

# Lattice-Based WOM Codebooks that Allow Two Writes

Brian M. Kurkoski

Japan Advanced Institute of Science and Technology

Nomi, Japan

kurkoski@jaist.ac.jp

**Abstract**—The continuous approximation is a technique to separate the shaping gain and coding gain of a channel code. In this paper, this technique is applied to codes for write-once memories (WOM codes) based upon lattices. For a lattice of arbitrary dimension  $n$ , a hyperbolic shaping region is optimal in the sense of maximizing the sum rate in the worst case, when there are two writes. Then, asymptotic results are obtained when the rates for two writes are equal. Under this condition, the sum rate assuming two equal rates closely approaches, but not achieve, the capacity which allows two unequal rates.

## I. INTRODUCTION

Flash memories physically deteriorate, and tolerate a limited number of erase cycles before they fail. With the motivation to extend the life of flash memories, recent coding-theoretic approaches allow multiple writes to flash memories without erasing [3], [4]. This has revived interest in the earlier work of Rivest and Shamir [5] on codes for “write-once memories,” called WOM codes.

While some WOM codes have been constructed considering error-correction [6], many constructions lack the capability to correct errors. Since flash memory read and write processes introduce noise, powerful error-correcting codes, typically BCH and LDPC codes, are used in practice. When an error-correcting code is applied to multi-level cells, typically using Gray coding, the codewords form a sphere packing [7].

An alternative approach is to perform error-correction using lattices. Lattices are linear sphere packings, and in particular, an  $n$ -dimensional lattice forms an additive subgroup of  $\mathbb{R}^n$ . While lattices have been studied for at least a century, only comparatively recently has it been shown that lattices can achieve the Shannon capacity of the AWGN channel [8].

However, the WOM-like properties of lattices are not well understood. A lattice-based construction with efficient encoding was proposed and evaluated from the perspective of average number of writes [9]. An earlier version of the present paper [1] first proposed using a continuous approximation of a lattice, and hypothesized that hyperbolic shaping regions were optimal for two writes for two-write lattices in arbitrary lattice dimension  $n$ . Based upon that work, Bhatia, Iyengar, and Siegel showed that hyperbolic shaping regions were optimal for an arbitrary number of writes, when the dimension is  $n = 2$  [2]. They also constructed some specific codes for  $n = 2$ .

The present paper continues this line of research. First, code constructions and their rates based on shaping regions

and lattices are given. Then, the continuous approximation is invoked and a corresponding definition of code rate is given. Stronger justification is given for the claim, for arbitrary dimension  $n$  and two writes, that hyperbolic shaping is optimal in the sense of maximizing the sum rate. Finally, applying this construction to the condition of two writes of equal rate, a lower bound on the sum rate is given. This sum rate closely approaches, but not achieve, neither the capacity (condition of two unequal rates) nor known upper bounds on achievable rates (condition of two equal rates).

## II. CODE CONSTRUCTION

### A. Lattices and Generalized Inequality

An  $n$ -dimensional lattice  $\Lambda$  is defined by an  $n$ -by- $n$  generator matrix  $G$ . The lattice consists of the discrete set of points  $x = (x_1, x_2, \dots, x_n)^t$  for which

$$x = G \cdot b, \quad (1)$$

where  $b = (b_1, \dots, b_n)^t$  is from the set of all possible integer vectors,  $b_i \in \mathbb{Z}$ . The Voronoi region is region of  $\mathbb{R}^n$  which is closer to  $x$  than to any other point, and the volume of this region  $\text{Vol}(\Lambda)$  is the determinant of  $G$ :

$$\text{Vol}(\Lambda) = |\det G|. \quad (2)$$

A partial ordering for two vectors  $x$  and  $y$  is:

$$x \preceq y \quad (3)$$

if and only if  $x_i \leq y_i$  for all  $i = 1, 2, \dots, n$ . A point  $y$  is said to be “accessible” from  $x$  if  $x \preceq y$ . This can be seen as a generalize inequality [10, Sec. 2.4], defined by a cone given by the unit vectors in  $n$  dimensions:  $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ .

### B. Codebook

The construction of the codebook concentrates on selection of a lattice  $\Lambda$ , and the partition of this codebook into sub-codebooks which correspond to WOM writes. This is distinct from any encoding, that is, any mapping from information to elements of the codebook.

