

# On the Best Lattice Quantizers

Erik Agrell, *Fellow, IEEE*, and Bruce Allen, *Member, IEEE*

**Abstract**—A lattice quantizer approximates an arbitrary real-valued source vector with a vector taken from a specific discrete lattice. The quantization error is the difference between the source vector and the lattice vector. In a classic 1996 paper, Zamir and Feder show that the globally optimal lattice quantizer (which minimizes the mean square error) has *white* quantization error: for a uniformly distributed source, the covariance of the error is the identity matrix, multiplied by a positive real factor. We generalize the theorem, showing that the same property holds (i) for any lattice whose mean square error cannot be decreased by a small perturbation of the generator matrix, and (ii) for an optimal product of lattices that are themselves locally optimal in the sense of (i). We derive an upper bound on the normalized second moment (NSM) of the optimal lattice in any dimension, by proving that any lower- or upper-triangular modification to the generator matrix of a product lattice reduces the NSM. Using these tools and employing the best currently known lattice quantizers to build product lattices, we construct improved lattice quantizers in dimensions 13 to 15, 17 to 23, and 25 to 48. In some dimensions, these are the first reported lattices with normalized second moments below the best known upper bound.

**Index Terms**—Dither autocorrelation, laminated lattice, lattice theory, mean square error, moment of inertia, normalized second moment, product lattice, quantization constant, quantization error, vector quantization, Voronoi region, white noise.

## I. INTRODUCTION

**L**ATTICES are regular arrays of points in  $\mathbb{R}^n$ . They are obtained as arbitrary linear combinations of (at most  $n$ ) linearly independent basis vectors, with integer coefficients. Hence, lattices are a countably infinite set of vectors, closed under addition. The remarkable book by Conway and Sloane [1] provides a comprehensive review of lattices and their properties.

As fundamental geometric structures, lattices have found applications in a variety of disciplines, including digital communications [2], experimental design [3], data analysis [4], and particle physics [5]. In each application, the problem of designing the best lattice for a given purpose arises. Such optimization challenges often reduce to familiar mathematical problems such as sphere-packing, sphere-covering, or quantization [1, Ch. 1–2].

In this paper, we are concerned with the *quantization problem*, which can be defined as follows. Random vectors in  $\mathbb{R}^n$  are drawn from some (source) probability distribution, and approximated by their closest lattice points. This approximation

(or quantization) process creates a round-off (or quantization) error: the difference between the vector and its closest lattice point. Among all lattices having the same number of lattice points per unit volume, the optimal lattice quantizer is the lattice with the minimum mean square error. This is equivalent to minimizing the *normalized second moment* (NSM), which is a scale-invariant measure of this mean square error.

As in most work on lattice quantization, we assume that the lattice is sufficiently dense so that the source probability distribution is approximately constant over each Voronoi region. In this case, the optimal lattice does not depend upon the source distribution of the random vectors.

Tables of the NSM, showing the best known lattices for quantization in various dimensions are listed in [6], [7], [1, p. 61], and the quantization performance of some additional lattices is computed in [8]–[13]. Yet, proofs of optimality are known only in dimensions up to three [6], [14].

In a pioneering 1996 paper [15], Zamir and Feder show that the optimal lattice quantizer in any dimension has a *white* quantization error. More precisely, the error defined above (vector difference between a random source vector and its closest lattice vector) has a covariance matrix which is the identity matrix, scaled by a positive real constant.

In this paper, we extend the Zamir and Feder result to *locally optimal* lattices. These are lattices whose NSM cannot be reduced by a small perturbation of the lattice generator matrix. We also consider *product lattices*, which are the Cartesian product of two or more lower-dimensional lattices. The NSM of a product lattice depends on the relative scaling between the component lattices. A closed-form expression for the optimal scale factors is derived, and we call a product lattice using such scale factors an *optimal product*. If each of the lower-dimensional lattices is locally optimal, then we prove that the optimal product is the one for which the quantization error is *white*.

Lastly, we apply these methods to explicitly design some product lattices and analytically optimize their scale factors. These provide constructive upper bounds on the quantization performance of the optimal lattices in their respective dimensions. This simple construction yields better lattice quantizers than previously reported in all dimensions above 12 except for 16 and 24. We also prove that further optimization is possible: the NSM of such product lattices is a saddle point in the space of generator matrices, and can be further reduced by certain perturbations of the generator matrix.

## II. MATHEMATICAL PRELIMINARIES AND METHOD

**Notation:** Bold lowercase letters  $\mathbf{x}$  denote row vectors, while bold uppercase letters  $\mathbf{X}$  denote either matrices or random vectors. An all-zero vector or matrix of an arbitrary

Manuscript received 18 February 2022; revised 25 April 2023; accepted 17 June 2023. The work of E. Agrell was supported by a Collaborating Scientist Grant from the Max Planck Institute for Gravitational Physics, Germany, which is gratefully acknowledged.

E. Agrell is with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden (e-mail: agrell@chalmers.se).

B. Allen is with the Max Planck Institute for Gravitational Physics, 30167 Hannover, Germany, and Leibniz Universität Hannover (e-mail: bruce.allen@aei.mpg.de).

size (inferred from the context) is denoted by  $\mathbf{0}$ , and identity matrices are denoted by  $\mathbf{I}$ . Sets are denoted by uppercase Greek letters  $\Omega$ , apart from the integers  $\mathbb{Z}$  and real numbers  $\mathbb{R}$ . Arithmetical operations on sets should be understood as operating per element, e.g.,  $\Omega + \lambda \triangleq \{\mathbf{x} + \lambda : \mathbf{x} \in \Omega\}$ . Definitions are indicated by  $\triangleq$ .

Without loss of generality, we consider  $n$ -dimensional lattices  $\Lambda$  that are generated by square invertible  $n \times n$  *generator matrices*  $\mathbf{B}$ . The lattice consists of the set of points  $\mathbf{u}\mathbf{B}$  for all row vectors  $\mathbf{u}$  with integer components. The all-zero row vector  $\mathbf{0}$  belongs to all lattices. The cubic lattice  $\mathbb{Z}^n$  is the special case for which  $\mathbf{B}$  is the identity matrix.

Until now, we have used “quantization” to denote the map from a vector in  $\mathbb{R}^n$  to the closest lattice point. However, for the proofs in this paper, we consider more general mappings. A *quantization rule* or *quantizer* for a lattice  $\Lambda$  is a function  $Q_\Lambda(\mathbf{x})$  such that for all  $\mathbf{x} \in \mathbb{R}^n$

$$Q_\Lambda(\mathbf{x}) \in \Lambda, \quad (1)$$

$$Q_\Lambda(\mathbf{x} + \lambda) = Q_\Lambda(\mathbf{x}) + \lambda \text{ for all } \lambda \in \Lambda. \quad (2)$$

The quantizer’s properties are completely determined by its behavior in the *fundamental decision region*

$$\Omega(Q_\Lambda) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Q_\Lambda(\mathbf{x}) = \mathbf{0}\}, \quad (3)$$

since (2) may then be used to determine the action anywhere. The translate  $\Omega(Q_\Lambda) + \lambda$  of the fundamental decision region is called the decision region of the lattice point  $\lambda$ . All have the same volume [2, Prop. 2.2.1]

$$V_\Lambda \triangleq \int_{\Omega(Q_\Lambda)} d^n \mathbf{x} = |\det \mathbf{B}|. \quad (4)$$

As indicated by the notation  $V_\Lambda$ , this volume depends upon the lattice  $\Lambda$ , but is independent of the quantization rule. Taken together, the decision regions of all lattice points cover  $\mathbb{R}^n$  without overlap.

As mentioned in the Introduction, the performance of lattice quantizers does not depend upon the source distribution if the lattice is sufficiently dense. To prove this, consider a source probability density function (pdf)  $p_{\mathbf{X}}(\mathbf{x})$ , normalized by  $\int p_{\mathbf{X}}(\mathbf{x}) d^n \mathbf{x} = 1$ . The mean square quantization error of the quantization rule is

$$\mathbb{E}[\|\mathbf{x} - Q_\Lambda(\mathbf{x})\|^2] = \int_{\mathbb{R}^n} p_{\mathbf{X}}(\mathbf{x}) \|\mathbf{x} - Q_\Lambda(\mathbf{x})\|^2 d^n \mathbf{x}. \quad (5)$$

Since the translates  $\Omega(Q_\Lambda) + \lambda, \forall \lambda \in \Lambda$  cover  $\mathbb{R}^n$  without overlap, the mean square error can be written as

$$\mathbb{E}[\|\mathbf{x} - Q_\Lambda(\mathbf{x})\|^2] = \int_{\Omega(Q_\Lambda)} \sum_{\lambda \in \Lambda} p_{\mathbf{X}}(\lambda + \mathbf{x}) \|\mathbf{x}\|^2 d^n \mathbf{x}, \quad (6)$$

where we use (2) to write this as an integral over the fundamental decision region and (1) to set  $Q_\Lambda(\mathbf{x}) = \mathbf{0}$  inside that region. If now  $\Lambda$  is sufficiently dense, then  $\sum_{\lambda \in \Lambda} p_{\mathbf{X}}(\lambda + \mathbf{x})$  is approximately constant, independent of  $\mathbf{x}$ . Such a probability distribution can be obtained by rescaling a smooth base pdf  $\tilde{p}$  of compact support, for example as  $p_{\mathbf{X}}(\mathbf{x}) = \alpha^n \tilde{p}(\alpha \mathbf{x})$  in the

limit as  $\alpha \rightarrow 0$ . In such a limit,  $\sum_{\lambda \in \Lambda} p_{\mathbf{X}}(\lambda + \mathbf{x}) = 1/V_\Lambda$  for any  $\mathbf{x}$ ,<sup>1</sup> and the mean square error (6) approaches [16]

$$E(Q_\Lambda) \triangleq \lim_{\alpha \rightarrow 0} \mathbb{E}[\|\mathbf{x} - Q_\Lambda(\mathbf{x})\|^2] \quad (7)$$

$$= \frac{1}{V_\Lambda} \int_{\Omega(Q_\Lambda)} \|\mathbf{x}\|^2 d^n \mathbf{x}. \quad (8)$$

Hence, in what follows, the pdf of the source does not appear.

