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
Notice: Accepted version subject to English editing.	

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: Experiments in functional magnetic resonance imaging (fMRI) are important to render a correct statistical inference on brain function, but the theoretical constructions of efficient designs for these crucial experiments are few in the literature. Recent work on the construction of circulant orthogonal arrays by using algebraic difference sets was promising, but it was of limited use assuming that any interactions between the effect of a hemodynamic response function (HRF) and its residual effects are negligible on the magnitude estimation of BOLD signals collected from fMRI. In this work, we proposed a theoretical construction of the circulant orthogonal array of high strength via an extension from the complete difference system. In the analysis of fMRI experiments that are conducted by our proposed designs, the main effects of the individual HRF towards the signal are unbiased from the main effects from other HRFs and the interaction with its residual effects. Some properties of this new class of designs are studied, and the statistical regression model associated with this class of designs is revealed.

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

1. Introduction

Functional magnetic resonance imaging (fMRI) experiments are widely conducted to provide guidelines for the prevention and treatment of some terrible brain disorders such as Alzheimer's disease. A highly efficient experimental design is an important step towards a successful study of functional brain images. The event-related fMRI (ER-fMRI) leads to the shape estimation of hemodynamic response functions (HRFs), which are associated with transient brain activation evoked by various mental stimuli. An ER-fMRI design is a sequence of stimuli administrated by an experimenter, and such design is regarded as a circulant design (Kao (2013)). In the study of an fMRI experiment, a design may contain tens to hundreds of stimuli from cerebral neuronal activities. They lead to a change, described in the form of HRFs, in the ratio of "oxy- to deoxy-blood" that was detected in the MRI scanner as a change in the strength of the magnetic field. After the onset of a stimulus, HRF takes several seconds to return to the baseline completely. 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 the use of the m-sequence in the fMRI experiments and its performance is good, as indicated in several studies (Liu and Frank (2004); Liu (2004); Jansma et al. (2013)), but its length was limited to be $n = (Q+1)^r - 1$, where Q stands for the total number of stimulus types and r is a positive integer. To relax this constraint, Liu (2004) and Kao (2013) proposed the uses of a truncated version and an extended version respectively. Since the former suffered from an efficiency loss (Kao (2014)) while the latter was universally optimal as only a few effects were estimated, a new class of highly efficient fMRI designs with flexible run sizes were of great interest. Kao (2014) proposed the H-sequence, which existed when its length $n \equiv 3 \pmod 4$ is a prime, twins prime or a power of 2, for ER-fMRI experiments with one stimulus type. Run size flexibility was the advantage of H-sequence, but it fitted 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 was a $k \times n$ circulant (± 1)-matrix such that $HH^T = nI_k$, and r was the row sum of H. A 0-$H(k \times n)$ was a two-symbol, n-run, k-factor *circulant orthogonal array* (COA) and it could be highly efficient for fMRI experiments (Kao (2015)) when $n \equiv 0 \pmod 4$. The optimal design properties of these classes were comprehensively described in (Buračas and Boynton (2002); Liu (2004); Kao (2013, 2015); Cheng and Kao (2015); Lin, Phoa, and Kao (2017a,b)).

Introduced in (Rao (1946)), an orthogonal array (OA) is a widely used class of experimental designs in various areas, such as medicine, agriculture, manufacturing, and many others. See (Hedayat, Sloane, and Stufken (1999)) for more details. Orthogonality and projectivity of effect estimates are the major advantages when an OA is used as an experimental plan (Cheng (1980, 1995); Raktoe, Hedayat, and Federer (1981)). However, an OA is also constrained for its inflexibility in run sizes (a multiple of s^t). Recently, Lin, Phoa, and Kao (2017b) proposed a unified method to obtain orthogonal arrays with circulant property, which can be used in fMRI experiments with any run sizes and two or higher number of levels. This generalized structure of fMRI designs, called *circulant (almost-)orthogonal array* (CAOA), guarantees the t-tuples to appear almost equally often. We provide its definition below.

