct effect if the residual effects are still significant. In the language of experimental designs, an OA of strength three or higher makes it possible to estimate the main effect in a way that is free of the bias of two-factor interactions. Therefore, a CAOA of strength three provides an adjusted estimation of the HRFs that is unbiased by its residual effects.

This study constructs good CAOAs of strength three, denoted as $CAOA(n, K, 2, 3, b)$, for $b = 0, 1$. Section 2 provides several properties of CAOAs, and connects them to the De Bruijn sequences introduced in (Bruijn (1946)). Section 3 is a theoretical study on the construction of CAOAs of high strength. A table of generating vectors for $CAOA(n, K, 2, 3, 0)$ is provided as a result of the theoretical derivations. Section 4 concludes the papaer. All proofs are available in the online Supplementary Material.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

2. Properties of CAOAs

We begin with a discussion of the properties of a $CAOA(n, k, s, t, b)$. A specific OA, $OA(s^t, t+1, s, t)$, can be constructed using the zero-sum array property that the levels in every run add up to zero. This provides a lower bound for the number of factors, but this method cannot be applied to COAs and CAOAs. Although an $OA(s^t, t, s, t)$ can be trivially obtained by finding all strings with s symbols, the structure of a $CAOA(n, t, s, t, b)$ is more difficult. When $b = 0$, the generating vector of a $CAOA(s^t, t, s, t, 0)$ is an s-ary *De Bruijn sequence* of order t, which is a cyclic s-ary sequence with the property that every s-ary t-tuple appears exactly once consecutively in the cycle (Bruijn (1946)). For example, (112233132) is a De Bruijn sequence for $s = 3$ and $t = 2$. De Bruijn sequences have been applied in studies on pseudo-random codes, cryptography, nonlinear shift registers, coding theory, and genome assembly (see Fredricksen (1982); Good (1946); MacWilliams and Sloane (1976); Compeau, Pevzner, and Tesler (2011)).

In graph theory, a t-dimensional *De Bruijn graph* of s symbols is a directed graph, where the vertices are sequences of symbols from some symbols, and the edges indicate the sequences that might overlap. An Eulerian circuit in a directed graph is a directed circuit that uses each edge exactly once. Please refer to (van Lint and Wilson (2001); West (2001, Sec. 1)) for

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

details. De Bruijn sequences can be constructed by taking an Eulerian circuit of a t-dimensional De Bruijn graph over s symbols (or, equivalently, a Hamiltonian cycle of a $(t+1)$-dimensional De Bruijn graph). Traditionally, a De Bruijn sequence requires that every t-tuple appear exactly once, which means its graph is a regular simple graph. We extend the De Bruijn graph to construct a $CAOA(n, t, s, t, b)$ below. Define Λ as a *frequency sequence* $(\lambda^{a_1 \dots a_t})_{a_i \in Z_s}$, which represents the frequency of an s-ary $(t-1)$-tuple (lexicographical order). In addition, the bandwidth of Λ, denoted as $B(\Lambda)$, is the difference between the maximum and the minimum entries in Λ.

Definition 2. A t-dimensional De Bruijn frequency graph of s symbols, based on a frequency sequence $(\lambda^{a_1 \dots a_t})_{a_i \in Z_s}$, is a directed multi-graph, the vertex set of which comprises all s-ary $(t-1)$-tuples $(d_1, d_2, \dots, d_{t-1})$, and there are m edges from $(d_1, d_2, \dots, d_{t-1})$ to $(d'_1, d'_2, \dots, d'_{t-1})$ if $d'_i = d_{i+1}$, for all $i = 1, 2, \dots, t-2$ and $\lambda^{d_1 d_2 \dots d_{t-1} d'_{t-1}} = m$.

In a t-dimensional De Bruijn frequency graph, each directed edge from $(d_1, d_2, \dots, d_{t-1})$ to $(d'_1, d'_2, \dots, d'_{t-1})$ is labeled as $d_1 d_2 \cdots d_{t-1} d'_{t-1}$. In general, each vertex $(d_1, d_2, \dots, d_{t-1})$ is also written as $d_1 d_2 \cdots d_{t-1}$, for short. A circuit in a t-dimensional De Bruijn frequency graph can be represented by a sequence $d_1 d_2 \dots d_n$ that comprises the edges $d_1 d_2 \dots d_t, d_2 d_3 \dots d_{t+1}, \cdots, d_n d_1 \dots d_{t-1}$. For example, a circuit $00 \longrightarrow 01 \longrightarrow 11 \longrightarrow 10$ can be simply

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

represented by 0011. By finding an Eulerian circuit in a t-dimensional De Bruijn frequency graph of s symbols based on Λ, the generating vector of a $COA(n,t,s,t,b)$ can be easily found, where n is the total sum of each component of Λ and $b = B(\Lambda)$.

Theorem 1. *Let $s,t \geq 2, b \geq 0$, and $\Lambda = (\lambda^{a_1 \ldots a_t})_{a_i \in Z_s}$ be a frequency sequence of s-ary t-tuples such that $B(\Lambda) = b$. If $\sum_{x=0}^{s-1} \lambda^{a_1 \ldots a_{t-1} x} = \sum_{x=0}^{s-1} \lambda^{x a_1 \ldots a_{t-1}}$ and $\lambda^{a_1 \ldots a_t} \geq 1$, for all $a_i \in Z_s$, then there exists a $CAOA(n,t,s,t,b)$, where $n = \sum_{a_i \in Z_s} \lambda^{a_1 \ldots a_t}$.*

The proof of this theorem is given in Section 4. Theorem 1 provides an easy construction with which to find the generating vector of a CAOA. This also guarantees that the lower bound for k in a CAOA of strength t is at least t. As in the demonstration, we construct a $CAOA(20,3,2,3,1)$ using a De Bruijn frequency graph.

