Optimal Algebraic Manipulation Detection Codes in the Constant-Error Model

Ronald Cramer[1], Carles Padró[2], and Chaoping Xing[3]

[1] CWI, Amsterdam and Mathematical Institute, Leiden University, The Netherlands
[2] Universitat Politècnica de Catalunya, Barcelona, Spain
[3] School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore

Abstract. Algebraic manipulation detection (AMD) codes, introduced at EUROCRYPT 2008, may, in some sense, be viewed as *keyless* combinatorial authentication codes that provide security in the presence of an *oblivious*, *algebraic* attacker. Its original applications included robust fuzzy extractors, secure message transmission and robust secret sharing. In recent years, however, a rather diverse array of additional applications in cryptography has emerged. In this paper we consider, for the first time, the regime of arbitrary positive constant error probability ϵ in combination with unbounded cardinality M of the message space. There are several applications where this model makes sense. Adapting a known bound to this regime, it follows that the binary length ρ of the tag satisfies $\rho \geq \log\log M + \Omega_\epsilon(1)$. In this paper, we shall call AMD codes meeting this lower bound *optimal*. Known constructions, notably a construction based on dedicated polynomial evaluation codes, are a multiplicative factor 2 *off* from being optimal. By a generic enhancement using error-correcting codes, these parameters can be further improved but remain suboptimal. Reaching optimality efficiently turns out to be surprisingly nontrivial. We propose a novel constructive method based on symmetries of codes. This leads to an explicit construction based on certain BCH codes that improves the parameters of the polynomial construction and to an efficient randomized construction of optimal AMD codes based on certain quasi-cyclic codes. In all our results, the error probability ϵ can be chosen as an arbitrarily small positive real number.

1 Introduction

Algebraic manipulation detection (AMD) codes, introduced at EUROCRYPT 2008 [5], may, in some sense, be viewed as *keyless* combinatorial authentication codes that provide security in the presence of an *oblivious*, *algebraic* attacker. Briefly, a systematic AMD encoding is a pair consisting of a message m and a tag τ. Given the message, the tag is sampled probabilistically from some given finite abelian group, according to a distribution depending on the details of the scheme. The attack model considers an adversary which substitutes an intercepted pair (m, τ) by a pair $(\widetilde{m}, \widetilde{\tau})$ with $\widetilde{m} \neq m$ such that it knows $\Delta := \widetilde{\tau} - \tau$ and such that Δ is independently distributed from τ. It may, however, depend on m. The

Y. Dodis and J.B. Nielsen (Eds.): TCC 2015, Part I, LNCS 9014, pp. 481–501, 2015.

error probability ϵ of an AMD code upper bounds the success probability of the best strategy to have a substitution accepted as a valid encoding.[1]

The original applications [5] of AMD codes included robust fuzzy extractors, secure message transmission, and robust secret sharing. During the last few years, however, several interesting new applications have emerged. Namely, AMD codes play a role in topics such as construction of non-malleable codes [7], codes for computationally bounded channels [10], unconditionally secure multiparty computation with dishonest majority [3], complete primitives for fairness [9], and public key encryption resilient against related key attacks [13].

In this paper we consider, for the first time, the regime of arbitrarily small positive constant error probability ϵ in combination with unbounded cardinality M of the message space. This model makes sense for most of the known information-theoretic applications of AMD codes. This is the case for secure message transmission, robust secret sharing and robust fuzzy extractors [5], and also for non-malleable codes [7, Theorem 4.1], unconditionally secure multiparty computation with dishonest majority [3, Theorem 8.3], and codes for computationally simple channels [10].[2]

Adapting a known bound to the constant-error model, it follows that the binary length ρ of the tag τ satisfies

$$\rho \geq \log\log M + \Omega_\epsilon(1),$$

where the hidden constant is about $-2\log\epsilon$. In this work, *optimal* AMD codes are those meeting this lower bound, i.e., their tag-length is $\log\log M + O_\epsilon(1)$. Known constructions, notably a construction based on dedicated polynomial evaluation codes [5], are a multiplicative factor 2 *off* from being optimal (Proposition 3). By a generic combination of these polynomial AMD codes with asymptotically good error-correcting codes, AMD codes with tag-length

$$\rho = \log\log M + \log\log\log M + O_\epsilon(1)$$

are obtained (Proposition 4), which is still suboptimal. Bridging the gap to optimality efficiently turns out to be surprisingly nontrivial.

Owing to our refinement of the mathematical perspective on AMD codes, which focuses on symmetries of codes, we propose novel constructive principles. As we show, this leads to the following results.

1. There is a straightforward Gilbert-Varshamov type *nonconstructive* proof of the existence of *optimal* AMD codes (Theorem 1).
2. There is an explicit construction of AMD codes based on cyclic codes (Theorem 2). A construction with equivalent parameters to the polynomial construction from [5] is retrieved immediately by instantiating the latter with Reed-Solomon codes. Instantiating it with narrow-sense primitive BCH codes, AMD codes with improved parameters are obtained (Theorem 4).

[1] The adversary is even allowed to dictate the original message m that occurs in the intercepted encoding.

[2] Nevertheless, other applications require negligible error probability.

3. There is an efficient randomized construction of *optimal* AMD codes, based on twists of asymptotically good quasi-cyclic codes of finite index (Theorem 3). As an aside, the hidden constant in this construction is actually quite small, namely about $-6 \log \epsilon$, which is roughly 3 times the hidden constant in the lower bound (Remark 3). Nevertheless, the dependence on the error probability ϵ is worse than in the polynomial construction in [5], for which the tag-length is roughly $2 \log \log M - 2 \log \epsilon$.

Note that in all our results, the error probability ϵ can be chosen as an arbitrarily small positive real number.

Related Work The reader is referred to the survey [6] for more information about known results, techniques and applications of AMD codes. A class of AMD codes with a stronger security requirement was recently introduced in [11,12]. Namely, *all* algebraic manipulations, even those that do not change the message m but only the tag τ, should be detected with high probability. This additional requirement is not needed in most of the applications of AMD codes. [3] Our novel constructions of AMD codes in this paper satisfy that stronger security requirement.[4] A variant of AMD codes achieving leakage resilience has been presented [1].

2 Best Previous Constructions

The following definition of systematic AMD code was introduced in [5,6]. A new, equivalent definition, which fits our refinement of the mathematical perspective on AMD codes, is given in Section 3.

Definition 1. *Let ϵ be a real number with $0 \leq \epsilon \leq 1$ and let M, n be integers with $M, n \geq 1$. A systematic (M, n, ϵ)-AMD code consists of a map $f : \mathcal{M} \times G \to V$, where $\mathcal{M}$ is a set and G, V are finite abelian groups such that $M = |\mathcal{M}|$ and $n = |G| \cdot |V|$, and*

$$|\{g \in G \ : \ f(m, g) + c = f(m', ge)\}| \leq \epsilon \cdot |G|.$$

for all $m, m' \in \mathcal{M}$ with $m \neq m'$ and for all $(e, c) \in G \times V$. The tag-length *of an (M, n, ϵ)-AMD code is the quantity $\rho = \log_2 n$.*

As discussed after Definition 3, a message $m \in \mathcal{M}$ is encoded by choosing $g \in G$ uniformly at random and adding the *tag* $\tau = (g, f(m, g)) \in G \times V$ to the message m.

A simple example of a systematic AMD code, the so-called *multiplication* AMD code, is given in Proposition 1. It is extracted from the robust secret sharing construction in [4]. The proof of this result is straightforward.

[3] It is nevertheless essential for the non-malleable secret sharing schemes introduced in [9].

[4] The only exceptions appearing in this paper are the *non-constructive* family in Corollary 3 and the (known) multiplication AMD code in Proposition 1. The AMD code from [5] also satisfies the stronger security requirement.

Proposition 1. *Let q be a positive prime power and k, ℓ positive integers with $k \geq \ell$, and take an embedding of $\mathbb{F}_q^\ell$ into $\mathbb{F}_{q^k}$. Then the map $f : \mathbb{F}_q^\ell \times \mathbb{F}_{q^k} \to \mathbb{F}_{q^k}$ given by $f(m, g) = mg$ (here the embedding of $\mathbb{F}_q^\ell$ into $\mathbb{F}_{q^k}$ is used to compute the product mg) defines a systematic $(q^\ell, q^{2k}, 1/q^k)$-AMD code.*

We present next the family of efficient AMD codes, with rather good parameters, that was introduced in [5]. The reader is referred to [5,6] for more details about this construction.

