Statistica Sinica **20** (2010), 1063-1075

AN APPROACH TO CONSTRUCTING NESTED SPACE-FILLING DESIGNS FOR MULTI-FIDELITY COMPUTER EXPERIMENTS

Ben Haaland and Peter Z. G. Qian

University of Wisconsin-Madison

Abstract: Multi-fidelity computer experiments are widely used in many engineering and scientific fields. Nested space-filling designs (NSFDs) are suitable for conducting such experiments. Two classes of NSFDs are currently available. One class is based on special orthogonal arrays of strength two, and the other consists of nested Latin hypercube designs; both of them assume all factors are continuous. We propose an approach to constructing new NSFDs based on powerful (t, s)-sequences. The method is simple, easy to implement, and quite general. For continuous factors, this approach produces NSFDs with better space-filling properties than existing ones. Unlike the previous methods, it can also construct NSFDs for categorical and mixed factors. Some illustrative examples are given. Other applications of the constructed designs are briefly discussed.

Key words and phrases: Computer experiments, design of experiments, finite element analysis models, nested space-filling designs, scrambled nets, space-filling designs.

1. Introduction

A large, expensive computer program, like a finite element analysis model, can be executed at various degrees of fidelity, resulting in computer experiments with multiple levels of cost and accuracy. Observations from such experiments are often used for building a statistical emulator to predict the output from the most accurate experiment involved (Kennedy and O'Hagan (2000); Qian et al. (2006); Qian and Wu (2008)). Efficient data collection is critical to conducting such experiments. Throughout, consider a situation involving u computer experiments $y_1, y_2, \ldots, y_u$, where y_1 is the most accurate, y_2 is the second most accurate, and so on.

Space-filling designs achieve uniformity in low dimensions: when projected into low dimensions, the design points are evenly scattered in the design region. Let $D_1 \subset \cdots \subset D_u$ denote a set of *nested space-filling designs* (NSFDs) with u components, where each D_i is a space-filling design. It is appealing to use a set of of such D_i's for obtaining observations from $y_1, \ldots, y_u$, with D_i associated with y_i. The space-filling properties of the designs enable the possibility of fully exploring the design space. The nesting among the D_i's makes it easier to model the systematic differences among the y_i's and implies more observations are taken for less expensive experiments.

Currently, NSFDs constructed by using nested Latin hypercubes (Qian (2009); Husslage, Dam and Hertog (2005)) and some special orthogonal arrays of strength two (Qian, Ai and Wu (2009); Qian, Tang and Wu (2009)) are available. These designs are for continuous factors and achieve uniformity in one- or two-dimensional projections. In this article we propose an approach to constructing NSFDs by exploiting structures in powerful (t, s)-sequences. This approach has several advantages over existing NSFD constructions. First, the constructed designs have better space-filling properties than existing ones; more discussion on this is given in Sections 2 and 3. Second, the constructions can produce designs for any number of continuous, categorical, or mixed factors. Third, the designs can be generated in a very flexible manner.

The remainder of the article is organized as follows. Section 2 gives some definitions, properties and useful notation. Sections 3−5 give the constructions of three types of NFSDs. We conclude with a short summary in Section 6.

2. Definitions and Notation

Let $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$. For an integer $b \geq 1$, $\mathbb{Z}_b$ denotes the set $\{0, 1, \ldots, b-1\}$ and Ω_b denotes the set $\{1, \ldots, b\}$. For a real number x, $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. Throughout, any continuous factor is assumed to take values in $[0, 1]$.

Definition 1. For $b \geq 2$, an elementary interval $E \subset [0, 1)^s$ in base b is any set of the form

$$E = \prod_{i=1}^{s} \left[\frac{a_i}{b^{d_i}}, \frac{a_i + 1}{b^{d_i}} \right), \tag{2.1}$$

where a_i, d_i are integers with $d_i \geq 0$, $0 \leq a_i < b^{d_i}$ for $i = 1, \ldots, s$. Note that the set E has vol $E = b^{-\sum_{i=1}^{s} d_i}$.

Definition 2. For $b \geq 2$ and $0 \leq t \leq m$, a (t, m, s)-net in base b is a set of b^m points $D \subset [0, 1)^s$ so that every elementary interval E in base b with vol $E = b^{t-m}$ contains exactly b^t points from D.

Definition 3. A sequence of points in $[0, 1)^s$ is called a (t, s)-sequence in base b if for all $k \geq 0$ and $m > t$, the set of points in the sequence with indices n, $kb^m < n \leq (k + 1)b^m$, form a (t, m, s)-net in base b.

Points from a (t,s)-sequence achieve what is widely believed to be the best possible uniformity in terms of the rate of decrease of the star discrepancy, $O((\log\ N)^s/N)$ (Niederreiter (1992)). The star discrepancy of a set of points in $[0,1)^s$ is the maximum difference between the empirical and Uniform$(0,1)^s$ distribution functions. Mullen, Mahalanabis and Niederreiter (1995) provide a survey of the known methods of constructing (t,s)-sequences. In general, a (t,s)-sequence exists and is not too difficult to construct for every $s \geq 1$, $b \geq 2$, and a minimal $t \geq 0$. The sequences used in the examples of this article are constructed using the method given in Section 4.5 of Niederreiter (1992).

