# Pushing the boundaries of group-based aggregation with zero-evading generators of low additive complexity

Ariel Gabizon[1] and Dmitry Krachun[2]

[1]Zeta Function Technologies
[2]Princeton University

July 17, 2026

**Abstract**

A *zero-evading generator* with error parameter $\lambda$ is a distribution $Z$ on $\mathbb{F}^n$ such that for any non-zero vector $x \in \mathbb{F}^n$ the probability that $\langle a, x\rangle = 0$ is at most $2^{-\lambda}$, when $a$ is chosen according to $Z$. We investigate the *number of additions* required to compute $\langle a, x\rangle$ given $x$. The traditional construction chooses a vector $a$ with random $\lambda$-bit elements. Pippenger's algorithm gives an additive complexity of at least $\Omega(\lambda n/\log(n))$ for this approach. We give a construction requiring only $O(n+\lambda)$ additions. We highlight the impact of reducing the number of additions on aggregation of group-based commitments, such as KZG commitments [KZG10].

## 1 Introduction

Recursive aggregation of zk-SNARKs is one of the main contributors to proving time in systems like the Aztec network[1]. When the proof system is elliptic-curve based, this aggregation boils down to computing a linear combination of given curve points where the coefficients of the combination are "random enough". Let us look at a concrete example to illustrate how "random enough" should be defined. In the following text we assume some familiarity with KZG commitments [KZG10]. We recommend, e.g. Section three of [GWC19] for an introduction.

Say verifier $\mathbf{V}$ has KZG commitments $C_1, \ldots, C_n$ to polynomials $f_1(X), \ldots, f_n(X)$, an evaluation point $r \in \mathbb{F}$ and values $s_1, \ldots, s_n$. Prover $\mathbf{P}$ is claiming that the values $\{s_i\}$ are indeed the evaluations of the $\{f_i\}$ at $r$; i.e. that $f_i(r) = s_i$ for each $i \in [n]$. $\mathbf{V}$ would like to somehow choose a vector of coefficients $a = (a_1, \ldots, a_n)$ to combine the commitments into $C = \sum_{i\in[n]} a_i C_i$, so that $\mathbf{P}$ only has to send a KZG proof for the

[1]https://aztec.network/

single value $f(r)$; where $f(X) := \sum_{i\in[n]} a_i f_i(X)$. In such a scheme, $\mathbf{V}$ only checks the batched value $f(r)$ against the analogous batching of the claimed evaluations, namely $\sum_{i\in[n]} a_i s_i$. Thus, the case where a malicious $\mathbf{P}$ has "gotten lucky" and can cheat, is when for some $j$ $f_j(r) \neq s_j$, and yet $\sum_{i\in[n]} a_i(f_i(r) - s_i) = 0$. Hence, the vector $a$ must be chosen so that except with negligible probability $\langle a, x\rangle \neq 0$, where $x$ is the vector with entries $x_i := (f_i(r) - s_i)$. Note that although $\mathbf{V}$ uses $a$ to combine the commitments $\{C_i\}$, the vector that $a$ must "zero-evade" is the vector $x$ defined in part by the values $f_i(r)$ not known to $\mathbf{V}$. Hence, the choice of $a$ must be non-adaptive.

This motivates the following definition.

**Definition 1.1** (Zero-Evading Generator)**.** *Fix a finite field $\mathbb{F}$ and positive integers $n, \lambda$. A* zero-evading generator (ZEG) for $\mathbb{F}^n$ with error parameter $\lambda$ *is a distribution $Z$ on $\mathbb{F}^n$ such that for any non-zero $x \in \mathbb{F}^n$ we have the following. Except with probability $2^{-\lambda}$ over the choice of $a \in \mathbb{F}^n$ according to $Z$,*

$$\langle a, x\rangle \neq 0.$$

Note that although the definition concerns field elements, $\mathbf{V}$'s task is actually to combine the *group elements* $C_i$ into $C$ via a multi-scalar multiplication (MSM). Obviously, when group elements are involved there is no "direct" multiplication with a scalar, and the MSM must be computed via an addition chain. Hence $\mathbf{V}$'s efficiency depends on the *number of group additions* needed to compute $C$ from the $\{C_i\}$. This motivates the following definition.

**Definition 1.2** (Additive complexity of ZEG)**.** *We say a ZEG $Z$ over $\mathbb{F}^n$ has* additive complexity $\tau$ *if there is an efficient randomized algorithm $\mathscr{add}$ that satisfies the following. Given any $x \in \mathbb{F}^n$, $\mathscr{add}$ computes $\langle a, x\rangle$ where $a$ is distributed according to $Z$ with at most $\tau$ additions or subtractions in $\mathbb{F}$.*

**Remark 1.3.** *We allow subtractions as our scheme will need it and inversion is efficient in the group setting by flipping the $x$-coordinate.*

There are two typical constructions of ZEGs.