An overview of the codebook construction follows. The code is described by a lattice  $\Lambda$  with a generator matrix  $G$ . Define a shaping region for the code as  $\mathbb{A} \subseteq \mathbb{R}^n$ , which forms the code  $\mathbb{L}$  as:

$$\mathbb{L} = \Lambda \cap \mathbb{A}. \quad (4)$$



(To be clear, adding the argument  $s$  to  $\mathbb{L}_i$  makes a distinct  $\mathbb{L}_i(s)$ , and  $\mathbb{L}_i(s)$  is a subset of  $\mathbb{L}_i$ .) Then, the minimum number of messages available for the second write is  $M_2$ :

$$M_2 = \min_{s \in \mathbb{L}_1} |\mathbb{L}_2(s)| \quad (9)$$

The normalized rate for the second write is:

$$\tilde{R}_2 = \frac{\log_2 M_2}{\log_2 |\mathbb{L}|}. \quad (10)$$

This definition results in the generally-accepted notion that the code rate for the second write should be independent of the data written for the first write. This applies in the worst-case scenario that is considered in this paper.

### III. CONTINUOUS APPROXIMATION

#### A. Continuous Approximation

From here, it is assumed that the lattice is sufficiently fine that the code rates can be approximated by continuous volumes. Forney and Wei introduced the continuous approximation as a method to separate the contribution of the shaping region and the lattice  $\Lambda$  to the total transmit power in coded modulation for AWGN channels [11]. Here, the same approximation is used with the goal of describing idealized shaping regions for WOM codes. Under the continuous approximation, the number of codewords  $|\mathbb{L}|$  is approximated as:

$$|\mathbb{L}| \approx \frac{\text{Vol}(\mathbb{A})}{\text{Vol}(\Lambda)} \quad (11)$$

Using this notion, code rates can be defined with respect to the volume of the shaping regions. As in the previous section, the entire space is partitioned into disjoint regions,  $\mathbb{A}_1$  and  $\mathbb{A}_2$ .

#### B. Code Rates Under the Continuous Approximation

It is assumed  $\ell = 1$ , and so that the entire space is the unit  $n$ -cube, that is  $\mathbb{A} = [0, 1]^n$  and  $\text{Vol}(\mathbb{A}) = 1$ . Because the continuous approximation has been made, this results in no loss of generality. Further, rather than working with discrete code points, continuous volumes are of interest.

Analogous to  $\mathbb{L}_i(s)$ , let  $\mathbb{A}_i(s)$  denote the subset of  $\mathbb{A}_i$  that remains accessible when the current state is  $s$ , that is:

$$\mathbb{A}_i(s) = \{x \in \mathbb{A}_i | x \succeq s\}, \quad (12)$$

and furthermore define  $V_i(s)$  as the volume of  $\mathbb{A}_i(s)$ :

$$V_i(s) = \text{Vol}(\mathbb{A}_i(s)). \quad (13)$$

Analogous to  $M_2$ , define  $V_2$  as:

$$V_2 = \min_{s \in \mathbb{A}_1} \text{Vol}(\mathbb{A}_2(s)). \quad (14)$$

Define  $V_1 = \text{Vol}(\mathbb{A}_1)$ .

For  $\mathbb{L}_i$ , which has  $|\mathbb{L}_i|$  codewords, the continuous approximation gives:

$$|\mathbb{L}_i| \approx \frac{V_i}{\text{Vol}(\Lambda)}. \quad (15)$$

Explicitly, the normalized rate under the continuous approximation is obtained by applying (11) and (15):

$$\tilde{R}_1 = \frac{\log_2 |\mathbb{L}_1|}{\log_2 |\mathbb{L}|} \approx \frac{\log_2 \frac{V_1}{\text{Vol}(\Lambda)}}{\log_2 \frac{1}{\text{Vol}(\Lambda)}} \text{ and} \quad (16)$$

$$\tilde{R}_2 = \frac{\log_2 M_2}{\log_2 |\mathbb{L}|} \approx \frac{\log_2 \frac{V_2}{\text{Vol}(\Lambda)}}{\log_2 \frac{1}{\text{Vol}(\Lambda)}}, \quad (17)$$

where  $\text{Vol}(\mathbb{A}) = 1$  was used.