Example 1. *Let the frequency sequence $\Lambda = (\lambda^{000}, \lambda^{001}, \lambda^{010}, \lambda^{011}, \lambda^{100}, \lambda^{101}, \lambda^{110}, \lambda^{111}) = (3,2,2,3,2,3,3,2)$. A three-dimensional De Bruijn frequency graph $G = (V,E)$ is given in Figure 1.*

The vertex set $V(G) = \{00, 01, 10, 11\}$, and each directed edge from x_1x_2 to x_2x_3 is labeled as $x_1x_2x_3$. For instance, $\lambda^{001} = 2$ implies that there are two directed edges, labeled as 001, from 00 to 01. It can be verified that $\lambda^{a_1a_2 0} + \lambda^{a_1 a_2 1} = \lambda^{0 a_1 a_2} + \lambda^{1 a_1 a_2}$, for all $a_1, a_2 \in Z_2$; for example,

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

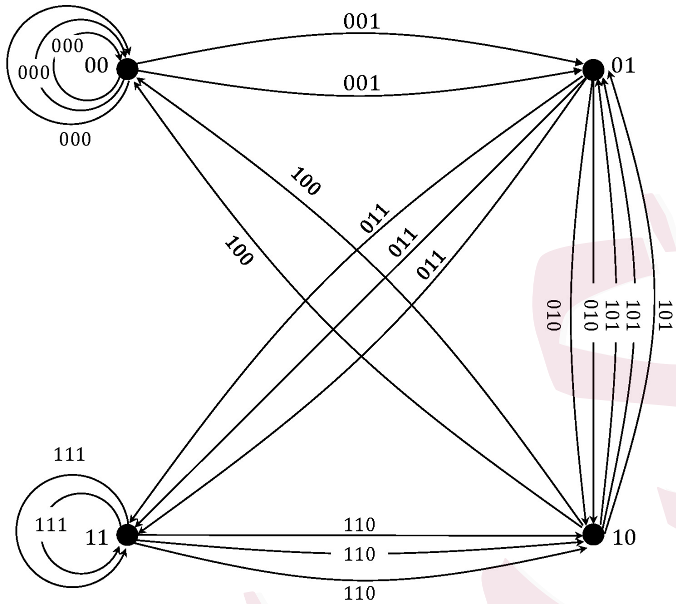

Figure 1: A three-dimensional De Bruijn frequency graph of two symbols based on $(3, 2, 2, 3, 2, 3, 3, 2)$

$\lambda^{010} + \lambda^{011} = \lambda^{101} + \lambda^{001} = 5$. *In addition,* $\lambda^{a_1a_2a_3} \geq 1$ *and* $\sum \lambda^{a_1a_2a_3} = 20$, *for all* $a_i \in Z_2$.

In Figure 1, each vertex has an equal in-degree and out-degree; thus, G is an Eulerian circuit. By Theorem 1, there exists a $CAOA(20, 3, 2, 3, 1)$, *and the Eulerian circuit corresponds to the generating vector of a CAOA. For instance, the Eulerian circuit* $000 \to 000 \to 001 \to 010 \to 101 \to \ldots \to 100 \to 000$ *can be represented by the sequence* 00001010110011110110. *This is a generating vector of a* $CAOA(20, 3, 2, 3, 1)$, *and each column is a triplet.*

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

The number of occurrences of each triplet (a_1, a_2, a_3) *is equal to its frequency* $\lambda^{a_1a_2a_3}$.

Corollary 1. *Let* k, s *be nonzero integers and* $s > 1$. *There exists a* $CAOA(s^k, k, s, t, 0)$, *for all* $2 \le t \le k$.

Corollary 1 can be trivially proved using Theorem 1. From Corollary 1, we find the generating vectors of a $CAOA(s^k, k, s, t, 0)$ when $8 \le s^k \le 1000$ and $2 \le s \le 10$. All designs are provided in Table 1 of the Supplementary Materials.

With an appropriate choice of frequency sequence, a CAOA of strength t might produce a CAOA of strength t', where $t' \le t$. For example, the generating vector of the $CAOA(20, 3, 2, 3, 1)$ in Example 1 can generate a $CAOA(20, 7,\ 2, 2, 0)$ with the maximal number of columns. Furthermore, for an Eulerian graph, there are many different Eulerian circuits. Every Eulerian circuit guarantees the existence of a $CAOA(n, t, s, t, b)$, but a good choice of an Eulerian circuit can construct a $CAOA(n, k, s, t, b)$ such that $k > t$.