Proposition 2. *Let $\mathbb{F}_q$ be a finite field of characteristic p. Let $d > 0$ be an integer such that $d + 1 < q$ and p is not a divisor of $d + 2$. Then the function $f : \mathbb{F}_q^d \times \mathbb{F}_q \to \mathbb{F}_q$ defined by*

$$f((m_1, \ldots, m_d), g) = g^{d+2} + \sum_{i=1}^{d} m_i g^i$$

determines a systematic $(q^d, q^2, (d + 1)/q)$-AMD code.

The following discussion, which is adapted from [6, Section 6], demonstrates the flexibility in the values of the parameters of this family of AMD codes. In addition, it proves Proposition 3.

Consider a prime p, a real number ϵ_0 with $0 < \epsilon_0 < 1$, and an integer $M_0 \geq 1/\epsilon_0$. Take the smallest integer d such that $d + 2$ is not divisible by p and $\log M_0 \leq d(\log(d + 1) - \log \epsilon_0)$,

$$k = \left\lceil \frac{\log(d + 1) - \log \epsilon_0}{\log p} \right\rceil,$$

and $q = p^k$. Then $M = q^d \geq M_0$ and $\epsilon = (d + 1)/q \leq \epsilon_0$. Therefore, there exists in the family introduced in Proposition 2 an (M, p^{2k}, ϵ_0)-AMD code, which can be trivially transformed into an $(M_0, p^{2k}, \epsilon_0)$-AMD code, with tag-length

$$\begin{aligned} \rho = 2k \log p &\leq -2 \log \epsilon_0 + 2 \log(d + 1) + 2 \log p \\ &\leq -2 \log \epsilon_0 + 2 \log \left(-\frac{\log M_0}{\log \epsilon_0} + 3 \right) + 2 \log p. \end{aligned}$$

We have used here that $(k - 1) \log p \leq \log(d + 1) - \log \epsilon_0$ and $(d - 2)(\log(d - 1) - \log \epsilon_0) \leq \log M_0$. The following two propositions are direct consequences of this discussion.

Proposition 3. *For every fixed value of ϵ with $0 < \epsilon < 1$ and arbitrarily large values of M, there exist systematic (M, n, ϵ)-AMD codes in the family introduced in Proposition 2 such that the asymptotic behavior of the tag-length is $\rho = 2 \log \log M + O(1)$.*

When comparing the result in Proposition 3 with the asymptotic lower bound in Corollary 1, we observe that the construction of AMD codes in [5] is a multiplicative factor 2 off from being optimal.

Finally, we observe here that it is possible to obtain an almost optimal construction by combining the AMD codes above with an asymptotically good family of $\mathbb{F}_q$-linear error-correcting codes. The idea is to encode the message $x \in \mathcal{M}$ with an error-correcting code C of length s in the family, take the tag $(g, C_g(x))$, where g is chosen uniformly at random from G_s, the cyclic group of order s, and $C_g(x) \in \mathbb{F}_q$ is the g-th component of the codeword $C(x)$, and then encode the tag $(g, C_g(x))$ with a suitable AMD code.

Proposition 4. *For every fixed value of ϵ with $0 < \epsilon < 1$ and arbitrarily large values of M, there exist systematic (M, n, ϵ)-AMD codes such that the asymptotic behavior of the tag-length is $\rho = \log\log M + \log\log\log M + O(1)$.*

Proof. Consider a family of $\mathbb{F}_q$-linear codes with constant rate $R > 0$ and constant relative minimum distance $\delta \geq 1 - \epsilon$. That is, for arbitrarily large values of s there is in the family a code $C : \mathbb{F}_q^k \to \mathbb{F}_q^s$ with length s, dimension $k \geq Rs$ and minimum distance at least δs. For every h in G_s and $x \in \mathcal{M} = \mathbb{F}_q^k$, let $C_g(x) \in \mathbb{F}_q$ be the g-th component of the codeword $C(x)$. We have seen before that one can find for these values of s AMD codes $f' : \mathcal{M}' \times G' \to V'$ with message space $\mathcal{M}' = G_s \times \mathbb{F}_q$, error probability ϵ, and tag-length $\log\log sq + O(1) = \log\log s + O(1)$. The proof is concluded by considering the AMD code

$$f : \mathbb{F}_q^k \times (G_s \times G') \to \mathbb{F}_q \times V'$$

defined by $f(x, (g, g')) = (C_g(x), f'((g, C_g(x)), g'))$.

3 Overview of Our Results

To enable a bird's eye view on our main results, we first briefly sketch our refinement of the mathematical perspective on AMD codes. Let V and G be finite abelian groups. Define the finite abelian group

$$V[G] = \bigoplus_{g \in G} V,$$

together with the group action denoted by "$\cdot$" that turns $V[G]$ into a so-called G-module by having G act on the coordinates. More precisely, if $x \in V[G]$ with "coordinates" $x(g) \in V$ $(g \in G)$, then

$$h \cdot x \in V[G]$$

is defined such that

$$(h \cdot x)(g) := x(-h + g),$$

for all $g \in G$.[5] In particular, the G-action permutes coordinates. A *G-submodule* C' is a subgroup of $V[G]$ that is invariant under the G-action, i.e.,

$$G \cdot C' = C',$$

[5] Note that, $(h' \cdot (h \cdot x))(g) = x(-h - h' + g) = (h' + h) \cdot x$, for all $h, h', g \in G$ and $x \in V$.

or equivalently, $h \cdot x \in C'$ for all $h \in G$, $x \in C'$. Let $\Gamma \subset V[G]$ denote the G-submodule of *constants*, i.e., it consists of the elements $x \in V[G]$ such that $x(g) = x(g')$ for all $g, g' \in G$. If $x \in V[G]$, then $G \cdot x$ is the G-*orbit* of x, i.e., it is the set of elements $\{h \cdot x \mid h \in G\}$. Note that, if $x \neq 0$, then this is *not* a G-submodule. Recall that, if A, B are subsets of an additive group, then $A + B$ is defined as $\{a + b : a \in A, b \in B\}$.

Definition 2. *For $x, y \in V[G]$, the* AMD-equivalence relation *in $V[G]$ is defined by*

$$x \sim y \text{ if and only if } x \in (G \cdot y + \Gamma).$$

For $x \in V[G]$, the equivalence class of x under the AMD-equivalence relation is denoted by $\mathrm{cl}(x)$.

Consider the set $V[G]/\sim$, i.e., $V[G]$ taken modulo this equivalence relation. Also consider the *induced* Hamming-distance $\overline{d}_{\mathrm{H}}$, which defines the distance between classes $\mathrm{cl}(x), \mathrm{cl}(x') \in V[G]/\sim$ as the minimum of the (regular) Hamming-distance $d_{\mathrm{H}}(y, y')$ taken over all $y \in \mathrm{cl}(x)$ and $y' \in \mathrm{cl}(x')$.[6] Observe that $\overline{d}_H(\mathrm{cl}(x), \mathrm{cl}(x')) = d_H(\{x\}, \mathrm{cl}(x'))$. For a subset $C \subset V[G]$, the image of C under reduction by the equivalence relation is denoted by $\overline{C} \subset V[G] \,/\sim$.

Our new perspective concerns the observation that "good" AMD codes correspond to codes $C \subset V[G]$ such that $|C| = |\overline{C}|$, the cardinality $|\overline{C}|$ is "large" and the minimum distance $\overline{d}_{\min}(\overline{C})$ of $\overline{C} \subset V[G] \,/\sim$ (i.e, in terms of the induced Hamming-distance) is "large" as well. Only systematic algebraic manipulation detection codes are considered in this paper. The reader is referred to [6] for additional definitions and results about this and other classes of algebraic manipulation detection codes. For completeness, we present in Appendix 2 the equivalent definition of asymptotic AMD code from [5].

