RANDOMIZED POLYNOMIAL LATTICE RULES FOR MULTIVARIATE INTEGRATION AND SIMULATION

CHRISTIANE LEMIEUX AND PIERRE L'ECUYER *

Abstract. Lattice rules are among the best methods to estimate integrals in a large number of dimensions. They are part of the *quasi-Monte Carlo* set of tools. A theoretical framework for a class of lattice rules defined in a space of polynomials with coefficients in a finite field is developed in this paper. A randomized version is studied, implementations and criteria for selecting the parameters are discussed, and examples of its use as a variance reduction tool in stochastic simulation are provided. Certain types of digital net constructions, as well as point sets constructed by taking all vectors of successive output values produced by a Tausworthe random number generator, are special cases of this method.

Key words. Numerical integration, lattice rules, variance reduction, quasi-Monte Carlo

AMS subject classifications. 11K45, 65C05, 65D30, 65D32

1. Introduction.

We are concerned with the problem of estimating μ, the integral of a function f over the t-dimensional unit hypercube:

$$\mu = \int_{[0,1)^t} f(\mathbf{u}) d\mathbf{u}. \tag{1.1}$$

The aim of most stochastic simulations is to estimate a mathematical expectation that can be expressed in this form. Indeed, randomness in a Monte Carlo simulation is imitated on a computer by taking a sequence of independent $U(0,1)$ (uniforms over $[0,1)$) "random numbers", which are transformed by some complicated function f to simulate a random variable whose expectation is to be estimated. If f depends on a *random* and unbounded number of uniforms, t can simply be viewed as infinite.

Classical numerical integration methods to approximate μ work fine when t is small and f is smooth [8], but are impractical if t exceeds 4 or 5. The *Monte Carlo* (MC) simulation method estimates μ by the sample average

$$Q_n = \frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{u}_i), \tag{1.2}$$

where $\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}$ are n independent random vectors uniformly distributed over $[0,1)^t$. One has $E[Q_n] = \mu$. Moreover, if

$$\sigma^2 = \int_{[0,1)^t} f^2(\mathbf{u}) d\mathbf{u} - \mu^2 \tag{1.3}$$

*C. Lemieux, Department of Mathematics and Statistics, University of Calgary, 2500 University Drive N.W., Calgary (Alberta), T2N 1N4, Canada. P. L'Ecuyer, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, C.P. 6128, Succ. Centre-Ville, Montréal, H3C 3J7, Canada. This research was supported by NSERC-Canada grant no. OGP0110050 and FCAR-Québec grant no. 00ER3218 to the second author, and by an NSERC-Canada grant to the first author. We owe many thanks to Raymond Couture, who provided several ideas for this paper but declined to be a co-author. Suggestions by Harald Niederreiter and an anonymous referee helped improving the manuscript. The first author wishes to thank Clifton Cunningham and John Watrous for helpful discussions.

is finite then $\mathrm{Var}[Q_n] = \sigma^2/n$, Q_n obeys the central-limit theorem, and the error $E_n = Q_n - \mu$ converges probabilistically as $|E_n| = O_p(\sigma/\sqrt{n})$, regardless of t.

The aim of *Quasi-Monte Carlo* (QMC) methods is to reduce the error by replacing the random points $\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}$ by a set of points $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ that covers the unit hypercube $[0,1)^t$ *more evenly* than typical random points. Two important classes of construction methods are the *digital nets* and the *integration lattices* [20, 32, 35, 38].

A *lattice rule* estimates μ by taking $P_n = L_t \cap [0,1)^t$, where L_t is an *integration lattice* in $\mathbb{R}^t$, i.e., a discrete subset of $\mathbb{R}^t$ closed under addition and subtraction, and that contains $\mathbb{Z}^t$. In this paper, we study a polynomial version of lattice rules which we call *polynomial lattice rules*, obtained by replacing $\mathbb{R}$ and $\mathbb{Z}$ in ordinary lattice rules by the field $\mathbb{L}$ of formal Laurent series over the finite field $\mathbb{F}_2$, and the ring $\mathbb{F}_2[z]$ of polynomials over $\mathbb{F}_2$, respectively. The point set P_n is obtained by defining an appropriate output mapping that associates to each element in $\mathbb{L}$ a number between 0 and 1. These rules turn out to be a special case of digital nets [25].

Our construction generalizes the special kind of digital net introduced by Niederreiter [32, Section 4.4] and Tezuka [42] and studied further in [20, 22, 21], which corresponds in our setting to a *polynomial lattice rule of rank 1*. Also, the focus of our work is different from that of those previous contributions, which consisted in deriving bounds on the rectangular star discrepancy of the point sets P_n, and convergence rates for these bounds when $n \to \infty$ for fixed t. These discrepancy bounds can in turn be used, via the Koksma-Hlawka inequality, to obtain worst-case bounds on $|E_n|$ that converge as $O(n^{-1}(\ln n)^t)$. See, e.g., [32] for the details. For fixed t, this gives a better convergence rate than MC, but the improvement turns out to be practically meaningful only for small t, and other justifications are needed to explain why QMC methods work for real-life applications.

In this paper, we analyze the quality of general polynomial lattice rules by considering instead a randomization of the point set P_n that preserves its uniformity in a certain sense, and we measure the quality of the resulting (unbiased) estimator by its variance. We also use a functional ANOVA decomposition of f, together with some heuristics, to argue that for large t, the quality of P_n should be measured by looking at a selected set of its projections over lower-dimensional subspaces, namely those whose corresponding terms in the ANOVA decomposition capture a large fraction of the variance σ^2. One way of measuring the uniformity of these projections is via the same equidistribution criteria that are used to assess the quality of random number generators based on linear recurrences modulo 2 [23, 43]. Ideally, this selected set of projections should depend on f, but typically f is very complicated and designing P_n specifically for it is not practical.

The paper is organized as follows. In Section 2 we recall some facts about ordinary lattice rules. Polynomial lattice rules and their basic properties are discussed in Section 3. In Section 4 we explain how we use the notion of *equidistribution* to measure the quality of polynomial lattice rules and we make some connections with the so-called (t, m, s)-nets. A randomization for polynomial lattice rules and the variance of the corresponding randomized estimator are studied in Section 5. Specific selection criteria are defined and compared in Section 6. A compromise must be made between choosing an easy-to-compute criterion and one that tests the uniformity from more viewpoints (e.g., by examining a larger number of projections). In Section 7, we discuss implementation issues and give examples of specific parameter choices that are optimal with respect to one of the criteria introduced in Section 6. The effect of additional linear output transformations, such as those proposed in [26, 28, 29], on the structure of the underlying lattice and on the quality of the point sets is studied in Section 8. Section 9 presents simulation examples where the polynomial lattice rules given in the previous sections provide estimators with a smaller empirical variance than those coming from the MC method.

Our development is based on arithmetic in $\mathbb{F}_2$, but it can be generalized easily to $\mathbb{F}_b$ for an arbitrary prime b.

2. Lattice Rules. We now briefly recall some facts about ordinary lattice rules (more can be found in [24, 38]). Analogous results will be developed later for their polynomial version. As mentioned in the introduction, a lattice rule estimates μ by taking $P_n = L_t \cap [0,1)^t$, where the integration lattice L_t is obtained as $L_t = \left\{\mathbf{v} = \sum_{j=1}^t z_j \mathbf{v}_j \text{ such that each } z_j \in \mathbb{Z}\right\}$, where $\mathbf{v}_1, \ldots, \mathbf{v}_t$ are linearly independent vectors in $\mathbb{R}^t$, and $\mathbb{Z}^t \subseteq L_t$. The latter condition is what makes L_t an *integration* lattice. If P_n has n points, then each coordinate of each vector of L_t is a multiple of $1/n$. A simple and convenient way to construct P_n is to take the set of all t-dimensional vectors of successive output values from a linear congruential generator (LCG) [10, 18, 24]; that is, take P_n as the set of all vectors $(u_0, \ldots, u_{t-1})$ where $x_0 \in \mathbb{Z}_n = \{0, \ldots, n-1\}$ and the u_i obey the recurrence

$$x_i = (a x_{i-1}) \bmod n, \qquad u_i = x_i/n,$$

for some positive integer a in $\mathbb{Z}_n$. The corresponding integration rule Q_n given by (1.2) was proposed by Korobov [19] and is called a *Korobov lattice rule.*

The *dual* of an integration lattice L_t is defined by $L_t^* = \{\mathbf{h} \in \mathbb{R}^t \text{ such that } \mathbf{v}\cdot\mathbf{h} \in \mathbb{Z} \text{ for all } \mathbf{v} \in L_t\}$ and is a subset of $\mathbb{Z}^t$. If we write the Fourier expansion of f as

$$f(\mathbf{u}) = \sum_{\mathbf{h}\in\mathbb{Z}^t} \hat{f}(\mathbf{h}) \exp(2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u}), \tag{2.1}$$

with *Fourier coefficients* $\hat{f}(\mathbf{h}) = \int_{[0,1)^t} f(\mathbf{u}) \exp(-2\pi\sqrt{-1}\,\mathbf{h}\cdot\mathbf{u}) d\mathbf{u}$, the integration error with the lattice rule is given explicitly by

$$E_n = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} \hat{f}(\mathbf{h}) \tag{2.2}$$

if f has an absolutely convergent Fourier expansion (2.1) [38]. Unfortunately, estimating the error via (2.2) is impractical because the absolute convergence rarely holds and this expression would be too hard to compute anyway.

An alternative is to randomize the point set P_n so that the integration error can be estimated statistically. One way of doing this is the Cranley-Patterson rotation [7]: Generate one point $\mathbf{U}$ uniformly over $[0,1)^t$ and replace each $\mathbf{u}_i$ in P_n by $\tilde{\mathbf{u}}_i = (\mathbf{u}_i + \mathbf{U}) \bmod 1$ where the reduction modulo 1 is coordinate-wise. The set P_n is thus replaced by $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$, and Q_n and E_n are replaced by the corresponding $\tilde{Q}_n$ and $\tilde{E}_n$. One can show that $E[\tilde{E}_n] = 0$ and

$$\mathrm{Var}[\tilde{E}_n] = \sum_{\mathbf{0}\neq\mathbf{h}\in L_t^*} |\hat{f}(\mathbf{h})|^2, \tag{2.3}$$

as long as f is square-integrable [27]. To estimate the error, compute m i.i.d. copies of $\tilde{Q}_n$ with the same P_n, using m independent uniform shifts $\mathbf{U}$, and compute their sample variance, which is an unbiased estimator of $\mathrm{Var}[\tilde{Q}_n] = E[\tilde{E}_n^2]$.

The variance expression (2.3) suggests discrepancy measures of the form

$$D(P_n) = \sum_{\mathbf{0} \neq \mathbf{h} \in L_t^*} w(\mathbf{h}) \quad \text{or} \quad D(P_n) = \sup_{\mathbf{0} \neq \mathbf{h} \in L_t^*} w(\mathbf{h}), \tag{2.4}$$

with weights $w(\mathbf{h})$ that decrease with $\|\mathbf{h}\|$ in a way that corresponds to how we think the squared Fourier coefficients $|\hat{f}(\mathbf{h})|^2$ decrease with $\|\mathbf{h}\|$, and where $\|\cdot\|$ is an arbitrary norm (see, e.g., [10, 14, 16, 27]). For a given choice of weights $w(\mathbf{h})$, either definition of $D(P_n)$ in (2.4) can be used as a selection criterion (to be minimized) over a given set of n-point lattice rules. Most selection criteria in the literature are of the form (2.4). Examples are $\mathcal{P}_\alpha$ and the Babenko-Zaremba index ρ (see, e.g., [10, 38]), as well as the $\tilde{\mathcal{P}}_\alpha$ defined in [16] and the criterion $M_{t_1,\ldots,t_d}$ proposed in [24].

3. Polynomial Lattice Rules. In this section, we define and study polynomial lattice rules, and give examples for which an easy implementation is available. We also discuss the functional ANOVA decomposition and the Walsh series expansion, which turns out to be the counterpart of the Fourier series expansion for polynomial lattice rules. We conclude by studying the projections of polynomial lattice rules in light of these functional decompositions.

3.1. Definition and basic properties. DEFINITION 3.1. A *polynomial lattice rule* estimates μ by taking a *polynomial lattice point set* $P_n = \varphi(\mathcal{L}_t) \cap [0,1)^t$, where $\mathcal{L}_t$ is a *polynomial integration lattice*, i.e., a set of the form

$$\mathcal{L}_t = \left\{ \mathbf{v}(z) = \sum_{j=1}^{t} q_j(z)\mathbf{v}_j(z) \;\; \text{such that } q_j(z) \in \mathbb{F}_2[z] \text{ for each } j \right\},$$

where $\mathbf{v}_1(z), \ldots, \mathbf{v}_t(z)$ are arbitrary vectors in $\mathbb{L}^t$, independent over $\mathbb{L}$, and such that the set $(\mathbb{F}_2[z])^t$ of all t-dimensional vectors of polynomials is contained in $\mathcal{L}_t$. The map $\varphi : \mathbb{L} \to \mathbb{R}$ is defined by

$$\varphi\left(\sum_{l=\omega}^{\infty} d_l z^{-l} \right) = \sum_{l=\omega}^{\infty} d_l 2^{-l}, \tag{3.1}$$

For a vector $\mathbf{v}(z) = (v_1(z), \ldots, v_t(z))$, we define $\varphi(\mathbf{v}(z)) = (\varphi(v_1(z)), \ldots, \varphi(v_t(z)))$.

DEFINITION 3.2. The *dual lattice* of $\mathcal{L}_t$ is defined as

$$\mathcal{L}_t^* = \{\mathbf{h}(z) \in \mathbb{L}^t \text{ such that } \mathbf{h}(z) \cdot \mathbf{v}(z) \in \mathbb{F}_2[z] \text{ for each } \mathbf{v}(z) \in \mathcal{L}_t\},$$

where $\mathbf{h}(z) \cdot \mathbf{v}(z) = \sum_{j=1}^t h_j(z) v_j(z)$. It is a subset of $(\mathbb{F}_2[z])^t$.

This dual lattice plays a role in providing error and variance expressions, as we explain in Subsection 3.3 and in Section 5. The *dual* of the basis $\{\mathbf{v}_1(z), \ldots, \mathbf{v}_t(z)\}$ is the set of vectors $\{\mathbf{h}_1(z), \ldots, \mathbf{h}_t(z)\}$ in $(\mathbb{F}_2[z])^t$ such that

$$\mathbf{h}_i(z) \cdot \mathbf{v}_j(z) = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$

It forms a basis of the dual lattice. If $\mathbf{V}$ is the matrix with rows $\mathbf{v}_1(z), \ldots, \mathbf{v}_t(z)$, then $\mathbf{h}_1(z), \ldots, \mathbf{h}_t(z)$ are the columns of $\mathbf{V}^{-1}$, the inverse of $\mathbf{V}$.

DEFINITION 3.3. The *determinant* of $\mathcal{L}_t$ is $\det(\mathcal{L}_t) = \det(\mathbf{V})$. Likewise, $\det(\mathcal{L}_t^*) = \det(\mathbf{V}^{-1}) = 1/\det(\mathcal{L}_t)$.

Lemmas A.1 and A.3 (in the appendix) state that these determinants do not depend on the choice of basis and that $\det(\mathcal{L}_t^*)$ is a polynomial, which we denote by $P(z)$. Also, Lemma A.5 implies that if $P(z)$ has degree k, then P_n has exactly $n = 2^k$ distinct elements. We call this number n the *density* of $\mathcal{L}_t$. This is analogous to the case of ordinary lattice rules, where $P_n = L_t \cap [0,1)^t$ has cardinality $1/|\det(\mathbf{V})|$ for any matrix $\mathbf{V}$ whose rows form a basis of L_t.

DEFINITION 3.4. The *rank* r of a polynomial lattice rule based on the lattice $\mathcal{L}_t$, denoted $\operatorname{rank}(\mathcal{L}_t)$, is the minimal value r in $\{1,\ldots,t\}$ such that a basis $\mathbf{v}_1(z),\ldots,\mathbf{v}_t(z)$ for $\mathcal{L}_t$ can be chosen with $\mathbf{v}_j(z) = \mathbf{e}_j$ for $j > r$, where $\mathbf{e}_j$ is the jth unit vector in t dimensions (a vector with a 1 in position j and zeros elsewhere). A basis is in *minimal form* if $\mathbf{v}_j(z) = \mathbf{e}_j$ for $j > \operatorname{rank}(\mathcal{L}_t)$.

With a slight abuse of notation, we denote the commutative ring of polynomials of degree less than k with coefficients in $\mathbb{F}_2$ by $\mathbb{F}_2[z]/(P)$, where the two basic operations are the addition and multiplication of polynomials modulo $P(z)$. For ordinary lattice rules, one can always find a basis whose vectors have coordinates of the form a/n, for some positive integer n and with $a \in \mathbb{Z}_n$ or $a = n$. The following proposition states an analogous result for polynomial integration lattices. Its proof is in the appendix.