Owen (1995) proposed the idea of scrambling (t,m,s)-nets. Let A be a (t,m,s)-net in base b with entries $\{A_i = (A_i^1,\ldots,A_i^s)' : i = 1,\ldots,b^m\}$, and let X denote the scrambled version of A with entries $\{X_i = (X_i^1,\ldots,X_i^s)' : i = 1,\ldots,b^m\}$. Here A_i^j and X_i^j denote element j of A_i and X_i, respectively. This use of superscripts does not cause any confusion since we do not use powers of A_i and X_i. Observe that each A_i^j can be written as

$$A_i^j = \sum_{k=1}^{\infty} a_{i,k}^j b^{-k} \tag{2.2}$$

where $a_{i,k}^j \in \mathbb{Z}_b$. The basic idea of the scrambling is as follows. Each dimension of A is randomized independent of the other dimensions. For each dimension, the interval $[0,1]$ is chopped into b equal length pieces and these are randomly shuffled. At each additional stage, each of the individual pieces from the previous stage is chopped into b more equal length pieces and those pieces are randomly shuffled. Now the details are developed, following Loh (2003). First generate a set of random permutations of $\mathbb{Z}_b$,

$$\{\pi_j, \pi_{j;a_1}, \pi_{j;a_1,a_2}, \pi_{j;a_1,a_2,a_3}, \ldots : 1 \leq j \leq s, a_k \in \mathbb{Z}_b, k = 1,2,...\}, \tag{2.3}$$

where these permutations are mutually independent and each of them is uniformly distributed over its $b!$ possible values. Then take the entry X_i^j to be

$$X_i^j = \sum_{k=1}^{\infty} x_{i,k}^j b^{-k}, \tag{2.4}$$

where $x_{i,1}^j = \pi_j(a_{i,1}^j)$ and $x_{i,k}^j = \pi_{j;a_{i,1}^j,\ldots,a_{i,k-1}^j}(a_{i,k}^j)$ for $k \geq 2$.

For a geometrical illustration of this scheme, see Owen (1997). Owen (1995) shows that X is still a (t,m,s)-net with probability one, and the points in X are marginally Uniform$(0,1)^s$.

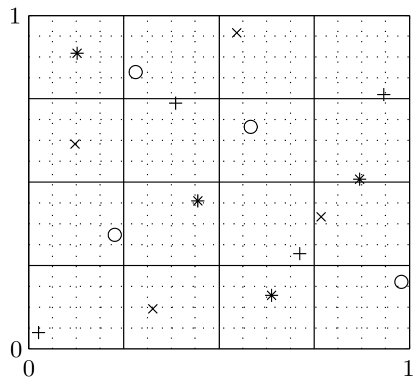


Figure 3.1. Four $(0, 1, 2)$-nets in base 4 contained in one $(0, 2, 2)$-net in base 4.

3. Construction of NSFDs for Continuous Factors

In this section we present the construction of NSFDs for continuous factors based on a (t, s)-sequence. This construction is simple, following the definition of (t, s)-sequences. It is known that it is possible to obtain nested designs from (t, s)-sequences Owen (1997). From the definition of a (t, s)-sequence in Section 2, we find the following.

Proposition 1. *Suppose there is a (t, s)-sequence in base b, and $t < m_1 < m_2$. Then, there are $b^{m_2-m_1}$ disjoint sets of points $A_k \subset B$ so that each A_k is a (t, m_1, s)-net in base b and B is a (t, m_2, s)-net in base b.*

Proof. Take B to be the collection of points in the sequence with indices n, $lb^{m_2} < n \leq (l+1)b^{m_2}$ for some $l \in \mathbb{Z}^+$. Then, B is a (t, m_2, s)-net in base b. Take A_k to be the collection of points in the sequence with indices n, $lb^{m_2} + kb^{m_1} < n \leq lb^{m_2} + (k+1)b^{m_1}$ for $0 \leq k < b^{m_2-m_1}$. Then, the A_k's are disjoint (t, m_1, s)-sequences in base b contained in B.

It is clear from the above that not only can we obtain a pair of nested nets with one containing the other but we can also *slice* a large net into several small nets.

Example 1. Figure 3.1 shows four $(0, 1, 2)$-nets in base 4 contained in one $(0, 2, 2)$-net in base 4. Notice that each elementary interval in base 4 with volume $1/4$ contains exactly one point from each of the subsets and every elementary interval in base 4 with volume $1/16$ contains exactly one point from the combined subsets.