Definition 3 (AMD Codes). *Let ϵ be a real number with $0 \leq \epsilon \leq 1$ and let M, n be integers with $M, n \geq 1$. A* systematic (M, n, ϵ)-algebraic manipulation detection (AMD) code *consists of finite abelian groups G, V and a subset $C \subset V[G]$ such that $|C| = |\overline{C}| = M$ and $|G| \cdot |V| = n$, and $\overline{d}_{min}(\overline{C}) \geq (1 - \epsilon) \cdot |G|$. The* tag-length *of an (M, n, ϵ)-AMD code is the quantity $\rho = \log_2 n$.*

We prove in the following that Definitions 3 and 1 are equivalent. First assume that $C \subset V[G]$ is a systematic AMD code given in Definition 3. Take $\mathcal{M} = C$ and consider the map $f : C \times G \to V$ defined by $(x, g) \mapsto x(g)$. Then it is easy to verify that this coincides with Definition 1. On the other hand, assume that we have a systematic AMD code given in Definition 1. Consider the set $C := \{\sum_{g \in G} f(m, g)g : m \in \mathcal{M}\}$. Then it is straightforward to verify that C is a systematic AMD code given in Definition 3.

In applications, a bijection $\phi : \mathcal{M} \to C$ between the message space $\mathcal{M}$ and the code C is fixed. To encode a message $m \in \mathcal{M}$, take $x = \phi(m)$, select $h \in G$ uniformly at random and set

[6] The (regular) Hamming-distance between two elements of $V[G]$ is, of course, the number of non-zero coordinates in their difference.

$$\tau := (h, x(h)) \in G \times V$$

as the tag.

AMD codes are a relaxation of combinatorial authentication codes. Their purpose is similar, namely ensuring *message integrity*. However, AMD codes are keyless and security is only guaranteed against a *non-adaptive, algebraic* adversary that has a priori knowledge of m and effectively replaces (m, τ) by $(m', \tau') \in \mathcal{M} \times (G \times V)$, under the following restrictions:

- $m' \neq m$.
- Effectively selects an *offset* $(e, c) \in G \times V$ and sets $\tau' = (h + e, x(h) + c) \in G \times V$.
- The selection of (m', e, c) may only depend on the message m and independent randomness chosen by the adversary. In particular, this selection does not depend on h.

Then the adversary is successful if and only if $x'(h + e) = x(h) + c$, where $x' = \phi(m') \in C$. It follows that success is equivalent to $((-e) \cdot x')(h) - c = x(h)$. Since x and x' are in distinct equivalence classes and since $h \in G$ is uniformly random and independent of x, x', e, c, the success probability of the adversary is at most ϵ because

$$1 - \frac{\overline{d}_H(\mathrm{cl}(x), \mathrm{cl}(x'))}{|G|} \leq 1 - \frac{\overline{d}_{\min}(\overline{C})}{|G|} \leq \epsilon.$$

In several specialized situations the adversary is effectively reduced to non-adaptive, algebraic attack. Moreover, authentication codes are typically not an option there: the secret key is susceptible to the same attack. Interestingly, the choice of the groups V, G is typically immaterial in applications.[7]

These observations motivate the following novel approaches to show existence of good AMD codes. Suppose, for now, that $C' \subset V[G]$ is such that

1. C' is a G-submodule.
2. $\Gamma \subset C'$.

Suppose that $|C'|$ is "large" and that the (regular) minimum distance $d_{\min}(C')$ is "large." In order to get a good AMD code out of this, it now suffices to develop an (efficient) method to select a subset $C \subset C'$ such that for each distinct $x, x' \in C$, the intersection between the orbits $G \cdot x$ and $G \cdot x'$ is empty (*orbit avoidance*). This way, one potentially achieves an AMD code C such that

$$|C| \geq \frac{|C'|}{|V| \cdot |G|},$$

where the denominator upper bounds the cardinality of a class, and such that the error probability ϵ satisfies

[7] Except perhaps that it is sometimes convenient if neither $|V|$ nor $|G|$ has a small prime divisor.

$$\epsilon = 1 - \frac{d_{\min}(C')}{|G|}.$$

This discussion is summarized in the following result.

Lemma 1. *Suppose* $C' \subset V[G]$ *is a* G*-submodule,* $\Gamma \subset C'$*, and* $d_{\min}(C') \geq (1-\epsilon) \cdot |G|$ *for some* ϵ *with* $0 < \epsilon < 1$*. Then there exists a systematic* $\left(\frac{|C'|}{|V| \cdot |G|}, |V| \cdot |G|, \epsilon\right)$*-AMD code* $C \subset C'$*.*

As we show, this approach immediately leads to a greedy, *non-constructive* proof of the existence of optimal AMD codes.[8]

Theorem 1. *For every real number* ϵ *with* $0 < \epsilon < 1$*, there exist AMD codes with unbounded message space cardinality* M *and error probability at most* ϵ *whose tag-length* ρ *satisfies*

$$\rho = \log\log M + O_\epsilon(1),$$

which is optimal.

*Remark 1 (*Locking trick). Suppose the condition that $\Gamma \subset C'$. This complicates the situation as the (regular) relative minimum distance of the code C' no longer gives a non-trivial upper bound on the error probability of the AMD code C, i.e., the code obtained after application of orbit avoidance. But if $|V|$ is *constant*, the situation can be reduced to the previous situation by means of our *locking trick*, without harming *asymptotic performance*: simply augment an AMD encoding with a standard AMD encoding (with appropriate error probability) of the value $x(h)$ in the tag $\tau = (x, x(h))$. This way, at the cost of an additive constant increase in tag-length, we may as well assume that the adversary does not change the V-component of the tag. This obviates the need for considerations involving the constants Γ. As a consequence, the relative minimum distance of C' once again governs the error probability ϵ of the AMD code C.[9]

Any AMD code with the suitable parameters can be used in the locking trick as, for example, the simple multiplication AMD code in Proposition 1.[10].

Hence, the remaining question is about effective construction. We apply the idea above to *cyclic* $\mathbb{F}_q$*-linear codes* and show an efficiently enforceable algebraic conditions on the generator polynomial to ensure orbit avoidance. If $V = K$ is a (finite) field, then $K[G]$ is a ring, where multiplication is defined from the G-action by convolution, and hence $K[G]$ is a K-algebra (since K is contained in a natural way). A cyclic $\mathbb{F}_q$-linear code is a G-submodule of $\mathbb{F}_q[G]$, where G

[8] In fact, a Gilbert-Varshamov style argument.

[9] Note that locking only makes sense if $|V|$ is very small compared to $|G|$; otherwise this is too costly!

[10] If an AMD code under the stronger security requirements from [11] is needed, then one should select for the lock an AMD code that also satisfies those requirements as, for instance, the polynomial AMD code from [5].

is a finite cyclic group. It is convenient, though, to work with the following more common, equivalent definition.

Let q be a positive prime power and let $\mathbb{F}_q$ be a finite field with q elements. Let s be a positive integer that is not a multiple of the characteristic p of $\mathbb{F}_q$. Set $V = \mathbb{F}_q$ and $G = G_s$, the cyclic group of order s. Let $\pi_q(s) = \mathrm{ord}_s^*(q)$ be the multiplicative order of q mod s. An $\mathbb{F}_q$*-linear cyclic code* C' of length s is an ideal of $\mathbb{F}_q[G_s] \simeq \mathbb{F}_q[X]/(X^s - 1)$, and hence it is generated by the class $\overline{a(X)} \in \mathbb{F}_q[X]/(X^s - 1)$ of some polynomial $a(X) \in \mathbb{F}_q[X]$ that divides $X^s - 1$. This polynomial is called the *generator* of the cyclic code C'.

Theorem 2. *Let $\mathbb{F}_q$ be a finite field and $s > 1$ an integer that is not a multiple of the characteristic p of $\mathbb{F}_q$. Let $C' \subset \mathbb{F}_q[X]/(X^s - 1)$ be an $\mathbb{F}_q$-linear cyclic code of length s with generator $a(X) \in \mathbb{F}_q[X]$. Let d be the minimum distance of C'. Suppose that the following conditions are satisfied.*

1. *The all-one vector is in C' or, equivalently, $(X - 1)$ does not divide $a(X)$.*
2. *There is a primitive s-th root of unity $\omega \subset \overline{\mathbb{F}}_q$ with $a(\omega) \neq 0$.*
3. *$\pi_q(s) < s - \deg a - 1$.*

Then there exists an explicit *construction of a $(q^{s-\deg a-\pi_q(s)-1}, sq, (s-d)/s)$-AMD code $C \subset C'$.*

Notice that the conditions imply $\Gamma \subset C'$, so the locking trick from Remark 1 is not necessary. For every real number ϵ with $0 < \epsilon < 1$, instantiation with Reed-Solomon codes defined over a large enough finite field leads to an explicit construction of AMD codes with arbitrarily large message space cardinality M and tag-length $2 \log\log M + O_\epsilon(1)$, which is the same as in the explicit construction from [5] (see Appendix 2). Instantiation with narrow-sense primitive BCH codes defined over a large enough finite field leads to an explicit construction of *almost*-optimal AMD codes, i.e., achieving tag-length $(1 + \delta) \log\log M + O_\epsilon(1)$ where δ is an arbitrary real constant with $0 < \delta < 1$.