PROPOSITION 3.5. *Any polynomial integration lattice $\mathcal{L}_t$ admits a basis $\mathbf{v}_1(z),\ldots,$ $\mathbf{v}_t(z)$ whose vectors have coordinates of the form $p(z)/P(z)$, where $p(z) \in \mathbb{F}_2[z]/(P)$ or $p(z) = P(z)$.*

If we let $\Xi_t = \mathcal{L}_t \cap \mathbb{L}_0^t$, where $\mathbb{L}_0 = \{s(z) \in \mathbb{L} : s(z) = \sum_{l=1}^{\infty} d_l z^{-l}\}$, then this result implies that the coordinates of each point of Ξ_t are of the form $p(z)/P(z)$ where $p(z) \in \mathbb{F}_2[z]/(P)$, so that $P(z)\Xi_t \subseteq (\mathbb{F}_2[z])^t$. From the next proposition, this implies in turn that the coefficients of these coordinates follow a linear recurrence in $\mathbb{F}_2$ with characteristic polynomial $P(z)$. Obviously, the successive bits of each coordinate of any point in $P_n = \varphi(\Xi_t)$ then follow the same recurrence. Note that $P(z)$ is not necessarily the *minimal* polynomial of this linear recurrence, so it is possible that the successive bits also follow a linear recurrence of order strictly less than k. Example 3.7 will illustrate such a case.

PROPOSITION 3.6. *If $v(z) = p(z)/P(z) = \sum_{j=1}^{\infty} x_j z^{-j}$ where $P(z) = \sum_{l=0}^{k} a_l z^{k-l}$ and $p(z) \in \mathbb{F}_2[z]/(P)$, then the sequence $\{x_j,\ j \geq 1\}$ follows the linear recurrence*

$$x_j = a_1 x_{j-1} + \cdots + a_k x_{j-k} \tag{3.2}$$

in $\mathbb{F}_2$, for which $P(z)$ is a characteristic polynomial. Moreover, assuming that $p(z) = \sum_{j=1}^{k} c_j z^{k-j}$, we have the one-to-one correspondence

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ a_1 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{k-1} & \ldots & a_1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}. \tag{3.3}$$

Proof. We have

$$p(z) = P(z)v(z) \ = \ \left(\sum_{\ell=0}^{k} a_\ell z^{k-\ell}\right)\left(\sum_{j=1}^{\infty} x_j z^{-j}\right).$$

If we multiply the latter two sums, regroup the terms, and equal the coefficient of each power of z with the corresponding coefficient in $p(z)$, we obtain that for $j > k$, $a_0 x_j + a_1 x_{j-1} + \cdots + a_k x_{j-k} = 0$, whereas for $1 \leq j \leq k$, $a_0 x_j + a_1 x_{j-1} + \cdots + a_{j-1} x_1 =$

c_j, where $a_0 = 1$. This gives (3.2) and (3.3), respectively. The correspondence (3.3) is clearly one-to-one. □

EXAMPLE 3.7. (Rectangular rule) Consider the lattice generated by the basis vectors $\mathbf{v}_j(z) = \mathbf{e}_j/Q(z)$, $1 \le j \le t$, for some polynomial $Q(z)$ of degree q. The corresponding lattice rule has rank t, $P(z) = \det(\mathcal{L}_t^*) = 1/\det(\mathbf{V}) = (Q(z))^t$, and order $n = 2^k = 2^{qt}$. For example, if $Q(z) = z$, then $k = t$, $P(z) = z^k$ and, if $t > 1$, all the coefficients $a_1, \ldots, a_k$ in the recurrence (3.2) are 0, in accordance with the fact that $p(z)/P(z)$ has no term z^{-j} with $j > k$ in this case. If $Q(z) = z^2 + z$ and $t > 1$, then $k = 2t$ and $P(z)$ has a coefficient a_k equal to 0, which means that the minimal polynomial of the recurrence (3.2) has a degree smaller than k.

EXAMPLE 3.8. (Rank-1 polynomial lattice rule) Let $P(z)$ be in $\mathbb{F}_2[z]$ and $\mathbf{g}(z)$ in $(\mathbb{F}_2[z]/(P))^t$ with no zero component. Let $\mathcal{L}_t = \{q(z)\mathbf{g}(z)/P(z), q(z) \in \mathbb{F}_2[z]\}$, and $P_n = \varphi(\mathcal{L}_t) \cap [0,1)^t$. This defines a polynomial lattice rule of rank 1. This construction was introduced in [42] and [32, Section 4.4], where it is presented as a special case of digital net.

EXAMPLE 3.9. (Polynomial LCG [41, 43, 44]) Consider the linear recurrence

$$p_i(z) = a(z)p_{i-1}(z) \bmod P(z), \tag{3.4}$$

where $P(z)$ is an arbitrary polynomial of degree k over $\mathbb{F}_2$, and $a(z) \neq 0$ in $\mathbb{F}_2[z]/(P)$. Dividing (3.4) by $P(z)$ yields

$$s_i(z) = a(z)s_{i-1}(z) \bmod \mathbb{F}_2[z], \tag{3.5}$$

where

$$s_i(z) = p_i(z)/P(z) = \sum_{j=1}^{\infty} d_{i,j} z^{-j} \tag{3.6}$$

is in $\mathbb{L}_0$ and the operator "mod $\mathbb{F}_2[z]$" discards the non-negative powers of z. In analogy with an ordinary LCG, (3.4)–(3.6) define a *polynomial LCG*, whose output at step i is the quotient of the state $p_i(z)$ by the modulus $P(z)$. Let

$$\Xi_t = \{(s_0(z), s_1(z), \ldots, s_{t-1}(z)), \text{ such that } p_0(z) \in \mathbb{F}_2[z]/(P)\}, \tag{3.7}$$

the set of all vectors of t successive (formal series) outputs of the polynomial LCG, from all initial states $p_0(z)$. Define the point set $P_n = \varphi(\Xi_t)$. Analogously to the LCG situation [18], we have the following proposition, stated without proof in [42], and which means that the corresponding rule can be interpreted as a *Korobov polynomial lattice rule*.

PROPOSITION 3.10. *The set Ξ_t defined in (3.7) satisfies $\Xi_t = \mathcal{L}_t \cap \mathbb{L}_0^t$, where $\mathcal{L}_t$ is the t-dimensional polynomial integration lattice with basis $\mathbf{v}_1(z) = (1, a(z), a^2(z) \bmod P(z), \ldots, a^{(t-1)}(z) \bmod P(z))/P(z)$, $\mathbf{v}_2(z) = \mathbf{e}_2$, ..., $\mathbf{v}_t(z) = \mathbf{e}_t$.*

Proof. If $\mathbf{v}(z) \in \Xi_t$, one can write $\mathbf{v}(z) = \sum_{j=1}^t q_j(z)\mathbf{v}_j(z)$ where $q_1(z) = p_0(z)$ and for $j > 1$, $q_j(z)$ is the polynomial part of $p_0(z)(a^{(j-1)}(z) \bmod P(z))/P(z)$, i.e., the polynomial such that subtracting $q_j(z)\mathbf{v}_j(z)$ from $p_0(z)\mathbf{v}_1(z)/P(z)$ reduces the jth coordinate of this vector modulo $\mathbb{F}_2[z]$. Conversely, any linear combination of $\mathbf{v}_1(z), \ldots, \mathbf{v}_t(z)$ over $\mathbb{F}_2[z]$, whose coordinates are reduced modulo $\mathbb{F}_2[z]$, clearly belongs to Ξ_t. To see that we have an integration lattice, just note that every vector $\mathbf{v}(z) = (v_1(z), \ldots, v_t(z)) \in (\mathbb{F}_2[z])^t$ can be written as $v_1(z)P(z)\mathbf{v}_1(z) + (v_2(z) - v_1(z)a(z))\mathbf{v}_2(z) + \cdots + (v_t(z) - v_1(z)a^{(t-1)}(z) \bmod P(z))\mathbf{v}_t(z)$. □

Thus, a lattice rule based on a polynomial LCG has $\mathrm{rank}(\mathcal{L}_t) = 1$, $\det(\mathcal{L}_t^*) = P(z)$, and $n = 2^k$ points.

EXAMPLE 3.11. (LFSR generator) As a special case of the polynomial LCG, suppose that the multiplier $a(z)$ can be written as $z^\nu \bmod P(z)$ for some positive integer ν, which we call the *step size*. When $P(z)$ is a primitive polynomial over $\mathbb{F}_2$, every polynomial nonzero $a(z)$ in $\mathbb{F}_2[z]/(P)$ can be written in this way. In this case, from (3.5), one has $d_{i,j} = d_{i-1,j+\nu}$ in (3.6), i.e., computing the right-hand side of (3.5) amounts to shifting the coefficients of $s_{i-1}(z)$ by ν positions and dropping the non-negative powers of z. If we define $x_j = d_{0,j}$, then $d_{i,j} = x_{i\nu+j}$ for all i, and therefore

$$u_i = \sum_{j=1}^{\infty} x_{i\nu+j} 2^{-j}$$

is the output at step i, where the x_j's obey (3.2). This defines a Tausworthe-type LFSR random number generator [23, 43]. In practice, the output u_i is truncated to its first w bits, for some positive integer w. In this case, P_n is the set of all vectors of t successive output values produced by the Tausworthe generator, over all of its cycles (including the trivial cycle that contains only 0).

EXAMPLE 3.12. *Combined Tausworthe generators* can be defined as follows [23, 43, 44]. Take m Tausworthe generators, the lth one based on a linear recurrence with characteristic polynomial $P_l(z)$ of degree k_l, step size ν_l, and output sequence $u_{l,0}, u_{l,1}, u_{l,2}, \cdots$. Each of these generators can be viewed as a polynomial LCG and defines a polynomial integration lattice $\mathcal{L}_t^l$, as explained in Example 3.9. Define $u_i = u_{1,i} \oplus \cdots \oplus u_{m,i}$, for $i \geq 0$, where $u \oplus v$ performs a bitwise exclusive-or on the binary expansions of u and v. If the $P_l(z)$'s are pairwise relatively prime, the sequence $\{u_i,\ i \geq 0\}$ produced by the combined generator turns out to be the output sequence of a Tausworthe generator with characteristic polynomial $P(z) = P_1(z) \cdots P_m(z)$ [44]. Hence its lattice structure can be analyzed as in Example 3.9. An alternative and more general approach to study point sets P_n obtained this way is to consider sums of polynomial lattice rules, as discussed in the next section.

REMARK 3.13. (Linear output transformations) The output mapping φ from $\mathbb{L}$ to $\mathbb{R}$ can be replaced by a more general linear transformation as follows: for $v(z) = \sum_{j=1}^{\infty} x_j z^{-j}$ in $\mathbb{L}_0$, replace $u_i = \varphi(v(z))$ by

$$u_i = \sum_{j=1}^{\infty} y_{i,j} 2^{-j}, \tag{3.8}$$

where

$$y_{i,j} = \sum_{\ell=1}^{k} b_{\ell,j} x_\ell, \tag{3.9}$$

and where the $b_{\ell,j}$'s are constants in $\mathbb{F}_2$. This additional linear transformation can be applied to improve the uniformity of the point set P_n, especially when important restrictions are imposed on the parameters to make the implementation more efficient [28, 29, 36]. Although the corresponding point set P_n is still a digital net, it does not generally yield a polynomial lattice rule, as discussed in Section 8.

REMARK 3.14. (From R. Couture.) For the polynomial LCG, one could also define the output u_i directly in terms of the coefficients c_j of the polynomial $p_i(z)$, e.g., replace d_l in (3.1) by c_l or by c_{k-l}, assuming that the sum is truncated to at most k terms. This can be convenient from the implementation viewpoint, as explained in [26, 36], and would also lead to a polynomial lattice structure, although different from the one analyzed here. The linear relationship (3.3) shows that this can be analyzed as a special case of the linear transformation discussed in Remark 3.13. In this paper, we do not use this formulation and work with (3.1) instead.

3.2. Sums of polynomial lattice rules. When restrictions on the choice of parameters are imposed for implementation purposes, alternative constructions can be obtained by choosing a few polynomial lattice rules that are easily implemented, and combine them via a *sum*. That is, given m rules with respective point sets $P_{n_1}, \ldots, P_{n_m}$, the *sum rule* uses the point set $P_n = P_{n_1} \oplus \ldots \oplus P_{n_m}$, where $P \oplus Q = \{\mathbf{u}_i \oplus \mathbf{u}_j : \mathbf{u}_i \in P \text{ and } \mathbf{u}_j \in Q\}$, and $\mathbf{u}_i \oplus \mathbf{u}_j$ denotes the bitwise exclusive-or of the binary expansions of the coordinates of $\mathbf{u}_i$ and $\mathbf{u}_j$. This idea has been used successfully to implement good LFSR generators [23, 44, 46], as discussed in Example 3.12. Those LFSR generators can in turn be used to construct point sets for polynomial lattice rules. More details on this are given in Section 7. In terms of the polynomial lattices $\mathcal{L}_t^1, \ldots, \mathcal{L}_t^m$ on which the m rules are based, performing a sum of lattice rules is equivalent to using a rule based on the lattice $\mathcal{L}_t = \mathcal{L}_t^1 \oplus \cdots \oplus \mathcal{L}_t^m$, the sum of $\mathcal{L}_t^1, \ldots, \mathcal{L}_t^m$.

The next proposition, proved in the appendix (see also [38, Theorem 3.25]), gives a necessary and sufficient condition for the number of points in the rule resulting from the sum to equal the product of the numbers of points in the component rules. In that case, we can say that the (finite) point set P_n is a *direct sum* of $P_{n_1}, \ldots, P_{n_m}$, in the sense that each point of P_n can be expressed *uniquely* as a sum (mod 1) involving one point from each component P_{n_j}. On the other hand, the corresponding sum of integration lattices is not a direct sum, because those lattices have a non-empty intersection (they all contain the polynomials).

PROPOSITION 3.15. *Let P_{n_l} be the point set of a polynomial lattice rule containing $n_l = 2^{k_l}$ distinct points, for $l = 1, \ldots, m$, for some integer $m \geq 2$. Then $P_{n_1} \oplus \cdots \oplus P_{n_m}$ contains $n_1 \cdots n_m$ distinct points if and only if for each $l = 1, \ldots, m$, the intersection of P_{n_l} with the sum of the remaining $m-1$ point sets is the point $\mathbf{0}$.*

For rank-1 rules, we have the following result, also proved in the appendix:

PROPOSITION 3.16. *Let $\mathcal{L}_t^1, \ldots, \mathcal{L}_t^m$ be lattices defining polynomial lattices rules of rank 1 in t dimensions, where for each l, $\mathcal{L}_t^l$ is generated by a basis formed by the vector $\mathbf{v}_1^l(z) = (1, v_2^l(z), \ldots, v_t^l(z))/P_l(z)$ together with the unit vectors $\mathbf{e}_2, \ldots, \mathbf{e}_t$, where $P_l(z)$ is a polynomial of degree k_l, and each $v_j^l(z)$ is in $\mathbb{F}_2[z]/(P_l)$. Consider the sum $\mathcal{L}_t = \mathcal{L}_t^1 \oplus \cdots \oplus \mathcal{L}_t^m$.*

If the polynomials $P_1(z), \ldots, P_m(z)$ are pairwise relatively prime, then $\mathcal{L}_t$ is a lattice of rank 1 that admits a basis formed by a vector $\mathbf{v}_1(z) = (1, v_2(z), \ldots, v_t(z))/P(z)$ together with the unit vectors $\mathbf{e}_2, \ldots, \mathbf{e}_t$, where $P(z) = P_1(z) \cdots P_m(z)$ is a polynomial of degree $k = k_1 + \cdots + k_m$, and each $v_j(z)$ is in $\mathbb{F}_2[z]/(P)$. The point set that corresponds to $\mathcal{L}_t$ thus has $n = 2^k$ distinct points.

REMARK 3.17. If the polynomials $P_1(z), \ldots, P_m(z)$ are not pairwise relatively prime, then $\mathcal{L}_t$ is not a lattice of rank 1, and its associated point set is not *fully projection-regular*; this property is defined in Subsection 3.4.

In Section 8, we discuss other alternative constructions that use linear transformations such as those discussed in Remark 3.13, which can also be used to improve the quality of P_n.

3.3. Walsh expansion and error expression. For any multivariate polynomial $\mathbf{h} \equiv \mathbf{h}(z) = (h_1, \dots, h_t) \in (\mathbb{F}_2[z])^t$ where $h_s \equiv h_s(z) = \sum_{j=0}^{\ell_s - 1} h_{s,j} z^j$ for some ℓ_s, and for $\mathbf{u} = (u_1, \dots, u_t)$ where $u_s = \sum_{j \geq 1} u_{s,j} 2^{-j} \in [0, 1)$ and $u_{s,j} \neq 1$ for infinitely many j, define

$$\langle \mathbf{h}, \mathbf{u} \rangle = \sum_{s=1}^{t} \sum_{j=1}^{\ell_s} h_{s,j-1} u_{s,j} \bmod 2.$$

The *Walsh expansion in base 2* of a function $f : [0,1)^t \to \mathbb{R}$ is defined as (e.g., [2]):

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in (\mathbb{F}_2[z])^t} \tilde{f}(\mathbf{h})(-1)^{\langle \mathbf{h}, \mathbf{u} \rangle}, \tag{3.10}$$

with Walsh coefficients

$$\tilde{f}(\mathbf{h}) = \int_{[0,1)^t} f(\mathbf{u})(-1)^{\langle \mathbf{h}, \mathbf{u} \rangle} d\mathbf{u}. \tag{3.11}$$

Each term in (3.10) represents a piecewise-constant periodic function of $\mathbf{u}$ with amplitude $\tilde{f}(\mathbf{h})$. Each vector $\mathbf{h}$ is a *bit selector.* It picks a finite number of bits from the binary expansion of $(u_1, \dots, u_t)$: The jth bit of u_s is selected if and only if $h_{s,j-1} = 1$.