Thus, the normalized rate for write  $i = 1, 2$  is:

$$\tilde{R}_i \approx 1 - \frac{\log_2 V_i}{\log_2 \text{Vol}(\Lambda)}. \quad (18)$$

The normalized sum rate  $\tilde{R}$  for two writes is:

$$\tilde{R} = \tilde{R}_1 + \tilde{R}_2 = 2 - \frac{\log_2 V_1 + \log_2 V_2}{\log_2 \text{Vol}(\Lambda)}. \quad (19)$$

Some of these sets and variables are expressed in Fig. 1 using the  $D_2$  lattice (the integral version of the  $D_2$  checkerboard lattice has points where the sum of the two coordinates is even).

#### C. Restriction to Boundary $\mathbb{B}$

It is shown that the boundary  $\mathbb{B}$  determines  $V_2$ :

$$\min_{s \in \mathbb{A}_1} \text{Vol}(\mathbb{A}_2(s)) = \min_{s \in \mathbb{B}} \text{Vol}(\mathbb{A}_2(s)), \quad (20)$$

and thus we can instead use  $\mathbb{B}$  for determining  $V_2$  in (14).

Note that the volume of  $\mathbb{A}_2(s)$  is:

$$\text{Vol}(\mathbb{A}_s(s)) = (1 - s_1)(1 - s_2) \cdots (1 - s_n), \quad (21)$$

and so (14) is a linear optimization problem. It is known that the solution to (14) can only occur at an extreme point of  $\mathbb{A}_2$  [12, Theorem 1.19], and the extreme points are only on the boundary,  $\mathbb{B}$ . Thus, it is sufficient to consider the boundary.

#### D. Characterization of $\mathbb{B}$ as a Hyperbola

This section shows that the boundary  $\mathbb{B}$  separating  $\mathbb{A}_1$  and  $\mathbb{A}_2$  which maximizes the normalized sum rate  $\tilde{R}_1 + \tilde{R}_2$  is a hyperbola. To do so, the following Lemma is key.

*Lemma Making  $V_2(s)$  equal to a constant  $V_2$  independent of  $s \in \mathbb{B}$  will maximize  $V_1 + V_2$ .*

*Proof sketch* The proof is by contradiction. Consider some boundary  $\mathbb{B}$  with two distinct  $u, v \in \mathbb{B}$  which satisfies the hypothesis, that is  $V_2(u) = V_2(v)$ . Then, consider an alternative boundary  $\mathbb{B}'$  for which the hypothesis does not hold, say:

$$V_2(u) < V_2(v'), \quad (22)$$

for distinct  $u, v' \in \mathbb{B}$ . Here,  $\mathbb{B}$  and  $\mathbb{B}'$  share a common  $u$ . In addition, let  $V'_1$  and  $V'_2$  be the volumes corresponding to region  $\mathbb{B}'$ . Refer to Fig. 1.

Since  $V_2 \leq \min(V_2(u), V_2(v))$ , clearly  $V_2 \leq V_2(u)$  alone. Since  $V_2(u) < V_2(v')$ , it must hold that  $V'_1 < V_1$  (intuitively, this can be shown as some incremental volume moved from

$\mathbb{A}_1$  to  $\mathbb{A}_2$  will decrease the volume of  $\mathbb{A}_1 = V_1$ .) Now, assume the opposite, that  $\mathbb{B}'$  maximizes  $V_1 + V_2$ , so that:

$$V_1 + V_2 < V_1' + V_2'. \quad (23)$$

Since  $V_2' \leq V_2$  and  $V_1' < V_1$ , this assumption cannot hold, and selecting  $\mathbb{B}$  such that  $\text{Vol}(\mathbb{A}_2(s))$  is independent of  $s$  will maximize  $V_1 + V_2$ .  $\square$

Let  $s = (s_1, s_2, \dots, s_n)$  be any point on  $\mathbb{B}$ . By the above lemma, for any such  $s$ :

$$V_2 = \prod_{i=1}^n (1 - s_i) \quad (24)$$

This describes a multidimensional hyperbola. It is convenient to characterize  $\mathbb{B}$  with a parameter  $\omega$ , where  $(\omega, 0, 0, \dots, 0)$ , or any permutation of positions, is a point on  $\mathbb{B}$ , so that:

$$V_2 = 1 - \omega, \quad (25)$$

for  $0 \leq \omega \leq 1$ .