Theorem 1 provides a necessary condition for the existence of CAOAs. If there exists a $CAOA(n, k, s, t, b)$ associated with $\Lambda = (\lambda^{a_1\cdots a_t})_{a_i \in Z_s}$, then $\sum_{x=0}^{s-1} \lambda^{a_1\ldots a_{t-1}x} = \sum_{x=0}^{s-1} \lambda^{xa_1\ldots a_{t-1}}$ and $\lambda^{a_1\cdots a_t} \ge 1$, for all $a_i \in Z_s$.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

3. CAOAs: Construction Method

In fMRI experiments, designs with the circulant property are required when estimating HRFs, and such designs are not well studied in the literature. We consider applying CAOAs (Definition 1) in fMRI experiments. Lin, Phoa, and Kao (2017b) revealed the mathematical structure of CAOAs of strength two using a *complete difference system* (CDS), which describes the entire matrix structure of a circulant design of strength two, and obtained many optimal circulant designs. However, the extension to designs of higher strength using a CDS is not trivial.

We present the difference structure of CAOAs of high strength using a *high-order complete difference system* (HCDS). In this paper, we consider designs of strength three only, but the structure can be extended easily to designs of strength greater than three. Let $V = \{V_0, V_1, \cdots, V_{s-1}\}$ be a partition of Z_n. The collection of differences is $\mathscr{S}^{\alpha,\beta} = \{\boldsymbol{S}_{a_i}^{\alpha,\beta}|, \, foralla_i \in V_\alpha\}$, where $\boldsymbol{S}_{a_i}^{\alpha,\beta} = \{a_i - b_j \pmod n)|$ for all $b_j \in V_\beta\}$ and $\alpha, \beta \in Z_s$. Let $\{r_1, \cdots, r_m\}$ be a subset of $Z_n \setminus \{0\}$, with m distinct differences coming from $\mathscr{S}^{\alpha,\beta}$. In addition, $\mathscr{S} = (\mathscr{S}^{\alpha,\beta})_{s\times s}$ is called a *difference matrix*. This is a Latin square; thus, each element appears exactly once in each row and column of $\mathscr{S}$.

We define the mth-order difference $\lambda_{r_1,\cdots,r_m}^{\alpha,\beta} = \#\{g \in V_\alpha|\{r_1, \cdots, r_m\} \subseteq$

$S_g^{\alpha,\beta}\}$, which counts the total number of $\boldsymbol{S}_g^{\alpha,\beta}$ containing the subset $\{r_1, \cdots, r_m\}$. As in the definition of a CDS (Lin, Phoa, and Kao (2017b)), $\lambda_r^{\alpha,\beta}$ in $\boldsymbol{\Lambda}$ of a $(n, k, s, \boldsymbol{\Lambda})$-CDS is the first-order difference.

Next, we define the HCDS that summarizes all information on mth-order $(1 \leq m \leq t-1)$ differences and captures the whole structure of strength t. An $(r_1, \cdots, r_m)$-frequency matrix of V is a matrix $\boldsymbol{\Lambda}_{r_1,\cdots,r_m} = (\lambda^{\alpha,\beta}_{r_1,\cdots,r_m})_{s\times s}$. An HCDS of V is a collection of ordered multi-tuples $(\boldsymbol{\Lambda}_{1,\cdots,m}, \ldots, \boldsymbol{\Lambda}_{n-m,\cdots,n-1})$, for $1 \leq m \leq t-1$, which describes all frequency matrices of V. Let $I_D(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \cdots, \boldsymbol{\Phi}_{t-1})$ be the largest index k such that $\boldsymbol{\Lambda}_{r_1,\cdots,r_m} = \boldsymbol{\Phi}_m$, for $1 \leq r_1 < \cdots < r_m \leq k$. We say V is an $(n, k, s; \boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_{t-1})$-HCDS if V contains s disjoint parts from Z_n, such that $I_D(\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_{t-1}) = k-1$. Its *incidence matrix* is defined as follows.

Definition 3. Let V be an $(n, k, s; \boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_{t-1})$-HCDS. The incidence matrix of V is a $k \times n$ matrix $\boldsymbol{A} = (a_{i,j})$, defined as

$$a_{i,j} = l \text{ if } j \in V_l + (i-1),$$

where $V_l + (i-1) = \{x + (i-1) | \text{for all } x \in V_l\}$, and all elements are reduced modulo n, for $i = 1, \ldots, k$, $j = 1, \ldots, n$, and $l = 0, \ldots, s-1$.

Example 2. Let $V = \{V_0, V_1\}$ be a partition of Z_8, where $V_0 = \{1, 2, 3, 5\}$ and $V_1 = \{4, 6, 7, 8\}$. Then, the collection of all differences can be represented by the difference matrix $\mathscr{S}$ in Table 1. Let $\alpha, \beta \in \{0, 1\}$ and