DEFINITION 3.18. (see e.g., [6]) We define the norm $\|\mathbf{h}\|_\infty = \max_{1 \leq s \leq t} |h_s|_{\mathrm{p}}$, where $|h_s|_{\mathrm{p}} = 2^m$ if $h_s(z)$ has degree $m \geq 0$, and $|h_s|_{\mathrm{p}} = 0$ if $h_s = 0$.

Intuitively, the $\mathbf{h}$'s whose sup norm $\|\mathbf{h}\|_\infty$ is small are more important because they test only the most significant bits of $\mathbf{u}$.

The next lemma gives a property of the Walsh coefficients of the functions f_I obtained from the *ANOVA decomposition of* f. See [9, 35] for details about that decomposition, which writes a square-integrable function f as a sum of orthogonal functions,

$$f(\mathbf{u}) = \sum_{I \subseteq \{1,\dots,t\}} f_I(\mathbf{u}),$$

where $f_I(\mathbf{u}) = f_I(u_1, \dots, u_t)$ depends only on $\{u_i,\, i \in I\}$, $f_\phi(\mathbf{u}) \equiv \mu$ for the empty set ϕ,

$$\int_{[0,1)^t} f_I(\mathbf{u}) d\mathbf{u} = 0 \quad \text{for } I \neq \phi,$$

$$\int_{[0,1)^t} f_I(\mathbf{u}) f_J(\mathbf{u}) d\mathbf{u} = 0 \quad \text{for } I \neq J,$$

The variance σ^2 decomposes as

$$\sigma^2 = \sum_{I \subseteq \{1,\dots,t\}} \sigma_I^2 = \sum_{\phi \neq I \subseteq \{1,\dots,t\}} \int_{[0,1)^t} f_I^2(\mathbf{u}) d\mathbf{u}.$$

The f_I's are defined recursively in an explicit way described in [35]. It is used in Section 6 to define selection criteria for polynomial lattice rules.

LEMMA 3.19. *Let* $f : [0,1)^t \to \mathbb{R}$ *be a square-integrable function, and for* $\mathbf{h} \in (\mathbb{F}_2[z])^t$*, let* $I_{\mathbf{h}}$ *denote the set of indices* j *such that* $h_j \neq 0$*. Then, for each non-empty subset* I *of* $\{1, \dots, t\}$*, the Walsh coefficients of* f_I *are given by*

$$\tilde{f}_I(\mathbf{h}) = \begin{cases} \tilde{f}(\mathbf{h}) & \textit{if } I = I_{\mathbf{h}}, \\ 0 & \textit{otherwise .} \end{cases}$$

Proof. Denote by I^c the complement of I in $\{1,\dots,t\}$. We have that

$$\begin{aligned}\tilde{f}_I(\mathbf{h}) &= \int_{[0,1)^t} f_I(\mathbf{u})(-1)^{\langle \mathbf{h},\mathbf{u}\rangle} d\mathbf{u} = \int_{[0,1)^{|I|}} f_I(\mathbf{u}_I)(-1)^{\langle \mathbf{h}_I,\mathbf{u}_I\rangle} d\mathbf{u}_I \int_{[0,1)^{|I^c|}} (-1)^{\langle \mathbf{h}_{I^c},\mathbf{u}_{I^c}\rangle} d\mathbf{u}_{I^c} \\ &= \begin{cases} 0 & \text{if } h_j \neq 0 \text{ for at least one } j \in I^c, \\ \int_{[0,1)^{|I|}} f_I(\mathbf{u}_I)(-1)^{\langle \mathbf{h}_I,\mathbf{u}_I\rangle} d\mathbf{u}_I & \text{otherwise,}\end{cases}\end{aligned}$$

since $\int_{[0,1)^{|I^c|}} (-1)^{\langle \mathbf{h}_{I^c},\mathbf{u}_{I^c}\rangle} d\mathbf{u}_{I^c}$ equals 0 if $\mathbf{h}_{I^c} \neq \mathbf{0}$ and 1 otherwise. So assume that $h_j = 0$ for each $j \in I^c$ and that $h_j = 0$ for at least one $j \in I$. Let $I_0 = \{j \in I : h_j = 0\}$ and $I_0^c = I \backslash I_0$. Then $I_0 \neq \phi$ and $\int_{[0,1)^{|I_0|}} f_I(\mathbf{u}_I) d\mathbf{u}_{I_0} = 0$ by definition, so

$$\tilde{f}_I(\mathbf{h}) = \int_{[0,1)^{|I_0^c|}} (-1)^{\langle \mathbf{h}_{I_0^c},\mathbf{u}_{I_0^c}\rangle} \int_{[0,1)^{|I_0|}} f_I(\mathbf{u}_I) d\mathbf{u}_{I_0} d\mathbf{u}_{I_0^c} = 0.$$

The last thing we need to show is that if $I = I_{\mathbf{h}}$, then $\tilde{f}_I(\mathbf{h}) = \tilde{f}(\mathbf{h})$. This follows since $\tilde{f}(\mathbf{h}) = \int_{[0,1)^t} \sum_J f_J(\mathbf{x})(-1)^{\langle \mathbf{h},\mathbf{u}\rangle} d\mathbf{u} = \sum_J \tilde{f}_J(\mathbf{h}) = \tilde{f}_{I_{\mathbf{h}}}(\mathbf{h})$, the third equality coming from the fact that for all $J \neq I_{\mathbf{h}}$, we proved that $\tilde{f}_J(\mathbf{h}) = 0$. □

This lemma implies that we can write the Walsh expansion of f_I as

$$f_I(\mathbf{u}) = \sum_{\mathbf{h} \in \mathbb{F}_2[z]_I} \tilde{f}(\mathbf{h})(-1)^{\langle \mathbf{h},\mathbf{u}\rangle}, \tag{3.12}$$

where $\mathbb{F}_2[z]_I = \{\mathbf{h} \in (\mathbb{F}_2[z])^t : h_j \neq 0 \text{ if and only if } j \in I\}$. Using (3.12), we can also rewrite the variance of f_I as

$$\sigma_I^2 = \sum_{\mathbf{h} \in \mathbb{F}_2[z]_I} |\tilde{f}(\mathbf{h})|^2 = \sum_{\mathbf{h} \in (\mathbb{F}_2[z])^t} |\tilde{f}_I(\mathbf{h})|^2. \tag{3.13}$$

This decomposes the variance of f as

$$\sigma^2 = \sum_{\phi \neq I \subseteq \{1,\dots,t\}} \sum_{\mathbf{h} \in \mathbb{F}_2[z]_I} |\tilde{f}(\mathbf{h})|^2. \tag{3.14}$$

The first equality in (3.13) provides an expression for σ_I^2 that does not require explicit knowledge of the ANOVA component f_I. Also, the expression (3.14) determines a partition of the coefficients $|\tilde{f}(\mathbf{h})|^2$ that will be useful in Section 6 to define selection criteria that rely both on the ANOVA and Walsh decompositions of f.

We now examine the interplay between Walsh expansions and polynomial lattice rules. The following lemma, proved in a more general setting in [25], is the analogue of the result given in [38, *Lemma 2.7*] for ordinary lattice rules. It is used to prove Propositions 3.21 and 5.1.

LEMMA 3.20. *If P_n is a polynomial lattice point set, then*

$$\sum_{\mathbf{u} \in P_n} (-1)^{\langle \mathbf{h},\mathbf{u}\rangle} = \begin{cases} n & \textit{if } \mathbf{h} \in \mathcal{L}_t^*, \\ 0 & \textit{otherwise.}\end{cases}$$

Using this lemma, an expression similar to (2.2) for the integration error associated with a polynomial lattice rule is easily obtained for functions that have an absolutely convergent Walsh series. Note that the latter is a very strong assumption.

In [22, Lemma 1], an expression is given for the error in terms of the Walsh coefficients $\tilde{f}(\mathbf{h})$ for a general point set P_n (not necessarily a lattice). It is shown there that

$$E_n = \frac{1}{n}\sum_{\mathbf{u}\in P_n} f(\mathbf{u}) - \mu = \frac{1}{n}\sum_{\mathbf{0}\neq\mathbf{h}\in(\mathbb{F}_2[z])^t} \tilde{f}(\mathbf{h}) \sum_{\mathbf{u}\in P_n} (-1)^{\langle \mathbf{h},\mathbf{u}\rangle}. \tag{3.15}$$

By combining this result with Lemma 3.20, we obtain the next proposition, for which we provide a complete proof.

PROPOSITION 3.21. *If f is such that $\sum_{\mathbf{h}\in(\mathbb{F}_2[z])^t} |\tilde{f}(\mathbf{h})| < \infty$ and P_n is a polynomial lattice point set, then*

$$E_n = Q_n - \mu = \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} \tilde{f}(\mathbf{h}).$$

Proof. We have that

$$\begin{aligned}\frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{u}_i) - \mu &= \frac{1}{n}\sum_{i=0}^{n-1}\sum_{\mathbf{h}\in(\mathbb{F}_2[z])^t} \tilde{f}(\mathbf{h})(-1)^{\langle \mathbf{h},\mathbf{u}_i\rangle} - \tilde{f}(\mathbf{0}) \\ &= \frac{1}{n}\sum_{\mathbf{h}\in(\mathbb{F}_2[z])^t} \tilde{f}(\mathbf{h})\sum_{i=0}^{n-1}(-1)^{\langle \mathbf{h},\mathbf{u}_i\rangle} - \tilde{f}(\mathbf{0}) \\ &= \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} \tilde{f}(\mathbf{h}),\end{aligned}$$

where the first equality is obtained by using the Walsh expansion of f and the fact that $\tilde{f}(\mathbf{0}) = \mu$; the second equality is obtained by changing the order of summation, which is allowed by Fubini's theorem [37] since $\sum_{\mathbf{h}\in(\mathbb{F}_2[z])^t} |\tilde{f}(\mathbf{h})| < \infty$; the third equality follows from Lemma 3.20. □

Error bounds for functions having sufficiently fast decaying Walsh coefficients are given in [21, 22] for different types of digital nets. This is in analogy with existing results for ordinary lattice rules that can be found in [38], for example. We do not explore this topic here, as we are rather interested in studying randomizations of P_n and their corresponding variance expressions. This is the subject of Section 5.

3.4. Projections of P_n over subsets of coordinates. The functional ANOVA decomposition described previously is very useful to understand what are the important features of a function, and in turn, this helps determining what properties of P_n should be considered more carefully. In particular, this decomposition can be used to define the *effective dimension* of a function [3, 35]. For instance, a function f is said to have an effective dimension of d in proportion ρ *in the superposition sense* if $\sum_{I:|I|\leq d}\sigma_I^2 \geq \rho\sigma^2$. When ρ is close to 1, this means that f is well approximated by a sum of d-dimensional (or less) functions. For example, linear, quadratic, and cubic t-dimensional functions have effective dimensions 1, 2, and 3 in proportion 1 in the superposition sense, respectively.

Real-life simulations often involve high-dimensional functions with low effective dimension, in some sense, in proportion ρ close to 1. *Smoothing* techniques can also be used to change f in order to reduce the effective dimension, without changing μ [11, 30, 35, 40]. Often, $\sum_{I\in\mathcal{I}}\sigma_I^2$ is large if $\mathcal{I}$ contains all the sets I formed either by successive indices or by a small number of indices that are not too far apart. What counts then is that for each of these important sets I, the projection $P_n(I)$ of the point set P_n over the subspace determined by I be well distributed. When $|I|$ is small, it becomes possible to cover the $|I|$-dimensional subspace quite well with the points of $P_n(I)$. For the sets I for which σ_I^2/σ^2 is very small, there is no need to care much about the quality of $P_n(I)$. Hence we need appropriate tools to analyze these projections and measure their quality in the case of polynomial lattice rules.

For an integration lattice $\mathcal{L}_t$, we denote by $\mathcal{L}_t(I)$ the projection of $\mathcal{L}_t$ over the subspace determined by I. The *dual lattice* to $\mathcal{L}_t(I)$ is defined as

$$\mathcal{L}_t^*(I) = \{\mathbf{h}(z) \in \mathbb{L}^{|I|} : \mathbf{h}(z) \cdot \mathbf{v}(z) \in \mathbb{F}_2[z] \text{ for each } \mathbf{v}(z) \in \mathcal{L}_t(I)\},$$

which is a subset of $(\mathbb{F}_2[z])^{|I|}$.

The two definitions that follow are taken from [24] and [27], respectively. They are given for general point sets P_n. We will construct polynomial lattice rules whose projected point sets have those nice properties.

DEFINITION 3.22. A point set P_n in $[0,1)^t$ is *fully projection-regular* if for each non-empty subset I of $\{1,\ldots,t\}$, the projection $P_n(I)$ has n distinct points.

It is certainly sensible to ask for the point sets P_n to be fully projection-regular if we are interested in highly uniform projections. Point sets defined by rectangular grids in $t \geq 2$ dimensions, for example, are *not* fully projection-regular, because for every projection, several points are superposed on each other.

DEFINITION 3.23. A point set P_n in $[0,1)^t$ is *dimension-stationary* if whenever $1 \leq i_1 < \ldots < i_s < t$ and $1 \leq j \leq t - i_s$, $P_n(\{i_1,\ldots,i_s\}) = P_n(\{i_1+j,\ldots,i_s+j\})$.

For a dimension-stationary point set, the projections $P_n(I)$ depend only on the spacings between the indices in I. In particular, the quality of $P_n(\{i_1,\ldots,i_s\})$ does not deteriorate as i_1 increases, assuming that the spacings $i_j - i_{j-1}$ remain the same, for $j = 2,\ldots,s$. This property does not hold for typical low-discrepancy point sets proposed in the literature.

The next proposition implies that *every* polynomial lattice rule defined via a polynomial LCG – i.e., every Korobov polynomial lattice rule – has the two enjoyable properties that we just defined, as long as $\gcd(a(z),P(z)) = 1$.

PROPOSITION 3.24. *Let $\mathcal{L}_t$ be a polynomial lattice rule of rank 1 which admits a basis of minimal form such that $\mathbf{v}_1(z) = (1, v_2(z),\ldots,v_t(z))/P(z)$, where $P(z)$ is a polynomial of degree k. Then, the corresponding point set P_n is fully projection-regular if and only if* $\gcd(v_j(z),P(z)) = 1$ *for $j = 2,\ldots,t$. If one can write $v_j(z) = a^{(j-1)}(z)$* mod *$P(z)$ for some polynomial $a(z)$ such that* $\gcd(a(z),P(z)) = 1$*, then P_n is dimension-stationary.*

Proof. Assume $\gcd(v_j(z),P(z)) = 1$ for $j = 2,\ldots,t$. To verify that P_n is fully projection-regular, it suffices to check that $P_n(\{j\})$ has n distinct points for each $j = 1,\ldots,t$. Now, $P_n(\{1\}) = \varphi(\{q(z)/P(z), q(z) \in \mathbb{F}_2[z]/(P)\})$ obviously has $n = 2^k$ distinct points, because there are exactly n polynomials in $\mathbb{F}_2[z]/(P)$. For $j = 2,\ldots,t$, we have that $P_n(\{j\}) = \varphi(\{q(z)v_j(z)/P(z) \bmod \mathbb{F}_2[z], q(z) \in \mathbb{F}_2[z]/(P)\}) = P_n(\{1\})$ if $\gcd(v_j(z),P(z)) = 1$. On the other hand, if $\gcd(v_j(z),P(z)) \neq 1$ for some $j = 2,\ldots,t$, then $q(z)v_j(z) = 1 \pmod{P(z)}$ has no solution $q(z)$ and therefore $P_n(\{j\})$ contains less than n points.

If the basis has the specified form, with $v_j(z) = a^{(j-1)}(z) \bmod P(z)$, then $P_n = \{(g(p_0(z)), g(p_1(z)),\ldots,g(p_{t-1}(z))), p_0(z) \in \mathbb{F}_2[z]/(P)\}$, where $p_j(z) = a(z)p_{j-1}(z) \bmod P(z)$ and $g(p(z)) = \varphi(p(z)/P(z))$. By [27, Proposition 2], P_n is dimension-stationary if the recurrence $p_j(z) = a(z)p_{j-1}(z) \bmod P(z)$ is invertible, and a sufficient condition for this is to have $\gcd(a(z),P(z)) = 1$. □

4. Equidistribution and Nets. To measure the quality of a polynomial lattice rule, we use a methodology generalizing one that is often used for testing the theoretical properties of LFSR generators [6, 12, 43, 45], based on the notion of equidistribution of the point set P_n. We first explain this notion and then briefly discuss its relationship with the concept of net. These ideas are used to define selection criteria for polynomial lattice rules in Section 6.

DEFINITION 4.1. Let $n = 2^k$. For a vector of non-negative integers $q_1, \dots, q_t$, partition the interval $[0, 1)$ along the jth axis into 2^{q_j} equal subintervals. This partitions $[0, 1)^t$ into 2^q rectangular boxes, where $q = q_1 + \cdots + q_t$. A point set P_n is $(q_1, \dots, q_t)$-*equidistributed* if it has exactly 2^{k-q} points in each of these boxes.

Let $H(q_1, \dots, q_t)$ denote the set of vectors $\mathbf{h} = (h_1, \dots, h_t) \in (\mathbb{F}_2[z])^t$ such that $h_j(z)$ has degree less than q_j (or $h_j(z) = 0$) for each j.