#### E. Computation of $V_1$

$\mathbb{A}_1$  is the region bounded by the hyperbola  $\mathbb{B}$  and the hyperplanes  $x_i \geq 0$ . The volume is  $V_1$ , which is expressed as an  $n - 1$ -fold integral. Fortunately, it is possible to find  $V_1$  in closed form. By a defining parameter  $z$ :

$$z = -\log_e(1 - \omega), \quad (26)$$

(equivalently  $\omega = 1 - e^{-z}$ ), the volume  $V_1$ , is:

$$V_1 = 1 - e^{-z} \sum_{m=0}^{n-1} \frac{z^m}{m!}. \quad (27)$$

This relationship can be found using symbolic mathematical software. The key is to solve an  $(n - 1)$ -fold integral recursively. The description is omitted.

### IV. ASYMPTOTIC RESULTS FOR EQUAL RATES

#### A. Equal rates

Of practical interest is the situation where the normalized rates  $\tilde{R}_1$  and  $\tilde{R}_2$  are equal:

$$1 - \frac{\log_2 V_1}{\log_2 \text{Vol}(\Lambda)} = 1 - \frac{\log_2 V_2}{\log_2 \text{Vol}(\Lambda)} \quad (28)$$

which is equivalent to:

$$V_1 = V_2. \quad (29)$$

The parameter  $z_n^*$  for which the two rates are equal is the root of:

$$1 - e^{-z} \sum_{m=0}^{n-1} \frac{z^m}{m!} = 1 - (1 - e^{-z}), \quad (30)$$

Unfortunately,  $z_n^*$  can be found only numerically when  $n \geq 2$ . However, a  $z_n^*$  upper bound can be found by using the series expansion of  $e^z$ . Eqn. (30) can be written as:

$$1 = \frac{z^n}{n!} + \frac{z^{n+1}}{(n+1)!} + \frac{z^{n+2}}{(n+2)!} + \dots \quad (31)$$

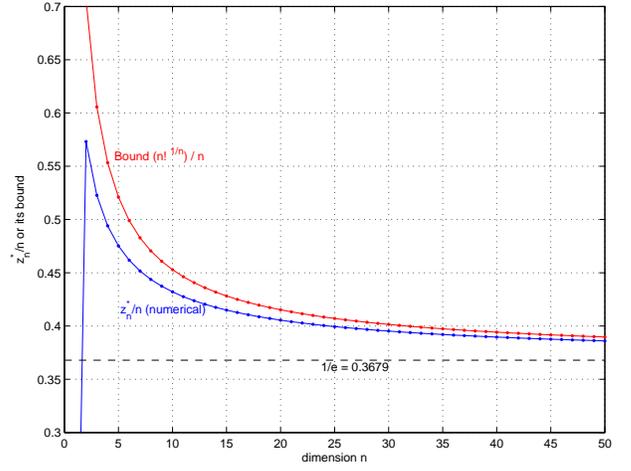


Fig. 2. Comparison of  $z_n^*/n$  and its bound,  $\frac{1}{n} \sqrt[n]{n!}$ .

Since all terms are positive, an upper bound on the root  $z_n^*$  is a root of  $1 = z^n/n!$ , that is:

$$z_n^* \leq (n!)^{\frac{1}{n}} \quad (32)$$

which appears to be reasonably tight. Noting that:

$$\frac{\sqrt[n]{n!}}{n} \geq \frac{1}{e} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{\sqrt[n]{n!}}{n} = \frac{1}{e}, \quad (33)$$

the numerical values for  $\frac{z_n^*}{n}$  and its bound are compared in Fig. 2 for small  $n$ . In addition, the following lower bound is hypothesized:

$$\frac{1}{e} \stackrel{\text{hyp}}{\leq} \frac{z_n^*}{n} \leq \frac{1}{n} \sqrt[n]{n!}, \quad (34)$$

for  $n \geq 2$ . Here “hyp” indicates the bound is hypothesized.