1. Choose a vector $a \in \mathbb{F}^n$ where each entry $a_i$ is sampled uniformly in the interval $\left\{0, \ldots, 2^\lambda - 1\right\}$.
2. Choose uniform $r \in \left\{0, \ldots, 2^\lambda - 1\right\}$ and output the vector $a = (1, r, r^2, \ldots, r^{n-1})$.

Both approaches require running Pippenger with at least[2] $\lambda$-bit scalars giving $\Omega(\lambda n/\log(n))$ additions. Furthermore, when implementing the MSM in a SNARK circuit, the complexities of Pippenger like sorting are typically avoided and the number of adds is closer to $O(\lambda n)$.

Can we break down the multiplicative dependence between $n$ and $\lambda$ and get additive complexity $O(n + \lambda)$? In this paper we answer this question in the affirmative. In the

[2]The second approach in fact has larger scalars when $p \gg 2^\lambda$, as powers of $r$ are full field elements.

following theorem, we give two results. The first having additive complexity $O(n^2 + \lambda)$ but with concrete constants; and the second giving the desired asymptotic complexity $O(n + \lambda)$, but without explicit constants, due to the reliance on the codes of Druk and Ishai [DI14] (based on Spielman [Spi96]).

**Theorem 1.** *Fix any positive integers $n, \lambda$ with $\lambda \geq 4$ and $n^2 < e^{\lambda/16}$. Fix any finite field $\mathbb{F}$ of prime size $p > n^2 \cdot 2^{2\lambda+2}$ such that $p-1$ is divisible by four. Then, there exists*

1. *a ZEG with additive complexity $(3/2)n(n-1) + 54\lambda$,*
2. *a ZEG with additive complexity $C(n+\lambda)$, for some universal constant $C$ independent of $n, \lambda$.*

### 1.1 Overview of construction

We give an informal sketch of the construction. We focus on the first item in Theorem 1. Given $x \in \mathbb{F}^n$, our ZEG proceeds in two main steps.

Step one: Construct many combinations out of which at least half are "good" We first use $O(n^2)$ adds to compute $2n$ linear combinations of $x$ such that no more than $n$ can be zero. For this purpose, we compute a product $A \cdot x$ for a $2n \times n$ matrix $A$ for which any $n$ rows are linearly independent. How can we obtain such a matrix $A$ where $A \cdot x$ can be computed in $O(n^2)$ additions? Let $f(X)$ be the polynomial of degree $< n$ with $f(i) = x_i$ for $i \in [n]$. We take $A$ to be matrix corresponding to the linear map $A \cdot x = (x_1, \ldots, x_n, f(n+1), \ldots, f(2n))$. The point is that for this $A$, $A \cdot x$ can be computed in $O(n^2)$ adds via forward-differencing! And obviously $A$ has the property that each $n$ rows are linearly independent as it is computing $2n$ distinct polynomial evaluations.

Step two: Combining the points via "double and random add" We finish step one with $m = 2n$ points $S = (s_1, \ldots, s_m)$ such that at most $n$ are zero, and more generally, at most $n$ are identical.[3] We must combine them into one point that is non-zero with high probability.

For this purpose we use $\lambda$ rounds where in each round we sample a random $s \in S$, add it to the current point, and then double. We show using Fourier analysis, that this indeed produces a non-zero point with high probability. The analysis roughly requires that a constant fraction of *pairs of points* from $S$ are far away from each other. This might not be true for $S$ itself, but we show that for most $c \in \mathbb{F}$, for at least a quarter of the indices $j \in [\lfloor \log p/2 \rfloor]$ it's true for $c2^j \cdot S$; and that suffices for the analysis.

### Organization of the paper

In Section 2 we go over required preliminaries. In Section 3 we show step one in detail. In Section 4 we investigate properties of the doubling map modulo $p$ that will be needed.

[3] The case of all points being the same non-zero constant is analyzed separately in Section 6.

In Section 5 we show using Fourier analysis that the second step produces a non-zero point with high probability. Finally, in Section 6 we summarize the full construction. The final Section 7 explains how the quadratic dependence on $n$ is eliminated to obtain the second item in Theorem 1 by using the linear time encodable codes of [DI14, Spi96].

## 2 Preliminaries

Notation: $p$ always denotes a prime with $4|(p-1)$, and $\mathbb{F}$ the field of size $p$. At times we identify elements of $\mathbb{F}$ with the integers $\{0, \ldots, p-1\}$ in the natural way. This allows us to use elements of $\mathbb{F}$ in expressions, for example, involving complex numbers in Section 2.1. For positive integer $M$ we denote $[M] := \{1, \ldots, M\}$.