PROPOSITION 4.2. (Generalization of results in [5, 6]) *A point set P_n based on a polynomial integration lattice $\mathcal{L}_t$ is $(q_1, \dots, q_t)$-equidistributed if and only if $\mathcal{L}_t^* \cap H(q_1, \dots, q_t) = \{\mathbf{0}\}$.*

Proof. Consider the class $\mathcal{F}$ of all real-valued functions that are constant on each of the 2^q rectangular boxes in Definition 4.1. Clearly, P_n is $(q_1, \dots, q_t)$-equidistributed if and only if the corresponding lattice rule integrates every function $f \in \mathcal{F}$ with zero error. But the Walsh expansion of $f \in \mathcal{F}$ is

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in H(q_1, \dots, q_t)} \tilde{f}(\mathbf{h})(-1)^{\langle \mathbf{h}, \mathbf{u} \rangle}. \tag{4.1}$$

To see this, note that any $f \in \mathcal{F}$ can be written as

$$f(\mathbf{u}) = \sum_{v_1=0}^{2^{q_1}-1} \cdots \sum_{v_t=0}^{2^{q_t}-1} c_{v_1, \dots, v_t} \prod_{j=1}^{t} \mathbf{1}_{v_j 2^{-q_j} \le u_j < (v_j+1)2^{-q_j}} \ , \tag{4.2}$$

where the $c_{v_1, \dots, v_t}$ are real numbers. If $\mathbf{h} \notin H(q_1, \dots, q_t)$, $h_j(z)$ has degree $w_j \ge q_j$ for some j. Let $d = w_j - q_j + 1$. When l goes from 0 to $2^d - 1$, $\langle h_j, v_j 2^{-q_j} + l 2^{-w_j - 1} \rangle$ is equal to each of 0 and 1 exactly 2^{d-1} times. Hence, if we first integrate $f(\mathbf{u})$ with respect to u_j when computing $\tilde{f}(\mathbf{h})$ via (3.11), in which $f(\mathbf{u})$ has been replaced by (4.2), any term of the sum will be 0 because

$$\int_0^1 c_{v_1, \dots, v_t} \mathbf{1}_{v_j 2^{-q_j} \le u_j < (v_j+1)2^{-q_j}} (-1)^{\langle h_j, u_j \rangle} du_j$$
$$= c_{v_1, \dots, v_t} \sum_{l=0}^{2^d - 1} (-1)^{\langle h_j, v_j 2^{-q_j} + l 2^{-w_j - 1} \rangle} = 0.$$

Thus $\tilde{f}(\mathbf{h}) = 0$ if $\mathbf{h} \notin H(q_1, \dots, q_t)$, and (4.1) follows. If $H(q_1, \dots, q_t) \cap \mathcal{L}_t^* = \{\mathbf{0}\}$, then $\sum_{\mathbf{0} \ne \mathbf{h} \in \mathcal{L}_t^*} \tilde{f}(\mathbf{h}) = 0$ for all $f \in \mathcal{F}$, so P_n is $(q_1, \dots, q_t)$-equidistributed. To prove the other direction, note that for any nonzero $\bar{\mathbf{h}} \in H(q_1, \dots, q_t)$, $g(\mathbf{u}) = (-1)^{\langle \bar{\mathbf{h}}, \mathbf{u} \rangle}$ is in $\mathcal{F}$ and $\tilde{g}(\mathbf{h}) = 1$ if $\mathbf{h} = \bar{\mathbf{h}}$, and 0 otherwise. Hence if P_n is $(q_1, \dots, q_t)$-equidistributed, then $\sum_{\mathbf{0} \ne \mathbf{h} \in \mathcal{L}_t^*} \tilde{g}(\mathbf{h}) = 0$ and $\bar{\mathbf{h}} \notin \mathcal{L}_t^*$. □

By taking $q_j = \ell$ for each j, we recover the result of [6]: P_n is t-distributed to ℓ bits of accuracy if and only if $\mathcal{L}_t^* \cap H(\ell, \dots, \ell) = \{\mathbf{0}\}$, i.e., if and only if the shortest nonzero vector $\mathbf{h}$ in $\mathcal{L}_t^*$ has length $\|\mathbf{h}\|_\infty \ge 2^\ell$. The largest value of ℓ for which P_n is t-distributed to ℓ bits of accuracy is called the t-dimensional *resolution* of P_n [12, 6, 23].

Definition 4.1 can be adapted to projections as follows.

DEFINITION 4.3. Let $n = 2^k$. For a subset $I = \{i_1, \ldots, i_s\}$ of $\{1, \ldots, t\}$, the projection $P_n(I)$ is $(q_{i_1}, \ldots, q_{i_s})$-*equidistributed* if each of the $2^{q(I)}$ rectangular boxes, where $q(I) = q_{i_1} + \ldots + q_{i_s}$, obtained by partitioning the interval $[0,1)$ along the i_jth axis into $2^{q_{i_j}}$ equal subintervals for $j = 1, \ldots, s$, contains $2^{k-q(I)}$ points from $P_n(I)$.

The previous definitions do not assume that the points in P_n or $P_n(I)$ are all distinct. When a point appears more than once in the set, it is counted as many times as it appears. In other words, P_n and $P_n(I)$ should be interpreted as *multisets.*

The definition of $(q_1, \ldots, q_t)$-equidistribution allows us to recover the definition of (q,k,t)-net introduced in [39] for base 2 and in [31] for general bases: a point set P_n in $[0,1)^t$ with $n = 2^k$ points is a *(q,k,t)-net in base 2* – usually called a (t,m,s)-net, using a different notation – if it is $(q_1, \ldots, q_t)$-equidistributed for every non-negative integer vector $(q_1, \ldots, q_t)$ such that $q_1 + \ldots + q_t = k - q$. We refer the reader to [25, Section 4] for more connections between this concept and the resolution.

5. Randomization and Variance Expression.

The counterpart of the Cranley-Patterson rotation for polynomial lattice rules over $\mathbb{F}_2$ is to generate a single t-dimensional vector of formal series $\mathbf{S}(z) = (S_1(z), \ldots, S_t(z))$ uniformly over $\mathbb{L}_0^t$, and add it to each vector $\mathbf{v}(z) \in \mathcal{L}_t \cap \mathbb{L}_0^t$ before applying φ in the definition of a polynomial lattice point set. In other words, P_n is replaced by

$$\tilde{P}_n = \{\mathbf{u} = \varphi(\mathbf{v}(z) + \mathbf{S}(z)) \text{ such that } \mathbf{v}(z) \in \mathcal{L}_t \cap \mathbb{L}_0^t.\}$$

This is equivalent to generating a random variable $\mathbf{U}$ uniformly over $[0,1)^t$ and replacing P_n by $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$, where $\tilde{\mathbf{u}}_i = \mathbf{u}_i \oplus \mathbf{U}$, the bitwise exclusive-or of the binary expansions of the coordinates of $\mathbf{u}_i$ and $\mathbf{U}$.

We define the random variables $\tilde{Q}_n$ and $\tilde{E}_n$ as in Section 2, but with this new $\tilde{P}_n$. To estimate the error, we can make m independent shifts and compute a confidence interval for μ from the m i.i.d. copies of $\tilde{Q}_n$.

PROPOSITION 5.1. *One has $E[\tilde{E}_n] = 0$ and, if f is square-integrable,*

$$\mathrm{Var}[\tilde{E}_n] = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{L}_t^*} |\tilde{f}(\mathbf{h})|^2. \tag{5.1}$$

Proof. Denote by $u_{i,j,k}$ the coefficient of 2^{-k} in the binary expansion of $u_{i,j}$, the jth coordinate of $\mathbf{u}_i$. Since $\mathbf{U}$ has the uniform distribution over $[0,1)^t$, its bits $U_{j,k}$, for $j = 1, \ldots, t$ and $k \geq 1$, are i.i.d. Bernoulli with parameter $p = 1/2$. Then the bits $(u_{i,j,k} + U_{j,k}) \bmod 2$ are also i.i.d. Bernoulli and each $\mathbf{u}_i \oplus \mathbf{U}$ has the uniform distribution over $[0,1)^t$. This implies that $E(f(\mathbf{u}_i \oplus \mathbf{U})) = \mu$ for each i, so $E[\tilde{E}_n] = 0$.

To show (5.1), we proceed as in the proof of [24, Proposition 4] for ordinary lattices. We define $g : [0,1)^t \to \mathbb{R}$ by $g(\mathbf{U}) = \sum_{i=0}^{n-1} f(\mathbf{u}_i \oplus \mathbf{U})/n$. Thus, $\mathrm{Var}(g(\mathbf{U})) = \mathrm{Var}(\tilde{E}_n)$. Parseval's equality holds for the Walsh series expansion [13], so

$$\mathrm{Var}(g(\mathbf{U})) = \sum_{\mathbf{0} \neq \mathbf{h} \in (\mathbb{F}_2[z])^t} |\tilde{g}(\mathbf{h})|^2, \tag{5.2}$$

and the coefficients $\tilde{g}(\mathbf{h})$ are given by:

$$\tilde{g}(\mathbf{h}) = \int_{[0,1)^t} g(\mathbf{u})(-1)^{\langle \mathbf{h}, \mathbf{u} \rangle} d\mathbf{u} = \int_{[0,1)^t} \left(\frac{1}{n} \sum_{i=0}^{n-1} f(\mathbf{u}_i \oplus \mathbf{u}) \right) (-1)^{\langle \mathbf{h}, \mathbf{u} \rangle} d\mathbf{u}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=0}^{n-1}\int_{[0,1)^t} f(\mathbf{u}_i \oplus \mathbf{u})(-1)^{\langle \mathbf{h},\mathbf{u}\rangle} d\mathbf{u} \\
&= \frac{1}{n}\sum_{i=0}^{n-1}\int_{[0,1)^t} f(\mathbf{v}_i)(-1)^{\langle \mathbf{h},\mathbf{u}_i\oplus\mathbf{v}_i\rangle} d\mathbf{v}_i \\
&= \frac{1}{n}\sum_{i=0}^{n-1}(-1)^{\langle \mathbf{h},\mathbf{u}_i\rangle}\int_{[0,1)^t} f(\mathbf{v}_i)(-1)^{\langle \mathbf{h},\mathbf{v}_i\rangle} d\mathbf{v}_i \\
&= \frac{1}{n}\sum_{i=0}^{n-1}(-1)^{\langle \mathbf{h},\mathbf{u}_i\rangle}\tilde{f}(\mathbf{h}) = \begin{cases} \tilde{f}(\mathbf{h}) & \text{if } \mathbf{h}\in\mathcal{L}_t^*, \\ 0 & \text{otherwise.}\end{cases}
\end{aligned}
$$

In the above display, the third equality is obtained by exchanging the sum and the integral, which is allowed by Fubini's theorem because f is integrable; the fourth equality is obtained by applying the change of variable $\mathbf{v}_i = \mathbf{u}_i \oplus \mathbf{u}$, which also permits us to rewrite $\mathbf{u}$ as $\mathbf{u}_i \oplus \mathbf{v}_i$; the fifth comes from the fact that $(-1)^{\langle \mathbf{h},\mathbf{v}_i\oplus\mathbf{u}_i\rangle} = (-1)^{\langle \mathbf{h},\mathbf{v}_i\rangle}(-1)^{\langle \mathbf{h},\mathbf{u}_i\rangle}$; and the last equality follows from Lemma 3.20. By replacing this in (5.2), the result (5.1) immediately follows. □

This variance expression suggests discrepancy measures of the form (2.4), with L_t^* replaced by $\mathcal{L}_t^*$. The weights $w(\mathbf{h})$ should be chosen in accordance with our knowledge (or intuition) of how the Walsh coefficients are likely to behave as a function of $\mathbf{h}$. This is discussed in the next section. Also, it is clear from (5.1) that bounds on the variance or on its convergence rate as a function of n can be obtained by making appropriate assumptions on the Walsh coefficients of f and on the dual lattice $\mathcal{L}_t^*$.

Other randomization techniques have been proposed for general digital nets. We refer the reader to [25, Section 6] and the references therein for more on this topic.

6. Selection Criteria.

We now examine and discuss specific selection criteria of the form (2.4) for choosing general-purpose polynomial lattice rules. Ideally, the weights $w(\mathbf{h})$ in a criterion of the form $D(P_n) = \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} w(\mathbf{h})$ should be proportional to the squared Walsh coefficients $|\tilde{f}(\mathbf{h})|^2$ that appear in the variance expression (5.1) for the function f of interest. But in practice, the polynomial lattice rules must be chosen without knowing these coefficients, and sometimes without any information at all on f (e.g., when selecting general-purpose lattice rules for numerical software).

This requires heuristic assumptions and arguments. Here, we take the usual approach of assuming that the large squared Walsh coefficients usually correspond to polynomial vectors $\mathbf{h}$ of small degree, and having a small or moderate number of nonzero components, with indices that are not too far from each other.

In the remainder of this section, we introduce a criterion based on the equidistribution of a selected set of projections. The specific rules used in Section 9 have been selected based on this criterion. We then discuss alternative criteria, based on different norms and/or weights.

6.1. Equidistribution of projections.

The following criterion computes the resolution ℓ_I of some specified low-dimensional projections $P_n(I)$, and makes sure that ℓ_I is close to its best possible value for each of those I. The choice of the subsets I for which ℓ_I is computed is arbitrary. Here we consider the same class of subsets as for the criterion $M_{t_1,\ldots,t_d}$ proposed in [24] for ordinary lattice rules. More specifically, suppose that P_n is fully projection-regular, dimension-stationary, and contains $n = 2^k$

points. We define

$$\Delta_{t_1,\ldots,t_d} = \max_{1\leq s\leq d} \max_{I\in S(t_s,s)} \left[\ell_s^*(n) - \ell_I\right], \tag{6.1}$$

where $\ell_s^*(n) = \lfloor k/s \rfloor$ is the maximal resolution for a set of $n = 2^k$ points in $[0,1)^s$, $S(t_1, 1) = \{\{1,\ldots,s\},\ 1 \leq s \leq t_1\}$, and $S(t_s, s) = \{\{i_1,\ldots,i_s\},\ 1 = i_1 < \ldots < i_s \leq t_s\}$ for $s \geq 2$. Efficient methods for computing the resolution are given in [5, 12, 23].

The criterion $\Delta_{t_1,\ldots,t_d}$ computes the difference between the maximal resolution and the actual resolution for all projections over successive indices for up to t_1 dimensions, then for all two-dimensional projections over pairs of non-successive indices (i_1, i_2) for $1 = i_1 < i_2 \leq t_2$, then for all three-dimensional projections over triples of non-successive indices (i_1, i_2, i_3) for $1 = i_1 < i_2 < i_3 \leq t_3$, and so on, up to the set of d-dimensional projections over the sets I of non-successive indices that belong to $S(t_d, d)$. Then it takes the worst case. We can fix $i_1 = 1$ without loss of generality because of our assumption that P_n is dimension-stationary, a property enjoyed by all the point sets proposed in Section 7. This criterion generalizes the property of being *maximally equidistributed* (ME) [23, 45]: P_n is ME if and only if $\Delta_k = 0$.

There are situations, especially when d and the t_s are large, where no rule can be found for which $\Delta_{t_1,\ldots,t_d} = 0$ and several rules can be found with the same minimal value of $\Delta_{t_1,\ldots,t_d}$ (e.g., 1 or 2). One may then use a second-level criterion to select among these rules. For example, take the one with the minimal value of the sum

$$\Sigma_{t_1,\ldots,t_d} = \sum_{s=1}^{d} \sum_{I\in S(t_s,s)} (\ell_s^*(n) - \ell_I). \tag{6.2}$$

The criterion $\Delta_{t_1,\ldots,t_d}$ is obviously not perfect. Just like any other criterion, it chooses to look more closely at some arbitrarily chosen features of P_n while it neglects other aspects. It also weights equally all the projections considered. It could be fine-tuned by introducing weights in the terms of (6.1), i.e., by multiplying these terms by different constants when defining the criterion.

6.2. From the sup norm to the product norm. Computing ℓ_I is equivalent to finding the length of the shortest nonzero vector in $\mathcal{L}_t^*(I)$ with respect to the sup norm $\|\cdot\|_\infty$. Now suppose that we replace this norm by the product norm $\|\mathbf{h}\|_\pi = \prod_{j=1}^t |h_j|_l$, with $|h_j|_l = |h_j|_{\mathrm{p}}$ if $h_j \neq 0$ and $|h_j|_l = 2^{-1}$ if $h_j = 0$. If P_n is dimension-stationary, the corresponding quantity $\Delta_{t_1,\ldots,t_d}$ with $d = t$ and $t_1 = \ldots = t_d = t$ is then equal to the parameter q defining a (q, k, t)-net. Recall that $q = 0$ cannot be reached if $t > 3$ [32, Corollary 4.21]. This may suggest a criterion of the form

$$\tilde{q} = \max_{\phi\neq I\subseteq\{1,\ldots,t\}} (q_I - q_{|I|}^*),$$

where q_I is the smallest q such that $P_n(I)$ is a $(q, k, |I|)$-net in base 2, and

$$q_{|I|}^* = \left\lceil \frac{|I|-1}{2} - \log_2\left(\frac{|I|+2}{2}\right) \right\rceil = \lceil (|I|+1)/2 - \log_2(|I|+2) \rceil$$

is the smallest possible value of q_I that can be attained for a $(q_I, k, |I|)$-net in base 2 [34]. This criterion can also be made more flexible by using parameters $t_1,\ldots,t_d$ as in the definition of $\Delta_{t_1,\ldots,t_d}$, in order to restrict the computation of q_I to the subsets $I \in S(t_s, s)$ for $s = 2,\ldots,d$, or of the form $I = \{1,\ldots,s\}$, for $1 \leq s \leq t_1$. A disadvantage of using q_I instead of ℓ_I is that its computation generally requires much more time than ℓ_I.