Definition 1. A circulant $k \times n$ array $\boldsymbol{A}$ with entries from Z_s is said to be a circulant almost orthogonal array (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 the CAOAs of strength two and compiled a table of universally optimal $CAOA(n, K, 2, 2, 1)$ when $n \leq 600$. These CAOAs helped the researchers to measure the magnitudes of HRFs at specific time points under an assumption that no interactions between the direct effect and the residual effects were detected in the signal output from the MRI scanner. Such an assumption was highly unlikely to be real, but the ignorance of these interaction effects was common in practice for analytical simplicity, and it might lead to a bias on the estimation of the direct effect if the residual effects were still significant. In the language of experimental designs, an orthogonal array of strength three or higher provided an ability to estimate the main effect that was free from the bias of two-factor interactions. Therefore, a CAOA of strength three provided an adjusted estimation on the HRFs so that it was unbiased from its residual effects.

This paper aims to propose a systematic construction of good CAOAs of strength three, denoted as $CAOA(n, K, 2, 3, b)$ for $b = 0, 1$. Section 2 provides some 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 of $CAOA(n, K, 2, 3, 0)$ is provided as a result of the theoretical derivations. Section 4 provides the proofs of the theorems. Some discussions on the analysis model are given together with a conclusion in the last section.

2. Circulant (Almost-)Orthogonal Arrays: Some Properties

We begin with a discussion on the properties of $CAOA(n, k, s, t, b)$. A specific orthogonal array, $OA(s^t, t+1, s, t)$, can be constructed via zero-sum array property that the levels in every run add up to zero, which provides a lower bound for the number of factors, but this method cannot apply to COAs and CAOAs. Although an $OA(s^t, t, s, t)$ can be trivially obtained by finding out 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 a $s-$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 to the study of 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 whose vertices are sequences of symbols from some symbols and whose 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 details. The 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). In tradition, a De Bruijn sequence requires every $t-$tuple to appear exactly once, so its graph is a regular simple graph. We extend the De Bruijn graph to construct $CAOA(n, t, s, t, b)$ below. Define Λ to be a *frequency sequence* $(\lambda^{a_1 \dots a_t})_{a_i \in Z_s}$, which represents the frequency of a $s-$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 whose vertex set 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 $(d_1, d_2, \ldots, d_{t-1})$ 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 $d_1 d_2 \ldots d_n$ that comprises the edges $d_1 d_2 \ldots d_t, d_2 d_3 \ldots d_{t+1}, \cdots, d_n d_1 \ldots d_{t-1}$. For example, a circuit $00 \longrightarrow 01 \longrightarrow 11 \longrightarrow 10$ can be simply represented by 0011. Through 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 $t-$tuple 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 will be given in Section 4. The above theorem provides an easy construction to find the generating vector of a CAOA. This also guarantees the lower bound for k in a CAOA of strength t is at least t. As in the demonstration, we construct a $CAOA(20, 3, 2, 3, 1)$ via 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 3-dimensional De Bruijn frequency graph $G = (V, E)$ is given in Figure 1.*

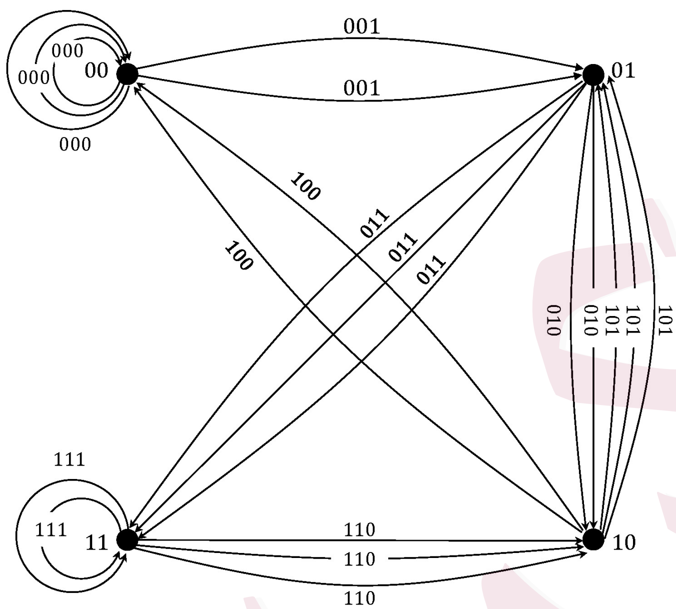

Figure 1: A 3-dimensional De Bruijn frequency graph of two symbols based on $(3, 2, 2, 3, 2, 3, 3, 2)$

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* $\lambda^{a_1a_20} + \lambda^{a_1a_21} = \lambda^{0a_1a_2} + \lambda^{1a_1a_2}$ *for all* $a_1, a_2 \in Z_2$, *such as* $\lambda^{010} + \lambda^{011} = \lambda^{101} + \lambda^{001} = 5$. *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 equal in-degree and out-degree, so G *is an Eulerian. 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. *It is a generating vector of a* $CAOA(20, 3, 2, 3, 1)$ *and each column is a triplet. The occurrence number 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 non-zero integers and* $s > 1$. *There exists a* $CAOA(s^k, k, s, t, 0)$ *for all* $2 \leq t \leq k$.