Important quantities that are closely related to the mean square error are the NSM or quantizer constant  $G(Q_\Lambda)$  and the *correlation matrix*  $\mathbf{R}(Q_\Lambda)$ , which are [6], [2, pp. 48, 71]

$$G(Q_\Lambda) \triangleq \frac{E(Q_\Lambda)}{nV_\Lambda^{2/n}}, \quad (9)$$

$$\mathbf{R}(Q_\Lambda) \triangleq \frac{1}{V_\Lambda} \int_{\Omega(Q_\Lambda)} \mathbf{x}^T \mathbf{x} d^n \mathbf{x}. \quad (10)$$

Note that the NSM  $G(Q_\Lambda)$  is “dimensionless” in the sense that it is invariant under uniform rescaling of the lattice. From (8) and (10), it follows that

$$E(Q_\Lambda) = \text{tr} \mathbf{R}(Q_\Lambda), \quad (11)$$

so the trace of the correlation matrix gives the mean square error.

It follows immediately from the definition (10) that the correlation matrix  $\mathbf{R}$  is real, symmetric, and positive definite. If the quantization error is not white, then  $\mathbf{R}$  provides “preferred directions” in the space, for example corresponding to the eigenvector with the largest or the smallest eigenvalue. In the case of white quantization error, however,  $\mathbf{R}$  is proportional to the identity, and does not generate preferred directions, since every vector is an eigenvector with the same positive real eigenvalue.

For a given lattice  $\Lambda$ , the most common and important rule is the *minimum-distance quantization rule*, denoted by a hat:

$$Q_\Lambda(\mathbf{x}) = \hat{Q}_\Lambda(\mathbf{x}) \triangleq \arg \min_{\lambda \in \Lambda} \|\mathbf{x} - \lambda\|^2. \quad (12)$$

For any vector  $\mathbf{x}$ , it returns the closest vector in the lattice. Ties can be broken by any criterion that respects condition (2). This quantization rule is special because, for a given lattice  $\Lambda$ , it minimizes  $E(Q_\Lambda)$  and  $G(Q_\Lambda)$ . This follows immediately from (7), because the expectation  $\mathbb{E}[\|\mathbf{x} - Q_\Lambda(\mathbf{x})\|^2]$  is minimal if  $\|\mathbf{x} - Q_\Lambda(\mathbf{x})\|^2$  is minimized for every  $\mathbf{x}$ . Hence,  $\hat{Q}_\Lambda$  is the *optimal decision rule* for a given lattice.

For this rule, the fundamental decision region (3) is the *Voronoi region*

$$\Omega(\hat{Q}_\Lambda) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|^2 \leq \|\mathbf{x} - \lambda\|^2, \forall \lambda \in \Lambda\}, \quad (13)$$

which geometrically consists of all points in  $\mathbb{R}^n$  whose closest lattice point is the origin.<sup>2</sup> An important property of the Voronoi region of any lattice is that it is symmetric about  $\mathbf{0}$ : the center of gravity  $\int_{\Omega(\hat{Q}_\Lambda)} \mathbf{x} d^n \mathbf{x} = \mathbf{0}$ . Hence, the correlation matrix  $\mathbf{R}(Q_\Lambda)$  is equal to the *covariance matrix* whenever  $Q_\Lambda = \hat{Q}_\Lambda$  (but not for arbitrary quantization rules  $Q_\Lambda$ ).

<sup>1</sup>This can be proved by writing  $\int \tilde{p}(\mathbf{x}) d^n \mathbf{x} = 1$  as a Riemann sum over a shifted and scaled lattice  $\mathbf{x} \in \alpha(\Lambda + \boldsymbol{\xi})$ .

<sup>2</sup>More precisely, the interiors of (3) and (13) are equal under the rule (12). With a slight abuse of notation, we disregard their boundaries, which have zero  $n$ -volume and do not contribute to any integral over a finite integrand.

Throughout this paper, the word “optimal” is used in several senses. For a given lattice, the *optimal decision rule* is the one which minimizes the NSM, i.e., (12). Among all lattices of given dimension, the *optimal lattice* is the one with the smallest NSM. The *optimal product* of given lattices is the one that minimizes the NSM among all Cartesian products of those lattices, by varying the relative scales between them.

Our main theorem-proving technique is, as in [2, Sec. 4.3], to construct different decision rules for a given lattice, exploiting the fact that their NSMs are equal to or greater than the NSM of the optimal decision rule  $\hat{Q}$ . For example, in Sec. III, if the quantization error of a lattice  $\Lambda$  is not white, then  $\mathbf{R}$  provides preferred directions (say, the eigenvectors with the largest eigenvalue)<sup>3</sup>. With these, we construct a family of lattices  $\tilde{\Lambda}$  with nonoptimal decision rules, whose NSM is smaller than that of the original lattice  $\Lambda$ . Since the NSM of  $\tilde{\Lambda}$  with an optimal decision rule cannot be larger, we have thus shown that the original lattice is not optimal. A similar proof technique is applied in Sec. V. Starting with a product lattice, we generate a new non-product lattice, with a nonoptimal decision rule, but whose NSM is equal to that of the original starting lattice. Hence, the optimal decision rule on the new non-product lattice must yield a smaller NSM than that of the original product.

### III. LOCALLY OPTIMAL LATTICES

Our starting point is the following theorem, which states that the globally optimal quantizer lattice has a white quantization error: a covariance matrix proportional to the identity.

**Theorem 1** (Zamir–Feder [15], [2, Sec. 4.3]): For the optimal lattice  $\Lambda$  in any dimension  $n$ ,

$$\mathbf{R}(\hat{Q}_\Lambda) = \frac{E(\hat{Q}_\Lambda)}{n} \mathbf{I}. \quad (14)$$

We now generalize this to the locally optimal case.

A *locally optimal lattice* is a lattice  $\Lambda$  whose NSM  $G(\hat{Q}_\Lambda)$  cannot be decreased by an infinitesimal perturbation of the generator matrix  $\mathbf{B}$  [14]. The extension to Theorem 1 is:

**Theorem 2:** Any locally optimal lattice satisfies (14).

**Proof:** Our proof is constructive. If the covariance matrix of  $\Lambda$  is not proportional to the identity matrix, we use it to build a nearby lattice  $\tilde{\Lambda}$  with a smaller NSM than the NSM of  $\Lambda$ .

Let  $\tilde{\Lambda} = \Lambda \mathbf{A}_\beta$ , where  $\mathbf{A}_\beta$  is an invertible  $n \times n$  matrix and  $\beta$  is a real parameter to be defined later. As in [15], we consider the minimum-distance quantizer  $\hat{Q}_\Lambda(\mathbf{x})$  on  $\Lambda$  and the suboptimal quantizer  $\tilde{Q}_\Lambda(\mathbf{x}) \triangleq \hat{Q}_\Lambda(\mathbf{x} \mathbf{A}_\beta^{-1}) \mathbf{A}_\beta$  on  $\tilde{\Lambda}$ .

It is straightforward to show that  $\tilde{Q}_\Lambda(\mathbf{x})$  satisfies (1)–(2) and has a fundamental decision region  $\Omega(\tilde{Q}_\Lambda) = \Omega(\hat{Q}_\Lambda) \mathbf{A}_\beta$ . Note that while  $\Omega(\hat{Q}_\Lambda)$  is the Voronoi region of  $\Lambda$ , the fundamental decision region  $\Omega(\tilde{Q}_\Lambda)$  is generally *not* the Voronoi region of  $\tilde{\Lambda}$ . By (4), it has volume  $V_{\tilde{\Lambda}} = V_\Lambda |\det \mathbf{A}_\beta|$ .

The covariance matrices of the two quantization rules are easily related using the change of variables (mapping) provided by  $\mathbf{A}_\beta$ . From (10), the covariance matrix of  $\tilde{Q}_\Lambda$  is

$\mathbf{R}(\tilde{Q}_\Lambda) = \mathbf{A}_\beta^T \mathbf{R}(\hat{Q}_\Lambda) \mathbf{A}_\beta$  [15, Eq. (15)]. Hence, the NSM is

$$G(\tilde{Q}_\Lambda) = \frac{E(\tilde{Q}_\Lambda)}{n V_{\tilde{\Lambda}}^{2/n}} = \frac{\text{tr}(\mathbf{A}_\beta \mathbf{A}_\beta^T \mathbf{R}(\hat{Q}_\Lambda))}{n (V_\Lambda |\det \mathbf{A}_\beta|)^{2/n}}, \quad (15)$$

where we have used the cyclic property of the trace.