6.3. A polynomial version of $\tilde{\mathcal{P}}_\alpha$. For ordinary lattice rules, Hickernell [16] introduced a measure of discrepancy denoted $\tilde{\mathcal{P}}_\alpha$. This criterion can be computed in $O(nt)$ and allows the different projections of the point set P_n to be weighted differently, e.g., according to the (anticipated) importance of the corresponding projections of the function f. Here we propose a similar criterion for polynomial lattice rules for the case $\alpha = 2$ and using *product-type weights*. More precisely, we define

$$\tilde{\mathcal{P}}_{2,\mathrm{PLR}} = \sum_{\mathbf{0} \neq \mathbf{h} \in \mathcal{L}_t^*} \beta_{I_{\mathbf{h}}}^2 \|\mathbf{h}\|_{\tilde{\pi}}^{-2}, \tag{6.3}$$

where $I_{\mathbf{h}} = \{j : h_j \neq 0\}$, $\|\mathbf{h}\|_{\tilde{\pi}} = \prod_{j \in I_{\mathbf{h}}} |h_j|_{\mathrm{p}}$, $\beta_I = \beta_0 \prod_{j \in I} \beta_j$, and $\beta_j > 0$ for $j = 0, \ldots, t$.

The next proposition, proved in the appendix, provides a function $\tilde{\psi}$ whose mean value over P_n equals $\tilde{\mathcal{P}}_{2,\mathrm{PLR}}$. This can be used to compute $\tilde{\mathcal{P}}_{2,\mathrm{PLR}}$ in $O(nt)$ time instead of having to deal with an infinite sum as in the representation (6.3).

PROPOSITION 6.1. *If P_n is a polynomial lattice point set, then one has*

$$\tilde{\mathcal{P}}_{2,\mathrm{PLR}} = \frac{\beta_0^2}{n} \sum_{i=0}^{n-1} \tilde{\psi}(\mathbf{u}_i), \ \textit{where } \tilde{\psi}(\mathbf{u}) = -1 + \prod_{j=1}^{t} \left[1 + 2\beta_j^2 \left(1 - 3 \cdot 2^{\lfloor \log_2 u_j \rfloor}\right)\right].$$

The relation between $\tilde{\mathcal{P}}_{2,\mathrm{PLR}}$ and other figures of merit such as the *dyadic diaphony* (see, e.g., [14]) is discussed in more details in [25]. The parameters reported in Section 7 were found using the criterion $\Delta_{t_1,\ldots,t_d}$ instead of $\tilde{\mathcal{P}}_{2,\mathrm{PLR}}$, mainly for computational efficiency reasons: for large values of n and t – say, $n \geq 2^{16}$ and $15 \leq t < 40$ – computer searches based on $\Delta_{t_1,\ldots,t_d}$ can be made rapidly, whereas searching with respect to $\tilde{\mathcal{P}}_{2,\mathrm{PLR}}$ becomes practically infeasible.

7. Implementations and Examples of Parameters. In this section, we give examples of parameters describing polynomial lattice rules based on simple and combined LFSR generators, chosen according to the criterion Δ_{t_1,t_2,t_3,t_4}, first with $(t_1, t_2, t_3, t_4) = (13, 13, 13, 13)$ and then with (40,40,30,20). Within each class of rules considered, we made an exhaustive search for the parameters that minimized this criterion, using the software package REGPOLY [36]. Ties were broken using the associated (secondary) criterion $\Sigma_{t_1,\ldots,t_4}$ given in (6.2).

Tezuka and L'Ecuyer [23, 44] provide an efficient implementation algorithm for a LFSR generator whose characteristic polynomial is a primitive trinomial of the form $P(z) = z^k + z^q + 1$, with $0 < 2q < k$, and $a(z) = z^\nu \bmod P(z)$ for some integer ν satisfying $0 < \nu \leq k - q < k \leq w$, $\gcd(\nu, 2^k - 1) = 1$, and w equal to the word length of the computer [23]. This method is also easy to generalize to the case where $P(z)$ has more than three nonzero coefficients [36], assuming that $\nu \leq k - q$, where q is the degree of $P(z) - z^k$ (e.g., $q = 3$ if $P(z) = z^7 + z^3 + 1$), although the computing cost increases with the number of coefficients. Table 7.1 gives the best parameters for LFSR generators that satisfy these conditions and for which $P(z)$ is either a trinomial or pentanomial, for three values of n. In this table, Δ and Σ are the values of the primary and secondary criteria, and $P(z)$ is represented by a vector containing the exponents of z whose coefficient are nonzero, e.g., (11,5,3,1,0) represents $z^{11} + z^5 + z^3 + z + 1$.

It is well recognized [23, 43] that LFSR generators having a polynomial $P(z)$ with too few nonzero coefficients must be avoided because of their bad high-dimensional properties. Therefore, in Table 7.2 we give search results for rules based on combined LFSR generators with two or three components, where the characteristic polynomials

Table 7.1
Best simple generators

	w.r.t. $\Delta_{13,13,13,13}$			w.r.t. $\Delta_{40,40,30,20}$		
n	$P(z)$	ν	Δ (Σ)	$P(z)$	ν	Δ (Σ)
2^{11}	(11,5,3,1,0)	4	1 (17)	(11,5,3,1,0)	3	2 (103)
2^{13}	(13,9,7,3,0)	4	2 (84)	(13,7,6,3,0)	4	2 (455)
2^{15}	(15,4,2,1,0)	8	2 (55)	(15,9,4,1,0)	5	3 (355)

$P_j(z)$ of the components are primitive trinomials or pentanomials of degrees k_j, with period lengths $2^{k_j} - 1$ that are pairwise relatively prime, and where the components satisfy the same conditions as in Table 7.1.

Table 7.2
Best combined generators

	w.r.t. $\Delta_{13,13,13,13}$			w.r.t. $\Delta_{40,40,30,20}$		
n	$P_j(z)$	ν_j	Δ (Σ)	$P_j(z)$	ν_j	$\Delta(\Sigma)$
2^{11}	(4,1,0) (7,3,2,1,0)	1 4	1 (17)	(4,1,0) (7,3,0)	1 4	2 (115)
2^{13}	(6,1,0) (7,3,0)	2 1	1 (87)	(3,1,0) (10,4,3,1,0)	2 4	2 (469)
2^{15}	(3,1,0) (5,2,0) (7,1,0)	1 3 2	2 (55)	(3,1,0) (5,2,0) (7,3,0)	2 1 2	3 (332)

As we see in Table 7.2, the combined generators do not always improve the two criteria. However, a more careful study of the equidistribution revealed that most of the simple generators from Table 7.1 have a *resolution gap* of 2 (i.e., $\ell^*_{|I|} - \ell_I = 2$) on the important two-dimensional projection $P(\{1, 2\})$, whereas the combined generators from Table 7.2 have a gap of 0 or 1 on this projection. We give in the next section examples illustrating how this hidden defect of simple generators can affect the quality of their associated estimator. This suggests that perhaps in future and more extensive searches, more weight should be given to certain projections in the criterion.

To construct P_n for these simple and combined LFSR generators, it suffices to implement the generator, run it over all of its cycles, and retain all the t-tuples of successive output values. In practice, one may just store the cycles into some data structure (this does not require the knowledge of t) and produce each t-tuple only when evaluating f for it. The case of a random t is nicely handled by this approach.

8. Linear output transformations for tempering. The idea here is that instead of applying the output mapping φ directly to $\Xi_t = \mathcal{L}_t \cap \mathbb{L}_0^t$ to produce P_n, one can apply an additional linear transformation $\tau : \Xi_t \to \mathbb{L}_0^t$ before applying φ, as in (3.8) and (3.9). That is, define $P_n = \varphi(\tau(\Xi_t))$ instead of $P_n = \varphi(\Xi_t)$. This can be convenient for improving the equidistribution on P_n if we decide to define it from a generator with a very simple recurrence, in order to make its implementation faster. Below we give two examples of such transformations, taken from [28, 29, 26]. Other ones can be defined easily.

We first examine how the transformation τ affects the structure of the set Ξ_t. Using the same notation as in (3.9), τ can be written in terms of a transformation on

$\mathbb{L}_0$, i.e., for $s(z) \in \mathbb{L}_0$,

$$\tau(s(z)) = \tau\left(\sum_{j=1}^{\infty} d_j z^{-j}\right) = \sum_{j=1}^{\infty} z^{-j} \sum_{i=1}^{k} b_{i,j} d_i, \tag{8.1}$$

where each $b_{i,j} \in \mathbb{F}_2$, $k = \deg P(z)$, and $P(z) = \det(\mathcal{L}_t^*)$. Let $\tau(s_0(z), \ldots, s_{t-1}(z)) = (\tau(s_0(z)), \ldots, \tau(s_{t-1}(z)))$. When $s(z)$ is a coordinate of a vector in Ξ_t, it can be written as $s(z) = p(z)/P(z)$, where $p(z) \in \mathbb{F}_2[z]/(P)$, and therefore, the d_j's in (8.1) follow a linear recurrence with characteristic polynomial $P(z)$. This implies that each d_i, for $i > k$, is a linear combination of $d_1, \ldots, d_k$, so replacing k by a larger value in the inner summation of (8.1) would bring no additional generality. In what follows, we denote by $\mathbf{B}$ the matrix whose entry on the jth row and ith column is $b_{i,j}$, for $j = 1, 2, \ldots$, $i = 1, \ldots, k$, and $\mathbf{B}_k$ represents the $k \times k$ matrix containing the first k rows of $\mathbf{B}$.

EXAMPLE 8.1. A *permutation* of the coordinates is an example of a simple linear output transformation [26]. The idea is to choose a permutation π of the set $\{1 \ldots k\}$, and to define $\mathbf{B}_k$ as the identity matrix with its rows permuted according to π. Permutations of the form $\pi(i) = (pi + q) \bmod k$ are suggested in [26]; they are easy to implement efficiently and require only two parameters p and q.

EXAMPLE 8.2. As a second example, we consider a slight modification of the *Matsumoto-Kurita (MK) tempering* introduced in [28]. This variant is described in [26]. It consists in choosing two integers s_1 and s_2 between 1 and k, and two k-bit vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. If $\mathbf{d}$ contains the first k coefficients of a coordinate $s(z)$ of a point in Ξ_t, and $\mathbf{y}$ contains the first k coefficients of the transformed point, we define

$$\begin{aligned}\tilde{\mathbf{d}} &= \mathbf{d} \oplus ((\mathbf{d} << s_1) \,\&\, \mathbf{b}_1)\\ \mathbf{y} &= \tilde{\mathbf{d}} \oplus ((\tilde{\mathbf{d}} << s_2) \,\&\, \mathbf{b}_2),\end{aligned}$$

where $\oplus$ means component-wise addition in $\mathbb{F}_2$, & means component-wise multiplication in $\mathbb{F}_2$, and $<< s$ means a left shift by s bits.

REMARK 8.3. The linear transformation (8.1) is natural for generators implemented in terms of the coefficients $d_{i,j}$ of the formal series $s_i(z)$, like the Tausworthe generators. But for generators implemented directly in terms of the coefficients $c_{i,j}$ of the polynomial $p_i(z)$, as discussed in Remark 3.14, it would be more natural to perform the tempering by applying the linear transformation directly to the vector of those $c_{i,j}$.

Although τ is linear over $\mathbb{F}_2^k$, in general it transforms Ξ_t (viewed as a subspace of $\mathbb{L}^t$) in a nonlinear way, and therefore the lattice structure of Ξ_t is not preserved under this transformation. Example 8.5 below illustrates this fact when transformations such as those discussed in Examples 8.1 and 8.2 are applied to a polynomial LCG. More precisely, the transformed point set $\tau(\Xi_t)$ retains the property of being closed under addition and subtraction, but it does not have the lattice structure described in Section 3.

The fact that these transformations do not generally preserve the lattice structure implies that a wider space of point sets is available when we search not only over some type of basis (e.g., Korobov rules), but also allow the possibility of applying transformations τ as described above. Recall that the main reason for doing this is to find point sets with better equidistribution properties, and that are easy to implement. Figure 8.1 in Example 8.5 illustrates how applying tempering can greatly improve the equidistribution of a point set based on a polynomial integration lattice. The techniques discussed in [12, 23] for measuring the equidistribution can be easily used for point sets to which tempering has been applied, and this is what is done in the search package REGPOLY [36]; the method described in [5] can also be used since it has been defined in a general setting that includes this type of point sets. Note that the transformed point set $P_n = \varphi(\tau(\Xi_t))$ is still a special case of digital net and therefore it could be analyzed under this more general setting.

In the next proposition, we give sufficent conditions on the $b_{i,j}$'s for τ to correspond to a bijection over $\mathbb{F}_2[z]/(P)$. This guarantees that if the initial point set $\varphi(\Xi_t)$ is dimension-stationary, then the transformed point set $\varphi(\tau(\Xi_t))$ also has this property. To do so, we rewrite τ in matrix form and examine how it transforms a coordinate $s(z)$ of a point in Ξ_t. Let $\tilde{s}(z) = \tau(s(z))$ and let $y_1, y_2, \ldots$ be the coefficients defining $\tilde{s}(z)$, i.e.,

$$\tilde{s}(z) = \sum_{j=1}^{\infty} y_j z^{-j}.$$

Hence, $y_j = \sum_{i=1}^{k} b_{i,j} d_i$ for each $j \geq 1$. Let $\mathbf{y}$ denote the column vector containing the coefficients y_j and $\mathbf{d} = (d_1, \ldots, d_k)^{\mathrm{T}}$. We then have $\mathbf{Bd} = \mathbf{y}$.

Now, remember that $s(z)$ is a rational function of the form $s(z) = p(z)/P(z)$. For the remaining part of this section, we assume for simplification that τ only modifies the numerator in the rational function that defines $s(z)$, i.e., we assume that the coefficients y_j follow a linear recurrence with characteristic polynomial $P(z)$. In terms of the matrix $\mathbf{B}$, this means that the successive row vectors of $\mathbf{B}$ follow the same recurrence. It then becomes clear that τ corresponds to a bijection applied to $p(z)$ if and only if $\mathbf{B}_k$ is invertible. The following proposition clarifies these observations.

PROPOSITION 8.4. *Let $\mathcal{L}_t$ be a polynomial integration lattice, $P(z) = \det(\mathcal{L}_t^*) = \sum_{j=0}^{k} a_j z^{k-j}$,*

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ a_1 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ a_{k-1} & \ldots & a_1 & 1 \end{pmatrix},$$

$\mathbf{B}_k$ an invertible $k \times k$ matrix with entries in $\mathbb{F}_2$, $\mathbf{B}$ the $\infty \times k$ matrix whose first k rows $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are those of $\mathbf{B}_k$ and whose jth row is defined by

$$\mathbf{b}_j = a_1 \mathbf{b}_{j-1} + \ldots + a_k \mathbf{b}_{j-k}$$

for $j > k$, and τ the linear transformation determined by $\mathbf{B}$ as in (8.1).

If $s(z) = p(z)/P(z)$ is a coordinate of a vector in Ξ_t, then $\tau(s(z)) = \chi(p(z))/P(z)$, where χ is a bijection on $\mathbb{F}_2[z]/(P)$ defined by

$$\chi(p(z)) = \sum_{j=1}^{k} \tilde{c}_j z^{k-j},$$

where $(\tilde{c}_1, \ldots, \tilde{c}_k)^{\mathrm{T}} = \mathbf{AB}_k\mathbf{A}^{-1}(c_1, \ldots, c_k)^{\mathrm{T}}$ and the c_j's are the coefficients of $p(z)$, i.e., $p(z) = \sum_{j=1}^{k} c_j z^{k-j}$. Also, if $\varphi(\Xi_t)$ is dimension-stationary, then $\varphi(\tau(\Xi_t))$ is dimension-stationary.

EXAMPLE 8.5. Consider the Korobov rule in 2 dimensions, with 2^7 points, defined by

$$\Xi_2 = \left\{ \frac{q(z)}{P(z)}(1, z) \bmod \mathbb{F}_2[z],\ q(z) \in \mathbb{F}_2[z]/(P) \right\},$$

where $P(z)$ is the primitive polynomial $P(z) = z^7 + z + 1$ (see Figure 8.1, left).

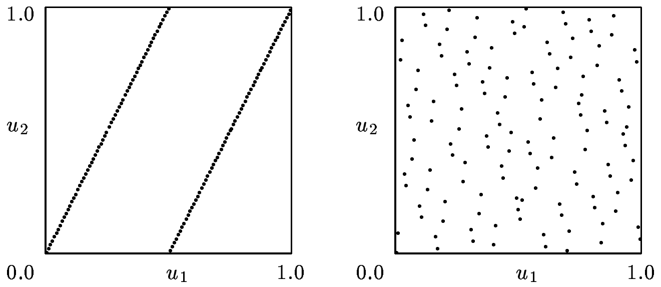

FIG. 8.1. Left: *Point set based on $P(z) = z^7 + z + 1$ and $a(z) = z$ without tempering.* Right: *Same parameters, but with tempering, namely a (1,5)-permutation followed by (2,4)-MK tempering with* $\mathbf{b}_1 = 0011000$ *and* $\mathbf{b}_2 = 0110000$.