Thus, for two writes with equal rates:

$$\exp\left(-\sqrt[n]{n!}\right) \leq V_1, V_2 \stackrel{\text{hyp}}{\leq} \exp\left(-\frac{n}{e}\right) \quad (35)$$

since  $V_1 = V_2 = e^{-z^*}$ .

#### B. Asymptotic Rate for Cubic Lattice

The lattice  $\Lambda$  considered is  $\frac{1}{q-1}\mathbb{Z}^n$ , which corresponds to the conventional coding schemes applied to cells with  $q$ -ary values. The volume of the Voronoi cell is  $\text{Vol}(\Lambda) = 1/(q-1)^n$ .

In the case of two equal rates, the bounds on the normalized sum rate (19) is obtained from (35):

$$2 - \frac{2 \log_2 \exp\left(-\sqrt[n]{n!}\right)}{-n \log_2(q-1)} \leq \tilde{R} \stackrel{\text{hyp}}{\leq} 2 - \frac{2 \log_2 \exp\left(-\frac{n}{e}\right)}{-n \log_2(q-1)} \quad (36)$$

where we should take care to note that  $\text{Vol}(\Lambda) < 1$  so  $\log_2 \text{Vol}(\Lambda) < 0$ . This reduces to:

$$2 - \frac{2}{\log_e 2 \log_2 q - 1} \left( \frac{\sqrt[n]{n!}}{n} \right) \leq \tilde{R} \stackrel{\text{hyp}}{\leq} 2 - \frac{2}{\log_e 2 \log_2 q - 1} \frac{1}{e}.$$

The main object of interest is the asymptotic  $n \rightarrow \infty$  normalized sum rate. Noting the the lower bound converges to the upper bound:

$$\lim_{n \rightarrow \infty} \tilde{R} = 2 - \frac{2}{e \log_e 2 \log_2(q-1)}, \quad (37)$$

using (33).

The capacity was found by Fu and Han Vinck [13]. For  $t$  writes with  $q$  levels, a (conventional) sum rate of:

$$R_{\text{cap}} = \log_2 \binom{q+t-1}{t}, \quad (38)$$

is achievable. The reader may recall that this paper deals with a normalized rate, see (7). Converting Fu and Han Vinck's capacity to a normalized sum rate when  $t = 2$  gives:

$$\tilde{R}_{\text{cap}} = \frac{R_{\text{cap}}}{\log_2 q} \quad (39)$$

$$= 1 - \frac{1}{\log_2 q} + \frac{\log_2 q - 1}{\log_2 q}, \quad (40)$$

so that  $0 \leq \tilde{R}_{\text{cap}} \leq 2$ , since  $t = 2$ .

In addition, Gabrys and Dolecek gave lower and upper bounds on the capacity of  $q$ -ary code for two writes, when the rate for the two rates are equal, that is  $R_1 = R_2$  [14]. The upper bound is:

$$R_{\text{cap}}^{\text{equal}} \leq \frac{2}{3} \log_q \left( \frac{q(q+1)(2q+1)}{6} \right) \quad (41)$$

The hyperbolic shaping bound (37), the capacity (40), as well as the upper (41) and lower bounds [14, Theorem 3] on  $R_{\text{cap}}^{\text{equal}}$  are plotted in Fig. 3 for various values of  $q$  (normalized rates are shown). It can be seen that the hyperbolic shaping bound gradually approaches, but does not achieve, the upper bound of Gabrys and Dolecek.

There are several possible explanations for this gap. One is that since (37) is based on a bound, the bound is not tight. Another possibility is that (41) is not tight. A third possibility concerns achievability of encoding methods. This paper only gave a codebook construction method, but encodings, that is mappings from information to codewords, were not discussed. While it is clear there is a bijective mapping from information to the first codebook  $\mathbb{L}_1$ , the second codebook  $\mathbb{L}_2$  requires multiple lattice points that correspond to the same information. For  $n = 2$ , a construction that effectively uses all the lattice points is possible [2], but for  $n \geq 3$ , this remains an open question.

## V. ACKNOWLEDGMENTS

The author extends thanks to Aman Bhatia, Aravind Iyengar, Paul Siegel, Eitan Yaakobi and Ryan Gabrys for helpful discussions. Thank you also to the anonymous reviewers who helped improve the exposition.