One quickly sees that achieving optimality along the lines of twists on cyclic codes as discussed above would require the existence of asymptotically good cyclic $\mathbb{F}_q$-linear codes, which is one of the central open problems in the theory of error correcting codes.[11] The remainder of our results is concerned with bypassing this major open problem.

Our final result is a randomized construction of optimal AMD codes.

[11] It would not even be enough if asymptotically good cyclic $\mathbb{F}_q$-linear codes exist for *some* value of q. Namely, to suit our purposes, such codes should exist for infinitely many values of q and, when sending q to infinity, the relative minimum distance achieved should tend to 1. Finally, a certain condition on the lengths of the codes should hold. Specifically, given such a value of q, the lengths $\ell(C')$ occurring must satisfy $\pi_q(\ell(C')) \leq \gamma \ell(C')$ for some absolute real constant $\gamma > 0$. Otherwise the orbit-avoidance eats away too many codewords, causing the rate to drop to 0. If the codes do not contain the all-one vector, the locking trick alluded below could be applied.

Theorem 3. *For every real number ϵ with $0 < \epsilon < 1$, there exist an* efficient, randomized *construction of explicit AMD codes with arbitrarily large message space cardinality M and tag-length $\rho = \log\log M + O_\epsilon(1)$, which is* optimal.

Relying on our AMD perspective as outlined above, it is achieved by a series of twists on a result in a beautiful paper by Bazzi and Mitter [2] on *asymptotically good quasi-cyclic codes of constant index ℓ*. Let $\ell \geq 2$ and define $V = \mathbb{F}_2^\ell$. One of their results (stated in our terminology here) is that there exists a randomized construction of G_s-submodules $C' \subset \mathbb{F}_2^\ell[G_s]$ of rate $\log|C'|/(\ell s) = 1/\ell$ achieving the Gilbert-Varshamov bound when s tends to infinity.[12] The error probability of this randomized construction is exponentially small in s if the lengths s are carefully selected.

We use four twists on their result to show our claim. First, we generalize it to work over *all* finite fields $\mathbb{F}_q$, with $\ell \geq 2$ an arbitrary integer constant. This ensures that relative minimum distance arbitrarily close to 1 can be achieved, and hence ϵ can be selected arbitrarily close to 0. This generalization is straightforward, using some results from [8]. Second, this time we need to resort to the locking trick from Remark 1. Third, we need an adaptation of the efficient orbit avoidance method alluded to above. This adaptation is necessary not only because of the shift from cyclic codes to quasi-cyclic ones, but also because of the probabilistic nature of the construction. Fourth, we need to craft the lengths s with additional care to ensure that the rate of the code drops by at most a multiplicative positive constant factor after application of orbit avoidance.

4 Nonconstructive Optimal AMD Codes

In this section, we present the proof of Theorem 1. We begin by presenting the asymptotic lower bound in Corollary 1, which is a consequence of the bound in Proposition 6. This bound is a refinement of similar bounds presented in [6,11]. We are going to use the following trivial result

Proposition 5. *Let ϵ be a real number with $0 < \epsilon < 1$ and let $C \subset V[G]$ be an (M, n, ϵ)-AMD code. Then $|G| \geq 1/\epsilon$ and $|V| \geq 1/\epsilon$. As a consequence, the tag-length ρ satisfies $\rho \geq -2\log\epsilon$.*

Proof. Consider $x, x' \in C$ with $x \neq x'$. Then there exists $c \in V$ such that the set $\{g \in G : x(g) - x'(g) = c\}$ has cardinality at least $|G|/|V|$.

$$|\{g \in G : x(g) - x'(g) = c\}| \geq \left\lceil \frac{|G|}{|V|} \right\rceil .$$

Therefore,

$$(1-\epsilon)\cdot|G| \leq \overline{d}_H(\mathrm{cl}(x), \mathrm{cl}(x')) \leq |G| - \left\lceil \frac{|G|}{|V|} \right\rceil \leq \min\left\{ |G| - 1, |G| - \frac{|G|}{|V|} \right\}$$

and the proof is concluded.

[12] i.e., the relative minimum distance $\delta > 0$ of these codes is such that $H_2(\delta) = 1/\ell$, where $H_2(\cdot)$ is the binary Shannon-entropy function.

Other lower bounds on the tag-length are obtained by applying some known classical bounds from coding theory, as in Proposition 6. As a corollary, we obtain a lower bound on the asymptotic behavior of the tag-length.[13]

Proposition 6. *Let ϵ be a real number with $0 < \epsilon < 1$. Suppose that $M \geq 1/\epsilon$. Then the tag-length ρ of a systematic (M, n, ϵ)-AMD code satisfies*

$$\rho \geq \log\log M - 2\log\epsilon - \max\{0, \log(-\log\epsilon)\}.$$

Proof. Let $C \subset V[G]$ be an (M, n, ϵ)-AMD code. From the definition of AMD code, $C + \Gamma$ is a code of size $M \cdot |V|$, length $|G|$ and minimum distance at least $(1-\epsilon)|G|$ over the alphabet V. Therefore, by the Singleton bound, $M \leq |V|^{\epsilon \cdot |G|}$, and hence

$$\log|G| \geq \log\log M - \log\epsilon - \log\log|V|.$$

By Proposition 5, $\log|G| \geq -\log\epsilon$ and $\log|V| \geq -\log\epsilon$. Take $x = \log|V|$, $y = \log|G|$, $A = \max\{1, -\log\epsilon\}$, and $B = \log\log M$. Since $B - \log A \geq 0$, the minimum value of $x + y$ under the constraints $x, y \geq A$ and $y \geq A + B - \log x$ is attained when $x = A$ and $y = A + B - \log A$.

Corollary 1. *For every real number ϵ with If $0 < \epsilon < 1$, the tag-length ρ of the AMD codes with arbitrarily large message space cardinality M and error probability at most ϵ satisfies*

$$\rho \geq \log\log M + \Omega_\epsilon(1).$$

Application of the Hamming and Plotkin bounds instead of the Singleton bound gives better results, but the asymptotic results are not improved. Next, we prove Theorem 1 by using a variation on the Gilbert-Varshamov bound.

Definition 4. *Consider a finite field $\mathbb{F}_q$, a real number ϵ with $1/q < \epsilon < 1$, and a positive integer s. Then the quantity $A'_q(s, \epsilon)$ is defined as the maximum cardinality M of an (M, n, ϵ)-AMD code $C \subset \mathbb{F}_q[G_s]$.*

Proposition 7. *With conditions as above,*

$$A'_q(s, \epsilon) \geq \left\lfloor \frac{q^s}{qs \cdot V_q(s, 1-\epsilon)} \right\rfloor,$$

where $V_q(s, 1-\epsilon)$ is the volume of a sphere in $\mathbb{F}_q[G_s]$ with radius $(1-\epsilon)s$.

Proof. Suppose that the result is false and take an (M, n, ϵ)-AMD code $C \subset \mathbb{F}_q[G_s]$ with $M = A'_q(s, \epsilon)$. Observe that $|\operatorname{cl}(x)| \leq qs$ for every $x \in \mathbb{F}_q[G_s]$. Therefore, the number of elements $y \in \mathbb{F}_q[G_s]$ such that $d_H(\{y\}, \operatorname{cl}(x)) < (1-\epsilon)s$

[13] The looser, but easier to prove, lower bound $\rho \geq \log\log M - \log\epsilon$ is enough to obtain the asymptotic lower bound in Corollary 1. Nevertheless, the bound in Proposition 6 provides a better description of the behavior of the tag-length ρ in relation to the error probability ϵ.

is at most $qs \cdot V_q(s, 1-\epsilon)$. Since $|C| \cdot qs \cdot V_q(s, 1-\epsilon) < q^s$ there exist a vector $y \in \mathbb{F}_q[G_s] \setminus C$ such that

$$\overline{d}_H(\mathrm{cl}(y), \mathrm{cl}(x)) = d_H(\{y\}, \mathrm{cl}(x)) \geq (1-\epsilon)s$$

for all $x \in C$. Therefore, C has not maximum cardinality among all codes with the required property, a contradiction.