Equivalently, one has

$$\Xi_2 = \{(s_0(z), s_1(z)) : s_i(z) = p_i(z)/P(z),\ p_0(z) \in \mathbb{F}_2[z]/(P)\},$$

where $p_i(z) = zp_{i-1}(z) \bmod P(z)$, for $i \geq 1$. For this example, $\mathbf{A}$ is almost equal to the 7×7 identity matrix, except that $\mathbf{A}_{7,1} = 1$. Also, $\mathbf{A}^{-1} = \mathbf{A}$. The application of an (1,5)-permutation followed by an (2,4)-MK tempering with $\mathbf{b}_1 = 0011000$ and $\mathbf{b}_2 = 0110000$ yields

$$\mathbf{B}_7 = \begin{pmatrix} 0&0&0&0&0&1&0\\ 0&0&0&0&1&0&1\\ 1&0&1&0&1&0&0\\ 0&1&0&1&0&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&1&0&0 \end{pmatrix} \text{ and } \mathbf{A}\mathbf{B}_7\mathbf{A}^{-1} = \begin{pmatrix} 0&0&0&0&0&1&0\\ 1&0&0&0&1&0&1\\ 1&0&1&0&1&0&0\\ 0&1&0&1&0&0&0\\ 0&0&1&0&0&0&0\\ 0&0&0&1&0&0&0\\ 0&0&0&0&1&1&0 \end{pmatrix}.$$

If we look at the first six terms of the recurrence $p_i(z) = zp_{i-1}(z) \bmod P(z)$ initialized at $p_0(z) = 1$, we get $1, z, z^2, z^3, z^4, z^5$. When the tempering is applied, these terms are transformed according to the mapping χ defined by $\mathbf{A}\mathbf{B}_7\mathbf{A}^{-1}$ and become $z^5, z^6 + 1, z^5 + z^4 + 1, z^3 + z, z^4 + z^2, z^3$.

It is easy to see that there is no $a(z) \in \mathbb{F}_2[z]/(P)$ such that the new terms follow the recurrence $p_i(z) = a(z)p_{i-1}(z) \bmod P(z)$. This means that the "tempered" point set is not in our search space if we limit our search to Korobov rules based on the usual multiplication. In fact, $\tau(\Xi_t)$ does not even have a lattice structure. For example, $(z^3 + z, z^4 + z^2) \in \tau(\Xi_2)$, but $z(z^3 + z, z^4 + z^2) = (z^4 + z^2, z^5 + z^3) \notin \tau(\Xi_2)$ because $(z^4 + z^2, z^3) \in \tau(\Xi_2)$, χ is one-to-one, and Ξ_2 is fully projection-regular.

The fact that τ does not preserve the lattice structure of Ξ_t means that Lemma 3.20, on which the error and variance results for polynomial lattice rules rely, has to be modified in order to hold for transformed point sets. More precisely, the dual lattice $\mathcal{L}_t^*$ used in this lemma has to be replaced by a more general dual space used for studying digital nets in [33]. Without giving all the details about the error and variance analysis in this more general setting (we refer the reader to [25] instead), a useful fact to mention is that when P_n is $(q_1, \dots, q_t)$-equidistributed, this dual space does not contain any non-zero vector in $H(q_1, \dots, q_t)$, which has been defined in Section 4. Hence for any square-integrable function f,

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=0}^{n-1} f(\mathbf{u}_i \oplus \mathbf{U})\right) \leq \sum_{\substack{0 \neq \mathbf{h} \in (\mathbb{F}_2[z])^t, \\ \mathbf{h} \notin H(q_1, \dots, q_t)}} |\tilde{f}(\mathbf{h})|^2,$$

if P_n is $(q_1, \dots, q_t)$-equidistributed. Therefore, transformations τ that maximize (in some sense) the values of $q_1, \dots, q_t$ for which $\varphi(\tau(\Xi_t))$ is $(q_1, \dots, q_t)$-equidistributed have smaller upper bounds on the variance of their associated estimator. This supports the use of criteria based on the $(q_1, \dots, q_t)$-equidistribution for choosing transformed polynomial lattice rules.

Based on these heuristic arguments, we searched for transformed point sets having a low value of $\Delta_{t_1, \dots, t_d}$. Table 8.1 contains the results of a search for rules based on LFSR generators with tempering. Using REGPOLY, we let the computer search randomly over a total of 100 different combinations of permutations and MK tempering applied to the individual components of each combined generator, for the same type of LFSR generators as in Section 7. The values of $\mathbf{b}_1$ and $\mathbf{b}_2$ are given in hexadecimal notation in this table. By allowing these output transformations, we were able to find point sets with a slightly improved equidistribution, but the improvement was not overwhelming, e.g., the value of Δ was improved only for $n = 2^{15}$.

TABLE 8.1
Best combined generators with tempering

n	$P_j(z)$	ν_j	perm	MK	$\mathbf{b}_1$	$\mathbf{b}_2$	Δ (Σ)
			w.r.t. $\Delta_{13,13,13,13}$				
2^{15}	(3,1,0)	2	(2,1)	(2,4)	0000	0000	1 (59)
	(5,2,0)	2	(4,1)	(1,2)	4000	1000	
	(7,3,0)	2	(5,6)	(1,2)	5400	6a00	
			w.r.t. $\Delta_{40,40,30,20}$				
2^{15}	(3,1,0)	1	(2,1)	(1,2)	0000	0000	2 (377)
	(5,2,0)	1	(3,4)	(2,4)	4800	1800	
	(7,3,0)	2	(6,5)	(1,2)	0400	2200	

9. Application to Simulation Models. In this section, we present two simulation problems on which we compare XOR-shifted polynomial lattice rules, randomly shifted Korobov rules, XOR-shifted Sobol' point sets [39], and the MC method. These two problems were also considered in [24]. We used the rules selected via $\Delta_{13,13,13,13}$ for the first example, a 13-dimensional stochastic activity network problem, and those selected via $\Delta_{40,40,30,20}$ for the second example, an Asian option problem having 40 dimensions. Korobov rules are chosen with the corresponding M_{t_1,t_2,t_3,t_4} criterion. Explicit expressions of the functions f for these problems are given in [24]. We denote the polynomial lattice rules from Tables 7.1, 7.2, and 8.1 by "simp.", "comb." and "temp.", respectively, Korobov rules by "Kor.", and Sobol' point sets by "Sob.".

Before presenting numerical results, we emphasize that even if our polynomial lattice point sets minimize $\Delta_{t_1, \dots, t_d}$ within certain families of constructions, they are not necessarily those yielding the estimators with the smallest variance for a given problem. From Proposition 5.1, it is clear that the point set with the smallest variance would be the one minimizing the sum of squared Walsh coefficients over the dual lattice. Our selection criteria are only based on a *tentative anticipation* of this sum for various problems. Hence two point sets with similar values of $\Delta_{t_1,\ldots,t_d}$ and $\Sigma_{t_1,\ldots,t_d}$ could conceivably produce estimators with significantly different variances. Examples illustrating this fact are given below. Nevertheless, the proposed rules outperform the MC method in our examples.

9.1. A stochastic activity network. This problem is taken from [1]. A *stochastic activity network* is a directed acyclic graph with a source and a sink, and in which each edge represents an activity that has a certain random duration. The *completion time* T of the network is the length of the longest path from the source to the sink. For a given threshold t_0, the goal is to estimate $Pr(T \leq t_0)$ by simulation. The number of dimensions of the corresponding function f is equal to the number of activities having a non-trivial probability distribution. *Conditional Monte Carlo* (CMC) can be used to reduce both the variance and the dimension for this problem [1]. In the example below, it reduces the dimension from 13 to 8. Denote by MC the naive Monte Carlo simulation, by MCc the CMC simulation, by QMC the naive quasi-Monte Carlo simulation, and by QMCc the quasi-Monte Carlo simulation that uses CMC. The latter amounts to replacing the random numbers by a quasi-Monte Carlo point set P_n in the CMC simulation. In the results reported in Table 9.1, for each pair (n, t_0), we give the estimated variance ratios MC/QMC and MC/QMCc, and the estimated relative errors on these ratios, in percentage (in parentheses).

The variance of the randomized quasi-Monte Carlo estimators is estimated by performing 200 independent randomizations, and the variance of the MC estimator based on n independent points is also estimated from 200 independent copies. We then use 250 bootstrap samples to estimate the relative error on the variance ratios. (The relative error is the standard error divided by the mean.)

TABLE 9.1
Variance reduction w.r.t. MC for the stochastic activity network; $t = 13$ for QMC and $t = 8$ for QMCc

n		$t_0 = 30$		$t_0 = 75$		$t_0 = 90$	
		MC/QMC	MC/QMCc	MC/QMC	MC/QMCc	MC/QMC	MC/QMCc
2^{11}	comb.	2.7 (15%)	187 (13%)	8.0 (16%)	479 (15%)	4.4 (13%)	422 (16%)
	simp.	1.6 (14%)	35 (14%)	3.3 (14%)	42 (13%)	3.0 (14%)	30 (12%)
	Kor.	1.8 (13%)	133 (14%)	8.6 (16%)	337 (15%)	8.1 (15%)	256 (13%)
	Sob.	1.4 (14%)	394 (13%)	5.5 (16%)	350 (15%)	6.2 (14%)	294 (13%)
2^{13}	comb.	2.3 (15%)	492 (13%)	10 (14%)	899 (14%)	8.9 (14%)	702 (14%)
	simp.	1.3 (14%)	9.5 (13%)	2.1 (13%)	11 (11%)	2.0 (15%)	8.5 (15%)
	Kor.	2.3 (15%)	81 (13%)	11 (12%)	449 (12%)	10 (14%)	395 (15%)
	Sob.	1.8 (13%)	1490 (13%)	12 (15%)	1016 (13%)	9.3 (16%)	366 (13%)
2^{15}	comb.	3.8 (14%)	3449 (14%)	15 (14%)	723 (12%)	12 (14%)	208 (13%)
	simp.	2.3 (14%)	80 (15%)	1.6 (12%)	74 (12%)	2.5 (13%)	106 (12%)
	temp.	2.5 (14%)	1058 (14%)	11 (14%)	1163 (12%)	6.1 (13%)	910 (13%)
	Kor.	2.4 (15%)	145 (15%)	12 (15%)	415 (15%)	8.5 (13%)	622 (13%)
	Sob.	3.0 (15%)	3093 (14%)	15 (14%)	3069 (14%)	10 (14%)	1209 (15%)

The polynomial lattice rules reduce the variance by important factors, especially when CMC is used. Also, even if the combined generator of size $n = 2^{11}$ from Table 7.2 has the same value of $\Delta_{13,13,13,13}$ and $\Sigma_{13,13,13,13}$ as the corresponding simple generator from Table 7.1, the variance reduction factors MC/QMCc obtained by the latter are much smaller. This also holds for $n > 2^{11}$. We suspect that this poor performance of simple generators is related to the comparatively bad quality of their two-dimensional projections $P_n(\{j, j+1\})$, for $j \geq 1$. The ratio MC/MCc is approximatively 14 for $t_0 = 30$, and 4 for $t_0 = 75$ and 90. Thus for instance, when $t_0 = 90$ and $n = 2^{13}$, using QMCc-comb instead of MCc [instead of MC] reduces the variance (or the computing time required for a given precision) by a factor of approximately 178 [702]. In most cases the combined generators do better than Korobov rules, which do better than simple generators. The Sobol' estimator is sometimes much better than combined generators and sometimes worse.

Using the tempered point set for $n = 2^{15}$ improves significantly over the one from Table 7.2 when $t_0 = 90$ and CMC is used. However, it does worse (but is at least better than with $n = 2^{13}$ points) when $t_0 = 30$, which suggests that the projections for which the tempered point set is better are not as important in this case.

9.2. Asian call options. An Asian call option is a financial contract whose value depends on an *underlying asset*. If $S(\tau)$ denotes the value of the underlying asset at time τ, the *payoff* $C(T)$ of the call at expiration time T is $C(T) = \max\left(0, \sum_{i=1}^{t} S(\tau_i)/t - K\right)$, where K is a constant called the *strike price*, and $\tau_1, \ldots, \tau_t$ are t distinct times between 0 and T. Under the no-arbitrage assumption, the value of this contract at time 0 is $C(0) = E(e^{-rT} C(T))$, where r is the risk-free rate in the economy, and the expectation E is taken under the *risk-neutral measure* [17]. Even if we use a simple model such as Black-Scholes for the price process $S(\tau)$, no analytical formula is known for $C(0)$, and therefore one must rely on simulation or numerical approximations. Other variance reduction techniques not considered here can be used on this problem (see e.g., [24] and the references therein). The dimension of the integral here is t.

Table 9.2 gives the estimated variance reduction factors of the randomized QMC estimators with respect to MC. These quantities were obtained similarly as for the previous example. The parameters of the option are $T = 120$ days, the average is taken over the last 40 days of the contract, $r = \ln 1.09$, the volatility parameter σ of the Black-Scholes model is set to 0.2, and $S(0) = 100$. The dimension is thus $t = 40$.

TABLE 9.2
Variance reduction w.r.t. MC for the Asian options problem, $t = 40$

		$K = 90$	$K = 100$	$K = 110$
$n = 2^{11}$	comb.	510 (15%)	220 (14%)	45 (11%)
	simp.	25 (12%)	13 (14%)	4.2 (11%)
	Kor.	177 (14%)	82 (13%)	25 (13%)
	Sob.	313 (14%)	209 (13%)	53 (14%)
$n = 2^{13}$	comb.	1500 (15%)	465 (13%)	124 (13%)
	simp.	33 (13%)	16 (13%)	5.4 (14%)
	Kor.	282 (14%)	105 (14%)	32 (14%)
	Sob.	906 (15%)	377 (15%)	158 (14%)
$n = 2^{15}$	comb.	302 (11%)	357 (12%)	13 (10%)
	simp.	28 (11%)	14 (11%)	4.7 (10%)
	temp.	1399 (11%)	506 (12%)	159 (14%)
	Kor.	605 (13%)	158 (12%)	55 (12%)
	Sob.	1741 (12%)	524 (13%)	243 (14%)

We see in Table 9.2 that polynomial lattice rules based on combined generators provide much more accurate estimators than MC in all cases, with variance reduction factors ranging between approximately 13 and 1500, depending mostly on the value of K. Note that when K is much larger than $S(0)$, the function f is zero over most of the unit hypercube, so its integral $C(0)$ is hard to estimate by simulation. The simple generators from Table 7.1 give significantly smaller variance reduction factors than the combined ones, but still improve over MC. Korobov rules always do better than simple generators and MC, but are often not as good as combined generators. The Sobol' estimator is clearly the best for $n = 2^{15}$, but otherwise it is comparable or worse than the combined generators. Using the tempered point set for $n = 2^{15}$ instead of the one given in Table 7.2 seems to make a difference. For example, when $K = 90$, the latter gave a reduction factor of about 300, which is even *less* than the variance reduction factor for $n = 2^{11}$. With tempering, we are able to do almost as well as Sobol's point sets.

10. Conclusion. We have developed a general theory of polynomial lattice rules analogous to the theory already available for ordinary lattice rules. Preliminary examples show that the polynomial rules can be competitive with other types of rules. Among the advantages of the proposed rules: Their point sets are easy to generate via linear recurrences, they are projection-regular and dimension-stationary under mild conditions, their equidistribution properties are easy to assess, and explicit variance expressions are available for simple randomizations. The choice of selection criteria for their parameters would require further study and experimentation on different classes of real-life problems.

REFERENCES

[1] A. N. Avramidis and J. R. Wilson, *Integrated variance reduction strategies for simulation*, Operations Research, 44 (1996), pp. 327–346.

[2] K. G. Beauchamp, *Applications of Walsh and related Functions*, Academic Press, London, 1984.

[3] R. E. Caflisch, W. Morokoff, and A. B. Owen, *Valuation of mortgage-backed securities using Brownian bridges to reduce effective dimension*, The Journal of Computational Finance, 1 (1997), pp. 27–46.

[4] J. Cassels, *An Introduction to the Geometry of Numbers*, Classics in Mathematics, Springer-Verlag, Berlin, 1997. Corrected reprint of the 1971 edition.

[5] R. Couture and P. L'Ecuyer, *Lattice computations for random numbers*, Mathematics of Computation, 69 (2000), pp. 757–765.

[6] R. Couture, P. L'Ecuyer, and S. Tezuka, *On the distribution of k-dimensional vectors for simple and combined Tausworthe sequences*, Mathematics of Computation, 60 (1993), pp. 749–761, S11–S16.

[7] R. Cranley and T. N. L. Patterson, *Randomization of number theoretic methods for multiple integration*, SIAM Journal on Numerical Analysis, 13 (1976), pp. 904–914.

[8] P. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, New York, second ed., 1984.

[9] B. Efron and C. Stein, *The jackknife estimator of variance*, Annals of Statistics, 9 (1981), pp. 586–596.

[10] K. Entacher, P. Hellekalek, and P. L'Ecuyer, *Quasi-Monte Carlo node sets from linear congruential generators*, in Monte Carlo and Quasi-Monte Carlo Methods 1998, H. Niederreiter and J. Spanier, eds., Berlin, 2000, Springer, pp. 188–198.

[11] B. L. Fox, *Strategies for Quasi-Monte Carlo*, Kluwer Academic, Boston, MA, 1999.

[12] M. Fushimi and S. Tezuka, *The k-distribution of generalized feedback shift register pseudorandom numbers*, Communications of the ACM, 26 (1983), pp. 516–523.

[13] B. Golubov, A. Efimov, and V. Skvortsov, *Walsh Series and Transforms: Theory and Applications*, vol. 64 of Mathematics and Applications: Soviet Series, Kluwer Academic Publishers, Boston, 1991.

[14] P. Hellekalek, *On the assessment of random and quasi-random point sets*, in Random and Quasi-Random Point Sets, P. Hellekalek and G. Larcher, eds., vol. 138 of Lecture Notes in Statistics, Springer, New York, 1998, pp. 49–108.