Using these constructions, the set of designs, $D_1 \subset \cdots \subset D_u$, can be generated in a very flexible manner. The nested designs should be able to accommodate more or less observations than expected while retaining good space-filling

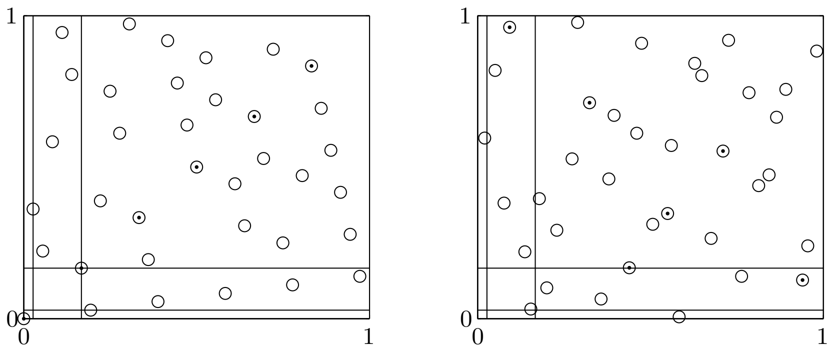


Figure 3.2. **Left Panel:** A $(0,2,2)$-net in base 6 with first 6 points and representative elementary intervals indicated. **Right Panel:** Randomized $(0,2,2)$-net in base 6 with first 6 points and representative elementary intervals indicated.

properties. Further, the set of nested designs should be able to accommodate additional designs with good space-filling properties at arbitrary locations within the nesting structure. In particular, suppose that the design D_i consists of the points from a fixed (t,s)-sequence with indices $k_i b^{m_i} < n \leq (k_i+1)b^{m_i}$. A design D_{i+1} with $D_i \subset D_{i+1}$ can be generated by taking the points from the (t,s)-sequence with indices $k_{i+1}b^{m_{i+1}} < n \leq (k_{i+1}+1)b^{m_{i+1}}$, where $m_{i+1} > m_i$ and $k_{i+1} = \lfloor k_i b^{m_i - m_{i+1}} \rfloor$. A design D_{i-1} with $D_{i-1} \subset D_i$ can be generated by taking the points from the (t,s)-sequence with indices $k_{i-1}b^{m_{i-1}} < n \leq (k_{i-1}+1)b^{m_{i-1}}$, where $m_{i-1} < m_i$ and $k_{i-1} = k_i b^{m_i - m_{i-1}} + j$ for some $j \in \{0, 1, \ldots, b^{m_i - m_{i-1}} - 1\}$.

Now we discuss how the randomization procedure given in Section 2 can affect nesting among the D_i's. It is apparent that if the D_i's are scrambled independently, the resulting designs would no longer be nested. A salient feature of the permutations in (2.3) is that they are defined with respect to the coefficients of X_i^j in the expansion (2.4), not the value of X_i^j directly. This property implies that if the A_i's are scrambled by using the same set of π's, the resulting designs will still be nested. Throughout, the same set of permutations is used whenever it is necessary to scramble nested (t,m,s)-nets. Similar ideas were used in Yue (1999) for scrambling a union of nets.

Example 2. Consider the $(0,2,2)$-net in base 6 from a $(0,2)$-sequence in base 6 shown in Figure 3.2. Representative elementary intervals in base 6 of volumes 1/6 and 1/36 are indicated. Notice that the first 6 points form a $(0,1,2)$-net in base 6 and all 36 points form a $(0,2,2)$-net in base 6. Now consider the randomized version of the net shown in Figure 3.2. Notice that the first 6 points are still a $(0,1,2)$-net in base 6, the set of all points is still a $(0,2,2)$ net in base 6, and the points of the randomized net have less pattern in their arrangement than those of the original un-randomized net.

As mentioned previously, NSFDs based on (t, s)-sequences have excellent space-filling properties. NSFDs constructed in Qian (2009), Qian, Ai and Wu (2009), and Qian, Tang and Wu (2009) have the property that one- or two-dimensional projections are uniform for all designs within the nesting structure. In contrast, the NSFDs constructed in this section can achieve uniformity in higher dimensions and have uniform low-dimensional projections for increasingly large sets of dimensions as one moves through the nesting structure. Also, it is well-known that Sobol' sequences have a nesting structure (Santner, Williams and Notz (2003)). Note that Sobol' sequences are (t, s)-sequences in base 2 with a larger than optimal value of t. All results here apply immediately to Sobol' sequences.

4. Construction of NSFDs for Categorical Factors

Computer models with only categorical factors appear in areas like combinatorial chemistry. This section presents the construction of NSFDs for categorical factors. This serves as a building block for the construction of NSFDs with mixed factors in Section 5.

For a c-level categorical factor, we denote the c levels by $0, 1, \ldots, c-1$. The basic idea of the construction is to map the values of nested designs constructed in Section 3 to form nested orthogonal arrays, i.e., large orthogonal arrays with high strength containing smaller ones with lower strength. Orthogonal arrays have good balance properties among the factors. An orthogonal array of N runs with s factors each with c levels and strength r, denoted $\mathrm{OA}(N, s, c, r)$, is an $N \times s$ matrix in which any r columns contain all possible combinations of symbols equally often. For $r \geq s$, $\mathrm{OA}(N, s, c, r)$ denotes a full factorial.

The following is a result which Owen (1995) pointed out. We present it here and provide a proof.