Corollary 2. *Let q be a positive prime power, let ϵ be a real number with $1/q < \epsilon < 1$, and let s be a positive integer. Then there exists a systematic*

$$\left(\left\lfloor \frac{q^s}{qs \cdot V_q(s, 1-\epsilon)} \right\rfloor, qs, \epsilon \right) \text{-AMD code}$$

with $V = \mathbb{F}_q$ and $G = G_s$, the cyclic group of order s.

Lemma 2. *With conditions as above,*

$$\lim_{s \to \infty} \frac{\log A'_q(s, \epsilon)}{s} \geq (1 - H_q(1-\epsilon)) \log q,$$

where H_q is the q-ary entropy function.

Proof. The result follows immediately from Corollary 2 by taking limits, taking into account that, by coding theory, $\lim_{s \to \infty} \dfrac{\log_q V_q(s, 1-\epsilon)}{s} = H_q(1-\epsilon)$.

Finally, Theorem 1 is an immediate consequence of the following result.

Corollary 3 (Non-constructive optimality). *For any real constant $c > 0$, fix a positive prime power q and a real number ϵ with $\epsilon = 1/q + 1/q^{1+c}$. Then there exist AMD codes with arbitrarily large message space cardinality M, error probability at most ϵ and tag-length*

$$\rho = \log\log|M| - (2+c)\log\epsilon + O(1).$$

Note that the tag-length is minimal up to an additive constant.

5 An Explicit Construction from Cyclic Codes

This section is devoted to prove Theorem 2. We present here an effective method to select, from any given cyclic code, a number of codewords in different AMD-equivalence classes. By Lemma 1, this provides an effective construction of systematic AMD codes.

5.1 General Construction

Let $\mathbb{F}_q$ be a finite field. Let $s > 1$ be an integer that is not a multiple of the characteristic p of $\mathbb{F}_q$. Then $\mathbb{F}_q[G_s]$ is a ring, where the product is defined from the G_s-action by convolution. So, $\mathbb{F}_q[G_s]$ is an $\mathbb{F}_q$-algebra. Since $X^s - 1$ is separable, it follows by the Chinese Remainder Theorem that $\mathbb{F}_q[G_s] \simeq \mathbb{F}_q[X]/(X^s - 1)$ is a product of finite extension fields of $\mathbb{F}_q$. Let $\omega \in \overline{\mathbb{F}}_q$ be a primitive s-th root of unity. Then the degree of $\mathbb{F}_q(\omega)$ over $\mathbb{F}_q$ equals

$$\pi_q(s) = \mathrm{ord}_s^*(q),$$

the multiplicative order of $q \bmod s$. Equivalently, it equals the degree of the minimal polynomial of ω over $\mathbb{F}_q$. It is also the *largest degree* occurring among the irreducible factors in the factorization of $X^s - 1$ over $\mathbb{F}_q[X]$ as each of the s roots sits in some intermediate extension of $\mathbb{F}_q(\omega) \supset \mathbb{F}_q$.

Definition 5. *Let $a(X) \in \mathbb{F}_q[X]$ be a polynomial such that $a(X)$ divides $X^s - 1$ and let $\omega \in \overline{\mathbb{F}}_q$ be a primitive s-th root of unity with $a(\omega) \neq 0$. We define $D(a(X), \omega) \subset \mathbb{F}_q[X]$ as the set of all polynomials $f(X) \in \mathbb{F}_q[X]$ such that*

1. *$f(\omega) = 1$, and*
2. *$\deg f < s - \deg a - \delta$, where $\delta = 0$ if $(X - 1)$ divides $a(X)$ and $\delta = 1$ otherwise.*

An *$\mathbb{F}_q$-linear cyclic code* C' of length s is a G_s-submodule of $\mathbb{F}_q[G_s]$. Equivalently, it is an ideal $C' \subset \mathbb{F}_q[X]/(X^s - 1)$ generated by the class of some polynomial $a(X) \in \mathbb{F}_q[X]$ that divides $X^s - 1$. This polynomial is called the *generator* of the cyclic code C'. Then

$$C' = \left\{ \overline{a(X)f(X)} \,:\, f(X) \in \mathbb{F}_q[X] \text{ and } \deg f < s - \deg a \right\} \subset \mathbb{F}_q[X]/(X^s - 1).$$

Let C' be an $\mathbb{F}_q$-linear cyclic code with generator $a(X)$ and suppose that $a(\omega) \neq 0$ for some primitive s-th root of unity $\omega \in \overline{\mathbb{F}}_q$. Let $C \subset C'$ be the set of all codewords $\overline{a(X)f(X)} \in \mathbb{F}_q[X]/(X^s - 1)$ with $f \in D(a(X), \omega)$.

Lemma 3. *No two distinct elements in C are in the same AMD-equivalence class.*

Proof. Suppose that there exist two different polynomials $f, g \in D$ such that the corresponding codewords in C are in the same AMD-equivalence class. Then

$$X^\ell a(X)f(X) + \lambda(X^{s-1} + \cdots + X + 1) \equiv a(X)g(X) \pmod{X^s - 1}$$

for some $\lambda \in \mathbb{F}_q$ and ℓ with $0 \leq \ell < s$. Therefore, $\omega^\ell a(\omega)f(\omega) = a(\omega)g(\omega)$, and hence $\omega^\ell = 1$ by the definition of C. Since ω is a primitive root, $\ell = 0$. Consequently,

$$a(X)(f(X) - g(X)) \equiv -\lambda(X^{s-1} + \cdots + X + 1) \pmod{X^s - 1}. \qquad (1)$$

Suppose that $(X-1)$ divides $a(X)$. Then $\lambda \neq 0$ because $\deg(f-g) < s - \deg a$, but this implies that $a(X)$ divides $X^{s-1} + \cdots + X + 1$, a contradiction. Suppose now that $(X-1)$ does not divide $a(X)$, and hence $\delta = 1$ and $\deg(f-g) < s - \deg a - 1$. But then (1) is impossible because $\deg(a(f-g))$ is too small, a contradiction again.

Lemma 4. *Suppose that* $\pi_q(s) < s - \deg a - \delta$. *Then* $|C| = q^{s-\deg a - \pi_q(s) - \delta}$.

Proof. Take $h = s - \deg a - \delta$ and let $\mathbb{F}_q[X]_{<h}$ be the $\mathbb{F}_q$-vector space of the polynomials in $\mathbb{F}_q[X]$ with degree at most $h-1$. Since $\pi_q(s) < h$, application of Lemma 7 implies that the kernel of the $\mathbb{F}_q$-linear map $\mathbb{F}_q[X]_{<h} \to \mathbb{F}_q[\omega]$, $f \mapsto f(\omega)$ has dimension $h - \pi_q(s)$.

The proof of Theorem 2 is now straightforward from Lemmas 1, 3 and 4.

5.2 Instantiations

Applying Theorem 2 to Reed-Solomon codes provides, for every real number ϵ with $0 < \epsilon < 1$, an effective construction of (M, n, ϵ)-AMD codes with unbounded message space cardinality M and tag-length $2 \log\log M + O_\epsilon(1)$, which is the same as in the polynomial construction from [5] (see Section 2). Indeed, consider a prime power q, a primitive element α of $\mathbb{F}_q^*$, an integer k with $1 \leq k \leq q-1$, and the polynomial $a(X) = (X-\alpha)(X-\alpha^2)\cdots(X-\alpha^{q-k-1})$. By applying Theorem 2 to the $\mathbb{F}_q$-linear cyclic code with length $q-1$ generated by $a(X)$, which is a Reed-Solomon code with minimum distance $d = q-k$, one obtains an effective AMD code with parameters $(q^{k-2}, q(q-1), (k-1)/(q-1))$. The proof of our claim is concluded by using a similar argument as for the polynomial construction from [5] (see Section 2).