To select  $\mathbf{A}_\beta$ , we follow the approach described earlier, using  $\mathbf{R}(\hat{Q}_\Lambda)$  to obtain preferred directions.<sup>4</sup> Let  $\bar{\mathbf{R}}$  denote the traceless part, which by assumption is nonzero:

$$\bar{\mathbf{R}} \triangleq \mathbf{R}(\hat{Q}_\Lambda) - \frac{\text{tr} \mathbf{R}(\hat{Q}_\Lambda)}{n} \mathbf{I}, \quad (16)$$

and let  $\mathbf{A}_\beta \triangleq \exp(\beta \bar{\mathbf{R}})$ . This choice of mapping is volume-preserving, since for any square matrix  $\mathbf{M}$ ,  $\det(\exp(\mathbf{M})) = \exp(\text{tr} \mathbf{M})$  [18, p. 16]. Thus,  $\det \mathbf{A}_\beta = \exp(\beta \text{tr} \bar{\mathbf{R}}) = 1$ . Note that because the covariance matrix is symmetric and real, both  $\bar{\mathbf{R}}$  and  $\mathbf{A}_\beta$  are symmetric and real.

For the proof, we only need  $\mathbf{A}_\beta$  for infinitesimal  $\beta$ :

$$\mathbf{A}_\beta = \mathbf{I} + \beta \bar{\mathbf{R}} + O(\beta^2). \quad (17)$$

Substituting  $\mathbf{A}_\beta$  from (17) and  $\mathbf{R}(\hat{Q}_\Lambda) = \mathbf{I} \text{tr} \mathbf{R}(\hat{Q}_\Lambda)/n + \bar{\mathbf{R}}$  from (16) into (15), the NSM becomes

$$\begin{aligned} G(\tilde{Q}_\Lambda) &= \frac{1}{n V_{\tilde{\Lambda}}^{2/n}} \text{tr} \left( [\mathbf{I} + \beta \bar{\mathbf{R}} + O(\beta^2)]^2 \left[ \frac{\text{tr} \mathbf{R}(\hat{Q}_\Lambda)}{n} \mathbf{I} + \bar{\mathbf{R}} \right] \right) \\ &= G(\hat{Q}_\Lambda) + 2\beta \frac{\text{tr} \bar{\mathbf{R}}^2}{n V_{\tilde{\Lambda}}^{2/n}} + O(\beta^2), \end{aligned} \quad (18)$$

where we have distributed the trace over additions and used  $\text{tr} \bar{\mathbf{R}} = 0$ .

It is clear from (18) that for negative  $\beta$  near zero, we have  $G(\tilde{Q}_\Lambda) < G(\hat{Q}_\Lambda)$ . This follows because, since  $\bar{\mathbf{R}}$  is a non-vanishing real symmetric matrix,  $\text{tr} \bar{\mathbf{R}}^2$  must be positive<sup>5</sup>. Since the NSM of the minimum-distance quantization rule (12) on  $\tilde{\Lambda}$  satisfies<sup>6</sup>  $G(\tilde{Q}_\Lambda) \leq G(\hat{Q}_\Lambda)$ , we have established that for negative  $\beta$  near zero,  $G(\tilde{Q}_\Lambda) < G(\hat{Q}_\Lambda)$ .  $\square$

To test Theorem 2, we examine a large number of numerically optimized lattice quantizers. These were designed in 1996 using an iterative algorithm, which converges to different locally optimal lattices [11]. A total of 90 locally optimal lattices are available as online supplementary material to the 1998 article [19]; we estimate the covariance matrices  $\mathbf{R}(\hat{Q}_\Lambda)$  of their quantization errors using Monte Carlo integration. In all cases, consistent with the theorem, the obtained covariance matrices are proportional to the identity matrix, apart from minor round-off errors.

Theorem 2 establishes that local optimality is a sufficient condition for a white quantization error (14). Is it also a necessary condition? In other words, is any lattice that satisfies (14) locally optimal? In the next two sections, we will show that this is false, using product lattices as a counterexample.

<sup>4</sup>From here on, our proof deviates from the corresponding proof in [15] for globally optimal lattices.

<sup>5</sup>To prove this, write  $\bar{\mathbf{R}} = \mathbf{U} \mathbf{D} \mathbf{U}^{-1}$  where  $\mathbf{U}$  is orthogonal and  $\mathbf{D}$  is real and diagonal, then use the cyclic property of the trace.

<sup>6</sup>We have no way to directly analyze the performance of  $\tilde{Q}_\Lambda$  because we have no simple expression for the Voronoi region  $\Omega(\tilde{Q}_\Lambda)$ .

<sup>3</sup>A similar argument leads to [17, Eq. 5.1.1].

#### IV. PRODUCT LATTICES

In this section, we study lattices that are formed as the Cartesian products of lower-dimensional lattices. Gersho applied this technique to obtain upper bounds on the optimal NSM for  $n = 5$  and  $n = 100$ , without formalizing the expressions [6, Sec. VII].

Let  $k$  lattices in dimensions  $n_1, \dots, n_k$  be denoted by  $\Lambda_1, \dots, \Lambda_k$ , and consider their product  $\Lambda_p = \Lambda_1 \times \dots \times \Lambda_k$ , whose dimension is  $n = n_1 + \dots + n_k$ . A generator matrix for  $\Lambda_p$  is

$$\mathbf{B}_p = \begin{bmatrix} \mathbf{B}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{B}_k \end{bmatrix}, \quad (19)$$

where  $\mathbf{B}_i$  is a generator matrix of  $\Lambda_i$  for  $i = 1, \dots, k$ . The Voronoi region, volume, and other properties of a product lattice are as follows.

*Proposition 3:* For any  $k \geq 1$  and  $\Lambda_p = \Lambda_1 \times \dots \times \Lambda_k$ ,

$$\Omega \triangleq \Omega(\hat{Q}_{\Lambda_p}) = \Omega_1 \times \dots \times \Omega_k, \quad (20)$$

$$V \triangleq V_{\Lambda_p} = V_1 V_2 \cdots V_k, \quad (21)$$

$$E \triangleq E(\hat{Q}_{\Lambda_p}) = E_1 + \dots + E_k, \quad (22)$$

$$G \triangleq G(\hat{Q}_{\Lambda_p}) = \frac{1}{nV^{2/n}} \sum_{i=1}^k n_i V_i^{2/n_i} G_i, \quad (23)$$

$$\mathbf{R} \triangleq \mathbf{R}(\hat{Q}_{\Lambda_p}) = \begin{bmatrix} \mathbf{R}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{R}_k \end{bmatrix}, \quad (24)$$

where  $\Omega_i$ ,  $V_i$ ,  $E_i$ ,  $G_i$ , and  $\mathbf{R}_i$  denote the corresponding properties of  $\Lambda_i$ .

An example of a 3-dimensional Voronoi region  $\Omega$ , constructed according to Proposition 3 as the Cartesian product of two lower-dimensional Voronoi regions  $\Omega_1$  and  $\Omega_2$ , is illustrated in Fig. 1.

*Proof:* Let  $\mathbf{x} = [\mathbf{x}_1 \cdots \mathbf{x}_k]$  and  $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_k]$ . By definition,  $\Omega$  in (13) is formed by all vectors  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  for  $i = 1, \dots, k$  such that

$$\|\mathbf{x}_1\|^2 + \dots + \|\mathbf{x}_k\|^2 \leq \|\mathbf{x}_1 - \boldsymbol{\lambda}_1\|^2 + \dots + \|\mathbf{x}_k - \boldsymbol{\lambda}_k\|^2, \quad (25)$$

for all  $\boldsymbol{\lambda}_i \in \Lambda_i$ .

If  $\mathbf{x}_i \in \Omega_i$  for all  $i$ , then  $\|\mathbf{x}_i\|^2 \leq \|\mathbf{x}_i - \boldsymbol{\lambda}_i\|^2$ . Summing over  $i = 1, \dots, k$  implies (25), so  $\Omega_1 \times \dots \times \Omega_k \subseteq \Omega$ . Conversely, if  $\mathbf{x} \in \Omega$ , then for a fixed  $i$  we set  $\boldsymbol{\lambda}_j = \mathbf{0}$  in (25) for all  $j \neq i$ . This implies  $\|\mathbf{x}_i\|^2 \leq \|\mathbf{x}_i - \boldsymbol{\lambda}_i\|^2$  for all  $\boldsymbol{\lambda}_i \in \Lambda_i$ , so that  $\mathbf{x}_i \in \Omega_i$ . Repeating for  $i = 1, \dots, k$  shows that  $\Omega \subseteq \Omega_1 \times \dots \times \Omega_k$ .

Taken together,  $\Omega_1 \times \dots \times \Omega_k \subseteq \Omega$  and  $\Omega \subseteq \Omega_1 \times \dots \times \Omega_k$  prove (20), which in turn proves (21).

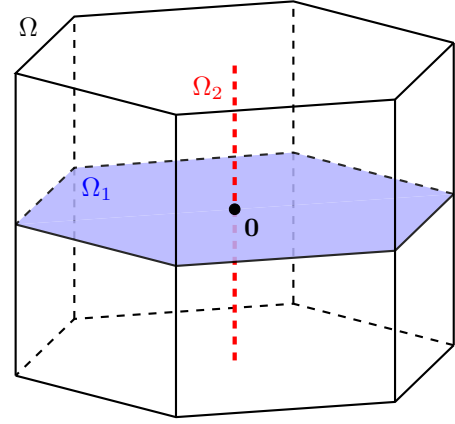


Fig. 1: The Voronoi region  $\Omega$  of the product lattice  $\Lambda_p = \Lambda_1 \times \Lambda_2$ , where  $\Lambda_1$  is the two-dimensional hexagonal lattice  $A_2$  and  $\Lambda_2$  is the one-dimensional integer lattice  $\mathbb{Z}$ . The origin  $\mathbf{0}$  belongs to all three lattices and is the centroid of all three Voronoi regions. The top and bottom facets of  $\Omega$  are shifted copies of  $\Omega_1$ , and the six vertical edges are shifted copies of  $\Omega_2$ .