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

Table 1: The collection of all differences of partition $V = \{\{1,2,3,5\},\{4,6,7,8\}\}$

$$\mathscr{S} = \left(\begin{array}{c|c} \mathscr{S}^{0,0} & \mathscr{S}^{0,1} \\ \hline \mathscr{S}^{1,0} & \mathscr{S}^{1,1} \end{array}\right) = \begin{array}{c} \boldsymbol{S}_1^{0,0} \\ \boldsymbol{S}_2^{0,0} \\ \boldsymbol{S}_3^{0,0} \\ \boldsymbol{S}_5^{0,0} \\ \boldsymbol{S}_4^{1,0} \\ \boldsymbol{S}_6^{1,0} \\ \boldsymbol{S}_7^{1,0} \\ \boldsymbol{S}_8^{1,0} \end{array} \left(\begin{array}{cccc|cccc} 0 & 7 & 6 & 4 & 5 & 3 & 2 & 1 \\ 1 & 0 & 7 & 5 & 6 & 4 & 3 & 2 \\ 2 & 1 & 0 & 6 & 7 & 5 & 4 & 3 \\ 4 & 3 & 2 & 0 & 1 & 7 & 6 & 5 \\ \hline 3 & 2 & 1 & 7 & 0 & 6 & 5 & 4 \\ 5 & 4 & 3 & 1 & 2 & 0 & 7 & 6 \\ 6 & 5 & 4 & 2 & 3 & 1 & 0 & 7 \\ 7 & 6 & 5 & 3 & 4 & 2 & 1 & 0 \end{array}\right) \begin{array}{c} \boldsymbol{S}_1^{0,1} \\ \boldsymbol{S}_2^{0,1} \\ \boldsymbol{S}_3^{0,1} \\ \boldsymbol{S}_5^{0,1} \\ \boldsymbol{S}_4^{1,1} \\ \boldsymbol{S}_6^{1,1} \\ \boldsymbol{S}_7^{1,1} \\ \boldsymbol{S}_8^{1,1} \end{array}$$

$r, r_1, r_2 \in Z_8 \setminus \{0\}$. The first-order difference $\lambda_r^{\alpha,\beta}$ is the frequency of element r in $\mathscr{S}^{\alpha,\beta}$. Furthermore, the second-order difference $\lambda_{r_1,r_2}^{\alpha,\beta}$ is the number of rows in $\mathscr{S}^{\alpha,\beta}$ that contain the elements r_1 and r_2 simultaneously. This implies that the frequency matrices $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = 2\boldsymbol{J}_2$ and $\boldsymbol{\Lambda}_{1,2} = \boldsymbol{J}_2$, but $\boldsymbol{\Lambda}_3 \neq 2\boldsymbol{J}_2$ and $\boldsymbol{\Lambda}_{1,3} \neq \boldsymbol{J}_2$. Let $\boldsymbol{\Phi}_1 = 2\boldsymbol{J}_2$ and $\boldsymbol{\Phi}_2 = \boldsymbol{J}_2$. Then, $I_D(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2) = I_D(2\boldsymbol{J}_2, \boldsymbol{J}_2) = 2$, the partition V is an $(8,3,2;2\boldsymbol{J}_2,\boldsymbol{J}_2)$-HCDS, and the incidence matrix of V is

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

The above matrix is a $CAOA(8,3,2,3,0)$, but the relationship between HCDSs and CAOAs of strength three is not trivial. In addition, high-order differences are difficult to count, which may increase the difficulty

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

in the construction of the design. For two-level CAOAs, we state a useful proposition that describes the relationship between the first-and second-order differences. This proposition also provides an easy way to calculate the second-order difference using the first-order differences.

Proposition 1. *Let* $(\lambda_{r_1}^{\alpha,\beta})_{2\times 2}$ *and* $(\lambda_{r_1,r_2}^{\alpha,\beta})_{2\times 2}$ *be the* r_1*- and* (r_1, r_2)*-frequency matrices of a partition* $V = \{V_0, V_1\} = Z_n$*, where* $1 \leq r_1 < r_2 \leq n-1$*. Then,*

(i) $\lambda_{r_1,r_2}^{\alpha,\beta} + \lambda_{r_1,r_2}^{\beta,\beta} = \lambda_{(r_2-r_1)}^{\beta,\beta}$ *and*

(ii) $\lambda_{r_1,r_2}^{\alpha,\beta} = |V_\alpha| - (\lambda_{r_1}^{\alpha,\alpha} + \lambda_{r_2}^{\alpha,\alpha} - \lambda_{r_1,r_2}^{\alpha,\alpha})$.

Furthermore, if $|V_\alpha| = |V_\beta| = n/2$ *and* $\lambda_r^{\alpha,\alpha} = n/4$*, for* $1 \leq r \leq k$*, then* $\lambda_{r_1,r_2}^{\alpha,\beta} = \lambda_{r_1,r_2}^{\alpha,\alpha}$ *and* $\lambda_{r_1,r_2}^{\alpha,\alpha} + \lambda_{r_1,r_2}^{\beta,\beta} = n/4$*, for* $1 \leq r_1 < r_2 \leq k$*.*

Because an $(n, k, s; \mathbf{\Phi}_1, \cdots, \mathbf{\Phi}_{t-1})$-HCDS is also a $(n, k, s, \mathbf{\Phi}_1)$-CDS, the existence of an HCDS is equivalent to the existence of a CAOA of strength two (by Corollary 3.3 in (Lin, Phoa, and Kao (2017b))). In a $(n, k, s, \mathbf{\Phi}_1)$-CDS, the frequency of two-factor combinations is described by the given frequency matrix $\mathbf{\Phi}_1$ only. However, the good properties do not hold for CAOAs of strength three. Now, let $c_0, c_1, \cdots, c_m$ be a level combination of $m+1$ factors. Following the definition of an HCDS, we define $\delta_{r_1,r_2,\cdots,r_m}^{c_0,c_1,\cdots,c_m} =$

$\#\{g \in V_{c_0} | r_j \in \boldsymbol{S}_g^{c_0,c_j}$ for $j = 1, 2, \cdots, m\}$. Then, the high-order factor combinations can be explicitly represented by the HCDS.