Proposition 2. *Using a (t, m, s)-net in base b, an $\mathrm{OA}(b^m, s, b^q, \lfloor (m-t)/q \rfloor)$ can be produced for $q = 1, \ldots, m-t$ by applying the element-wise mapping $\lfloor b^q X_i^j \rfloor$.*

Proof. Let $\tilde{X}$ denote the matrix consisting of the mapped X_i^j's. Note that $\tilde{X}_i^j \in \mathbb{Z}_{b^q}$. Arbitrarily take $r = \lfloor (m-t)/q \rfloor$ columns from $\tilde{X}$, call them $\mathcal{R}$, and consider a partition of $[0,1)^s$ into b^{qr} elementary intervals each of which is of the form

$$\left(\prod_{j \in \mathcal{R}^c} [0, 1) \right) \times \left(\prod_{j \in \mathcal{R}} \left[\frac{a_j}{b^q}, \frac{a_j + 1}{b^q} \right) \right), \tag{4.1}$$

where a_k ranges over $0, 1, \ldots, b^q - 1$ for $k = 1, \ldots, r$. Each of these elementary intervals contains exactly b^m/b^{qr} points from the net that correspond to the b^m/b^{qr} rows of $\tilde{X}$ where the r selected columns have the level combination $a_1, \ldots, a_r$.

Proposition 3. *Using a (t, s)-sequence in base b, $b^{m_2 - m_1}$ disjoint* $\mathrm{OA}(b^{m_1}, s, b^q, \lfloor (m_1 - t)/q \rfloor)$*'s* $\subset \mathrm{OA}(b^{m_2}, s, b^q, \lfloor (m_2 - t)/q \rfloor)$ *can be produced for $q = 1, \ldots, m_1 - t$, where $m_2 > m_1$.*

Proof. By Proposition 1, there exist $b^{m_2 - m_1}$ (t, m_1, s)-nets in base b contained in a (t, m_2, s)-net in base b. Apply the element-wise mapping $\lfloor b^q X_i^j \rfloor$ to each point. The conclusion follows from Proposition 2.

Similar to the constructions following Proposition 1, it is easy to construct NSFDs for categorical factors based on Proposition 3.

Example 3. Consider a computer model with four categorical factors, each with 4 levels. Suppose that the model needs to be conducted at three levels of accuracy, with 4 runs for y_1, 16 runs for y_2 and 64 runs for y_3. Figure 4.3 shows an $\mathrm{OA}(64, 4, 4, 3)$ constructed by Proposition 3. This array is divided into four $\mathrm{OA}(16, 4, 4, 2)$'s given in four blocks, each of which is further divided into four $\mathrm{OA}(4, 4, 4, 1)$'s as separated by the solid lines. The design sets for the y_i's are as follows: for y_1 use Runs 1-4 in block 1; for y_2 use all runs in block 1; for y_3 use all runs in the four blocks.

2	0	3	3	2	1	0	1	2	3	2	2	2	2	1	0
0	1	2	0	0	0	1	2	0	2	3	1	0	3	0	3
3	3	1	1	3	2	2	3	3	0	0	0	3	1	3	2
1	2	0	2	1	3	3	0	1	1	1	3	1	0	2	1
2	1	1	2	2	0	2	0	2	2	0	3	2	3	3	1
0	0	0	1	0	1	3	3	0	3	1	0	0	2	2	2
3	2	3	0	3	3	0	2	3	1	2	1	3	0	1	3
1	3	2	3	1	2	1	1	1	0	3	2	1	1	0	0
2	3	0	0	2	2	3	2	2	0	1	1	2	1	2	3
0	2	1	3	0	3	2	1	0	1	0	2	0	0	3	0
3	0	2	2	3	1	1	0	3	3	3	3	3	2	0	1
1	1	3	1	1	0	0	3	1	2	2	0	1	3	1	2
2	2	2	1	2	3	1	3	2	1	3	0	2	0	0	2
0	3	3	2	0	2	0	0	0	0	2	3	0	1	1	1
3	1	0	3	3	0	3	1	3	2	1	2	3	3	2	0
1	0	1	0	1	1	2	2	1	3	0	1	1	2	3	3

Figure 4.3. The designs used for Example 3, $\mathrm{OA}(4, 4, 4, 1)$'s $\subset \mathrm{OA}(16, 4, 4, 2)$'s $\subset \mathrm{OA}(64, 4, 4, 3)$.

5. Construction of NSFDs for Mixed Factors

In this section we discuss how to construct NSFDs for mixed factors. Recent work of Han et al. (2009) and Qian, Wu and Wu (2008) demonstrates that computer codes can include categorical factors as well. For example, the computational fluid dynamics code used for the data-center experiment in Schmidt, Cruz and Iyengar (2005) includes categorical variables "diffuser location" and "return air vent location." Computer models in marketing and social sciences often involve qualitative factors such as gender and commuting method.

Suppose that $y_1, \ldots, y_u$ involve both continuous factors w and categorical factors z. For D_i, the design for y_i, let D_i^w and D_i^z denote the continuous and categorical parts of D_i. The designs, $D_1 \subset \cdots \subset D_u$, should have several properties. First, the D_i^w's and D_i^z's should be respectively nested, as done in Propositions 1 and 3. Second, it is important to consider the interplay between w and z in obtaining observations from the experiment y_i. Note that y_i may perform distinctly under different level combinations of z or with different value ranges of w. Therefore, for every level combination of z, the corresponding points in D_i^w should have good space-filling properties. This idea was also used in Qian and Wu (2009) for constructing sliced space-filling designs based on orthogonal arrays. Similarly, for every possible range of values of w, the corresponding points in D_i^z should achieve some uniformity.