The instantiation to narrow-sense BCH codes is not so immediate. We refer to Appendix B for the background on BCH codes.

Let $e \geq 1$ be an integer. Let $s = q^e - 1$. Choose an element α of $\mathbb{F}_{q^e}$ of order $s = q^e - 1$. Let $m^{(i)}(x) \in \mathbb{F}_q[x]$ denote the minimal polynomial of α^i with respect to $\mathbb{F}_q$. For $0 < \epsilon < 1$, put $d = (1-\epsilon)s$ and consider the BCH code B of length $s = q^e - 1$ with the generator polynomial $a(X) := \mathrm{lcm}\{m^{(1)}(X), m^{(2)}(X), \ldots, m^{(d-1)}(X)\}$. Then the minimum distance of B is at least d. Let $f_{a_i,j}(X)$ be the polynomials defined in (2) of Appendix B.

By Lemma 11, the dimension $s - \deg a$ of B is equal to the dimension of the $\mathbb{F}_q$-span $V_{s-d} = V_{\epsilon s}$ of $\{f_{a_i,j}(X) : \ 1 \leq i \leq t, 1 \leq j \leq s_{a_i}, \ \deg f_{a_i,j} \leq \epsilon n\}$. Hence, $\dim(B) = s - \deg a \geq (\epsilon(q-1)+1)^e \approx e + 1 + (\epsilon q)^e$. Note that in this case $\pi_q(s) = e$. Applying Theorem 2, we obtain the following AMD codes.

Theorem 4. *For any* $\epsilon \in (0,1)$, *any integer* $e \geq 1$ *and prime power* q, *there exists an effective* $(q^{(\epsilon q)^e}, (q^e-1)q, \epsilon)$*-AMD code. Thus, the tag-length equals to*

$$\frac{e+1}{e} \log\log M - (e+1)\log\epsilon + O(1).$$

Proof. Note that the message size $M = q^{(\epsilon q)^e}$ satisfies $\log\log M \approx e \cdot \log\epsilon + e \cdot \log q$. The tag-length satisfies $\log q + \log(q^e - 1) \leq (e+1)\log q \leq \frac{e+1}{e}\log\log M - (e+1)\log\epsilon$. This completes the proof.

Remark 2. When $e = 1$ in Theorem 4, we get almost the same result as in the one in [5]. If we choose $e = (\log\log M)^{0.5}$ in Theorem 4, then the tag-length is $\log\log M + O((\log\log M)^{0.5})$.

6 Monte-Carlo Construction of Optimal AMD Codes

In this section we prove Theorem 3. Namely, we present an efficient randomized construction of explicit optimal AMD codes. We proceed as follows. We begin by presenting in Theorem 5 a randomized construction of G_s-submodules $C' \subset \mathbb{F}_q^\ell[G_s]$. By considering the codes C' over the alphabet $\mathbb{F}_q$, they have rate $\log_q |C'|/(\ell s) = 1/\ell$ and minimum relative distance δ arbitrarily close to 1 achieving the Gilbert-Varshamov bound when s tends to infinity. This is an extension of the corresponding result by Bazzi and Mitter [2] for the case $q = 2$. This extension is based on some results from [8]. The error probability of this randomized construction is exponentially small in s if the lengths s are carefully selected. Then we apply the general method derived from Lemma 1 to those G_s-submodules $C' \subset \mathbb{F}_q^\ell[G_s]$. Since $\Gamma \not\subset C'$, we have to use the locking trick in Remark 1. Furthermore, we have to adapt orbit avoidance to this probabilistic scenario involving quasi-cyclic codes. In addition, we need to craft the lengths s with additional care to ensure that the rate of the AMD code remains positive after application of orbit avoidance. Finally, in Remark 3, we use a simple modification to reduce the size of the hidden constant in the tag-length.

Let $\mathbb{F}_q$ be a finite field and $s > 1$ an integer such that the characteristic p of $\mathbb{F}_q$ does not divide s. As before, G_s denotes the cyclic group of order s. Recall that, if $\omega \in \overline{\mathbb{F}}_q$ is a primitive s-th root of unity, then the degree of $\mathbb{F}_q(\omega)$ over $\mathbb{F}_q$ equals $\pi_q(s) = \mathrm{ord}_s^*(q)$, the multiplicative order of q mod s. The smallest degree of an extension of $\mathbb{F}_q$ containing some (not necessarily primitive) s-th root of unity different from 1 equals

$$\alpha_q(s) = \min_{p'|s} \mathrm{ord}_{p'}^*(q) = \min_{p'|s} \pi_q(p'),$$

where the minimum ranges over all prime divisors p' of s. Equivalently, this equals the *smallest degree* occurring among the irreducible factors in the factorization of $X^{s-1} + \cdots + X + 1$ over $\mathbb{F}_q[X]$.

Let $\mathbb{F}_q$ be a finite field and let s, ℓ be positive integers such that s is coprime with q. An $\mathbb{F}_q$-linear (s,ℓ)-*quasi-cyclic code* C' is of the form $C' = \{(fa_1, \ldots, fa_\ell) : f \in \mathbb{F}_q[G_s]\} \subset (\mathbb{F}_q[G_s])^\ell$, for some fixed $a_1, ..., a_\ell \in \mathbb{F}_q[G_s]$. In particular, C' is an $\mathbb{F}_q$-linear code of length $s\ell$.

Let $R \subset \mathbb{F}_q[G_s]$ be the set formed by all $a \in \mathbb{F}_q[G_s]$ with $\sum_{g\in G_s} a(g) = 0$. Equivalently, R is the set of all $\overline{a(X)} \in \mathbb{F}_q[X]/(X^s - 1)$ with $a(1) = 0$. Recall that H_q denotes the q-ary entropy function. The following theorem is a consequence of the results in [2,8].

Theorem 5. *For a finite field $\mathbb{F}_q$, an integer $\ell > 1$, and an integer s that is not a multiple of the characteristic p of $\mathbb{F}_q$, consider the randomized construction of quasi-cyclic codes*

$$C' = \{(fa_1, \ldots, fa_\ell) \,:\, f \in \mathbb{F}_q[G_s]\} \subset (\mathbb{F}_q[G_s])^\ell,$$

where $a_1, \ldots, a_\ell$ are selected uniformly at random from R. Now consider C' as an $\mathbb{F}_q$-linear code of length $s\ell$. If δ is a real number with $0 < \delta < 1 - 1/q$ and

$$H_q(\delta) \leq 1 - \frac{1}{\ell} - \frac{\log_q s}{\ell \alpha_q(s)},$$

Then the probability that the relative minimum distance of the code C' is below δ or the rate of C' is below $\frac{1}{\ell} - \frac{1}{\ell s}$ is at most $q^{-\beta}$, where

$$\beta = \ell \alpha_q(s) \left(1 - \frac{1}{\ell} - H_q(\delta)\right) - (\ell + 2) \log_q s - \ell(1 + \log_q \ell)$$

As a consequence, for fixed values of δ, q and ℓ, if $\alpha_q(s)$ grows asymptotically faster than $\log s$, this code achieves the Gilbert-Varshamov (GV) bound for rate $1/\ell$ with high probability.

There is a natural identification between $\mathbb{F}_q^\ell[G_s]$ and $(\mathbb{F}_q[G_s])^\ell$. Indeed, every element $\mathbf{x} \in \mathbb{F}_q^\ell[G_s]$ is of the form $(\mathbf{x}(g))_{g \in G_s}$, where $\mathbf{x}(g) = (x_1(g), \ldots, x_\ell(g)) \in \mathbb{F}_q^\ell$ for every $g \in G_s$. Then $\mathbf{x} \in \mathbb{F}_q^\ell[G_s]$ can be identified with $(x_1, \ldots, x_\ell) \in (\mathbb{F}_q[G_s])^\ell$. By this identification, every $\mathbb{F}_q$-linear (s, ℓ)-quasi-cyclic code C' is a G_s-submodule of $\mathbb{F}_q^\ell[G_s]$.