Lemma 1. *Let $V = \{V_0, V_1, \cdots, V_{s-1}\}$ be an $(n, k, s; \boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_{t-1})$-HCDS and $\boldsymbol{A}$ be its incidence matrix. Then, $\delta_{r_1,r_2,\cdots,r_m}^{c_0,c_1,\cdots,c_m}$ is equal to the total number of level combinations $c_0, c_1, \cdots, c_m$ as column vectors of $\boldsymbol{A}_{g_0,g_1,\cdots,g_m}$, which comprises the g_0th, g_1th, $\cdots$, g_mth rows of $\boldsymbol{A}$, where $g_i - g_0 = r_i$.*

In Example 2, $\delta_{1,2}^{0,0,1}$ denotes the number of rows in $(\mathscr{S}^{0,0}|\mathscr{S}^{0,1})$, such that $\mathscr{S}^{0,0}$ and $\mathscr{S}^{0,1}$ contain 1 and 2, respectively. Because only the second row satisfies $1 \in \boldsymbol{S}_2^{0,0}$ and $2 \in \boldsymbol{S}_2^{0,1}$, we have $\delta_{1,2}^{0,0,1} = 1$. This implies that the triple 001 occurs exactly once as a column vector of $CAOA(8, 3, 2, 3, 0)$.

The definition of $\delta_{r_1,r_2,\cdots,r_m}^{c_0,c_1,\cdots,c_m}$ can lead to the mth-order differences, or lower. We show only the relationship between $\delta_{r_1,r_2}^{c_0,c_1,c_2}$ and $\lambda_{r_1,r_2}^{\alpha,\beta}$ for two-level CAOAs of strength three, but the results are readily extended to high-strength multi-level CAOAs.

Lemma 2. *Let V be an $(n, k, 2; \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$-HCDS, and its $(r_1, \cdots, r_m)$-frequency matrix $\boldsymbol{\Lambda}_{r_1,\cdots,r_m} = (\lambda_{r_1,\cdots,r_m}^{\alpha,\beta})_{2\times 2}$, where $m = 1, 2$ and $\alpha, \beta \in \{0, 1\}$. Then,* $\boldsymbol{\delta}_{r_1,r_2} =$

$$\begin{pmatrix} \delta_{r_1,r_2}^{0,0,0} & \delta_{r_1,r_2}^{0,1,1} & \delta_{r_1,r_2}^{0,0,1} & \delta_{r_1,r_2}^{0,1,0} \\ \delta_{r_1,r_2}^{1,0,0} & \delta_{r_1,r_2}^{1,1,1} & \delta_{r_1,r_2}^{1,0,1} & \delta_{r_1,r_2}^{1,1,0} \end{pmatrix} = \begin{pmatrix} \lambda_{r_1,r_2}^{0,0} & \lambda_{r_1,r_2}^{0,1} & \lambda_{r_1}^{0,0} - \lambda_{r_1,r_2}^{0,0} & \lambda_{r_2}^{0,0} - \lambda_{r_1,r_2}^{0,0} \\ \lambda_{r_1,r_2}^{1,0} & \lambda_{r_1,r_2}^{1,1} & \lambda_{r_2}^{1,1} - \lambda_{r_1,r_2}^{1,1} & \lambda_{r_1}^{1,1} - \lambda_{r_1,r_2}^{1,1} \end{pmatrix}.$$

From Lemmas 1 and 2, if V is an $(n, k, 3; \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$-HCDS and $\boldsymbol{A}$ is its

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

incidence matrix, then $\boldsymbol{A}_{1,1+r_1,1+r_2}$ is the $3 \times n$ submatrix of $\boldsymbol{A}$, and the frequency of the level combination associated with $\boldsymbol{A}_{1,1+r_1,1+r_2}$ is equal to $\boldsymbol{\delta}_{r_1,r_2}$. Even if $\boldsymbol{A}$ is a CAOA of strength three with bandwidth one, the pattern of distinct $3 \times n$ submatrices of $\boldsymbol{A}$ satisfies Lemma 2, although the combinations might be different. This implies that the bandwidths of $\boldsymbol{\delta}_{r_1,r_2}$ and $\boldsymbol{\delta}_{r_1',r_2'}$ are less than one, but that $\boldsymbol{\delta}_{r_1,r_2}$ is not equal to $\boldsymbol{\delta}_{r_1',r_2'}$. If the patterns of $\boldsymbol{\delta}_{r_1,r_2}$ are all the same, then this helps in our analysis. We say a $CAOA(n,k,s,t,b)$ is *uniform* if the frequency of the level combination associated with each $t \times n$ submatrix is equal to a fixed pattern $\boldsymbol{\delta}$. The existence of a uniform $CAOA(n,k,2,3,1)$ is given by the following theorem.