Here is some additional notation. Let card $\mathcal{S}$ denote the cardinality of a finite set $\mathcal{S}$. Recall that, for an integer $s \geq 1$, Ω_s denotes the set $\{1, \ldots, s\}$. For a subset $\mathcal{S}_i$ of Ω_s and an s-dimensional vector a, let $a|_{\mathcal{S}_i}$ denote the sub-vector of a consisting of the entries in the dimensions specified by $\mathcal{S}_i$.

Lemma 1. *Suppose that X is a (t, m, s)-net in base b, and that Ω_s is divided into two components $\mathcal{S}_1$ and $\mathcal{S}_2$, with $s_1 = \text{card } \mathcal{S}_1$ and $s_2 = \text{card } \mathcal{S}_2$. Then we have:*

(a) *$X|_{\mathcal{S}_1}$ and $X|_{\mathcal{S}_2}$ are (t, m, s_1) and (t, m, s_2)-nets in base b, respectively;*

(b) *for $k = 1$ or 2 and each fixed elementary interval E^* in $\mathbb{R}^{s_k}$ in base b with volume b^{-r} and $r \in \{0, 1, \ldots, m-t\}$, the points $\{x_i|_{\mathcal{S}_k^c} : x_i|_{\mathcal{S}_k} \in E^*\}$ form a $(t, m-r, s_k^c)$-net in base b where $\mathcal{S}_k^c = \Omega_s \setminus \mathcal{S}_k$ and $s_k^c = \text{card } \mathcal{S}_k^c$.*

Proof. (a) For $k = 1$ or 2, let $E = \prod_{j \in \mathcal{S}_k} \left[a_j/b^{d_j}, (a_j+1)/b^{d_j}\right)$ be an elementary interval in $\mathbb{R}^{s_k}$ in base b with volume $b^{-\sum_{j \in \mathcal{S}_k} d_j} = b^{t-m}$. Then,

$$\left(\prod_{j \in \mathcal{S}_k^c} [0,1)\right) \times \left(\prod_{j \in \mathcal{S}_k} \left[\frac{a_j}{b^{d_j}}, \frac{a_j+1}{b^{d_j}}\right)\right) \tag{5.1}$$

is an elementary interval in $\mathbb{R}^s$ in base b with volume b^{t-m} and so contains exactly b^t points from X.

(b) Write $E^* = \prod_{j \in \mathcal{S}_k} \left[a_j/b^{d_j}, (a_j+1)/b^{d_j}\right)$. Then, $b^{-\sum_{j\in\mathcal{S}_k} d_j} = b^{-r}$. Consider an arbitrary elementary interval in $\mathbb{R}^{s_k^c}$ in base b, $F^* = \prod_{j\in\mathcal{S}_k^c} \left[a_j/b^{d_j}, (a_j+1)/b^{d_j}\right)$, with volume $b^{-\sum_{j\in\mathcal{S}_k^c} d_j} = b^{-(m-t-r)}$. Then $F^* \times E^*$ is an elementary interval in $\mathbb{R}^s$ in base b with volume b^{t-m} and so contains exactly b^t points from X.

Proposition 4. *Suppose that X, $\mathcal{S}_1$ and $\mathcal{S}_2$ are as defined in Lemma 1. We obtain a new design, Y, by applying the element-wise mapping $\lfloor b^q X_i^j \rfloor$ for all entries of X with $j \in \mathcal{S}_1$, where $q \in \{1, \ldots, m-t\}$. Then, we have:*

(a) *$Y|_{\mathcal{S}_1}$ is an $\mathrm{OA}(b^m, s_1, b^q, \lfloor (m-t)/q \rfloor)$ and $Y|_{\mathcal{S}_2}$ is a (t, m, s_2)-net in base b;*

(b) *for any elementary interval E^* in $\mathbb{R}^{s_2}$ in base b with volume b^{-r}, where $r \in \{1, \ldots, m-t-q\}$, the points $\{y_i|_{\mathcal{S}_1} : y_i|_{\mathcal{S}_2} \in E^*\}$ form an $\mathrm{OA}(b^{m-r}, s_1, b^q, \lfloor (m-t-r)/q \rfloor)$;*

(c) *for any fixed set of v components with $v \in \{1, \ldots, \lfloor (m-t)/q \rfloor\}$ of $Y|_{\mathcal{S}_1}$, call them $\mathcal{S}_*$, and any fixed level combination $E^* \in \mathbb{Z}_{b^q}^v$, the points $\{y_i|_{\mathcal{S}_2} : y_i|_{\mathcal{S}_*} = E^*\}$ form a $(t, m-vq, s_2)$-net in base b.*

Proof. (a) It follows immediately from Lemma 1 and Proposition 3.

(b) From Lemma 1, $\{x_i|_{\mathcal{S}_1} : x_i|_{\mathcal{S}_2} \in E^*\}$ form a $(t, m-r, s_1)$-net in base b. Applying the element-wise mapping to this design gives an $\mathrm{OA}(b^{m-r}, s_1, b^q, \lfloor (m-t-r)/q \rfloor)$.