Corollary 1 can be trivially proved via Theorem 1. According to Corollary 1, we found the generating vectors of a $CAOA(s^k, k, s, t, 0)$ when $8 \leq s^k \leq 1000$ and $2 \leq s \leq 10$. All the designs are listed in Table 1 of supplementary materials.

With an appropriate choice of a frequency sequence, a CAOA of strength t might produce a CAOA of strength t' where $t' \leq t$. For example, the generating vector of a $CAOA(20, 3, 2, 3, 1)$ 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 the necessary condition of 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

$\sum_{x=0}^{s-1} \lambda^{a_1 \ldots a_{t-1} x} = \sum_{x=0}^{s-1} \lambda^{x a_1 \ldots a_{t-1}}$ and $\lambda^{a_1 \ldots a_t} \geq 1$ for all $a_i \in Z_s$.

3. Circulant (Almost-)Orthogonal Arrays: Construction Method

In the aspect of fMRI experiments, designs with circulant property are required for 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 via 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 onto designs of higher strength via a CDS is not trivial.

We present the difference structure of CAOAs of high strength via the *high-order complete difference system* (HCDS). In this work we only consider the designs of strength three, but it can be easily extended to designs of strength higher than three. Let $V = \{V_0, V_1, \cdots, V_{s-1}\}$ be a partition of Z_n. The collection of differences is $\mathscr{S}^{\alpha,\beta} = \{\boldsymbol{S}_{a_i}^{\alpha,\beta} |$ for all $a_i \in V_\alpha\}$ 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 $Z_n \setminus \{0\}$, there are 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*. It is a Latin square, so each element appears exactly once in each row and column of

$\mathscr{S}$.

We define the mth-order difference $\lambda^{\alpha,\beta}_{r_1,\cdots,r_m} = \#\{g \in V_\alpha | \{r_1, \cdots, r_m\} \subseteq S^{\alpha,\beta}_g\}$, which counts the total number of $\boldsymbol{S}^{\alpha,\beta}_g$ containing the subset $\{r_1, \cdots, r_m\}$. As in the definition of CDS (Lin, Phoa, and Kao (2017b)), the $\lambda^{\alpha,\beta}_r$ in the $\boldsymbol{\Lambda}$ of a $(n, k, s, \boldsymbol{\Lambda})$-CDS is the first-order difference.

We define the high-order complete difference system (HCDS) that summarizes all information on mth-order $(1 \le m \le t-1)$ 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-tuple $(\boldsymbol{\Lambda}_{1,\cdots,m}, \ldots, \boldsymbol{\Lambda}_{n-m,\cdots,n-1})$ for $1 \le m \le t-1$, 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 $1 \le r_1 < \cdots < r_m \le k$. 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; $i = 1, \ldots, k$, $j = 1, \ldots, n$ and $l = 0, \ldots, s-1$.

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

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

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 $r, r_1, r_2 \in Z_8 \setminus \{0\}$. The first-order difference $\lambda_r^{\alpha,\beta}$ is the frequency of the 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 contains 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 in the construction of the design. For 2-level CAOAs, we state a useful proposition that describes the relationship between the first-order and the second-order differences. It also provides an easy way to calculate the second-order difference via 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* $1 \leq r_1 < r_2 \leq k$.

Since 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 HCDS, we define

$\delta_{r_1,r_2,\cdots,r_m}^{c_0,c_1,\cdots,c_m} = \#\{g \in V_{c_0} | r_j \in \boldsymbol{S}_g^{c_0,c_j} \text{ for } j = 1,2,\cdots,m\}$, then the high-order factor combinations can be explicitly represented by 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 combination $c_0, c_1, \cdots, c_m$ 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, the $\delta_{1,2}^{0,0,1}$ is the number of rows in $(\mathscr{S}^{0,0}|\mathscr{S}^{0,1})$ such that $\mathscr{S}^{0,0}$ and $\mathscr{S}^{0,1}$ contains 1 and 2, respectively. Since 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 a $CAOA(8,3,2,3,0)$.