Theorem 2. *Let $V = \{V_0, V_1\}$ be an $(n,k,2;\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2)$-HCDS, where $|V_0| = \lfloor n/2 \rfloor$. Let $\boldsymbol{\Lambda}_r = (\lambda_r^{\alpha,\beta})_{2\times 2}$ and $\boldsymbol{\Lambda}_{r_1,r_2} = (\lambda_{r_1,r_2}^{\alpha,\beta})_{2\times 2}$ be its r- and (r_1,r_2)-frequency matrices, respectively. A uniform $CAOA(n,k,2,3,1)$ exists if and only if*

(i) $\lambda_r^{0,0} = n/4$ *and* $\lambda_{r_1,r_2}^{0,0} = n/8$ *when* $n \equiv 0 \pmod 8$,

(ii) $\lambda_r^{0,0} = \lfloor n/4 \rfloor$ *and* $\lambda_{r_1,r_2}^{0,0} = \lfloor n/8 \rfloor$ *when* $n \equiv 1,3,6,7 \pmod 8$,

(iii) $\lambda_r^{0,0} = \lceil n/4 \rceil$ *and* $\lambda_{r_1,r_2}^{0,0} = \lceil n/8 \rceil$ *when* $n \equiv 2 \pmod 8$,

(iv) $\lambda_r^{0,0} = n/4$ *and* $\lfloor n/8 \rfloor \le \lambda_{r_1,r_2}^{0,0} \le \lceil n/8 \rceil$ *when* $n \equiv 4 \pmod 8$,

(v) $\lambda_r^{0,0} = \lfloor n/4 \rfloor$ *and* $\lambda_{r_1,r_2}^{0,0} = \lceil n/8 \rceil$ *when* $n \equiv 5 \pmod 8$,

for $1 \leq r \leq k-1$ *and* $1 \leq r_1 < r_2 \leq k-1$. *Furthermore, a uniform* $CAOA(n,k,2,3,0)$ *exists if and only if condition (i) holds.*

Theorem 2 provides a good strategy with which to find a uniform $CAOA(n,k,2,3,1)$ without an exhaustive search. Moreover, it helps us to explore the maximum value of k of a $CAOA(n,k,2,3,0)$, in practice. Using the difference variance algorithm (DVA) proposed by Lin, Phoa, and Kao (2017a), we find all $CAOA(n,k,2,3,1)$ that possess maximum values of k when $8 \leq n \leq 27$; see Table 2. The maximum value k of a $CAOA(n,k,2,3,1)$ is approximately $n/4$.

Table 2: $CAOA(n,k,2,3,1)$, for all $8 \leq n \leq 27$.

n	8 – 13	14 – 21	22	23	24	25	26	27
k	3	4	5	5	6	6	6	5

We propose a method for constructing a $CAOA(n,k,2,3,0)$ with a maximum empirical value of $k = n/4$. According to Theorem 2, if $n \equiv 0 \pmod 8$, then we need an $(n,k,2;\mathbf{\Phi}_1,\mathbf{\Phi}_2)$-HCDS, $V = \{V_0, V_1\}$, where $|V_0| = |V_1| = n/2$, $\mathbf{\Phi}_1 = (n/4)\boldsymbol{J}_2$, and $\mathbf{\Phi}_2 = (n/8)\boldsymbol{J}_2$. Let D be a $(n/2, n/4-1; \lambda_1, \cdots, \lambda_{n/2-1})$- GDS, where all λ are equal to $n/8-1$, except $\lambda_{n/4} = 0$. Then, there exist two elements $g, g' \in D^c$, such that $g' - g = n/4$,

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

where D^c is the complement of D. By the *square principle* in Lin, Phoa, and Kao (2017b), D^c is a $(n/2, n/4-1; \lambda_1, \cdots, \lambda_{n/2-1})$-GDS, where all λ are equal to $n/8+1$, except $\lambda_{n/4} = 1$. Let $V_0 = D \cup (D+n/2) \cup \{g, g+n/4\}$ and $V_1 = (D+n/4) \cup (D+3n/4) \cup \{g+n/2, g+3n/4\}$. Then, $V = \{V_0, V_1\}$ is the required HCDS, and its incidence matrix is the required CAOA of strength three. The following theorem states the existence of a uniform $CAOA(n, k, 2, 3, 0)$, and shows that $V = \{V_0, V_1\}$ is the required HCDS.

Theorem 3. *Let $n \equiv 0 \pmod 8$. If $n/4-1$ is an odd prime power, there exists a uniform $CAOA(n, n/4, 2, 3, 0)$.*

Table 3: The maximal value of k of a $CAOA(n, k, 2, 3, 1)$, obtained from m-sequences.

n	7	15	31	63	127	255	511	1023
k	3	4	8	11	21	27	61	83

The m-sequence is popular in ER-fMRI experiments. It is equivalent to a $CAOA(n-1, n-1, 2, 2, 1)$, where n is a power of two (Lin, Phoa, and Kao (2017b)); thus, the CAOAs of strength two obtained from an m-sequence have a maximum value of k. However, this good property does not hold for CAOAs of strength three. The maximal values of k of a $CAOA(n, k, 2, 3, 1)$ obtained from m-sequences are shown in Table 3, and are smaller than

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

those of the CAOAs obtained by Theorem 3 when their run sizes are close. For example, if one needs a CAOA of strength three with $k \geq 83$, we recommend using the $CAOA(336, 84, 2, 3, 0)$, which can be found in the appendix, instead of the $CAOA(1023, 83, 2, 3, 1)$ in Table 3. Although their k is very close, the run size of the m-sequence is much higher than that of the CAOAs in Theorem 3. In addition, the CAOA of bandwidth zero always has a higher priority than the CAOA of bandwidth one. Therefore, the CAOAs described Theorem 3 are effective and efficient.