This research was supported in part by the Ministry of Education, Science, Sports and Culture; Grant-in-Aid for Scientific Research (C) number 23560439.

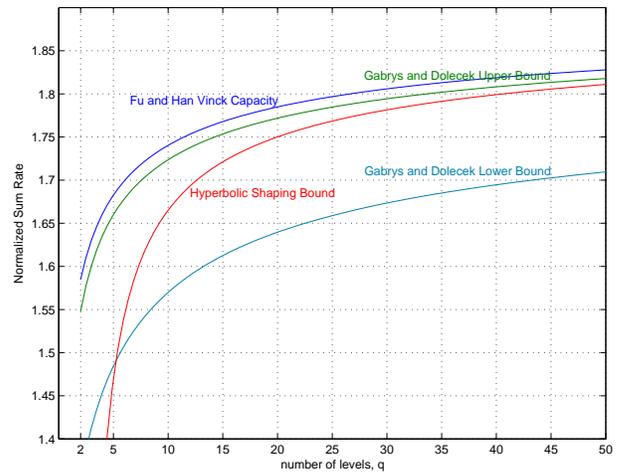


Fig. 3. The proposed hyperbolic shaping bound for two writes of equal rate approaches, but does not achieve, neither the capacity nor the upper bound on the capacity of two equal rates.

## REFERENCES

- [1] B. Kurkoski, "Notes on a lattice-based WOM construction that guarantees two writes," in *Proceedings of the 34th Symposium on Information Theory and Its Applications*, (Ousyuku, Iwate, Japan), pp. 520–524, SITA, November–December 2011.
- [2] A. Bhatia, A. R. Iyengar, and P. H. Siegel, "Multilevel 2-cell  $t$ -write codes," to appear in *Information Theory Workshop 2012*, Lausanne.
- [3] A. Jiang, V. Bohossian, and J. Bruck, "Floating codes for joint information storage in write asymmetric memories," in *Proceedings of IEEE International Symposium on Information Theory*, pp. 1166–1170, June 2007.
- [4] E. Yaakobi, A. Vardy, P. H. Siegel, and J. K. Wolf, "Multidimensional flash codes," in *Proceedings 46th Annual Allerton Conference on Communication, Control, and Computing*, (Monticello, IL, USA), pp. 392–399, September 2008.
- [5] R. L. Rivest and A. Shamir, "How to reuse a "write-once" memory," *Information and Control*, vol. 55, pp. 1–19, December 1982.
- [6] E. Yaakobi, P. H. Siegel, A. Vardy, and J. K. Wolf, "Multiple error-correcting WOM-codes," in *Proceedings of IEEE International Symposium on Information Theory*, (Austin, Texas, USA), pp. 1933–1937, June 2010.
- [7] B. M. Kurkoski, "Asymptotic rates for lattice-based WOM codes." Presented at Non-Volatile Memories Workshop 2012, March 2011. San Diego, USA.
- [8] U. Erez and R. Zamir, "Achieving  $\frac{1}{2} \log(1 + \text{SNR})$  on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, pp. 2293–2314, October 2004.
- [9] B. M. Kurkoski, "Rewriting codes for flash memories based upon lattices, and an example using the E8 lattice," in *Proceedings IEEE Global Telecommunications Conference*, (Miami, USA), pp. 1923–1927, IEEE, December 2010.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [11] G. D. Forney, Jr. and L.-F. Wei, "Multidimensional constellations—part I: Introduction, figures of merit, and generalized cross constellations," *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 877–891, August 1989.
- [12] R. Horst, P. M. Pardalos, and N. V. Thoai, *Introduction to Global Optimization*. Kluwer Academic Publishers, 1995.
- [13] F.-W. Fu and A. J. H. Vinck, "On the capacity of generalized write-once memory with state transitions described by an arbitrary directed acyclic graph," *IEEE Transactions on Information Theory*, vol. 45, pp. 308–313, January 1999.
- [14] R. Gabrys and L. Dolecek, "Characterizing capacity achieving write once memory codes for multilevel flash memories," in *Proceedings of IEEE International Symposium on Information Theory*, (St. Petersburg, Russia), pp. 2517–2521, IEEE, August 2011.