The definition (8), applied to a product lattice using (20) and (21), implies that

$$\begin{aligned} E &= \frac{1}{V} \int_{\Omega_1} \cdots \int_{\Omega_k} (\|\mathbf{x}_1\|^2 + \dots + \|\mathbf{x}_k\|^2) d^{n_1} \mathbf{x}_1 \cdots d^{n_k} \mathbf{x}_k \\ &= \frac{1}{V_1 V_2 \cdots V_k} ((V_1 E_1) V_2 \cdots V_k + \dots + V_1 V_2 \cdots (V_k E_k)), \end{aligned} \quad (26)$$

which proves (22). Equation (23) follows by substituting  $E = nV^{2/n}G$  and the corresponding expressions for  $E_i$  into (22), and simplifying using (21) and  $n = n_1 + \dots + n_k$ .

Lastly, to prove (24), we use (20) in (10) to obtain

$$\mathbf{R} = \frac{1}{V} \int_{\Omega_1} \cdots \int_{\Omega_k} \begin{bmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \cdots & \mathbf{x}_1^T \mathbf{x}_k \\ \vdots & \ddots & \vdots \\ \mathbf{x}_k^T \mathbf{x}_1 & \cdots & \mathbf{x}_k^T \mathbf{x}_k \end{bmatrix} d^{n_1} \mathbf{x}_1 \cdots d^{n_k} \mathbf{x}_k. \quad (27)$$

For the submatrices on the diagonal, whose integrands have the form  $\mathbf{x}_i^T \mathbf{x}_i$ , the volumes cancel as in (26), leaving  $\mathbf{R}_i$ . The off-diagonal submatrices with integrands  $\mathbf{x}_i^T \mathbf{x}_j$  for  $i \neq j$  are separable into products of two integrals such as  $\int_{\Omega_i} \mathbf{x}_i d^{n_i} \mathbf{x}_i$ . These vanish because (as pointed out after (13)) the Voronoi region  $\Omega_i$  is symmetric about zero and thus has its center of gravity at the origin.  $\square$

We now generalize the product construction by introducing a list of real positive scale factors  $\mathbf{a} = [a_1, \dots, a_k]$  to build a family of product lattices  $\Lambda(\mathbf{a}) = a_1 \Lambda_1 \times \dots \times a_k \Lambda_k$ . A generator matrix for  $\Lambda(\mathbf{a})$  is

$$\mathbf{B}(\mathbf{a}) = \begin{bmatrix} a_1 \mathbf{B}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & a_k \mathbf{B}_k \end{bmatrix}. \quad (28)$$



The properties of  $\Lambda(\mathbf{a})$  follow by replacing  $\Omega_i$  by  $a_i\Omega_i$ ,  $V_i$  by  $a_i^{n_i}V_i$ ,  $E_i$  by  $a_i^2E_i$ , and  $\mathbf{R}_i$  by  $a_i^2\mathbf{R}_i$  in Proposition 3, while  $G_i$ , due to its scale-invariant definition (9), remains unchanged for  $i = 1, \dots, k$ . These substitutions result in

$$\Omega(\mathbf{a}) = a_1\Omega_1 \times \dots \times a_k\Omega_k, \quad (29)$$

$$V(\mathbf{a}) = a_1^{n_1} \dots a_k^{n_k} V_1 \dots V_k, \quad (30)$$

$$E(\mathbf{a}) = a_1^2 E_1 + \dots + a_k^2 E_k, \quad (31)$$

$$G(\mathbf{a}) = \frac{1}{nV(\mathbf{a})^{2/n}} \sum_{i=1}^k n_i a_i^2 V_i^{2/n_i} G_i, \quad (32)$$

$$\mathbf{R}(\mathbf{a}) = \begin{bmatrix} a_1^2 \mathbf{R}_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & a_k^2 \mathbf{R}_k \end{bmatrix}. \quad (33)$$

For given  $\Lambda_1, \dots, \Lambda_k$ , what scale factors  $\mathbf{a}$  produce the optimal product  $\Lambda(\mathbf{a})$ , in the sense of minimizing  $G(\mathbf{a})$ ? We note that  $G(\mathbf{a})$  has at least one minimum for finite and positive  $a_1, \dots, a_k$ , because if one of these scale factors is varied while keeping the others fixed, then (32) diverges to infinity as the scale factor tends to either zero or infinity. This minimum is unique up to a linear scale factor, and has a closed form as follows.

**Theorem 4:** For given lattices  $\Lambda_1, \dots, \Lambda_k$ , varying only  $\mathbf{a}$ ,  $\Lambda(\mathbf{a})$  is an optimal product if and only if

$$a_i = \frac{C}{V_i^{1/n_i} \sqrt{G_i}}, \quad i = 1, \dots, k \quad (34)$$

for an arbitrary real constant  $C > 0$ . An equivalent condition is

$$a_i = C \sqrt{\frac{n_i}{E_i}}, \quad i = 1, \dots, k. \quad (35)$$

The optimal NSM  $G(\mathbf{a})$  is given by

$$G^n(\mathbf{a}) = G_1^{n_1} \dots G_k^{n_k} \quad (36)$$

independently of  $V_1, \dots, V_k$ .

*Proof:* The derivative of  $G(\mathbf{a})$  in (32) with respect to  $a_i$  is

$$\begin{aligned} \frac{\partial G(\mathbf{a})}{\partial a_i} &= \frac{1}{nV(\mathbf{a})^{2/n}} \cdot 2n_i a_i V_i^{2/n_i} G_i \\ &\quad - \frac{2}{n^2 V(\mathbf{a})^{1+2/n}} \frac{\partial V(\mathbf{a})}{\partial a_i} \cdot \sum_{j=1}^k n_j a_j^2 V_j^{2/n_j} G_j. \end{aligned} \quad (37)$$

Substituting

$$\frac{\partial V(\mathbf{a})}{\partial a_i} = \frac{n_i}{a_i} V(\mathbf{a}), \quad (38)$$

which follows from (30), into (37) and simplifying yields

$$\frac{\partial G(\mathbf{a})}{\partial a_i} = \frac{2n_i}{nV(\mathbf{a})^{2/n}} \left( a_i V_i^{2/n_i} G_i - \frac{1}{na_i} \sum_{j=1}^k n_j a_j^2 V_j^{2/n_j} G_j \right). \quad (39)$$

Equating (39) to zero for  $i = 1, \dots, k$  reveals that  $a_i^2 V_i^{2/n_i} G_i$  is constant for all  $i$ , which gives (34). Substituting (34) back into (39) confirms that (34) is not only necessary for (39) being

zero but also sufficient, regardless of  $C$ . Then (35) follows from (34) and (9). Substitute (34) into (30) and (32), and use  $n_1 + \dots + n_k = n$ . Finally,  $V_i$  and  $C$  cancel out and (36) emerges.  $\square$

An interesting special case is when the sublattices  $\Lambda_i$  are locally optimal for all  $i = 1, \dots, k$ . If the scale factors  $\mathbf{a}$  are optimally chosen according to Theorem 4, which we denote by  $\mathbf{a}_{\text{opt}}$ , then  $\Lambda(\mathbf{a}_{\text{opt}})$  also has white quantization error. We state this in a way similar to Theorems 1 and 2, but with different conditions.

**Corollary 5:** If  $\Lambda_1, \dots, \Lambda_k$  are locally optimal lattices,  $\Lambda(\mathbf{a}) = a_1\Lambda_1 \times \dots \times a_k\Lambda_k$ , and  $[a_1, \dots, a_k] = \mathbf{a}_{\text{opt}}$ , then

$$\mathbf{R}(\mathbf{a}_{\text{opt}}) = \frac{E(\mathbf{a}_{\text{opt}})}{n} \mathbf{I}. \quad (40)$$

Furthermore,  $G(\mathbf{a}_{\text{opt}})$  is locally minimal with respect to any perturbations in the submatrices on the diagonal of (28).

*Proof:* From Theorem 2,  $\mathbf{R}_i = (E_i/n_i)\mathbf{I}$  for  $i = 1, \dots, k$ . Using this together with (35) in (33) yields  $\mathbf{R}(\mathbf{a}_{\text{opt}}) = C^2 \mathbf{I} = (E(\mathbf{a}_{\text{opt}})/n)\mathbf{I}$ , where the latter equality is proved by substituting (35) into (31).