4. Conclusion

In this work, we propose using CAOAs of strength three as an experimental plan for an ER-fMRI experiment, where the estimates of the direct effects in HRFs are biased by their residual effects. Although this work is an extension of that of (Lin, Phoa, and Kao (2017b)), the core tool, namely the CDS is not sufficient for characterizing and constructing CAOAs of strength three. Thus, we propose a generalized version, called an HCDS. We theoretically study several properties of CAOAs of high strength, and establish an equivalence relation between HCDSs and CAOAs of high strength as an efficient construction method.

As mentioned before, it is efficient to search for generating vectors that

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

can be used to construct the class of $CAOA(n, k, 2, 3, 0)$. In the Supplementary Material, we provide the generating vectors when $n/4 - 1$ is an odd prime power and $n \leq 392$. Practitioners can simply use this table to find an fMRI experimental plan. For example, if an experimenter administers an ER-fMRI experiment with two stimuli and 40 time points, and if the residual effects are assumed to bias the estimates of the direct effects in the HRFs, one may use the generating vector with $n = 40$ and $b = 0$, which is (1100001011101111010001000010110011110100). Then, a CAOA with 40 time points (number of columns) circulating 10 times (number of rows) enjoys the orthogonality property among rows, as evidenced by the all-zero off-diagonal entries in its information matrix.

As mentioned in the introduction, the primary goal of this work is to provide a cost-efficient experimental plan for fMRI experiments when the interactions between the direct effect and its residual effects are nonnegligible. Following the notation in Cheng, Kao, and Phoa (2017), let $y_1, \ldots, y_N$ be the BOLD signals collected using an fMRI scanner to repeatedly scan a voxel of the subject's brain, while a stimulus sequence $d = (d_1, \ldots, d_N)^T$ is presented to the subject. The traditional model considers the aggregated

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

magnitude of the signals from the main effects of the individual HRF:

$$y_n = \gamma + \sum_{k=1}^{K}(x_{1,n-k+1}h_{1k} + (x_{2,n-k+1}h_{2k}) + \epsilon_n$$

where $x_{i,j}$, for $i = 1, 2$, and $j = n - K + 1, \ldots, n$, indicates the stimulus choice ($x_{i,j} = 1$ if the ith stimulus is assigned in the jth entry of the fMRI sequence, and zero otherwise). A CAOA of strength two guarantees that any two selected sequences (rows) from the CAOA are independent. Combined with the circulant property, this independency ensures that the magnitude measurement in a BOLD signal within a certain range of length K is independent of the others.

However, this signal independency assumption may be valid only when the time interval between two stimuli is sufficiently long. When two stimuli are given without an appropriate time interval, it is possible for memory effects to appear in the brain's recognition to the stimuli. Such effects can be expressed as an interaction effect between the current main-effect signal and its past main-effect signals . In order to guarantee that the main-effect signals in a certain sequence segment of length K are unbiased by other main-effect signals or the interaction effects of two signals, we need a CAOA of strength three or higher. In this case, any row selected from the CAOA is independent of both other rows and the interaction between two selected rows. The following model considers the aggregated magnitude of

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

BOLD signals from the main effects of the individual HRF and the memory (interaction) effects for $n = 1, \ldots, N$:

$$\begin{aligned} y_n = \gamma + \sum_{k=1}^{K} ((x_{1,n-k+1} h_{1k} + \sum_{k'=k+1}^{K} x_{1,n-k+1} x_{1,n-k'+1} h_{1k} h_{1k'}) \\ +(x_{2,n-k+1} h_{2k} + \sum_{k'=k+1}^{K} x_{2,n-k+1} x_{2,n-k'+1} h_{2k} h_{2k'})) + \epsilon_n. \end{aligned}$$

This argument solidifies our contribution to developing the theory of CAOAs of strength three. However, the detailed analysis and optimality studies of the fMRI experiments conducted using our proposed design are still under investigation; thus, we view these analysis methods as important future work to be built on our theoretical findings.

There are several other potential future areas of interest that can be extended from this work. First, it is of great interest to propose a systematic construction method for CAOAs of strength greater than three. An OA of high strength prevents the main effects from being biased by the effects of significant higher-order interactions, and provides an opportunity to study significant lower-order interaction effects. In the case of fMRI experiments, a CAOA of high strength is able to disentangle the bias on the estimates of the direct effects of an HRF from the multi-step residual effects. In addition, if some interaction effects are significant, and perhaps meaningful, from expert viewpoints, we can use these high-strength CAOAs to conduct fMRI experiments. Second, it is of great interest to consider fMRI experiments with more than two stimuli. Lin, Phoa, and Kao (2017b) proposed three- and four-level CAOAs of strength two. It is nontrivial to extend such results to CAOAs of strength three or higher using the HCDS method, but such experimental plans can further expand the applications of CAOAs in fMRI experiments.

Supplementary Material The online Supplemetary Materials consists of two parts. The first part includes the proofs of Theorem 1, Proposition 1, Lemma 1, Lemma 2, Theorem 2, and Theorem 3. The second part provides a list of the generating vectors of $CAOA(s^k, k, s, t, 0)$ when $8 \leq s^k \leq 1000$ and $2 \leq s \leq 10$, according the Corollary 1, and a list of the generating vectors of $CAOA(n, k, 2, 3, 0)$ when $n/4-1$ is a prime power and $8 \leq n \leq 392$, according to Theorem 3.