[15] P. Hellekalek and H. Leeb, *Dyadic diaphony*, Acta Arithmetica, 80 (1997), pp. 187–196.

[16] F. J. Hickernell, *A generalized discrepancy and quadrature error bound*, Mathematics of Computation, 67 (1998), pp. 299–322.

[17] J. Hull, *Options, Futures, and Other Derivative Securities*, Prentice-Hall, Englewood-Cliff, N.J., fourth ed., 2000.

[18] D. E. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison-Wesley, Reading, Mass., third ed., 1998.

[19] N. M. Korobov, *The approximate computation of multiple integrals*, Dokl. Akad. Nauk SSSR, 124 (1959), pp. 1207–1210. In Russian.

[20] G. Larcher, *Digital point sets: Analysis and applications*, in Random and Quasi-Random Point Sets, P. Hellekalek and G. Larcher, eds., vol. 138 of Lecture Notes in Statistics, Springer, New York, 1998, pp. 167–222.

[21] G. Larcher, H. Niederreiter, and W. C. Schmid, *Digital nets and sequences constructed over finite rings and their application to quasi-Monte Carlo integration*, Monatshefte für Mathematik, 121 (1996), pp. 231–253.

[22] G. Larcher and C. Traunfellner, *The numerical integration of Walsh series*, Mathematics of Computation, 63 (1994), pp. 277–291.

[23] P. L'Ecuyer, *Maximally equidistributed combined Tausworthe generators*, Mathematics of Computation, 65 (1996), pp. 203–213.

[24] P. L'Ecuyer and C. Lemieux, *Variance reduction via lattice rules*, Management Science, 46 (2000), pp. 1214–1235.

[25] ———, *Recent advances in randomized quasi-Monte Carlo methods*, in Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications, M. Dror, P. L'Ecuyer, and F. Szidarovszki, eds., Kluwer Academic Publishers, Boston, 2002, pp. 419–474.

[26] P. L'Ecuyer and F. Panneton, *A new class of linear feedback shift register generators*, in Proceedings of the 2000 Winter Simulation Conference, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds., Pistacaway, NJ, Dec 2000, IEEE Press, pp. 690–696.

[27] C. Lemieux and P. L'Ecuyer, *Selection criteria for lattice rules and other low-discrepancy*

point sets, Mathematics and Computers in Simulation, 55 (2001), pp. 139–148.

[28] M. Matsumoto and Y. Kurita, *Twisted GFSR generators II*, ACM Transactions on Modeling and Computer Simulation, 4 (1994), pp. 254–266.

[29] M. Matsumoto and T. Nishimura, *Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator*, ACM Transactions on Modeling and Computer Simulation, 8 (1998), pp. 3–30.

[30] W. J. Morokoff, *Generating quasi-random paths for stochastic processes*, SIAM Review, 40 (1998), pp. 765–788.

[31] H. Niederreiter, *Point sets and sequences with small discrepancy*, Monatshefte für Mathematik, 104 (1987), pp. 273–337.

[32] ———, *Random Number Generation and Quasi-Monte Carlo Methods*, vol. 63 of SIAM CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[33] H. Niederreiter and G. Pirsic, *Duality for digital nets and its applications*, Acta Arithmetica, 97 (2001), pp. 173–182.

[34] H. Niederreiter and C. Xing, *Nets, (t, s)-sequences, and algebraic geometry*, in Random and Quasi-Random Point Sets, P. Hellekalek and G. Larcher, eds., vol. 138 of Lecture Notes in Statistics, Springer, New York, 1998, pp. 267–302.

[35] A. B. Owen, *Latin supercube sampling for very high-dimensional simulations*, ACM Transactions of Modeling and Computer Simulation, 8 (1998), pp. 71–102.

[36] F. Panneton, *Générateurs pseudo-aléatoires utilisant des récurrences linéaires modulo 2*, master's thesis, Université de Montréal, 2000.

[37] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, second ed., 1974.

[38] I. H. Sloan and S. Joe, *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford, 1994.

[39] I. M. Sobol', *The distribution of points in a cube and the approximate evaluation of integrals*, U.S.S.R. Comput. Math. and Math. Phys., 7 (1967), pp. 86–112.

[40] J. Spanier and E. H. Maize, *Quasi-random methods for estimating integrals using relatively small samples*, SIAM Review, 36 (1994), pp. 18–44.

[41] S. Tezuka, *Lattice structure of pseudorandom sequences from shift-register generators*, in Proceedings of the 1990 Winter Simulation Conference, IEEE Press, 1990, pp. 266–269.

[42] ———, *A new family of low-discrepancy point sets*, Tech. Report RT-0031, IBM Research, Tokyo Research Laboratory, Jan. 1990.

[43] ———, *Uniform Random Numbers: Theory and Practice*, Kluwer Academic Publishers, Norwell, Mass., 1995.

[44] S. Tezuka and P. L'Ecuyer, *Efficient and portable combined Tausworthe random number generators*, ACM Trans. on Modeling and Computer Simulation, 1 (1991), pp. 99–112.

[45] J. P. R. Tootill, W. D. Robinson, and D. J. Eagle, *An asymptotically random Tausworthe sequence*, Journal of the ACM, 20 (1973), pp. 469–481.

[46] D. Wang and A. Compagner, *On the use of reducible polynomials as random number generators*, Mathematics of Computation, 60 (1993), pp. 363–374.

Appendix A. Lemmas and proofs.

This appendix contains the proofs that are missing in the core of the paper, as well as technical lemmas needed in these proofs.

LEMMA A.1. *The determinant of the matrix* $\mathbf{V}$ *whose lines are a set of basis vectors for* $\mathcal{L}_t$ *does not depend on the choice of basis.*

Proof. For any two bases $\mathbf{V}$ and $\mathbf{W}$ for $\mathcal{L}_t$, there is a matrix $\mathbf{D}$ with all entries in $\mathbb{F}_2[z]$ and with determinant 1, such that $\mathbf{V} = \mathbf{DW}$ (see [4, section I.2]). To see this, observe that there are polynomials $v_{ij}(z)$ and $w_{ij}(z)$ such that $\mathbf{w}_i(z) = \sum_j w_{ij}(z)\mathbf{v}_j(z)$ and $\mathbf{v}_i(z) = \sum_j v_{ij}(z)\mathbf{w}_j(z)$. Thus $\mathbf{w}_i(z) = \sum_j w_{ij}(z)(\sum_l v_{jl}(z)\mathbf{w}_l(z))$ and by the linear independence of the $\mathbf{w}_j(z)$, we must have

$$\sum_j w_{ij}(z)v_{jl}(z) = \begin{cases} 1 & \text{if } i = l, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the two matrices with elements $v_{ij}(z)$ and $w_{ij}(z)$ are matrices of polynomials whose product is the identity. Hence, these two matrices have determinant 1, which implies that $\det(\mathbf{V}) = \det(\mathbf{W})$. □

LEMMA A.2. *Let* $\mathcal{L}_t$ *be t-dimensional a polynomial integration lattice. Then one can construct a basis* $\{\mathbf{v}_1(z), \ldots, \mathbf{v}_t(z)\}$ *for* $\mathcal{L}_t$ *such that*

$$\begin{aligned} \mathbf{e}_1 &= h_{11}(z)\mathbf{v}_1(z) \\ \mathbf{e}_2 &= h_{21}(z)\mathbf{v}_1(z) + h_{22}(z)\mathbf{v}_2(z) \\ &\ldots \\ \mathbf{e}_t &= h_{t1}(z)\mathbf{v}_1(z) + \ldots + h_{tt}(z)\mathbf{v}_t(z), \end{aligned} \tag{A.1}$$

where $\mathbf{e}_j$ *is the* j*th unit vector in* t *dimensions, and the* $h_{ij}(z)$ *are in* $\mathbb{F}_2[z]$, *with* $h_{jj}(z) \neq 0$ *for* $j = 1, \ldots, t$.

Proof. The proof follows by adapting Part B of Theorem I in [4, section I.2.2] in the following way. Let $P(z)$ denote $(\det(\mathcal{L}_t))^{-1} = \det(\mathcal{L}_t^*)$, as in Definition 3.3. First, notice that since $(\mathbb{F}_2[z])^t \subseteq \mathcal{L}_t$, any basis $\mathbf{V}$ for $\mathcal{L}_t$ must be such that $\mathbf{V}^{-1}$ has entries that are all in $\mathbb{F}_2[z]$. But $\mathbf{V} = (\det \mathbf{V})(\text{adj}(\mathbf{V}^{-1})) = (\text{adj}(\mathbf{V}^{-1}))/P(z)$, where all entries in $\text{adj}(\mathbf{V}^{-1})$ (the adjunct of $\mathbf{V}^{-1}$) are in $\mathbb{F}_2[z]$, which implies that $P(z)\mathbf{V}$ also has all its entries in $\mathbb{F}_2[z]$, and therefore $P(z)\mathcal{L}_t \subseteq (\mathbb{F}_2[z])^t$.

For each $j = 1, \ldots, t$, there is a non-zero polynomial $w_{jj}(z)$ of smallest degree such that for some other polynomials $w_{j1}(z), \ldots, w_{j,j-1}(z)$, the vector

$$\tilde{\mathbf{v}}_j(z) = w_{j1}(z)\mathbf{e}_1 + \ldots + w_{jj}(z)\mathbf{e}_j \tag{A.2}$$

is in $P(z)\mathcal{L}_t$, a sublattice of $(\mathbb{F}_2[z])^t$. Without loss of generality, we can assume that for each j, $w_{jl}(z) \in \mathbb{F}_2[z]/(P)$ or $w_{jl}(z) = P(z)$, for $l = 1, \ldots, j$.

We now show that $\tilde{\mathbf{v}}_1(z), \ldots, \tilde{\mathbf{v}}_t(z)$ form a basis for $P(z)\mathcal{L}_t$ by adapting the proof of [4, Theorem I]. These vectors are clearly linearly independent. Because $\tilde{\mathbf{v}}_j(z) \in P(z)\mathcal{L}_t$, by construction, so is every linear combination

$$y_1(z)\tilde{\mathbf{v}}_1(z) + \ldots + y_t(z)\tilde{\mathbf{v}}_t(z), \tag{A.3}$$

where $y_1(z), \ldots, y_t(z)$ are in $\mathbb{F}_2[z]$. Suppose that there exists a vector $\mathbf{c}(z)$ in $P(z)\mathcal{L}_t$ not of the form (A.3). This $\mathbf{c}(z)$ is in $(\mathbb{F}_2[z])^t$, so it can be expressed as a linear combination of the canonical basis vectors of $(\mathbb{F}_2[z])^t$:

$$\mathbf{c}(z) = \tau_1(z)\mathbf{e}_1 + \ldots + \tau_s(z)\mathbf{e}_s,$$

where $1 \le s \le t$, $\tau_s(z) \neq 0$, and $\tau_1(z), \dots, \tau_s(z) \in \mathbb{F}_2[z]$. If there are several such $\mathbf{c}(z)$, we take one for which the integer s is as small as possible. Because $w_{ss}(z) \neq 0$, we can choose a polynomial $u(z)$ such that

$$\deg(\tau_s(z) - u(z)w_{ss}(z)) < \deg(w_{ss}(z)) \tag{A.4}$$

(just use the $u(z)$ that reduces $\tau_s(z)$ modulo $w_{ss}(z)$). The vector of polynomials

$$\mathbf{c}(z) - u(z)\tilde{\mathbf{v}}_s(z) = (\tau_1(z) - u(z)w_{s1}(z))\mathbf{e}_1 + \ldots + (\tau_s(z) - u(z)w_{ss}(z))\mathbf{e}_s$$

is in $P(z)\mathcal{L}_t$, because $\mathbf{c}(z)$ and $\tilde{\mathbf{v}}_s(z)$ are, but it is not of the form (A.3) because $\mathbf{c}(z)$ is not. Hence, $\tau_s(z) - u(z)w_{ss}(z) \neq 0$ by the assumption that s was chosen as small as possible. But then, (A.4) contradicts the assumption that the non-zero polynomial $w_{ss}(z)$ was chosen with the smallest possible degree in (A.2). This contradiction shows that every $\mathbf{c}(z) \in P(z)\mathcal{L}_t$ can be expressed in the form (A.3), which implies that $\tilde{\mathbf{v}}_1(z), \dots, \tilde{\mathbf{v}}_t(z)$ form a triangular basis for $P(z)\mathcal{L}_t$, so the vectors $\mathbf{v}_j(z) = \tilde{\mathbf{v}}_j(z)/P(z)$, $j = 1, \dots, t$, are a triangular basis for $\mathcal{L}_t$.

The last thing we need to show is that the basis $\mathbf{v}_1(z), \dots, \mathbf{v}_t(z)$ satisfies (A.1) with each $h_{ij}(z)$ in $\mathbb{F}_2[z]$ and $h_{jj}(z) \neq 0$ for $j = 1, \dots, t$. Let $\mathbf{W}$ be the lower triangular matrix whose jth row is $\tilde{\mathbf{v}}_j(z)$, and let $w_{ij}^{-1}(z)$ denote the element in position (i,j) of $\mathbf{W}^{-1}$. We have that

$$\begin{aligned}
\mathbf{e}_1 &= P(z)w_{11}^{-1}(z)\mathbf{v}_1(z)\\
\mathbf{e}_2 &= P(z)w_{21}^{-1}(z)\mathbf{v}_1(z) + P(z)w_{22}^{-1}(z)\mathbf{v}_2(z)\\
&\cdots\\
\mathbf{e}_t &= P(z)w_{t1}^{-1}(z)\mathbf{v}_1(z) + \ldots + P(z)w_{tt}^{-1}(z)\mathbf{v}_t(z),
\end{aligned}$$

and therefore (A.1) holds with $h_{ij}(z) = P(z)w_{ij}^{-1}(z)$, which is necessarily in $\mathbb{F}_2[z]$, because $\mathbf{v}_1(z), \dots, \mathbf{v}_t(z)$ is a basis for $\mathcal{L}_t$, the vectors $\mathbf{e}_j$ are in $\mathcal{L}_t$, and the representation

$$\mathbf{e}_j = g_1(z)\mathbf{v}_1(z) + \ldots + g_t(z)\mathbf{v}_t(z)$$

is unique by the independence of the $\mathbf{v}_j(z)$. Also, $\det(\mathbf{W}^{-1}) \neq 0$ implies that $w_{jj}^{-1}(z) \neq 0$ and hence $h_{jj}(z) \neq 0$ for each j. □

LEMMA A.3. *Let $\mathcal{L}_t$ be a polynomial integration lattice. Then $(\det \mathcal{L}_t)^{-1} = \det(\mathcal{L}_t^*)$ is in $\mathbb{F}_2[z]$.*

Proof. This follows from the proof of Lemma A.2, in which we saw that any basis for $\mathcal{L}_t^*$ is formed by vectors whose components are in $\mathbb{F}_2[z]$, and therefore $\det(\mathcal{L}_t^*) \in \mathbb{F}_2[z]$. □

PROOF OF PROPOSITION 3.5. From the proof of Lemma A.2, we know that if $P(z) = \det(\mathcal{L}_t^*)$, we can find a basis $\mathbf{v}_1(z), \dots, \mathbf{v}_t(z)$ such that for each $j = 1, \dots, t$,

$$\mathbf{v}_j(z) = (w_{j1}(z), \dots, w_{jj}(z), 0, \dots, 0)\,/P(z),$$

where the $w_{jj}(z)$ are non-zero polynomials and $w_{jl}(z) \in \mathbb{F}_2[z]/(P)$ or $w_{jl}(z) = P(z)$. □

LEMMA A.4. *Let $\mathcal{L}_t$ be a polynomial integration lattice with basis $\mathbf{v}_1(z), \dots, \mathbf{v}_t(z)$ of the form (A.1), with the polynomials $h_{jj}(z)$ also taken from this representation, for $j = 1, \dots, t$. Then, every $\mathbf{c}(z)$ in $\Xi_t = \mathcal{L}_t \cap \mathbb{L}_0^t$ is equal, modulo $\mathbb{F}_2[z]$, to exactly one vector in the set*

$$\{q_1(z)\mathbf{v}_1(z) + \ldots + q_t(z)\mathbf{v}_t(z) : q_j(z) \in \mathbb{F}_2[z]/(h_{jj}),\ j = 1, \dots, t\}.$$

Proof. Let $\mathbf{c}(z) \in \Xi_t$. Since $\mathbf{v}_1(z), \ldots, \mathbf{v}_t(z)$ is a basis of $\mathcal{L}_t$, we know that there are *unique* polynomials $w_1(z), \ldots, w_t(z)$ such that $\mathbf{c}(z) = \sum_{l=1}^t w_l(z)\mathbf{v}_l(z)$. We want to show that there exists a unique vector $(q_1(z), \ldots, q_t(z))$ such that $q_j(z) \in \mathbb{F}_2[z]/(h_{jj})$ for each j and $(\sum_{l=1}^t q_l(z)\mathbf{v}_l(z)) - \mathbf{c}(z) \in (\mathbb{F}_2[z])^t$. By using the representation (A.1) for the vectors $\mathbf{e}_1, \ldots, \mathbf{e}_t$ (which form a basis for $(\mathbb{F}_2[z])^t$), we have that $\sum_{l=1}^t (q_l(z) - w_l(z))\mathbf{v}_l(z)$ is in $(\mathbb{F}_2[z])^t$ if and only if there exist polynomials $g_1(z), \ldots, g_t(z)$ such that

$$\begin{aligned} &g_1(z)h_{11}(z)\mathbf{v}_1(z) + g_2(z)(h_{21}(z)\mathbf{v}_1(z) + h_{22}(z)\mathbf{v}_2(z)) \\ &\quad + \ldots + g_t(z)(h_{t1}(z)\mathbf{v}_1(z) + \ldots + h_{tt}(z)\mathbf{v}_t(z)) = \sum_{l=1}^t (q_l(z) - w_l(z))\mathbf{v}_l(z). \end{aligned} \tag{A.5}$$

Now, for (A.5) to hold, we must have

$$g_t(z)h_{tt}(z) = q_t(z) - w_t(z), \tag{A.6}$$

or equivalently, $q_t(z) = g_t(z)h_{tt}(z) + w_t(z)$, by the linear independence of the $\mathbf{v}_j(z)$. This can be achieved by taking $g_t(z) = \lfloor w_t(z)/h_{tt}(z) \rfloor$, where $\lfloor w_t(z)/h_{tt}(z) \rfloor$ is the polynomial part of $w_t(z)/h_{tt}(z)$. Note that $w_t(z)/h_{tt}(z)$ is defined because $h_{tt}(z) \neq 0$. Then, we obtain that $q_t(z) = w_t(z) \bmod h_{tt}(z)$, which is in $\mathbb{F}_2[z]/(h_{tt})$. It is easy to see that $q_t(z)$ is the unique element in $\mathbb{F}_2[z]/(h_{tt})$ satisfying (A.6). The same procedure can be applied to find the other values of $q_j(z)$ and $g_j(z)$ allowing (A.5) to hold: for $j = t-1, t-2, \ldots, 1$, simply set

$$a_j(z) = w_j(z) + \sum_{l=j+1}^{t} g_l(z)h_{lj}(z)$$

and $g_j(z) = \lfloor a_j(z)/h_{jj}(z) \rfloor$. Hence $q_j(z) = a_j(z) \bmod h_{jj}(z)$ is the unique element in $\mathbb{F}_2[z]/(h_{jj})$ satisfying $q_j(z) = a_j(z) + g_j(z)h_{jj}(z)$, which is equivalent to having equal coefficients for $\mathbf{v}_j(z)$ on both sides of (A.5). □

The following lemma states a result similar to the one for ordinary lattice rules to the effect that any unit hypercube in t dimensions that has its sides aligned along the axes of $\mathbb{R}^t$ always contains n points of the integration lattice L_t, where $n = \det(L_t^*)$.