(c) Note that the points $\{y_i|_{\mathcal{S}_1} : y_i|_{\mathcal{S}_*} = E^*\}$ correspond to the points $\{x_i|_{\mathcal{S}_1} : x_i|_{\mathcal{S}_1} \in E\}$, where E is an elementary interval in $\mathbb{R}^{s_1}$ in base b given by

$$E = \left(\prod_{j\in\mathcal{S}_*} \left[\frac{a_j}{b^q}, \frac{a_j+1}{b^q}\right) \right) \times \left(\prod_{j\in\mathcal{S}_1\cap\mathcal{S}_*^c} [0,1) \right). \tag{5.2}$$

Because of Lemma 1, $\{y_i|_{\mathcal{S}_2} : y_i|_{\mathcal{S}_*} = E^*\}$ form a $(t, m-vq, s_2)$-net in base b.

Now we are ready to present the construction of NSFDs with mixed factors. Suppose there is a (t, s)-sequence in base b, $\mathcal{S}_1$ and $\mathcal{S}_2$ are as defined in Proposition 4, $m_2 > m_1 > t$, and $q \in \{1, \ldots, m_1-t\}$. Apply the element-wise mapping $\lfloor b^q X_i^j \rfloor$ for $j \in \mathcal{S}_1$. Let Y denote the resulting design.

Proposition 5. *For the constructed Y,*

(a) $Y|_{\mathcal{S}_1}$ *consists of* $b^{m_2-m_1}$ *distinct* $\mathrm{OA}(b^{m_1},\ s_1,\ b^q,\ \lfloor (m_1-t)/q\rfloor)$*'s* $\subset \mathrm{OA}(b^{m_2},\ s_1,\ b^q,\ \lfloor (m_2-t)/q\rfloor)$*;*

(b) $Y|_{\mathcal{S}_2}$ *consists of* $b^{m_2-m_1}$ *disjoint* (t, m_1, s_2)*-nets in base* b *contained in a* (t, m_2, s_2)*-net in base* b*;*

(c) *for any fixed set of* v *components of* $Y|_{\mathcal{S}_1}$ *with* $v \in \{1,\ldots,\lfloor (m_1-t)/q\rfloor\}$*, call them* $\mathcal{S}_*$*, and any fixed level combination* $E^* \in \mathbb{Z}^v_{b^q}$*, the points*

$$\{y_i|_{\mathcal{S}_2} : \ y_i|_{\mathcal{S}_*} = E^*,\ (k-1)b^{m_1} < i \le kb^{m_1}\} \tag{5.3}$$

form $b^{m_2-m_1}$ *disjoint* (t, m_1-vq, s_2)*-nets in base* b *as* k *ranges over* $\{1,\ldots, b^{m_2-m_1}\}$ *contained in the* (t, m_2-vq, s_2)*-net in base* b *given by* $\{y_i|_{\mathcal{S}_2} : \ y_i|_{\mathcal{S}_*} = E^*\}$*;*

(d) *for any elementary interval* E^* *in* $\mathbb{R}^{s_2}$ *in base* b *with volume* b^{-r} *and* $r \in \{1,\ldots,m_1-t-q\}$*, the points*

$$\{y_i|_{\mathcal{S}_1} : \ y_i|_{\mathcal{S}_2} \in E^*,\ (k-1)b^{m_1} < i \le kb^{m_1}\} \tag{5.4}$$

form $b^{m_2-m_1}$ *distinct* $\mathrm{OA}(b^{m_1}, s_1, b^q, \lfloor (m_1-t-r)/q\rfloor)$*'s as* k *ranges over* $\{1,\ldots,b^{m_2-m_1}\}$ *contained in the* $\mathrm{OA}(b^{m_2}, s_1, b^q, \lfloor (m_2-t-r)/q\rfloor)$ *given by* $\{y_i|_{\mathcal{S}_1} : \ y_i|_{\mathcal{S}_2} \in E^*\}$.

Proof. The statements follow from Propositions 1, 3, and 4.

Example 4. As an illustration of Proposition 5, consider a computer model with two categorical factors, each at 3 levels, and two continuous factors. Suppose that it is necessary to run three versions of the model, y_1, y_2, y_3, with 9 runs for y_1, 27 runs for y_2 and 81 runs for y_3. Figure 5.4 gives the design constructed by Proposition 5 based on a randomized $(1, 4)$-sequence in base 3. We use the first 9 runs in block 1, all 27 runs in block 1, and all 81 runs in the three blocks of this design for y_1, y_2, and y_3, respectively.

According to Proposition 5, the design in Figure 5.4 has the following properties. First, the part for the categorical factors is nine $\mathrm{OA}(9, 2, 3, 2)$'s $\subset$ three $\mathrm{OA}(27, 2, 3, 3)$'s $\subset$ one $\mathrm{OA}(27, 2, 3, 4)$. Second, the part for the continuous factors is nine $(1, 2, 2)$-nets in base 3 $\subset$ three $(1, 3, 2)$-nets in base 3 $\subset$ one $(1, 4, 2)$-net in base 3. Third, the part for the continuous factors and that for the categorical factors have good balance with respect to one another. Of the first 27 runs in Figure 5.4, consider only those where the first categorical variable is at level 2. These 9 runs are indicated with right arrows in the figure and their continuous part forms three $(1, 1, 2)$-nets in base 3 $\subset$ one $(1, 2, 2)$-net in base 3 as shown in Figure 5.5. Next, of all the runs in Figure 5.4 consider only those whose continuous points lie in the elementary interval $E^* = [0, 1/3) \times [0, 1)$. These 27 runs are indicated with left arrows and their categorical part forms three $\mathrm{OA}(9, 2, 3, 2)$'s $\subset$ one $\mathrm{OA}(9, 2, 3, 3)$ as shown in Figure 5.5.