Acknowledgments The authors would like to thank Dr. Yasmeen Akhtar for her suggestions, and Ms. Ula Tuz-Ning Kung for her professional English editing on the revision of this article. This work was supported by Career Development Award of Academia Sinica (Taiwan) grant number 103-CDA-M04 and Ministry of Science and Technology (Taiwan) grant numbers 105-2118-M-001-007-MY2, 107-2118-M-001-011-MY3, and 107-2321-B-001-

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

038.

References

Bruijn, d. N. (1946). A combinatorial problem. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen. Series A* **49(7)**:758.

Buračas, G. T. and Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using M-sequences. *Neuroimage* **16(3)**, 801–813.

Cheng, C.-S. (1980). Orthogonal arrays with variable numbers of symbols. *The Annals of Statistics* **8(2)** 447–453.

Cheng, C.-S. (1995). Some projection properties of orthogonal arrays. *The Annals of Statistics*, **23(4)** 1223–1233.

Cheng, C.-S. and Kao, M.-H. (2015). Optimal experimental designs for fMRI via circulant biased weighing designs. *The Annals of Statistics* **43(6)**, 2565–2587.

Cheng, C.-S., Kao, M.-H. and Phoa, F.K.H. (2017). Optimal and efficient designs for functional brain imaging experiments. *Journal of Statistical Planning and Inference* **181**, 71–80.

Compeau, P. E. and Pevzner, P. A. and Tesler, G. (2011). How to apply de

Bruijn graphs to genome assembly. *Nature biotechnology* **29(11)**, 987–991.

Craigen, R., Faucher, G., Low, R., and Wares, T. (2013). Circulant partial Hadamard matrices. *Linear Algebra and its Applications* **439**, 3307–3317.

Elliott, J. and Butson, A. et al. (1966). Relative difference sets. *Illinois Journal of Mathematics* **10(3)**, 517–531.

Fredricksen, H. (1982). A survey of full length nonlinear shift register cycle algorithms. *SIAM review* **24(2)**, 195–221.

Good, I. J. (1946). Normal recurring decimals. *Journal of the London Mathematical Society* **1(3)**, 167–169.

Hedayat, A. S. and Sloane, N. J. A. and Stufken, J. (1999). Orthogonal arrays: theory and applications. *Springer.*

Jansma, J. M., de Zwart, J. A., van Gelderen, P., Duyn, J. H., Drevets, W. C., Furey, M. L. (2013). In vivo evaluation of the effect of stimulus distribution on FIR statistical efficiency in event-related fMRI. *Journal of neuroscience methods* **215(2)**, 190–195.

Kao, M.-H. (2013). On the optimality of extended maximal length linear feedback shift register sequences. *Statistics & Probability Letters* **83**, 1479–1483.

Kao, M.-H. (2014). A new type of experimental designs for event-related fMRI via Hadamard matrices. *Statistics & Probability Letters* **84**, 108–112.

Kao, M.-H. (2015). Universally optimal fMRI designs for comparing hemodynamic response functions. *Statistica Sinica* **25**, 499–506.

Lazar, N. (2008). The statistical analysis of functional MRI data. *Springer Science & Business Media*.

Lin Y.-L., Phoa, F. K. H., and Kao, M.-H. (2017a). Circulant partial Hadamard designs: construction via general difference sets and its application to fMRI experiments. *Statistica Sinica* **27(4)**, 1715–1724.

Lin Y.-L., Phoa, F. K. H., and Kao, M.-H. (2017b). Optimal design of fMRI experiments using circulant (almost-)orthogonal arrays. *The Annals of Statistics*, **45(6)**, 2483 — 2510.

Liu, T. T. and Frank, L. R. (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types: Part I: Theory. *Neuroimage* **21(1)**, 387–400.

Statistica Sinica: Preprint
doi:10.5705/ss.202017.0503

Liu, T. T. (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types: Part II: design of experiments. *Neuroimage* **21(1)**, 401–413.

Low, R., Stamp, M., Craigen, R., and Faucher, G. (2005). Unpredictable binary strings. *Congressus Numerantium* **177**, 65–75.

MacWilliams, F. J. and Sloane, N. J. (1976). Pseudo-random sequences and arrays. *Proceedings of the IEEE* **64(12)**, 1715–1729.

Pott, A., Reuschling, D., and Schmidt, B. (1997). A multiplier theorem for projections of affine difference sets. *Journal of statistical planning and inference* **62(1)**, 63–67.

Raktoe, B. L., Hedayat, A., and Federer, W. T. (1981). Factorial designs. *Wiley New York*.

Rao, C. R. (1946). Hypercubes of strength "d" leading to confounded designs in factorial experiments. *Bulletin of the Calcutta Mathematical Society* **38**, 67–78.

van Lint, J. H. and Wilson, R. M. (2001). A course in combinatorics. *Cambridge university press*.

West, D. B. (2001). Introduction to graph theory. 2nd Edition. *London: Prentice Hall*.

Institute of Statistical Science, Academia Sinica, Taiwan, R.O.C.

E-mail: gaussla@stat.sinica.edu.tw

E-mail: fredphoa@stat.sinica.edu.tw