To prove the local optimality of  $G(\mathbf{a}_{\text{opt}})$ , we let  $a'_i \triangleq a_i V_i^{1/n_i}$  and  $\mathbf{B}'_i \triangleq \mathbf{B}_i V_i^{-1/n_i}$ , where as before  $V_i = \det \mathbf{B}_i$ . Then the submatrices on the diagonal of (28) can be written as  $a_i \mathbf{B}_i = a'_i \mathbf{B}'_i$  for all  $i$ , where  $V'_i \triangleq \det \mathbf{B}'_i = 1$ . We will consider variations in  $a'_i$  and  $\mathbf{B}'_i$  separately. First, if  $a'_i$  is varied for any fixed (not necessarily optimal)  $\mathbf{B}'_i$ , then Theorem 4 applies and the minimal NSM is attained when  $a'_i = C/\sqrt{G_i}$ . Consequently, the optimal scale factors  $a_i = a'_i/V_i^{1/n_i}$  follow (34), and (39) is zero. Second, we consider variations in  $\mathbf{B}'_i$  for any fixed (not necessarily optimal)  $a'_i$ , keeping  $\det \mathbf{B}'_i = 1$ . Then the NSM in (32) is locally minimal when  $G_i$  is locally minimal, i.e., when  $\Lambda_i$  is a locally optimal lattice.  $\square$

Theorem 2 and Corollary 5 are curiously related to each other: both give sufficient (but not necessary) conditions for a white quantization error. Is Corollary 5 perhaps a special case of Theorem 2? In other words, is the optimal product  $\Lambda(\mathbf{a}_{\text{opt}})$ , to which Theorem 4 and Corollary 5 apply, also a locally optimal lattice, as defined in Sec. III? We will see in the next section that the answer is “no”; for any lattice of the form (28) with  $\mathbf{a} = \mathbf{a}_{\text{opt}}$ , the NSM can be decreased by perturbing any of the off-diagonal submatrices written as  $\mathbf{0}$  in (28).

## V. UPPER BOUND

In this section, we show that the NSM of any lattice is bounded above by that of a product lattice, which is given by (23) or (32). For brevity, we develop the case  $k = 2$  explicitly and then show that  $k > 2$  follows by induction.

Consider a square generator matrix of the form

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{H} & \mathbf{B}_2 \end{bmatrix}, \quad (41)$$

where  $\mathbf{B}_1$  is  $n_1 \times n_1$  and  $\mathbf{B}_2$  is  $n_2 \times n_2$ . Let  $\Lambda_1$ ,  $\Lambda_2$ , and  $\Lambda$  be the lattices generated by  $\mathbf{B}_1$ ,  $\mathbf{B}_2$ , and  $\mathbf{B}$ , respectively, and let  $\Lambda_p = \Lambda_1 \times \Lambda_2$  as in Sec. IV.

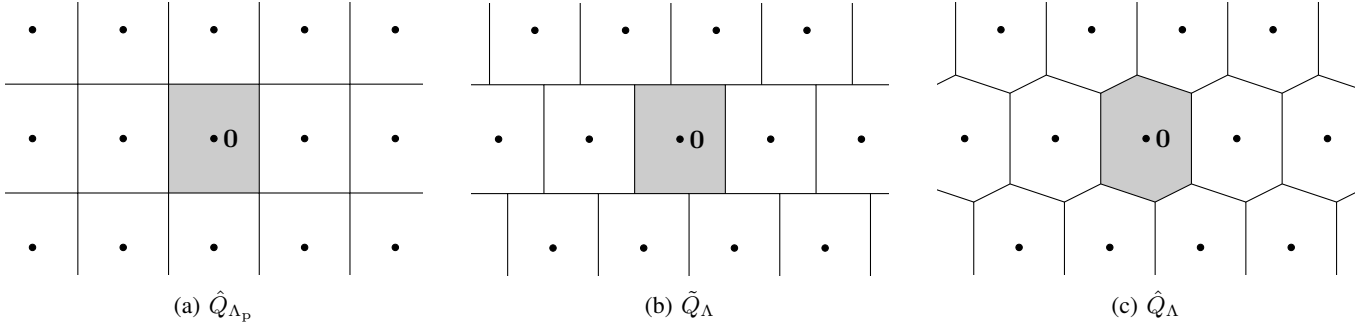


Fig. 2: An example of Theorem 7 with  $n_1 = n_2 = 1$ . Each cell is the decision region of the lattice point it contains, and the shaded cells are the fundamental decision regions. Comparing (a) and (b) shows that the NSMs of  $\hat{Q}_{\Lambda_p}$  and  $\tilde{Q}_{\Lambda}$  are equal, because their fundamental decision regions are identical. Comparing (b) and (c) shows that  $\tilde{Q}_{\Lambda}$  cannot have a smaller NSM than  $\hat{Q}_{\Lambda}$ , because the lattices are identical and  $\hat{Q}_{\Lambda}(x)$  minimizes the quantization error for every  $x$ .

**Lemma 6:** For any  $H$ ,  $G(\hat{Q}_{\Lambda}) \leq G(\hat{Q}_{\Lambda_p})$ , with equality if  $H = 0$ .

*Proof:* The method of proof is to define a suboptimal quantizer  $\tilde{Q}_{\Lambda}$  for which  $G(\tilde{Q}_{\Lambda}) = G(\hat{Q}_{\Lambda_p})$ . The lemma then follows, because by the definition of  $\hat{Q}_{\Lambda}$ ,  $G(\hat{Q}_{\Lambda}) \leq G(\tilde{Q}_{\Lambda})$ . The decision regions for the three different quantizers are illustrated in Fig. 2.

We construct the suboptimal quantization rule from the optimal quantization rules for  $\Lambda_1$  and  $\Lambda_2$ . As earlier, write the source vector as  $x = [x_1 \ x_2]$ , where  $x_1$  and  $x_2$  have respective dimensions  $n_1$  and  $n_2$ . Our suboptimal quantization rule is defined by the following four-step algorithm:

$$\hat{x}_2 \triangleq \hat{Q}_{\Lambda_2}(x_2), \quad (42)$$

$$z_1 \triangleq \hat{x}_2 B_2^{-1} H, \quad (43)$$

$$\hat{x}_1 \triangleq \hat{Q}_{\Lambda_1}(x_1 - z_1), \quad (44)$$

$$\tilde{Q}_{\Lambda}(x) \triangleq [\hat{x}_1 + z_1 \ \hat{x}_2]. \quad (45)$$

Quantities with subscript “1” lie in the subspace spanned by the first  $n_1$  coordinates (horizontal in Fig. 2) and the quantities with subscript “2” lie in the subspace spanned by the final  $n_2$  coordinates (vertical in Fig. 2). We first show that this is a quantization rule: it satisfies the conditions (1)–(2).

**Condition (1):** Since  $\hat{Q}_{\Lambda_1}(x_1 - z_1) \in \Lambda_1$  and  $\hat{Q}_{\Lambda_2}(x_2) \in \Lambda_2$ , there exist integer vectors  $u_1$  and  $u_2$  such that  $\hat{x}_1 = u_1 B_1$  and  $\hat{x}_2 = u_2 B_2$ . Then by (43),  $z_1 = u_2 H$  and by (45),  $\tilde{Q}_{\Lambda}(x) = [u_1 B_1 + u_2 H \ u_2 B_2] = [u_1 \ u_2] B$ , which shows that  $\tilde{Q}_{\Lambda}(x) \in \Lambda$ .

**Condition (2):** Consider  $\tilde{Q}_{\Lambda}(x + \lambda)$  for an arbitrary  $\lambda \in \Lambda$ . Let  $\lambda \triangleq [v_1 \ v_2] B = [v_1 B_1 + v_2 H \ v_2 B_2]$ , where  $v_1$  and  $v_2$  are integer vectors. Let  $x' \triangleq [x'_1 \ x'_2] \triangleq x + \lambda$ , and let  $\hat{x}'_2$ ,  $z'_1$ ,  $\hat{x}'_1$ , and  $\tilde{Q}_{\Lambda}(x')$  denote the quantities obtained via the algorithm (42)–(45) when  $x$  is replaced by  $x'$ . Since  $\hat{Q}_{\Lambda_1}$  and

$\hat{Q}_{\Lambda_2}$  satisfy (2), one finds

$$\begin{aligned} \hat{x}'_2 &= \hat{Q}_{\Lambda_2}(x_2 + v_2 B_2) \\ &= \hat{x}_2 + v_2 B_2, \end{aligned} \quad (46)$$

$$\begin{aligned} z'_1 &= (\hat{x}_2 + v_2 B_2) B_2^{-1} H \\ &= z_1 + v_2 H, \end{aligned} \quad (47)$$

$$\begin{aligned} \hat{x}'_1 &= \hat{Q}_{\Lambda_1}(x_1 + v_1 B_1 + v_2 H - z_1 - v_2 H) \\ &= \hat{x}_1 + v_1 B_1, \end{aligned} \quad (48)$$

$$\begin{aligned} \tilde{Q}_{\Lambda}(x') &= [\hat{x}_1 + v_1 B_1 + z_1 + v_2 H \ \hat{x}_2 + v_2 B_2] \\ &= \tilde{Q}_{\Lambda}(x) + \lambda. \end{aligned} \quad (49)$$

Thus, since it satisfies both conditions,  $\tilde{Q}_{\Lambda}$  is a valid quantization rule.

The NSM of  $\tilde{Q}_{\Lambda}$  is determined by its fundamental decision region, defined by (3). From (45) and (43),  $\tilde{Q}_{\Lambda}(x) = 0$  if and only if  $\hat{x}_1 = 0$  and  $\hat{x}_2 = 0$ . Hence, (3) yields

$$\begin{aligned} \Omega(\tilde{Q}_{\Lambda}) &= \{x \in \mathbb{R}^n : \hat{Q}_{\Lambda_1}(x_1) = 0 \text{ and } \hat{Q}_{\Lambda_2}(x_2) = 0\} \\ &= \Omega(\hat{Q}_{\Lambda_1}) \times \Omega(\hat{Q}_{\Lambda_2}) \\ &= \Omega(\hat{Q}_{\Lambda_p}). \end{aligned} \quad (50)$$

Thus, the fundamental decision region of the (suboptimal) quantization rule for  $\Lambda$  is identical to the fundamental decision region of the optimal quantization rule for  $\Lambda_p$ . This can be intuitively understood by comparing Figs. 2(a) and 2(b).