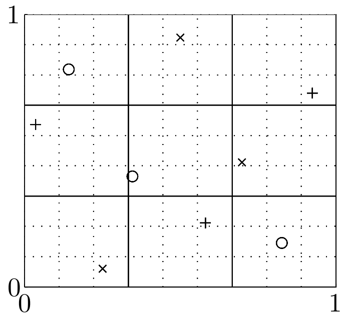
	1	0	0.3238	0.8999	$\leftarrow +$	1	1	0.8999	0.5172		1	2	0.4352	0.0867	
$+ \rightarrow$	2	2	0.9243	0.7107		2	0	0.4078	0.6358		2	1	0.3046	0.2919	$\leftarrow \circ$
	0	1	0.4281	0.8107		0	2	0.3208	0.3688	$\leftarrow \times$	0	0	0.9107	0.1455	
	1	1	0.7307	0.0388		1	2	0.5802	0.9746		1	0	0.1578	0.5252	$\leftarrow \circ$
$+ \rightarrow$	2	0	0.5817	0.2354		2	1	0.1847	0.7454	$\leftarrow \times$	2	2	0.7231	0.5625	
	0	2	0.1631	0.1575	$\leftarrow +$	0	0	0.7072	0.8306		0	1	0.5662	0.4347	
	1	2	0.5446	0.4779		1	0	0.0225	0.0137	$\leftarrow \times$	1	1	0.8601	0.9446	
$+ \rightarrow$	2	1	0.0362	0.5955	$\leftarrow +$	2	2	0.8885	0.3061		2	0	0.5305	0.6790	
	0	0	0.8662	0.3883		0	1	0.5332	0.1960		0	2	0.0106	0.8848	$\leftarrow \circ$
	1	2	0.8058	0.7798		1	0	0.4854	0.3432		1	1	0.0505	0.1222	$\leftarrow \circ$
$\times \rightarrow$	2	1	0.5000	0.9148		2	2	0.0483	0.4965	$\leftarrow \times$	2	0	0.7988	0.0803	
	0	0	0.0667	0.7331	$\leftarrow +$	0	1	0.7822	0.6585		0	2	0.5112	0.2660	
	1	0	0.3763	0.1758		1	1	0.2284	0.8261	$\leftarrow \times$	1	2	0.9387	0.4191	
$\times \rightarrow$	2	2	0.2511	0.0672	$\leftarrow +$	2	0	0.9348	0.9905		2	1	0.3827	0.5519	
	0	1	0.9542	0.2309		0	2	0.4003	0.7690		0	0	0.2349	0.5686	$\leftarrow \circ$
	1	1	0.2188	0.3719	$\leftarrow +$	1	2	0.6891	0.2045		1	0	0.6564	0.8673	
$\times \rightarrow$	2	0	0.6979	0.4569		2	1	0.6362	0.0042		2	2	0.1983	0.9526	$\leftarrow \circ$
	0	2	0.6533	0.6205		0	0	0.1897	0.3301	$\leftarrow \times$	0	1	0.6715	0.6740	
	1	1	0.5965	0.7203		1	2	0.1190	0.6438	$\leftarrow \times$	1	0	0.7473	0.2735	
$\circ \rightarrow$	2	0	0.1422	0.7987	$\leftarrow +$	2	1	0.7716	0.3486		2	2	0.6068	0.1244	
	0	2	0.7537	0.9107		0	0	0.6194	0.4925		0	1	0.1279	0.1035	$\leftarrow \circ$
	1	2	0.0943	0.2496	$\leftarrow +$	1	0	0.8421	0.7567		1	1	0.4716	0.5895	
$\circ \rightarrow$	2	1	0.8267	0.1608		2	2	0.4473	0.8470		2	0	0.1043	0.4199	$\leftarrow \circ$
	0	0	0.4611	0.0558		0	1	0.0827	0.9801	$\leftarrow \times$	0	2	0.8359	0.5310	
	1	0	0.9683	0.6094		1	1	0.3668	0.3112		1	2	0.2827	0.6946	$\leftarrow \circ$
$\circ \rightarrow$	2	2	0.3471	0.4053		2	0	0.2621	0.2121	$\leftarrow \times$	2	1	0.9816	0.8617	
	0	1	0.2941	0.4512	$\leftarrow +$	0	2	0.9934	0.0332		0	0	0.3353	0.9358	

Figure 5.4. The designs for Example 4. The first 27 runs where the first categorical variable is at level 2 are indicated with right arrows and runs where the continuous variables are in the elementary interval $[0, 1/3) \times [0, 1)$ are indicated with left arrows.

6. Summary

By exploiting the underlying structure in (t, s)-sequences, some new NSFDs have been constructed. They can accommodate various types of factors and have better space-filling properties than existing NSFDs. These designs can be used for obtaining observations from multi-fidelity computer experiments to build a flexible prediction model. They can also be used to obtain observations for estimating the means of the outputs of such experiments given a distribution of inputs.