We will use the following simple claim on the distance from uniform when taking one larger interval modulo a smaller one.

**Claim 2.1.** *Fix integers $0 < s < t$. Let $A$ be an interval of non-negative integers of size $t$, i.e. $A = [a, a+1, \ldots, a+t-1]$ for some integer $a \geq 0$. Look at the random variable $X$ consisting of choosing random $x \in A$ and outputting $x \mod s$.*

*Then $X$ has distance at most $(s-1)/t$ from the uniform distribution on $[0, \ldots, s-1]$.*

*Proof.* It's enough to argue about $A = [0, \ldots, t-1]$. Given the claim for such $A$ the one for $A' = A + a$ follows from the fact that the modified random variable $X'$ simply shifts $X$ by $a$ mod $s$, which doesn't change the distance of the distribution from uniform.

Now write $t = q \cdot s + r$ for integers $q > 0$ and $0 < r < s$.

Let $G$ be the event that $x$ chosen was in $[0, \ldots, qs-1]$. We have $Pr(G) = qs/t$. Fix any subset $S \subset [0, \ldots, s-1]$. We can write the probability of the event $X \in S$ as

$$\mathbf{Pr}(X \in S) = \mathbf{Pr}(X \in S|G) \cdot (qs/t) + \mathbf{Pr}(X \in S|\neg G) \cdot (r/t).$$

Let $U$ be a uniform random variable on $[0, \ldots, s-1]$. As conditioning on $G$ makes $X$ uniform on $[0, \ldots, s-1]$, we have

$$|\mathbf{Pr}(X \in S) - \mathbf{Pr}(U \in S)| \leq r/t \leq (s-1)/t.$$

□

### 2.1 Bounding probabilities with Fourier Analysis

Let $h = (h_0, \ldots, h_{p-1}) \in \mathbb{R}^p$ be such that

- For each $0 \leq i < p$, $h_i \geq 0$.
- $\sum_{i=0}^{p-1} h_i = 1$.

We overload $h$ to denote the probability distribution on $\mathbb{F}$ described by the vector. Namely, the distribution such that for each $i \in \mathbb{F}$,

$$\mathbf{Pr}(h = i) = h_i.$$

Let $\omega \in \mathbb{C}$ be a primitive $p$'th root of unity. For each $a \in \mathbb{F}$, we define

$$\hat{h}_a = \sum_{i=0}^{p-1} h_i \omega^{a \cdot i} = \mathop{\mathbb{E}}_{i \leftarrow h} \omega^{a \cdot i}.$$

A well-known calculation shows

**Claim 2.2.**

$$h_0 = \frac{1}{p} \sum_{i=0}^{p-1} \hat{h}_i.$$

*Therefore in particular,*

$$h_0 \leq 1/p + \frac{1}{p} \sum_{i=1}^{p-1} |\hat{h}_i|$$

## 3 Computing polynomial evaluations with few additions

Given vector $x \in \mathbb{F}^n$, let $f$ be the univariate polynomial of degree $< n$ such that $f(i) = x_i$ for $i \in [n]$. Let $A : \mathbb{F}^n \to \mathbb{F}^{2n}$ be the linear map $A(x) := (x, f(n+1), \ldots, f(2n)) = (f(1), \ldots, f(2n))$.

**Claim 3.1.** *Given $x \in \mathbb{F}^n$,*

- *$A(x)$ can be computed in $n(n-1)$ additions and $n(n-1)/2$ subtractions.*
- *If $A(x)$ is not a constant vector then at most $n-1$ of its coordinates can be identical.*

*Proof.* The second item is immediate: if $f$ is nonconstant, then for every $a \in \mathbb{F}$ the polynomial $f - a$ is nonzero of degree $< n$, and hence has at most $n - 1$ roots.

It remains to describe the computation (see also Figure 1). Define the backward-difference polynomials

$$f_0(X) := f(X), \qquad f_{r+1}(X) := f_r(X) - f_r(X-1).$$

Then, by induction $\deg f_r < n - r$, so $f_{n-1}$ is constant. First, compute the values $f_r(i)$ for $r = 0, \ldots, n-1$ and $i = r+1, \ldots, n$ by the usual triangular difference table: start with $d_i = f(i)$ for $i = 1, \ldots, n$, and for $r = 1, \ldots, n-1$ replace, for $i = n, n-1, \ldots, r+1$,

$$d_i \leftarrow d_i - d_{i-1}.$$

After the $r$'th pass, $d_i = f_r(i)$ for all $i = r+1, \ldots, n$. This therefore computes $b_r := f_r(n)$ for all $r$, using

$$\sum_{r=1}^{n-1} (n - r) = \frac{n(n-1)}{2}$$

subtractions.