Since the fundamental decision regions of  $\tilde{Q}_{\Lambda}$  and  $\hat{Q}_{\Lambda_p}$  are identical, so are all properties derived from these regions, e.g.,  $E(\tilde{Q}_{\Lambda}) = E(\hat{Q}_{\Lambda_p})$ ,  $R(\tilde{Q}_{\Lambda}) = R(\hat{Q}_{\Lambda_p})$ , and  $G(\tilde{Q}_{\Lambda}) = G(\hat{Q}_{\Lambda_p})$ . But since the optimal decision rule for  $\Lambda$  satisfies  $G(\hat{Q}_{\Lambda}) \leq G(\tilde{Q}_{\Lambda})$ , our proof is complete:  $G(\hat{Q}_{\Lambda}) \leq G(\hat{Q}_{\Lambda_p})$ . Equality if and only if  $H = 0$  follows, since  $B$  in (41) and  $B_p$  in (19) are equal if and only if  $H = 0$ .  $\square$

In digital communications, the suboptimal quantization rule  $\tilde{Q}_{\Lambda}$  is known as *successive interference cancellation* [20]. In that scenario,  $x_1$  and  $x_2$  represent information received on two parallel communication channels, which interfere with each other. If  $\hat{x}_2$  is detected first (42), its effect on  $x_1$  can be calculated (43) and cancelled (44) before  $\hat{x}_1$  is detected in (45). If (42)–(45) are extended to  $n$  steps of one-dimensional quantization, then the resulting suboptimal quantization rule yields the *Babai point* [21].

The extension to  $k > 2$  follows immediately. Let  $\Lambda$  be the lattice generated by

$$B = \begin{bmatrix} B_1 & \mathbf{0} & \cdots & \mathbf{0} \\ H_{2,1} & B_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ H_{k,1} & H_{k,2} & \cdots & B_k \end{bmatrix}, \quad (51)$$

$\Lambda_i$  be the lattice generated by  $B_i$  for  $i = 1, \dots, k$ , and  $\Lambda_p = \Lambda_1 \times \cdots \times \Lambda_k$ .

**Theorem 7:** For given  $B_1, \dots, B_k$  and any  $H_{i,j}$ ,  $G(\hat{Q}_\Lambda) \leq G(\hat{Q}_{\Lambda_p})$ , with equality if  $H_{i,j} = \mathbf{0}$  for all  $i, j$ .

*Proof:* By induction. If the theorem holds for the upper-left  $(k-1) \times (k-1)$  part of (51), then it extends to  $k \times k$  by Lemma 6.  $\square$

Like Proposition 3, Theorem 7 can also be extended by scale factors  $\mathbf{a}$ . Specifically, if  $B_1, \dots, B_k$  in (51) are multiplied by scale factors  $\mathbf{a} = [a_1, \dots, a_k]$ , then the NSM  $G(\mathbf{a})$  of the resulting lattice is bounded by (32) for arbitrary given  $\mathbf{a}$  or by (36) for optimal  $\mathbf{a}$ .

We now return to the question of whether the optimal product  $\Lambda(\mathbf{a})$  generated by  $B(\mathbf{a})$  in (28) is always locally optimal at  $\mathbf{a} = \mathbf{a}_{\text{opt}}$ . It was observed in Corollary 5 that if  $\Lambda_1, \dots, \Lambda_k$  are locally optimal, then  $G(\mathbf{a}_{\text{opt}})$  is locally *minimal* with respect to perturbations in any submatrix  $a_i B_i$  about its local optimum. On the other hand, it follows from Theorem 7 that  $G(\mathbf{a}_{\text{opt}})$  is locally *maximal* with respect to perturbations about  $\mathbf{0}$  in any submatrix below the block diagonal of  $B(\mathbf{a}_{\text{opt}})$ . The same theorem holds if (51) is replaced by an upper-triangular matrix  $B$ . This proves that  $G(\mathbf{a}_{\text{opt}})$  is also locally maximal with respect to perturbations about  $\mathbf{0}$  *above* the block diagonal of  $B(\mathbf{a}_{\text{opt}})$ . Since the NSM increases for any perturbations in the submatrices on the diagonal of (28) and decreases for any perturbations in submatrices either below or above the block diagonal, the first derivative of  $G$  with respect to these entries must vanish: the NSM has a *saddle point* at  $B(\mathbf{a}_{\text{opt}})$ . In conclusion, all lattices that fulfill (14) are *not* locally optimal.

One way to construct lattices is by *lamination* of a lower-dimensional lattice. To build an  $n$ -dimensional laminated lattice  $\Lambda$ , take a generator matrix  $B_1$  for an  $(n-1)$ -dimensional lattice  $\Lambda_1$  and an arbitrary  $(n-1)$ -dimensional vector  $\mathbf{h}$ . Then construct the  $n \times n$  generator matrix

$$B = \begin{bmatrix} B_1 & \mathbf{0} \\ \mathbf{h} & a \end{bmatrix}. \quad (52)$$

Here  $a > 0$  is a real number, which is the distance between the shifted lattice copies in the direction orthogonal to their subspace, and the vector  $\mathbf{h}$  is the stacking offset in the  $(n-1)$ -plane. Fig. 1 illustrates the Voronoi region of a laminated lattice with  $n = 3$  and  $\mathbf{h} = \mathbf{0}$ .

In the classical lattice literature, “laminated lattices” are built recursively in this way, to maximize the packing density, starting from  $\mathbb{Z}$ . So  $B_1$  is itself the generator for a laminated lattice, and in each recursive iteration,  $\mathbf{h}$  and  $a$  in (52) are selected to maximize the packing density. This construction gives rise to the well-studied  $\Lambda$  and  $K$  series in [22], [23], [1, Sec. 4 of Ch. 5 and Ch. 6].

In this paper, our focus is the NSM rather than the packing density, so we use “laminated lattice” more broadly for *any* lattice generated via (52). This broader meaning is consistent with [13], [24]. Note that to maximize the packing density, the optimal choice of  $\mathbf{h}$  is a “deep hole” in  $\Lambda_1$  (a vertex of the Voronoi region most distant from the origin). This is also (intuitively) a good choice to minimize the NSM, although it may not be optimal.

An upper bound on the quantization performance of a laminated lattice follows directly from the results of Secs. IV and V, as follows.

**Corollary 8:** An  $n$ -dimensional lattice  $\Lambda$  obtained by lamination of an  $(n-1)$ -dimensional lattice  $\Lambda_1$  satisfies, for an arbitrary offset  $\mathbf{h}$  and the optimal layer separation  $a$ ,

$$G(\hat{Q}_\Lambda) \leq \frac{G(\hat{Q}_{\Lambda_1})^{1-\frac{1}{n}}}{12^{\frac{1}{n}}}. \quad (53)$$

*Proof:* Setting  $n_2 = 1$ ,  $\Lambda_2 = a\mathbb{Z}$ , and  $\mathbf{H} = \mathbf{h}$  in Lemma 6 yields

$$G(\hat{Q}_\Lambda) \leq G(\hat{Q}_{\Lambda_p}), \quad (54)$$

where  $\Lambda_p = \Lambda_1 \times a\mathbb{Z}$ . Furthermore, setting  $n_2 = 1$ ,  $\Lambda_2 = a\mathbb{Z}$ ,  $G_1 = G(\hat{Q}_{\Lambda_1})$ , and  $G_2 = G(\hat{Q}_{\mathbb{Z}}) = 1/12$  in Theorem 4 yields

$$G(\hat{Q}_{\Lambda_p}) \leq \frac{G(\hat{Q}_{\Lambda_1})^{1-\frac{1}{n}}}{12^{\frac{1}{n}}}. \quad (55)$$

Combining (54) and (55) completes the proof.  $\square$

As a curiosity, we note that (53) would have been more appealing if  $G$  had been defined a factor of 12 larger than the standard definition (9). With that alternative definition,  $\mathbb{Z}^n$  would have an NSM of 1 in any dimension and the denominator would disappear from expressions like (53) and (55).

## VI. BEST KNOWN LATTICE QUANTIZERS

The upper bound in Corollary 8 has interesting implications. These call to mind an observation made by Cohn in the context of sphere packing [27]. Referring to a plot of sphere-packing density as a function of dimension, he comments that “*Certain dimensions, most notably 24, have packings so good that they seem to pull the entire curve in their direction. The fact that this occurs is not so surprising, since one expects cross sections and stackings of great packings to be at least good, but the effect is surprisingly large.*” Indeed, the sphere-packing density in a given dimension  $n$  is bounded by a function of the sphere-packing density in dimension  $n-1$  according to *Mordell’s inequality* [1, Eq. (19) of Ch. 6]. Here, in the context of lattice quantizers, Corollary 8 does precisely this for NSMs. A lattice with particularly small NSM  $G$  in dimension  $n-1$  makes it possible to also obtain a small NSM in dimension  $n$ , and hence “pulls down the NSM curve” for larger dimensions. More generally, Theorem 7 can pull down the curve over intervals of more than one dimension.