We proceed next with the detailed description of our efficient randomized construction of explicit optimal AMD codes. Given a real number ϵ with $0 < \epsilon < 1$, take a large enough prime power q such that $1/q < \epsilon$ and a large enough integer ℓ such that $1/\ell < 1 - H_q(1 - \epsilon)$. Note that this means that if an $\mathbb{F}_q$-linear code is on the GV-bound and it has rate $1/\ell$, then its relative minimum distance is at least $1 - \epsilon$.

Next, we select arbitrarily large values of s such that the following conditions are satisfied.

1. The characteristic p of $\mathbb{F}_q$ does not divide s.
2. The value $\alpha_q(s)$ grows asymptotically faster than $\log s$. By Theorem 5, this ensures that the relative minimum distance or the code C' is at least $1 - \epsilon$, except with exponentially small (in s) probability.
3. Finally, $\pi_q(s) \leq s/(\ell + 1)$. This condition is needed to ensure that the rate of the code drops by at most a multiplicative positive constant factor after application of orbit avoidance.

We describe next how to efficiently select arbitrarily large values of s satisfying those conditions. Take s a product of 2 distinct odd primes, In addition, we require that these primes are different from the characteristic p of $\mathbb{F}_q$, they

have roughly the same size, and they satisfy $\pi_q(p') > \log^2 p'$. Then $\alpha_q(s)$ grows asymptotically faster than $\log s$. Indeed, since the primes p' are of similar size, $\alpha_q(s) = \Omega(\log^2 s)$. By the Prime Number Theorem, a random prime satisfies $\pi_q(p') > \log^2 p'$ with quite high probability. We can efficiently check that the condition is satisfied by simply factoring $p' - 1$ over a factor basis consisting of the primes up to $\log^2 p'$ (brute-force suffices as the factor basis is so small). Moreover, by the Chinese Remainder Theorem, it is straightforward to verify that the exponent of the group $(\mathbb{Z}/s\mathbb{Z})^*$ is at most $s/2$ if s is the product of 2 distinct odd primes. Therefore, $\pi_q(s) \leq s/2$.

Given a large enough integer s sampled as above, take a primitive s-th root of unity ω and a code $C' = \{(fa_1, \ldots, fa_\ell) : f \in \mathbb{F}_q[G_s]\} \subset \mathbb{F}_q^\ell[G_s]$ such that $a_1, \ldots, a_\ell \in R$ are selected independently and uniformly at random. By Theorem 5, the relative minimum distance of C' is at least $1 - \epsilon$ except with probability exponentially small in s. In addition, we require that $a_i(\omega) \neq 0$ for every $i = 1, \ldots, \ell$ and that there is no s-th root of unity $\eta \neq 1$ with $a_i(\eta) = 0$ for every $i = 1, \ldots, \ell$. The first property is used in Lemma 5 and the second property is used in Lemma 6. By using a similar argument as in the proof of Lemma 4, the probability that $a_i(\omega) = 0$ for some $i = 1, \ldots, \ell$ is at most $\ell q^{-\pi_q(s)}$. The probability that there is some s-th root of unity different from 1 that is a root of each $a_i(X)$ is at most $(s-1)q^{-\ell\pi_q(s)}$. Therefore, these two additional requirements do not substantially decrease the success probability (use union bound) if $\alpha_q(s)$ is much larger than $\log s$.

Let $D \subset \mathbb{F}_q[X]$ be the subset of polynomials $f(X) \in \mathbb{F}_q[X]$ such that $\deg f < s - 1$ and $f(\omega) = 1$. The code $C \subset C'$ is now formed by the codewords $(\overline{f(X)a_1(X)}, \ldots, \overline{f(X)a_\ell(X)}) \in C'$ such that $f(X) \in D$

The following two lemmas are conditioned on the "bad events" described above not happening.

Lemma 5. *$G_s \cdot \mathbf{x}$ and $G_s \cdot \mathbf{x}'$ have empty intersection for every $\mathbf{x}, \mathbf{x}' \in C$ with $\mathbf{x} \neq \mathbf{x}'$.*

Proof. Assume that the result is false. Then there exist polynomials $f(X), f'(X) \in D$ such that $X^i \cdot f(X) \cdot a_j(X) \equiv f'(X) \cdot a_j(X) \bmod (X^s - 1)$ for some integers i, j with $1 \leq i < s$ and $1 \leq j \leq \ell$. Then this implies the identity $\omega^i = 1$, which is nonsense since ω is a primitive s-th root of unity.

Lemma 6. *$|C| \geq q^{s-1-\ell\pi_q(s)} = q^{\Omega(s)}$.*

Proof. Consider the map $\phi : \mathbb{F}_q[X]_{<s} \to (\mathbb{F}_q[X]/(X^s - 1))^\ell$ defined by $\phi(f) = (fa_1, \ldots, fa_\ell)$. Then the kernel of this map is spanned by the polynomial $X^{s-1} + \cdots + X + 1$. Since the degrees of the polynomials in D are smaller than $s - 1$, it follows that $|D| = |\phi(D)| = |C|$. It now suffices to lower bound $|D|$. By Lemma 7, the kernel of the map $\psi : \mathbb{F}_q[X]_{<s} \to \mathbb{F}_q$, $f \mapsto f(\omega)$ has dimension $s - \pi_q(s)$. Hence, $|D| \geq q^{s-1-\pi_q(s)}$. The claim follows since $\pi_q(s) \leq s/2$ by hypothesis.

The final ingredient in our construction is the locking trick in Remark 1, that is, we use the multiplication AMD code described in Proposition 1 to encode

$\mathbf{x}(g) \in \mathbb{F}_q^\ell$. Since $\epsilon > 1/q$, we can take $k = \ell$, and hence we add to the tag two elements from $\mathbb{F}_{q^\ell}$.[14] This increases the tag-length by an additive constant.

This concludes the proof of Theorem 3.

Remark 3 (Achieving a smaller hidden constant). Even though this randomized construction of AMD codes is optimal, the hidden constant is very large because so is the value of ℓ. This drawback can be avoided with a simple modification to our construction. Namely, instead of the tag $(g, \mathbf{x}(g)) \in G_s \times \mathbb{F}_q^\ell$ with a lock for $\mathbf{x}(g)$, use the tag $(g, h, x_h(g)) \in G_s \times G_\ell \times \mathbb{F}_q$ with locks for h and $x_h(g)$. In this way, the tag-length is reduced from $\log s + 3\ell \log q$ to $\log s + 3 \log \ell + 3 \log q$, which is around $\log s - 6 \log \epsilon$.

References

1. Ahmadi, H., Safavi-Naini, R.: Detection of Algebraic Manipulation in the Presence of Leakage. In: Padró, C. (ed.) ICITS 2013. LNCS, vol. 8317, pp. 238–258. Springer, Heidelberg (2014)
2. Bazzi, L.M.J., Mitter, S.K.: Some Randomized Code Constructions from Group Actions. IEEE Trans. Inf. Theory 52, 3210–3219 (2006)
3. Broadbent, A., Tapp, A.: Information-theoretic security without an honest majority. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 410–426. Springer, Heidelberg (2007), `http://arxiv.org/abs/0706.2010`
4. Cabello, S., Padró, C., Sáez, G.: Secret sharing schemes with detection of cheaters for a general access structure. Des. Codes Cryptogr. 25, 175–188 (2002)
5. Cramer, R., Dodis, Y., Fehr, S., Padró, C., Wichs, D.: Information-theoretic security without an honest majority. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 471–488. Springer, Heidelberg (2008)
6. Cramer, R., Fehr, S., Padró, C.: Algebraic manipulation detection codes. Sci. China Math. 56, 1349–1358 (2013)
7. Dziembowski, S., Pietrzak, K., Wichs, D.: Non-Malleable Codes. In: Innovations in Computer Science, ICS 2010, pp. 434–452 (2010)
8. Fan, Y., Lin, L.: Thresholds of random quasi-abelian codes (2013), `http://arxiv.org/pdf/1306.5377.pdf`
9. Gordon, D., Ishai, Y., Moran, T., Ostrovsky, R., Sahai, A.: On Complete Primitives for Fairness. In: Micciancio, D. (ed.) TCC 2010. LNCS, vol. 5978, pp. 91–108. Springer, Heidelberg (2010)
10. Guruswami, V., Smith, A.: Codes for Computationally Simple Channels: Explicit Constructions with Optimal Rate. In: FOCS 2010, pp. 723–732 (2010), Full version available at arXiv.org, arXiv:1004.4017 [cs.IT]
11. Karpovski, M., Wang, Z.: Algebraic Manipulation Detection Codes and Their Applications for Design of Secure Communication or Computation Channels (2011) (manuscript), `http://mark.bu.edu/papers/226.pdf`
12. Wang, Z., Karpovsky, M.: Algebraic manipulation detection codes and their applications for design of secure cryptographic devices. In: IEEE 17th International On-Line Testing Symposium, IOLTS 2011, pp. 234–239 (2011)
13. Wee, H.: Public Key Encryption against Related Key Attacks. In: Fischlin, M., Buchmann, J., Manulis, M. (eds.) PKC 2012. LNCS, vol. 7293, pp. 262–279. Springer, Heidelberg (2012)

[14] If the additional security requirement introduced in [11] is required, a polynomial AMD code [5] with suitable parameters can be used instead.