Now suppose we know $b_r = f_r(t)$ for all $r$ at some $t \geq n$. Since

$$f_r(t+1) = f_r(t) + f_{r+1}(t+1),$$

we can advance from $t$ to $t+1$ by updating, from top to bottom,

$$b_r \leftarrow b_r + b_{r+1} \qquad (r = n-2, n-3, \ldots, 0),$$

while $b_{n-1}$ stays fixed because $f_{n-1}$ is constant. After this update, $b_0 = f(t+1)$. Repeating this $n$ times outputs $f(n+1), \ldots, f(2n)$ and uses $n(n-1)$ additions. Together with the original input values $f(1), \ldots, f(n)$, this gives $A(x)$ with the claimed operation counts. □

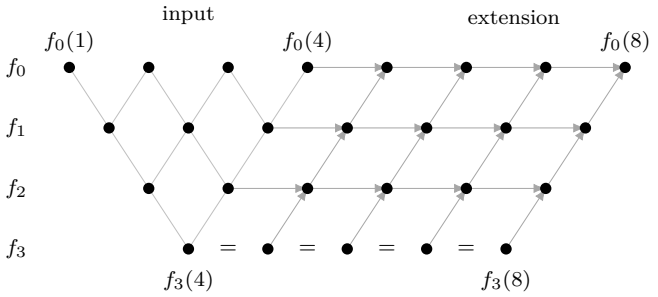



Figure 1: The backward-difference table for $n = 4$. The top row is $f_0 = f$, and the lower rows are $f_1, f_2, f_3$. The left triangle is filled by subtractions; after the right edge values $f_r(4)$ are known, the table is extended to the right by Pascal-style additions. Thin gray lines without arrows correspond to subtractions, and thin gray lines with arrows correspond to additions. Since $\deg f_3 = 0$, all values of $f_3$ are the same.

## 4 Properties of the doubling map

Let $\ell : \mathbb{F} \to \mathbb{F}$ be the map $\ell(x) := 2x$ **mod** $p$. We investigate properties of $\ell$ that are needed for our analysis in the next section.

Conventions and terminology: Operations and equalities are by default **mod** $p$. Otherwise, we explicitly say what modulus is being used or explicitly state "as integers" when operations/equalities are being done/stated as integers. For integer $M > 0$, we denote by $\mathbb{I}_M$ the interval/set $\{0, \ldots M-1\}$. When we use the term even/odd for residues modulo $M$ we mean their representatives in $\mathbb{I}_M$ are even/odd. When we say we are doing operations on elements of $\mathbb{F}$ "as integers", we mean, naturally, that we are identifying the elements of $\mathbb{F}$ with the integers $\mathbb{I}_p = \{0, \ldots, p-1\}$ and doing the operations on those integers.

Define the intervals

- $C_1 := (0, p/4)$,
- $C_2 := (p/4, p/2)$,
- $C_3 := (p/2, 3p/4)$,
- $C_4 = (3p/4, p)$.

We refer to these intervals as *quarters*, thinking of elements mod $p$ on a circle, made up of the four quarters. To be precise, note that actually $C_1 \cup C_2 \cup C_3 \cup C_4 = \mathbb{F}^* = \mathbb{F} \setminus \{0\}$.

Note that the even (as integers) elements of $\mathbb{F}^*$ are the image $\ell(C_1 \cup C_2)$; and the odd elements of $\mathbb{F}^*$ are the image $\ell(C_3 \cup C_4)$. This is because elements of $C_3 \cup C_4$ will wrap once around $p$ during doubling and $p$ is odd.

For $i \in [4]$, denote the even/odd elements of $C_i$ by $C_{i,0}$ and $C_{i,1}$ respectively. In more detail, we have

- $\ell(C_1) = C_{1,0} \cup C_{2,0}$.
- $\ell(C_2) = C_{3,0} \cup C_{4,0}$.
- $\ell(C_3) = C_{1,1} \cup C_{2,1}$.
- $\ell(C_4) = C_{3,1} \cup C_{4,1}$.

We denote by $e_i, o_i$ the interval in which the preimages of $C_{i,0}$ and $C_{i,1}$ respectively reside. Namely, $e_1 = 1, e_2 = 1, e_3 = 2, e_4 = 2, o_1 = 3, o_2 = 3, o_3 = 4, o_4 = 4$.

Let $I := (p/4, 3p/4) = C_2 \cup C_3$. We think of $I$ as the "good region" of elements far away from zero. We define $\mathbf{b} : [4] \to \{0, 1\}$ to indicate if a quarter is in the good region. Namely, $\mathbf{b}(1) = \mathbf{b}(4) := 0$ and $\mathbf{b}(2) = \mathbf{b}(3) := 1$.