We designed product lattices in dimension  $n$  by applying Theorem 4 with  $k = 2$  to the best known lattices in dimensions  $n_1$  and  $n_2 = n - n_1$ , for  $n_1$  ranging from 1 to  $n-1$ . There is no need to explicitly consider  $k > 2$ , although

TABLE I: The best known lattice quantizers. Columns 2–3 list the smallest normalized second moments (NSM)  $G$  previously reported in dimension  $n$ , while columns 4–5 list the conjectured lower and upper bounds, respectively. Columns 6–7 list our best product lattices. Columns 8–9 indicate if the product lattice in Column 7 provides a smaller NSM than previously reported in Column 2 ( $<G$ ) and/or is below the upper bound of Column 5 ( $<U$ ). NSMs that are known exactly are listed with nine decimals, whereas NSMs with five decimals are derived from numerical estimates in [8].

$n$	Best previously reported		Generic bounds		Best product		
	NSM	Lattice	Lower [25]	Upper [26]	NSM	Lattice	Better?
1	0.083333333	$\mathbb{Z}$	0.083333333	0.083333333			
2	0.080187537	$A_2$	0.080187537	0.080267223	0.083333333	$\mathbb{Z} \otimes \mathbb{Z}$	
3	0.078543281	$A_3^*$	0.077874985	0.079723711	0.081222715	$A_2 \otimes \mathbb{Z}$	
4	0.076603235	$D_4$	0.076087080	0.078822518	0.079714343	$A_3^* \otimes \mathbb{Z}$	
5	0.075625443	$D_5^*$	0.074654327	0.078731261	0.077904301 <sup>7</sup>	$D_4 \otimes \mathbb{Z}$	$<U$
6	0.074243697	$E_6^*$	0.073474906	0.077779975	0.076858706	$D_5^* \otimes \mathbb{Z}$	$<U$
7	0.073116493	$E_7^*$	0.072483503	0.076858058	0.075478834	$E_6^* \otimes \mathbb{Z}$	$<U$
8	0.071682099	$E_8$	0.071636064	0.075654034	0.074321725	$E_7^* \otimes \mathbb{Z}$	$<U$
9	0.071622594	$AE_9$	0.070901661	0.075552237	0.072891732	$E_8 \otimes \mathbb{Z}$	$<U$
10	0.070813818	$D_{10}^+$	0.070257874	0.074856027	0.072715487	$AE_9 \otimes \mathbb{Z}$	$<U$
11	0.070426259	$A_{11}^3$	0.069688002	0.074026177	0.071869620	$D_{10}^+ \otimes \mathbb{Z}$	$<U$
12	0.070095600	$K_{12}$	0.069179323	0.073098569	0.071420842	$A_{11}^3 \otimes \mathbb{Z}$	$<U$
13	0.074873919	$D_{13}^*$	0.068721956	0.072400247	0.071034583	$K_{12} \otimes \mathbb{Z}$	$<G$ $<U$
14	0.074954492	$D_{14}^*$	0.068308096	0.071672217	0.071455542	$K_{12} \otimes A_2$	$<G$ $<U$
15	0.075039738	$D_{15}^*$	0.067931488	0.071008692	0.071709124	$K_{12} \otimes A_3^*$	$<G$
16	0.06830	$\Lambda_{16}$	0.067587055	0.070399705	0.071668753	$K_{12} \otimes D_4$	
17	0.075216213	$D_{17}^*$	0.067270625	0.069886791	0.06910	$\Lambda_{16} \otimes \mathbb{Z}$	$<G$ $<U$
18	0.075304924	$D_{18}^*$	0.066978741	0.069403282	0.06953	$\Lambda_{16} \otimes A_2$	$<G$
19	0.075392902	$D_{19}^*$	0.066708503	0.068958664	0.06982	$\Lambda_{16} \otimes A_3^*$	$<G$
20	0.075479665	$D_{20}^*$	0.066457468	0.068548490	0.06988	$\Lambda_{16} \otimes D_4$	$<G$
21	0.075554858	$A_{21}^*$	0.066223553	0.068173104	0.06998	$\Lambda_{16} \otimes D_5^*$	$<G$
22	0.075577414	$A_{22}^*$	0.066004976	0.067826205	0.06987	$\Lambda_{16} \otimes E_6^*$	$<G$
23	0.075601888	$A_{23}^*$	0.065800200	0.067506055	0.06973	$\Lambda_{16} \otimes E_7^*$	$<G$
24	0.06577	$\Lambda_{24}$	0.065607893	0.067208741	0.06941	$\Lambda_{16} \otimes E_8$	
25	0.075655156	$A_{25}^*$	0.065426891	0.066939780	0.06640	$\Lambda_{24} \otimes \mathbb{Z}$	$<G$ $<U$
26	0.075683386	$A_{26}^*$	0.065256179	0.066685133	0.06678	$\Lambda_{24} \otimes A_2$	$<G$
27	0.075712385	$A_{27}^*$	0.065094858	0.066446178	0.06708	$\Lambda_{24} \otimes A_3^*$	$<G$
28	0.075741975	$A_{28}^*$	0.064942137	0.066222221	0.06722	$\Lambda_{24} \otimes D_4$	$<G$
29	0.075772009	$A_{29}^*$	0.064797312	0.066012092	0.06737	$\Lambda_{24} \otimes D_5^*$	$<G$
30	0.075802366	$A_{30}^*$	0.064659756	0.065814436	0.06738	$\Lambda_{24} \otimes E_6^*$	$<G$
31	0.075832940	$A_{31}^*$	0.064528911	0.065628191	0.06736	$\Lambda_{24} \otimes E_7^*$	$<G$
32	0.075863646	$A_{32}^*$	0.064404271	0.065452387	0.06720	$\Lambda_{24} \otimes E_8$	$<G$
33	0.075894409	$A_{33}^*$	0.064285386	0.065286162	0.06732	$\Lambda_{24} \otimes AE_9$	$<G$
34	0.075925169	$A_{34}^*$	0.064171846	0.065128681	0.06722	$\Lambda_{24} \otimes D_{10}^+$	$<G$
35	0.075955874	$A_{35}^*$	0.064063282	0.064979253	0.06720	$\Lambda_{24} \otimes A_{11}^3$	$<G$
36	0.075986480	$A_{36}^*$	0.063959359	0.064837252	0.06718	$\Lambda_{24} \otimes K_{12}$	$<G$
37	0.076016949	$A_{37}^*$	0.063859771	0.064702115	0.06757	$\Lambda_{24} \otimes K_{12} \otimes \mathbb{Z}$	$<G$
38	0.076047252	$A_{38}^*$	0.063764240	0.064573335	0.06781	$\Lambda_{24} \otimes K_{12} \otimes A_2$	$<G$
39	0.076077363	$A_{39}^*$	0.063672511	0.064450455	0.06799	$\Lambda_{24} \otimes K_{12} \otimes A_3^*$	$<G$
40	0.076107259	$A_{40}^*$	0.063584352	0.064333062	0.06677	$\Lambda_{24} \otimes \Lambda_{16}$	$<G$
41	0.076136923	$A_{41}^*$	0.063499548	0.064220781	0.06713	$\Lambda_{24} \otimes \Lambda_{16} \otimes \mathbb{Z}$	$<G$
42	0.076166341	$A_{42}^*$	0.063417902	0.064113272	0.06736	$\Lambda_{24} \otimes \Lambda_{16} \otimes A_2$	$<G$
43	0.076195500	$A_{43}^*$	0.063339234	0.064010223	0.06753	$\Lambda_{24} \otimes \Lambda_{16} \otimes A_3^*$	$<G$
44	0.076224390	$A_{44}^*$	0.063263376	0.063911352	0.06761	$\Lambda_{24} \otimes \Lambda_{16} \otimes D_4$	$<G$
45	0.076253004	$A_{45}^*$	0.063190174	0.063816399	0.06770	$\Lambda_{24} \otimes \Lambda_{16} \otimes D_5^*$	$<G$
46	0.076281336	$A_{46}^*$	0.063119483	0.063725126	0.06770	$\Lambda_{24} \otimes \Lambda_{16} \otimes E_6^*$	$<G$
47	0.076309381	$A_{47}^*$	0.063051171	0.063637315	0.06768	$\Lambda_{24} \otimes \Lambda_{16} \otimes E_7^*$	$<G$
48	0.076337136	$A_{48}^*$	0.062985115	0.063552764	0.06577	$\Lambda_{24} \otimes \Lambda_{24}$	$<G$

<sup>7</sup>Proposed as an upper bound in [6, Tab. I].



recursive application of Theorem 4 with  $k = 2$  can generate product lattices with larger  $k$ . The minimal NSM obtained in each dimension provides a constructive upper bound on the optimal NSM. It follows from Theorem 7 that better lattice quantizers can be found among lattices of the form (51), where  $B_1, \dots, B_k$  are the best known lattices in their dimensions, for example by lamination if  $n_i = 1$  for any  $i$ .

Tab. I summarizes the best known lattice quantizers in dimensions  $n \leq 48$ . The first such list was compiled in 1979 for  $n = 1$  to 5 [6]. It was extended to  $n \leq 10$  in 1982 [7], which also analytically calculated the NSMs of the classical lattices  $A_n$ ,  $A_n^*$ ,  $D_n$ , and  $D_n^*$  for any  $n$ . Since then, progress has been much slower. Better lattices were reported for  $n = 6$  and 7 in [8] and for  $n = 9$  and 10 in [11], although the NSMs of these lattices were only computed numerically. Reference [8] also gave the best known lattice quantizers for  $n = 12, 16$ , and 24 with numerically computed NSMs. Later, the corresponding exact NSMs were calculated for  $n = 6$  [9],  $n = 7$  [10],  $n = 9$  [13], and  $n = 10$  and 12 [12]. In the latter reference, a new best known lattice quantizer was identified for  $n = 11$ . NSM results for  $n \leq 15$  were summarized in [4].