A A Generalization of Lagrange's Interpolation Theorem

It is convenient to recall a simple extension of the usual version of Lagrange Interpolation.

Lemma 7. *Let K be a field. Fix an algebraic closure $\overline{K}$ of K. Suppose $\alpha_1, \ldots, \alpha_m \in \overline{K}$ satisfy the property that if $m > 1$ then their respective minimal polynomials $h_i(X) \in K[X]$ are pair-wise distinct. Equivalently, α_i, α_j are not Galois-conjugate over K if $i \neq j$. For $i = 1, \ldots, m$, define*

$$\delta_i = \deg h_i \ (= \dim_K K(\alpha_i)).$$

Moreover, define

$$M = \sum_{i=1}^{m} \delta_i.$$

Let $K[X]_{\leq M-1}$ denote the K-vector space of polynomials $f(X) \in K[X]$ such that $\deg f \leq M - 1$.

Then the evaluation map

$$\mathcal{E} : K[X]_{\leq M-1} \longrightarrow \bigoplus_{i=1}^{m} K(\alpha_i)$$

$$f(X) \mapsto (f(\alpha_i))_{i=1}^{m}$$

is an isomorphism of K-vector spaces.

B On BCH Codes

Let q be a prime power and let $e \geq 1$ be a positive integer. Put $s = q^e - 1$.

For any $a \in \mathbb{Z}_s$, we define a q-cylotomic coset modulo s

$$S_a := \{a \cdot q^i \bmod s : \ i = 0, 1, 2, \ldots\}.$$

It is a well-know fact that all q-cyclotomic cosets partition the set $\mathbb{Z}_s$. Let $S_{a_1}, S_{a_2}, \ldots, S_{a_t}$ stand for all distinct q-cyclotomic cosets modulo s. Then, we have that $\mathbb{Z}_s = \cup_{i=1}^{t} S_{a_i}$ and $s = \sum_{i=1}^{t} |S_{a_i}|$. We denote by s_a the size of the q-cyclotomic coset S_a. The following fact can be easily derived.

Lemma 8. *For every $a \in \mathbb{Z}_s$, the size s_a of S_a divides e which is the order of q modulo s.*

Proof. It is clear that s_a is the smallest positive integer such that $a \equiv aq^{s_a} \bmod s$, i.e, s_a is the smallest positive integer such that $s/\gcd(s, a)$ divides $q^{s_a} - 1$. Since $s/\gcd(s, a)$ also divides $q^e - 1$, we have $e \equiv 0 \bmod s_a$ by applying the long division.

Now for each S_a, we form s_a polynomials in the following way. Let $\alpha_1, \ldots, \alpha_{s_a}$ be an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^{s_a}}$ (note that $\mathbb{F}_{q^{s_a}}$ is a subfield of $\mathbb{F}_{q^e}$). Define the polynomials

$$f_{a,j}(X) := \sum_{i=0}^{s_a-1} (\alpha_j X^a)^{q^i} \tag{2}$$

for $j = 1, 2, \ldots, s_a$.

Lemma 9. *For every $a \in \mathbb{Z}_s$, we have the following facts.*

(i) *The polynomials $f_{a,j}(X)$ for $j = 1, 2, \ldots, s_a$ are linearly independent over $\mathbb{F}_q$.*
(ii) *$f_{a,j}(\beta)$ belongs to $\mathbb{F}_q$ for all $\beta \in \mathbb{F}_{q^e}$.*

Proof. The first statement is clear since the coefficients of X^a in $f_{a,j}(X)$ are α_j and $\alpha_1, \alpha_2, \ldots, \alpha_{s_a}$ form an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^{s_a}}$. To prove (ii), it is sufficient to prove that $(f_{a,j}(\beta))^q = f_{a,j}(\beta)$ for every $\beta \in \mathbb{F}_{q^e}$. Consider

$$\begin{aligned}(f_{a,j}(\beta))^q &= \left(\sum_{i=0}^{s_a-1} (\alpha_j \beta^a)^{q^i}\right)^q = \sum_{i=0}^{s_a-1} (\alpha_j \beta^a)^{q^{i+1}} \\ &= \sum_{i=1}^{s_a-1} (\alpha_j \beta^a)^{q^i} + \alpha_j^{q^{s_a}} \beta^{aq^{s_a}} = \sum_{i=1}^{s_a-1} (\alpha_j \beta^a)^{q^i} + \alpha_j \beta^a = f_{a,j}(\beta).\end{aligned}$$

This completes the proof.

Lemma 10. *The following properties hold.*

(i) *The set $\{f_{a_i,j}(X) : \ j = 1, 2, \ldots, s_{a_i}, i = 1, 2, \ldots, t\}$ is linearly independent over $\mathbb{F}_q$.*
(ii) *Let V be the $\mathbb{F}_q$-span of the set $\{f_{a_i,j}(X) : \ j = 1, 2, \ldots, s_{a_i}, i = 1, 2, \ldots, t\}$. Then the map*

$$\pi : \ V \to \mathbb{F}_q^s; \quad f(X) \mapsto ((f(\alpha))_{\alpha \in \mathbb{F}_{q^e}^*} \tag{3}$$

is an $\mathbb{F}_q$-isomorphism.

Proof. (i) The degrees of $f_{a_{i_1},j_1}(X)$ and $f_{a_{i_2},j_2}(X)$ are distinct for any $i_1 \neq i_2$. Thus, the desired result follows from Lemma 9(ii).

Since both V and $\mathbb{F}_q^s$ have the same dimension, it is sufficient to prove that π is injective. This is clear since all polynomials in V has degree at most $s-1$.

Choose an element α of $\mathbb{F}_{q^e}$ of order $s = q^e - 1$. Let $m^{(i)}(X) \in \mathbb{F}_q[X]$ denote the minimal polynomial of α^i with respect to $\mathbb{F}_q$. For $1 \leq d \leq s$, consider the BCH code B of length $s = q^e - 1$ with the generator polynomial $\text{l.c.m}\{m^{(1)}(X), m^{(2)}(X), \ldots, m^{(d-1)}(X)\}$. Then the minimum distance of B is at least d.

Lemma 11. *With notations defined above, we have $B = \pi(V_{n-d})$, where V_{n-d} is the $\mathbb{F}_q$-span of the set $\{f_{a_i,j}(X) : \ \deg(f_{a_i,j}(X)) \leq n-d\}$.*

Proof. It is clear that $f(X) = \sum_{i=0}^{s-1} f_i X^i \in \mathbb{F}_q[X]/(X^s - 1)$ belongs to B if and only if $f(\alpha^i) = 0$ for $i = 1, 2, \ldots, d-2$. This means that $(f_0, f_1, \ldots, f_{s-1})$ belongs to the dual code of the following code

$$\{(a(1), a(\alpha), \ldots, a(\alpha^{s-1})) : \ a(X) \in \mathbb{F}_q[X];\ 1 \leq \deg(a(X)) \leq d-1\}.$$

On the other hand, the dual of the above code is in fact the generalized Reed-Solomon code

$$GRS(s-d) := \{(a(1), a(\alpha), \ldots, a(\alpha^{s-1})) : \ a(X) \in \mathbb{F}_q[X];\ \deg(a(X)) \leq s-d\}.$$

This means that $B = \mathbb{F}_q^s \cap GRS(s-d)$. The desired result follows from Lemma 10(ii).