A simple observation is that

**Claim 4.1.** *For each $i \in [4]$, exactly one of $\mathbf{b}(e_i)$ and $\mathbf{b}(o_i)$ is one.*

For our analysis, we need more detailed definitions regarding how the quarters interact with repeated doubling:

For $i \in [4], 0 \le j \le k$ and $r \in \mathbb{I}_{2^j}$ define $C_{i,j,r}$ to be the subset of $C_i$ equal to $r$ modulo $2^j$. Formally,

$$C_{i,j,r} := \left\{ a \in C_i | a \bmod 2^j = r \right\}.$$

**Claim 4.2.** *1. When $r$ is even*

$$C_{i,j,r} = \ell(C_{e_i,j-1,r/2}) \cap C_i.$$

*2. When $r$ is odd*

$$C_{i,j,r} = \ell(C_{o_i,j-1,r'}) \cap C_i,$$

*where $r' := (p + r)/2 \bmod 2^{j-1}$.*

*Proof.* Fix $c \in C_{i,j,r}$.

1. Suppose $r \in \mathbb{I}_{2^j}$ is even.

   This means $c$ is even as for some $q$ as integers

   $$c = 2^j q + r$$

   Hence as integers

   $$c = 2(2^{j-1}q + r/2) := 2c'$$

   For $c' \in [1, \ldots, p-1]$.

   So we have $c' \in C_{e_i, j-1, r/2}$.

2. Fix odd $r \in \mathbb{I}_{2^j}$.

   $c \in C_{i,j,r}$ implies for some $q$ as integers:

   $$c = 2^j q + r,$$

   and thus $c$ is odd. Let $c' = c/2 \mod p$. Recall the definition of $o_i$ as the index of the quarter in which the $\ell$-preimages of odd elements of $C_i$ reside. We thus have $c' \in C_{o_i}$. We have as integers

   $$2c' = p + 2^j q + r$$

   Hence as integers $c' = 2^{j-1}q + (p+r)/2$. Note that $c' \in [1, \ldots, p-1]$. Let $r' := (p+r)/2 \mod 2^{j-1}$. So we have $c' = r' \mod 2^{j-1}$. So $c' \in C_{o_i, j-1, r'}$.

□

### 4.1 The sets $B_{i,v}$

Given $c \in \mathbb{F}^*$ we define the binary vector $v(c) \in \{0,1\}^{k+1}$ that describes $c$'s "journey" through application of $\ell$ in and out of the good interval $I$. Namely, for $0 \le j \le k$, $v(c)_j = 1$ if and only if $c \cdot 2^j \in I \mod p$.

Fix $1 \le j \le k$ and vector $v \in \{0,1\}^{\le j+1}$. We define $B_{i,v}$ to be the set of $c \in \mathbb{F}^*$ with $v(c) = v$ and $c2^j \in C_i$. Note that given $v$, $B_{i,v}$ is non-empty at most for two values of $i$ according to $v_j$. Namely, if $v_j = 0$, $B_{4,v}, B_{1,v}$ can be non-empty; and when $v_j = 1$ $B_{2,v}, B_{3,v}$ can be non-empty.

**Claim 4.3.** *Fix $i \in [4], 0 \le j \le k$ and $v \in \{0,1\}^{j+1}$ with $v_j = \mathbf{b}(i)$. There exists $r \in \mathbb{I}_{2^j}$ such that $c \in B_{i,v}$ if and only if $c \cdot 2^j \in C_{i,j,r}$*

*Proof.* We prove the claim by induction on $j$. For $j = 0$, and the vector $v$ is just the bit $\mathbf{b}(i)$. And $B_{i,v} = C_{i,0,0} = C_i$.

Assume the claim for $j-1$. Fix $i \in [4]$ and $v \in \{0,1\}^{j+1}$. Let $v' = (v_0, \ldots, v_{j-1})$. From Claim 4.1, exactly one of $\mathbf{b}(e_i) = v_{j-1}$ and $\mathbf{b}(o_i) = v_{j-1}$ holds. Denote $i' = e_i$ or $i' = o_i$ according to the case. We have

$$2^j \cdot B_{i,v} = 2^j \cdot B_{i',v'} \cap C_i.$$

From the induction hypothesis, $2^{j-1} \cdot B_{i',v'} = C_{i',j-1,r'}$ for some $r' \in \mathbb{I}_{2^{j-1}}$. From Claim 4.2 $\ell \cdot C_{i',j-1,r'} \cap C_i = C_{i,j,r}$ for some $r \in \mathbb{I}_{2^j}$. Thus,

$$2^j \cdot B_{i,v} = 2^j \cdot B_{i',v'} \cap C_i = \ell(2^{j-1} B_{i',v'}) \cap C_i = C_{i,j,r}.$$

□

With the above terminology in place we state and prove the main claim of this section.