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 only show the relationship between $\delta_{r_1,r_2}^{c_0,c_1,c_2}$ and $\lambda_{r_1,r_2}^{\alpha,\beta}$ for 2-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 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 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, but the combinations might be different. It implies that the bandwidth of $\boldsymbol{\delta}_{r_1,r_2}$ and $\boldsymbol{\delta}_{r_1',r_2'}$ are all less than one, but $\boldsymbol{\delta}_{r_1,r_2}$ 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 it helps in our analysis. We call a $CAOA(n,k,s,t,b)$ is *uniform* if the frequency of the level-combination associated with each $t \times n$ submatrix equals 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. 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 the condition (i) holds.*

Theorem 2 provides a good strategy 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 a 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 $8 \leq n \leq 27$ and we summarize the results in 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 to construct 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 λs are equal to $n/8-1$ except $\lambda_{n/4} = 0$. Then there exist two elements $g, g' \in D^c$ such that $g' - g = n/4$, 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 λs 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 commonly used and 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 m-sequence have the 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, which are smaller than those of 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 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 m-sequence is much higher than the CAOAs in Theorem 3. In addition, the CAOA of bandwidth zero always has a higher priority than the CAOA of bandwidth one. Therefore, CAOAs described Theorem 3 is effective as well as economical.

4. Discussion and Conclusion

In this work, we introduce the use of circulant (almost-)orthogonal arrays (CAOAs) of strength three as an experimental plan for an ER-fMRI experiment, where the estimates of the direct effects in HRFs are biased from their residual effects. Although this work is an extension from (Lin, Phoa, and Kao (2017b)), the core tool, namely the complete difference system (CDS), is not enough for characterizing and constructing the CAOAs of strength three. Thus, it leads to our introduction of its generalized version called high-order complete difference system (HCDS). We not only theoretically study some properties of CAOAs of high strength, but also 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 the generating vectors that 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 $n \leq 392$. The practitioners can simply use this table and find their required fMRI experimental plans. For example, if an experimenter administrates 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 HRFs, one may use the generating vector with $n = 40$ and $b = 0$, which is (11000010111011110100010000101100111110100). Then a CAOA with 40 time points (number of columns) circulating 10 times (number of rows) enjoys the orthogonality property among rows, which is reflected from 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 non-negligible. Following the notations in Cheng, Kao and Phoa (2017), let $y_1, \ldots, y_N$ be the BOLD signals collected by 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 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 $j = n - K + 1, \ldots, n$, is the indicator of the stimulus choice ($x_{i,j} = 1$ if the ith stimulus is assigned in the jth entry of the fMRI sequence, or 0 otherwise). A CAOA of strength two guarantees that any two selected sequences (rows) from that CAOA are independent. Combining 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 long enough. When two stimuli are given without an appropriate amount of interval time, it is possible for memory effects to appear in the brain's recognition to the stimuli. This memory effect can be expressed as the interaction effect between the current main-effect signal and its past main-effect signals . In order to guarantee the main-effect signals in a certain sequence segment of length K to be unbiased from other main-effect signals and these interaction effects of two signals, we need a CAOA with strength three or higher, so that any selected row from that CAOA is independent from both any other rows and the interaction between two selected rows. The following model considers the aggregated magnitude of 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 fMRI experiments conducted via our proposed design are still under investigation, and we view these analysis methods as important future work 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 higher than 3. An orthogonal array of a high strength avoids the bias estimates of the main effects from the effects of significant higher-order interactions and opens an opportunity to study significant lower-order interaction effects. In the case of fMRI experiments, a CAOA of high strength possesses the disentanglement ability towards the bias on the estimates of the direct effects of HRF from the multi-steps residual effects. In addition, if some interaction effects are significant and perhaps meaningful from expert viewpoints, one may use these high-strength CAOAs to conduct the fMRI experiments. Second, it is of great interest to consider fMRI experiments with more than two stimuli. Lin, Phoa, and Kao (2017b) proposed the 3-level and 4-level CAOAs of strength two. It is non-trivial to extend such results to CAOAs of strength 3 or higher using the HCDS method, but such experimental plans can further expand the applications of CAOAs in the fMRI experiment.

Acknowledgements 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-001-038.

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.

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