The constructed designs have other applications in statistics. They are useful for sequential experimentation in computer models, calibration and validation of computer models (Kennedy and O'Hagan (2001); Oberkampf and Trucano (2007)), optimization under uncertainty methods (Nemirovski and Shapiro (2006); Ruszczynski and Shapiro (2003), multi-level function estimation (Fasshauer



1	0
0	2
2	1
0	0
2	2
1	1
2	0
1	2
0	1

0	2
2	1
1	0
2	2
1	1
0	0
1	2
0	1
2	0

2	1
1	0
0	2
1	1
0	0
2	2
0	1
2	0
1	2

Figure 5.5. **Left Panel:** The continuous part of the first 27 runs where the first categorical variable is at level 2 forms three $(1,1,2)$-nets in base 3 $\subset$ one $(1,2,2)$-net in base 3. **Right Panel:** The categorical part of the runs where the continuous variables are in the elementary interval $[0,1/3)\times[0,1)$ forms three OA$(9,2,3,2)$'s $\subset$ one OA$(9,2,3,3)$.

(2007)), linking parameters (Husslage et al. (2003)) and sequential evaluations (Husslage, Dam and Hertog (2005)).

Acknowledgements

The authors thanks the Editor, an associate editor, and two referees for their comments that led to improvement in the article. Haaland is supported by Award Number T32HL083806 from the National Heart Lung and Blood Institute. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Heart Lung and Blood Institute or the National Institutes of Health. Qian is supported by NSF Grant DMS-0705206 and a faculty award from IBM.

References

Fasshauer, G. E. (2007). *Meshfree Approximation Methods with MATLAB®*. New Jersey, World Scientific Publishing.

Han, G., Santner, T. J., Notz, W. I. and Bartel, D. L. (2009). Prediction for computer experiments having quantitative and qualitative input variables. *Technometrics* **51**, 278-288.

Husslage, B. G. M., Dam, E. R. V. and Hertog, D. D. (2005). Nested maximin Latin hypercube designs in two dimensions. CentER Discussion Paper 2005-79. Tilburg University, Tilburg, The Netherlands.

Husslage, B., Dam, E. V., Hertog, D. D., Stehouwer, P. and Stinstra, E. (2003). Collaborative metamodeling: coordinating simulation-based product design. *Concurrent Engineering* **11**, 267-278.

Kennedy, M. C. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87**, 1-13.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *J. Roy. Statist. Soc. Ser. B* **63**, 425-464.

Loh, W. L. (2003). On the asymptotic distribution of scrambled net quadrature. *Ann. Statist.* **31**, 1282-1324.

Mullen, G. L., Mahalanabis, A. and Niederreiter, H. (1995). Tables of (t, m, s)-Net and (t, s)-Sequence Parameters. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Lecture Notes in Statistics* **106**, 58-86 Springer-Verlag, New York.

Nemirovski, A. and Shapiro, A. (2006). Convex approximations of chance constrained programs. *SIAM J. Optim.* **17**, 969-996.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods.* Society for Industrial Mathematics, Philadelphia.

Oberkampf, W. L. and Trucano, T. (2007). Verification and validation benchmarks. Sandia National Laboratories (SAND 2007-0853), Albuquerque.

Owen, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. Lecture Notes in Statist.* **106**, 299-317. Springer, New York.

Owen, A. B. (1997). Monte Carlo variance of scrambled net quadrature. *SIAM J. Numer. Anal.* **34**, 1884-1910.

Qian, P. Z. G. (2009). Nested Latin hypercube designs. *Biometrika* **96**, 957-970.

Qian, P. Z. G., Ai, M. and Wu, C. F. J. (2009). Construction of nested space-filling designs. *Ann. Statist.* **37**, 3616-3643.

Qian, P. Z. G., Tang, B. and Wu, C. F. J. (2009). Nested space-filling designs for computer experiments with two levels of accuracy. *Statist. Sinica* **19**, 287-300.

Qian, P. Z. G. and Wu, C. F. J. (2008). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics* **50**, 192-204.

Qian, P. Z. G. and Wu, C. F. J. (2009). Sliced space-filling designs. *Biometrika* **96**, 945-956.

Qian, P. Z. G., Wu, H. and Wu, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **50**, 383-396.

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K. and Wu, C. F. J. (2006). Building surrogate models based on detailed and approximate simulations. *ASME Transactions, J. Mechanical Design* **128**, 668-677.

Ruszczynski, A. and Shapiro, A. (eds) (2003). *Stochastic Programming. Handbooks in Operations Research and Management Science* **10**. Elsevier, Amsterdam.

Santner, T. J., Williams, B. J. and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments.* Springer, New York.

Schmidt, R. R., Cruz, E. E. and Iyengar, M. K. (2005). Challenges of data center thermal management. *IBM J. Research and Development* **49**, 709-723.

Yue, R.-X. (1999). Variance of quadrature over scrambled unions of nets. *Statist. Sinica* **9**, 451-473.

Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA.

E-mail: haaland@stat.wisc.edu

Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA.

E-mail: peterq@stat.wisc.edu

(Received November 2008; accepted April 2009)