LEMMA A.5. *Let $\mathcal{L}_t$ be a polynomial integration lattice and $\mathbf{y}(z)$ be a vector in $\mathbb{L}^t$. Let $H_y = \{\mathbf{w}(z) + \mathbf{y}(z) : \mathbf{w}(z) \in \mathbb{L}_0^t\}$, n be the cardinality of $H_y \cap \mathcal{L}_t$, and k be the degree of* $\det(\mathcal{L}_t^*)$. *Then $n = 2^k$.*

Proof. The proof proceeds by adapting arguments used in [4] for lattices in $\mathbb{R}^t$. Following [4, section I.2.2], we say that $\mathbf{c}(z), \mathbf{d}(z)$ in $\mathcal{L}_t$ *are in the same class with respect to* $(\mathbb{F}_2[z])^t$ if $\mathbf{c}(z) - \mathbf{d}(z) \in (\mathbb{F}_2[z])^t$. We then show, as a first step toward our result, that n_c, the number of different classes in $\mathcal{L}_t$, is equal to 2^k by using an adaptation of the proof of Lemma 1 [4, section I.2.2].

Let $P(z) = \det(\mathcal{L}_t^*)$ and let $\{\mathbf{v}_j(z), j = 1, \ldots, t\}$ be a basis for $\mathcal{L}_t$, in the shape (A.1). By our definition of $P(z)$ and (A.1), we must have

$$P(z) = h_{11}(z) \cdots h_{tt}(z).$$

Then, if we denote by d_j the degree of $h_{jj}(z)$, we must have $d_1 + \cdots + d_t = k$.

Now, by Lemma A.4, we know that every $\mathbf{c}(z) \in \mathcal{L}_t \cap \mathbb{L}_0^t$ is in the same class as precisely one of the vectors in the set

$$\{q_1(z)\mathbf{v}_1(z) + \ldots + q_t(z)\mathbf{v}_t(z) : q_j(z) \in \mathbb{F}_2[z]/(h_{jj}), j = 1, \ldots, t\}. \tag{A.7}$$

Note that any element in $\mathcal{L}_t$ is in the same class as one of the $2^{d_1} \cdots 2^{d_t} = 2^k$ elements in (A.7), and therefore $n_c \leq 2^k$. Also, using the fact that $(\mathbb{F}_2[z])^t \subseteq \mathcal{L}_t$, it is easy to show that each element in (A.7) is in the same class as one of the elements in Ξ_t. Therefore, we cannot have $n_c < 2^k$ since it would mean that two different points in Ξ_t would be in the same class, which is impossible since they are both in $\mathbb{L}_0^t$. We now show that $n_c = n$. Let $\mathbf{c}(z), \mathbf{d}(z) \in \mathcal{L}_t$ be in the same class. This implies that $\mathbf{c}(z) - \mathbf{d}(z) \in (\mathbb{F}_2[z])^t$ and thus, $\mathbf{c}(z)$ and $\mathbf{d}(z)$ cannot be both in H_y. This means that $n \leq n_c$. Now assume $n < n_c$. For this to hold, there must exist a class C_1 such that $C_1 \cap H_y = \phi$. However there must be an element $\mathbf{x}_1(z) \in (\mathbb{F}_2[z])^t$ such that $C_1 \cap \{H_y + \mathbf{x}_1(z)\} \neq \phi$, for otherwise C_1 would be empty. Hence we can assume that there exists $\mathbf{w}_1(z) \in C_1 \cap \{H_y + \mathbf{x}_1(z)\}$ with $\mathbf{w}_1(z) \in (\mathbb{F}_2[z])^t$. This implies that $\mathbf{w}_1(z) - \mathbf{x}_1(z) \in (\mathbb{F}_2[z])^t$, since they are both in $(\mathbb{F}_2[z])^t$. In addition, $\mathbf{w}_1(z) - \mathbf{x}_1(z) \in H_y$ and is in the same class as $\mathbf{w}_1(z)$, which implies that $\mathbf{w}_1(z) - \mathbf{x}_1(z) \in H_y \cap C_1$. This contradicts the statement that $C_1 \cap H_y$ is empty, which means we must have $n = n_c$. □

Proof of Proposition 3.15. The proof uses an adaptation of [38, Theorem 3.25]. We prove the result for $m = 2$. The proof for $m > 2$ follows easily by induction on m. Let $P_{n_l} = \{\mathbf{u}_{l,0} = \mathbf{0}, \mathbf{u}_{l,1}, \ldots, \mathbf{u}_{l,n_l-1}\}$, for $l = 1, 2$. $P_{n_1} \oplus P_{n_2}$ contains $n_1 n_2$ distinct points if and only if, for $0 \leq i_1, i_2 < n_1$ and $0 \leq j_1, j_2 < n_2$, the equality

$$\mathbf{u}_{1,i_1} + \mathbf{u}_{2,j_1} = \mathbf{u}_{1,i_2} + \mathbf{u}_{2,j_2}$$

holds if and only if $i_1 = i_2$ and $j_1 = j_2$. This means that

$$\mathbf{u}_{1,i_1} + \mathbf{u}_{1,i_2} = \mathbf{u}_{2,j_1} + \mathbf{u}_{2,j_2} \tag{A.8}$$

holds if and only if $i_1 = i_2$ and $j_1 = j_2$, and thus no other point than $\mathbf{0}$ can be in both point sets, for otherwise (A.8) would hold with $i_1 = j_1 = 0$ and i_2, j_2 each taken equal to the index of this common point, which in both cases would be different from 0. To prove the inverse implication, if $\mathbf{0}$ is the only common point in P_{n_1} and P_{n_2}, then (A.8) cannot hold with $i_1 \neq i_2$ or $j_1 \neq j_2$ unless both sides of the equality equal $\mathbf{0}$, but this cannot happen because each point is its own unique additive inverse. □

Proof of Proposition 3.16. The proof slightly generalizes the one used in [44, Theorem 1]. We will show that $\mathcal{L}_t^1 \oplus \ldots \oplus \mathcal{L}_t^m = \mathcal{M}_t$, where $\mathcal{M}_t$ is the lattice whose basis vectors are $(1, v_2(z), \ldots, v_t(z))/P(z), \mathbf{e}_2, \ldots, \mathbf{e}_t$, with $P(z) = \prod_{j=1}^m P_j(z)$, $v_j(z) = \sum_{l=1}^m v_j^l(z)\eta_l(z)P_{-l}(z) \bmod P(z)$, for $j = 2, \ldots, t$, $P_{-l}(z) = \prod_{j=1, j\neq l}^m P_j(z)$, and $\eta_l(z)$ such that

$$\eta_l(z)P_{-l}(z) \bmod P_l(z) = 1 \tag{A.9}$$

for $l = 1, \ldots, m$. The fact that the $P_l(z)$ are pairwise relatively prime guarantees that a unique $\eta_l(z)$ in $\mathbb{F}_2[z]/(P_l)$ satisfies (A.9), i.e., that $P_{-l}(z)$ has a unique inverse modulo $P_l(z)$.

Firstly, it is easy to show that $\mathcal{M}_t$ contains $(\mathbb{F}_2[z])^t$, because all the coordinates of the dual basis vectors are in $\mathbb{F}_2[z]$. To prove that $\mathcal{M}_t = \mathcal{L}_t^1 \oplus \ldots \oplus \mathcal{L}_t^m$, it suffices to show that $\mathcal{M}_t \cap \mathbb{L}_0^t = (\mathcal{L}_t^1 \oplus \ldots \oplus \mathcal{L}_t^m) \cap \mathbb{L}_0^t$ because both $\mathcal{M}_t$ and $\mathcal{L}_t^1 \oplus \ldots \oplus \mathcal{L}_t^m$ contain $(\mathbb{F}_2[z])^t$.

Let $\mathbf{x}(z) = \sum_{l=1}^m \mathbf{x}_l(z)$ where $\mathbf{x}_l(z) \in \mathcal{L}_t^l \cap \mathbb{L}_0^t$ for $l = 1, \ldots, m$. We want to show that $\mathbf{x}(z) \in \mathcal{M}_t \cap \mathbb{L}_0^t$. From our hypothesis on $\mathcal{L}_t^l$, there exists a polynomial $a_l(z)$ such that

$$\mathbf{x}_l(z) = \frac{1}{P_l(z)} \left(a_l(z)(1, v_2^l(z), \ldots, v_t^l(z)) \bmod P_l(z) \right).$$

Now,

$$\begin{aligned}\mathbf{x}(z) &= \sum_{l=1}^{m} \frac{1}{P_l(z)} \left(a_l(z)(1, v_2^l(z), \ldots, v_t^l(z)) \bmod P_l(z) \right) \\ &= \sum_{l=1}^{m} \frac{1}{P(z)} \left(P_{-l}(z) a_l(z)(1, v_2^l(z), \ldots, v_t^l(z)) \bmod P(z) \right). \qquad \text{(A.10)}\end{aligned}$$

Let $a(z) = (\sum_{l=1}^{m} P_{-l}(z) a_l(z)) \bmod P(z)$. Showing that (A.10) equals

$$\frac{1}{P(z)} \left(a(z)(1, v_2(z), \ldots, v_t(z)) \bmod P(z) \right), \qquad \text{(A.11)}$$

is sufficient to prove that $\mathbf{x}(z) \in \mathcal{M}_t \cap \mathbb{L}_0^t$ because it implies that $\mathbf{x}(z)$ can be obtained by multiplying the first vector of the basis by $a(z)$, and using the other basis vectors to perform the "mod $P(z)$" operation in (A.11). Now, for each $j = 2, \ldots, t$, we have, for some polynomial $k(z)$,

$$\begin{aligned}a(z)v_j(z) \bmod P(z) &= \left(\sum_{l=1}^{m} P_{-l}(z) a_l(z) \right) \left(\sum_{l=1}^{m} v_j^l(z) \eta_l(z) P_{-l}(z) \right) \bmod P(z) \\ &= \sum_{l=1}^{m} P_{-l}^2(z) a_l(z) v_j^l(z) \eta_l(z) \bmod P(z) \\ &= \sum_{l=1}^{m} P_{-l}(z) a_l(z) v_j^l(z)(k(z)P_l(z) + 1) \bmod P(z) \\ &= \sum_{l=1}^{m} P_{-l}(z) a_l(z) v_j^l(z) \bmod P(z), \qquad \text{(A.12)}\end{aligned}$$

where the second equality comes from the fact that for $l \neq k$, $P_{-l}(z)P_{-k}(z) \bmod P(z) = 0$, the third one is obtained by the definition of $\eta_l(z)$, and the last one follows from the fact that $P_{-l}(z)P_l(z) = P(z)$. This shows that (A.10) equals (A.11).

Now take an arbitrary $\mathbf{x}(z) \in \mathcal{M}_t \cap \mathbb{L}_0^t$, of the form

$$\frac{1}{P(z)} \left(a(z)(1, v_2(z), \ldots, v_t(z)) \bmod P(z) \right).$$

If we define $a_l(z) = a(z) \bmod P_l(z) \in \mathbb{F}_2[z]/(P_l)$ and $\mathbf{x}_l(z) = (a_l(z)/P_l(z))(1, v_2^l(z), \ldots, v_t^l(z)) \bmod \mathbb{F}_2[z]$, we obtain that $\mathbf{x}(z) = \sum_{l=1}^{m} \mathbf{x}_l(z)$, where $\mathbf{x}_l(z) \in \mathcal{L}_t^l \cap \mathbb{L}_0^t$ for $l = 1, \ldots, m$. □

Proof of Proposition 6.1. Let $g(\mathbf{u}) = \tilde{\psi}(\mathbf{u})$. If we expand $g(\mathbf{u})$ as a Walsh series, we get $g(\mathbf{u}) = \sum_{\mathbf{h}} \tilde{g}(\mathbf{h})(-1)^{\langle \mathbf{h}, \mathbf{u} \rangle}$ where, for $\mathbf{h} \neq \mathbf{0}$, we have

$$\begin{aligned}\tilde{g}(\mathbf{h}) &= \int_{[0,1)^t} g(\mathbf{u})(-1)^{\langle \mathbf{h}, \mathbf{u} \rangle} d\mathbf{u} \\ &= \int_{[0,1)^t} \left(-1 + \prod_{j=1}^{t} (1 + 2\beta_j^2(1 - 3 \cdot 2^{\lfloor \log_2 u_j \rfloor})) \right) (-1)^{\langle \mathbf{h}, \mathbf{u} \rangle} d\mathbf{u} \\ &= \prod_{j=1}^{t} \left(\int_0^1 \left(1 + 2\beta_j^2(1 - 3 \cdot 2^{\lfloor \log_2 u_j \rfloor}) \right) (-1)^{\langle h_j, u_j \rangle} du_j \right) \\ &= \prod_{j \in I_{\mathbf{h}}} \beta_j^2 |h_j|_{\mathrm{p}}^{-2},\end{aligned}$$

where $I_{\mathbf{h}} = \{j : h_j \neq 0\}$. The last equality follows from [15, Corollary 4.4]. For $\mathbf{h} = \mathbf{0}$,

$$\begin{aligned}\tilde{g}(\mathbf{0}) = \int_{[0,1)^t} g(\mathbf{u})d\mathbf{u} &= -1 + \prod_{j=1}^{t} \left(\int_0^1 (1 + 2\beta_j^2(1 - 3 \cdot 2^{\lfloor \log_2 u_j \rfloor}))du_j \right) \\ &= -1 + \prod_{j=1}^{t} \left(1 + 2\beta_j^2(1 - 3 \cdot (2/3))\right) = 0,\end{aligned}$$

where the third equality again follows from [15, Corollary 4.4]. Thus, we have that

$$\begin{aligned}\frac{\beta_0^2}{n}\sum_{i=0}^{n-1} \tilde{\psi}(\mathbf{u}_i) &= \frac{\beta_0^2}{n}\sum_{i=0}^{n-1} \left(\sum_{\mathbf{h}\neq\mathbf{0}} (-1)^{\langle \mathbf{h},\mathbf{u}_i \rangle} \prod_{j\in I_{\mathbf{h}}} \beta_j^2 |h_j|_{\mathrm{p}}^{-2} \right) \\ &= \frac{1}{n}\sum_{\mathbf{h}\neq\mathbf{0}} \beta_{I_{\mathbf{h}}}^2 \|\mathbf{h}\|_{\tilde{\pi}}^{-2} \left(\sum_{i=0}^{n-1} (-1)^{\langle \mathbf{h},\mathbf{u}_i \rangle} \right) \\ &= \sum_{\mathbf{0}\neq\mathbf{h}\in\mathcal{L}_t^*} \beta_{I_{\mathbf{h}}}^2 \|\mathbf{h}\|_{\tilde{\pi}}^{-2} = \tilde{\mathcal{P}}_{2,\mathrm{PLR}},\end{aligned}$$

where the next-to-last equality follows from Lemma 3.20, and it is easy to see that the sum

$$\sum_{\mathbf{h}\neq\mathbf{0}} (-1)^{\langle \mathbf{h},\mathbf{u}_i \rangle} \prod_{j\in I_{\mathbf{h}}} |h_j|_{\mathrm{p}}^{-2}$$

converges absolutely (it is actually given by $3^t - 1$), which validates the change in the summation order from the first to the second line. □