**Lemma 4.4.** *Let $p$ be an odd prime such that $4|(p-1)$. Fix positive integer $k$ with $2^k < p/4$. Define the following random variable $V = (V_0, \ldots, V_k)$: We choose random $c \in \mathbb{F}^*$ and let $V = V(c)$ as defined above.*

*Then $V$ is $2^{k+2}/(p-1)$ close to uniform on $\{0,1\}^{k+1}$.*

*Proof.* Fix a subset $S \subset \{0,1\}^{k+1}$. We want to estimate the probability that $V \in S$.

Write $S = S_0 \cup S_1$ where for $i \in \{0,1\}$, $S_i$ are the vectors $V$ in $S$ with $V_k = i$.

Now, we look at the four events $A_i := (V \in S | c2^k \in C_i)$.

According to Claim 4.3, $A_i$ is equivalent to the event $c2^k \bmod 2^k \in D_i$ for a subset $D_i \in \mathbb{I}_{2^k}$ with $D_i = S_{b(i)}$.

Thus, for each $i \in [4]$

$$\mathbf{Pr}(A_i) = \mathbf{Pr}_{c \in C_i}(c \bmod 2^k \in D_i).$$

Using Claim 2.1

$$|\mathbf{Pr}(A_i) - |D_i|/2^k| \leq 2^k/|C_i| = 2^{k+2}/(p-1)$$

Since $p-1$ is divisible by four we have $\mathbf{Pr}(c \in C_i) = 1/4$ for $i \in [4]$.

Thus

$$|\mathbf{Pr}(V \in S) - |S|/2^{k+1}| = \left|\frac{1}{4}\left(\mathbf{Pr}(A_1) + \mathbf{Pr}(A_2) + \mathbf{Pr}(A_3) + \mathbf{Pr}(A_4)\right) - |S|/2^{k+1}\right|$$

$$= \frac{1}{4}\left|\mathbf{Pr}(A_1) - |S_0|/2^k + \mathbf{Pr}(A_2) - |S_1|/2^k + \mathbf{Pr}(A_3) - |S_1|/2^k + \mathbf{Pr}(A_4) - |S_0|/2^k\right| \leq 2^{k+2}/(p-1).$$

□

**Corollary 4.5.** *The probability over uniform $c \in \mathbb{F}^*$ that $v(c)$ has less than $k/4$ ones is at most*

$$e^{-k/8} + \frac{2^{k+2}}{p-1}.$$

*Proof.* Let $H$ be the binary entropy function. One can show that a uniform vector in $\{0,1\}^{k+1}$ has less than $k/4$ ones with probability at most

$$2^{k \cdot (H(1/4)-1)} \leq e^{-k/8}.$$

The claim now follows from Lemma 4.4. □

## 5 Analyzing "double and random add" via Fourier

In this section we focus on analyzing the second step of our ZEG. To do it in a self-contained way we consider the following setting. Let $m = 2n$. We assume we are given a sequence of elements $S = (s_1, \ldots, s_m) \in \mathbb{F}^m$ such that no more than $n-1$ elements of $S$ agree with each other. That is, for any $c \in \mathbb{F}$, $|\{i|s_i = c\}| < n$.

We apply the following randomized procedure given $S$.

double-and-random-add(S,k):

1. Initialize $P = 0$.
2. For $i = 1, \ldots, k$
   (a) Set $P = \ell(P)$.
   (b) Sample random $j \in [m]$ and set $P = P + s_j$.
3. Output $P$.

We need to upper bound the probability that double-and-random-add outputs zero.

**Lemma 5.1.** *Let $k$ be a positive integer. Assume that $S$ has the form described above. Then the output $P$ of* double-and-random-add$(S, k)$ *is zero with probability at most*

$$1/p + n^2 e^{-k/8} + e^{-k/18} + n^2 \frac{2^{k+2}}{p-1}.$$

*Proof.* Note that $P$ is a random variable of the form

$$\sum_{j \in [k]} 2^{j-1} s_j,$$

where $s_1, \ldots, s_k$ are chosen from $S$ uniformly and independently. Let $h = (h_0, \ldots, h_{p-1}) \in \mathbb{R}^p$ be the distribution vector of $P$.

Now fix $c \in \mathbb{F}^*$. We have

$$\hat{h}_{2c} = \mathop{\mathbb{E}}_{s_1,\ldots,s_k \leftarrow S} \omega^{c \cdot \sum_{j \in [k]} 2^j s_j} = \mathop{\mathbb{E}}_{s_1,\ldots,s_k \leftarrow S} \prod_{j \in [k]} \omega^{c \cdot 2^j s_j}.$$

Using the independence of the $\{s_j\}$:

$$= \prod_{j \in [k]} \mathop{\mathbb{E}}_{s \leftarrow S} \omega^{c \cdot 2^j s}.$$

Denoting for $j \in [k]$

$$D_j := \mathop{\mathbb{E}}_{s \leftarrow S} \omega^{c \cdot 2^j s},$$

we have

$$|\hat{h}_{2c}| = \prod_{j\in[k]} |D_j|.$$

Note that $D_j = \frac{1}{m}\sum_{i\in[m]} \omega^{c\cdot 2^j s_i}$.