In Tab. I, we also include the best known lower and upper bounds on the NSM. The lower bound is a conjecture by Conway and Sloane [25], which we evaluated numerically by high-resolution trapezoidal integration. The upper bound is by Torquato in [26, Eq. (105)]<sup>8</sup> and improves on a well-known bound by Zador [16, Lemma 5]. The Torquato upper bound is for arbitrary quantizers, which might not be lattices. We conjecture that there is always at least one lattice quantizer satisfying this bound. For the best known packing densities, which are needed to evaluate [26, Eq. (105)], we use [1, Tab. I.1]. Taken together, these two results provide conjectured lower and upper bounds on the NSM of optimal lattice quantizers. While intuitively plausible, these remain unproven.

For each dimension, the last four columns of Tab. I list the best product lattice that can be constructed from the lattices given on the preceding rows of the table. While Theorem 7 establishes that these lattices are not even locally optimal, they nevertheless improve significantly on the previously lowest reported NSMs. For brevity, we use  $\otimes$  to denote the Cartesian product of lattices with the optimal choice of relative scale. More precisely,  $\Lambda_1 \otimes \dots \otimes \Lambda_k \triangleq a_1 \Lambda_1 \times \dots \times a_k \Lambda_k$  with  $[a_1, \dots, a_k]$  given by Theorem 4.

The first dimension in which this construction provides a better lattice quantizer than previously reported is  $n = 13$ , where our best product lattice is  $K_{12} \otimes \mathbb{Z}$ . With an NSM of 0.0710, it is the first reported 13-dimensional lattice whose NSM falls below the generic upper bound 0.0724. It is significantly better than the currently best known 13-dimensional lattice quantizer  $D_{13}^*$ , whose NSM is 0.0749. Our optimized product lattices are also the first reported lattices which lie below the upper bound in dimensions  $n = 14, 17$ , and 25.

## VII. CONCLUSIONS

As far as we know, Table I is currently the most extensive published table of record NSM lattices.<sup>9</sup> Previous tables have not gone beyond  $n = 16$ , apart from a numerical estimate for  $n = 24$ . It is disconcerting that in many dimensions the smallest known NSMs are for product lattices, since we have proven that these cannot be optimal. This reflects the complexity of designing good lattices and of evaluating their NSMs. The theoretical results in this paper provide some properties of optimal lattice quantizers, which in combination with the tabulated product lattices may hopefully provide guidance and benchmarks for further progress in lattice quantization. An analytic or algorithmic method to optimize  $H$  in (41) or  $h$  in (52) would be highly desirable.

## REFERENCES

- [1] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed. New York, NY: Springer, 1999. [Online]. Available: <https://doi.org/10.1007/978-1-4757-6568-7>
- [2] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge, UK: Cambridge University Press, 2014. [Online]. Available: <https://doi.org/10.1017/CBO9781139045520>
- [3] F. A. Hamprecht and E. Agrell, "Exploring a space of materials: Spatial sampling design and subset selection," in *Experimental Design for Combinatorial and High Throughput Materials Development*, J. N. Cawse, Ed. New York, NY: Wiley, 2003, ch. 13. [Online]. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.20.6178&rep=rep1&type=pdf>
- [4] B. Allen, "Optimal template banks," *Phys. Rev. D*, vol. 104, p. 042005, Aug 2021. [Online]. Available: <https://doi.org/10.1103/PhysRevD.104.042005>
- [5] W. Lerche, A. N. Schellekens, and N. P. Warner, "Lattices and strings," *Physics Reports*, vol. 177, no. 1–2, pp. 1–140, May 1989. [Online]. Available: [https://doi.org/10.1016/0370-1573\(89\)90077-X](https://doi.org/10.1016/0370-1573(89)90077-X)
- [6] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 4, pp. 373–380, July 1979. [Online]. Available: <https://doi.org/10.1109/TIT.1979.1056067>
- [7] J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 211–226, Mar. 1982. [Online]. Available: <https://doi.org/10.1109/TIT.1982.1056483>
- [8] —, "On the Voronoi regions of certain lattices," *SIAM J. Alg. Disc. Meth.*, vol. 5, no. 3, pp. 294–305, Sept. 1984. [Online]. Available: <https://doi.org/10.1137/0605031>
- [9] R. T. Worley, "The Voronoi region of  $E_6^*$ ," *J. Austral. Math. Soc. (Series A)*, vol. 43, no. 2, pp. 268–278, Oct. 1987. [Online]. Available: <https://doi.org/10.1017/S1446788700029402>
- [10] —, "The Voronoi region of  $E_7^*$ ," *SIAM J. Disc. Math.*, vol. 1, no. 1, pp. 134–141, Feb. 1988. [Online]. Available: <https://doi.org/10.1137/0401015>
- [11] E. Agrell and T. Eriksson, "Optimization of lattices for quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1814–1828, Sept. 1998. [Online]. Available: <https://doi.org/10.1109/18.705561>
- [12] M. Dutour Sikirić, A. Schürmann, and F. Vallentin, "Complexity and algorithms for computing Voronoi cells of lattices," *Mathematics of Computation*, vol. 78, no. 267, pp. 1713–1731, July 2009. [Online]. Available: <https://doi.org/10.1090/S0025-5718-09-02224-8>
- [13] B. Allen and E. Agrell, "The optimal lattice quantizer in nine dimensions," *Annalen der Physik*, vol. 533, no. 12, p. 2100259, Dec. 2021. [Online]. Available: <https://doi.org/10.1002/andp.202100259>
- [14] E. S. Barnes and N. J. A. Sloane, "The optimal lattice quantizer in three dimensions," *SIAM J. Alg. Disc. Meth.*, vol. 4, no. 1, pp. 30–41, Mar. 1983. [Online]. Available: <https://doi.org/10.1137/0604005>
- [15] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, July 1996. [Online]. Available: <https://doi.org/10.1109/18.508838>

<sup>8</sup>See [28, Note 31] for corrections.

<sup>9</sup>Several improved lattice quantizers in dimensions  $n \leq 24$  are reported in [29], which appeared while this paper was in review.

- [16] P. L. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, vol. IT-82, no. 2, pp. 139–149, Mar. 1982. [Online]. Available: <https://doi.org/10.1109/TIT.1982.1056490>
- [17] R. M. Wald, *General Relativity*. Chicago: Chicago University Press, 1984. [Online]. Available: <https://doi.org/10.7208/chicago/9780226870373.001.0001>
- [18] L. Perko, *Differential Equations and Dynamical Systems*, 3rd ed. New York, NY: Springer, 2001. [Online]. Available: <https://doi.org/10.1007/978-1-4613-0003-8>
- [19] E. Agrell *et al.*, "Online supplements." [Online]. Available: <https://codes.se/supplements>
- [20] D. Wübben, D. Seethaler, J. Jaldén, and G. Matz, "Lattice reduction," *IEEE Sig. Proc. Mag.*, vol. 28, no. 3, pp. 70–91, May 2011.
- [21] L. Babai, "On Lovász' lattice reduction and the nearest lattice point problem," *Combinatorica*, vol. 6, no. 1, pp. 1–13, Mar. 1986. [Online]. Available: <https://doi.org/10.1007/BF02579403>
- [22] J. H. Conway and N. J. A. Sloane, "Laminated lattices," *Annals of Mathematics*, vol. 116, no. 3, pp. 593–620, Nov. 1982. [Online]. Available: <https://doi.org/10.2307/2007025>
- [23] W. Plesken and M. Pohst, "Constructing integral lattices with prescribed minimum. II," *Mathematics of Computation*, vol. 60, no. 202, pp. 817–825, Apr. 1993. [Online]. Available: <https://doi.org/10.1090/S0025-5718-1993-1176715-1>
- [24] M. Dutour Sikirić, A. Schürmann, and F. Vallentin, "A generalization of Voronoi's reduction theory and its application," *Duke Mathematical Journal*, vol. 142, no. 1, pp. 127–164, Mar. 2008. [Online]. Available: <https://doi.org/10.1215/00127094-2008-003>
- [25] J. H. Conway and N. J. A. Sloane, "A lower bound on the average error of vector quantizers," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 1, pp. 106–109, Jan. 1985. [Online]. Available: <https://doi.org/10.1109/TIT.1985.1056993>
- [26] S. Torquato, "Reformulation of the covering and quantizer problems as ground states of interacting particles," *Phys. Rev. E*, vol. 82, no. 5, p. 056109, Nov. 2010. [Online]. Available: <https://doi.org/10.1103/PhysRevE.82.056109>
- [27] H. Cohn, "A conceptual breakthrough in sphere packing," *Notices Amer. Math. Soc.*, vol. 64, no. 2, pp. 102–115, Feb. 2017. [Online]. Available: <https://doi.org/10.1090/noti1474>
- [28] B. Allen, "Performance of random template banks," *Phys. Rev. D*, vol. 105, no. 10, p. 102003, May 2022. [Online]. Available: <https://doi.org/10.1103/PhysRevD.105.102003>
- [29] S. Lyu, Z. Wang, C. Ling, and H. Chen, "Better lattice quantizers constructed from complex integers," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 7932–7940, Dec. 2022. [Online]. Available: <https://doi.org/10.1109/TCOMM.2022.3215685>