Using $z\cdot\bar{z} = |z|^2$ for $z\in\mathbb{C}$, we can write

$$|D_j|^2 = \frac{1}{m^2}\sum_{i\in[m]} \omega^{c2^j s_i} \sum_{i\in[m]} \omega^{-c2^j s_i} = \frac{1}{m^2}\sum_{(i_1,i_2)\in[m]^2} \omega^{c2^j(s_{i_1}-s_{i_2})}.$$

Denote $\ell := m(m-1)/2$ to be the number of unordered distinct pairs in $[m]$, i.e. $(i_1,i_2)$ with $1\le i_1 < i_2 \le m$. Now let $U$ be the subset of such pairs such that $s_{i_1} - s_{i_2} \neq 0$. By our assumption on $S$, $M := |U| \ge \ell/2$. For convenience we assume $M = \ell/2 \le n^2$ (otherwise take a subset of $U$of size $\ell/2$ as $U$). Let $A := \{s_{i_1} - s_{i_2}\}_{(i_1,i_2)\in U}$.

For $a\in A$, let $G_a$ be the set of $j\in[k]$ such that $ca2^j \in I$ for the set $I = (p/4, 3p/4)$ defined in Section 4. Now, let us call $c\in\mathbb{F}^*$ *good* if for each $a\in A$, the set $|G_a| \ge k/4$. According to Corollary 4.5 the fraction of bad $c$ is at most

$$M\cdot\left(e^{-k/8} + \frac{2^{k+2}}{p-1}\right) \le n^2 e^{-k/8} + n^2\frac{2^{k+2}}{p-1}.$$

Assume $c$ is good. Fix $(i_1,i_2)\in U$ and let $a := s_{i_1} - s_{i_2}$. For $j\in G_a$

$$\omega^{c2^j a} + \omega^{-c2^j a} = 2\cos(2\pi(c2^j a)/p) \le 0,$$

That is, in the indices $G_a$ the pairs $(i_1,i_2),(i_2,i_1)$ don't contribute a positive value to $|D_j|^2$.

So if we write for $j\in[k]$

$$|D_j|^2 = (1-\delta_j)$$

We must have $\sum_{j<k}\delta_j \ge \frac{2}{m^2}\cdot k/4\cdot(\ell/2) = km(m-1)/(8m^2)$. Let $\alpha := m(m-1)/(8m^2) \ge (1/9)$. It follows from the AM-GM inequality that

$$\prod_{j\in[k]} |D_j|^2 \le (1-\alpha)^k \le e^{-k/9}$$

Thus, when $c$ is good $\hat{h}_{2c} \le \sqrt{e^{-k/9}} = e^{-k/18}$.

Using the fact that $\hat{h}_{2c}\le 1$ also for bad $c\in\mathbb{F}$, using Claim 2.2 we have

$$h_0 \le 1/p + \mathbf{Pr}_{c\leftarrow\mathbb{F}}(c \text{ is bad}) + \mathbf{Pr}_{c\leftarrow\mathbb{F}}(c \text{ is good})\cdot e^{-k/18} \le 1/p + n^2 e^{-k/8} + e^{-k/18} + n^2\frac{2^{k+2}}{p-1}.$$

□

### 5.1 Error amplification:

To get the error strictly down to $2^{-\lambda}$ we can use a constant number of repetitions

**amplified**-double-and-random-add (S,k,t):

1. For $i = 1, \ldots, t$ run double-and-random-add$(S, k)$ to get output $P_i$.
2. Choose $\lambda + 1$ bit random scalars $a_1, \ldots, a_t$ and output
$$\sum_{i \in [t]} a_i P_i.$$

**Claim 5.2.** *Assume that $2^{\lambda+1} < p$. Let $S$ be as described above. Let $\gamma := 1/p + n^2 e^{-k/8} + e^{-k/18} + n^2 \frac{2^{k+2}}{p-1}$.* **amplified**-double-and-random-add$(S, k, t)$ *requires $2kt + \lambda t$ additions and outputs zero with probability at most*
$$2^{-(\lambda+1)} + \gamma^t$$

*Proof.* We know from Lemma 5.1 that the probability that all $P_i$ are zero is at most $\gamma^t$. Assuming one $P_i$ is non-zero, $\sum_{i \in [t]} a_i P_i = 0$ with probability at most $2^{-(\lambda+1)}$ as the $a_i$ are random independent $\lambda + 1$ bit scalars. □

## 6 Putting it all together

The $\mathscr{add}$ algorithm of the final ZEG to prove Theorem 1 is
$\mathscr{add}(x)$

1. Compute $z := A(x)$ where $A$ is the map in Section 3.
2. Let $S$ be the multiset of $z$'s values. Let $k := \lambda$. By our requirements on $n, \lambda, p$ in the theorem statement
$$n^2 \frac{2^{k+2}}{p-1} \leq 2^{-\lambda}, n^2 e^{-k/8} \leq e^{-k/16}.$$
Hence, we can choose $t := 18$. It can be checked that when $\lambda \geq 4$
$$\left(1/p + n^2 e^{-k/8} + e^{-k/18} + n^2 \frac{2^{k+2}}{p-1}\right)^t \leq 2^{-(\lambda+1)}.$$
3. Output $P :=$ **amplified**-double-and-random-add$(S, k, t)$.

*proof of first item of Theorem 1.* We wish to show the above procedure satisfies the requirements of Theorem 1. In terms of the additive complexity, it's clear from Claim 3.1 the first step requires $3n(n-1)/2$ additions or subtractions. And from Claim 5.2 given our settings of $k = \lambda, t = 18$ it's clear the second step requires at most $2kt + \lambda t = 54\lambda$ additions. Now fix non-zero $x \in \mathbb{F}^n$. A technicality is that we must deal separately with the case that $A(x)$ is constant.

- Suppose $A(x)$ is a non-zero constant vector $(b, \ldots, b)$. In this case double-and-random-add always outputs some non-zero multiple of $b$ - as the scalars appearing in double-and-random-add never reach $p$. Hence **amplified**-double-and-random-add outputs zero with probability at most $2^{-(\lambda+1)}$.

- Suppose $A(x)$ is non-constant. Then $S$ satisfies the requirements in Section 5 and so from Claim 5.2 **amplified**-double-and-random-add outputs zero with probability at most
$$2 \cdot 2^{-\lambda-1} = 2^{-\lambda}.$$

□

# 7 Getting to additive complexity $O(n + \lambda)$

We proceed to proving the second item of Theorem 1. Recall that in Section 5 we assumed the set $S$ of size $m = 2n$ doesn't repeat an element $n$ or more times. It is easy to see that a similar analysis would work if for some constants $d_1, d_2$, $S$ had size $m = d_2 n$ and no element repeats $d_1 n$ or more times. Using Lemma 4 of [DI14] we can replace our first phase by applying the error correcting code described there, and obtain precisely such a set $S$ (or a set $S$ consisting of repetitions of the same non-zero element). Since the encoding procedure in that lemma requires only $O(n)$ additions ($k$ in their notation) we obtain the second item of Theorem 1.

# 8 Open Problems

1. The non-explicit aspect of our proof of the second item of Theorem 1 is not completely satisfying. This ultimately originates from the use of lossless expanders in the code constructions of [DI14, Spi96]. Since our encoding can be randomized, inspecting the proof of lemma 4 of [DI14], this motivates the following question: Can we come up with an explicit, simple randomized "compression" function $c : \mathbb{F}^n \to \mathbb{F}^{n/2}$ with the following property. For some constant $0 < \delta < 1$ (ideally a constant like $1/2$) except with probability $2^{-n}$ (or similar), for any non-zero $x \in \mathbb{F}^n$ of weight at most $\delta n$, $c(x) \neq 0$.

2. The constant 54 in front of $\lambda$ in our first item of Theorem 1 is also not very satisfying, especially if thinking to use our method in practice instead of the traditional method of random $\lambda$-bit scalars. This constant can be somewhat improved with better choices of parameters, but ideally we would want to get it close to one or two. This would require a significantly more insightful analysis than the current Fourier-based one in Section 5.

## Acknowledgements:

We thank the Aztec Foundation for supporting this work. We thank Li Quan for a correction. We thank Ron D. Rothblum for discussions on linear time encodable codes.

## References

[DI14] E. Druk and Y. Ishai. Linear-time encodable codes meeting the gilbert-varshamov bound and their cryptographic applications. In *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 169–182. ACM, 2014.

[GWC19] A. Gabizon, Z. J. Williamson, and O. Ciobotaru. PLONK: permutations over lagrange-bases for oecumenical noninteractive arguments of knowledge. *IACR Cryptology ePrint Archive*, 2019:953, 2019.

[KZG10] A. Kate, G. M. Zaverucha, and I. Goldberg. Constant-size commitments to polynomials and their applications. pages 177–194, 2010.

[Spi96] D. A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Trans. Inf. Theory*, 42(6):1723–1731, 1996.