#A54 INTEGERS 16 (2016)

INTEGER ADDITION AND HAMMING WEIGHT

John Y. Kim[1]
Department of Mathematics, Rutgers University, New Brunswick, NJ 08901, USA
jonykim@math.rutgers.edu

Received: 12/22/14, Revised: 2/10/16, Accepted: 7/4/16, Published: 7/22/16

Abstract

We study the effect of addition on the Hamming weight of a positive integer. Consider the first 2^n positive integers, and fix an α among them. We show that if the binary representation of α consists of $\Theta(n)$ blocks of zeros and ones, then addition by α causes a constant fraction of low Hamming weight integers to become high Hamming weight integers. This result has applications in complexity theory to the hardness of computing powering maps using bounded-depth arithmetic circuits over $\mathbb{F}_2$. Our result implies that powering by α composed of many blocks requires exponential-size, bounded-depth arithmetic circuits over $\mathbb{F}_2$.

1. Introduction

We begin with a natural, but largely unstudied question: How does the Hamming weight of an integer (written in base 2) change under addition? To make this precise, we take $0 < \alpha < 2^n$ to be a fixed integer and let S be chosen uniformly at random from $\{0, 1, \cdots, 2^n - 1\}$. Write S in binary, and take X to be its Hamming weight. Let Y be the Hamming weight of the translation $S + \alpha$. Then what can we say about the joint distribution of initial and final weights, (X, Y)?

Our question is motivated by the problem of determining the complexity of powering maps in $\mathbb{F}_{2^n}$. This problem has been studied extensively in complexity theory [7, 2, 3, 4, 5, 6, 9]. Recently, Kopparty [7] showed that the powering map $x \to x^{\frac{1}{3}}$ from $\mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ cannot be computed with a polynomial-size, bounded-depth arithmetic circuit over $\mathbb{F}_2$ (also known as an $\mathrm{AC}^0(\oplus)$ circuit). Recall that arithmetic circuits are only allowed addition and multiplication gates which may have an unbounded number of inputs. A major advantage of working in $\mathrm{AC}^0(\oplus)$ is that it is basis invariant. That is, determining the $\mathrm{AC}^0(\oplus)$ complexity of powering does not depend on the choice of basis for F_{2^n}. Define $\mathbb{Z}_k := \mathbb{Z}/k\mathbb{Z}$. At the core of

[1] Research supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1433187.

Kopparty's argument was the following shifting property of $\frac{1}{3}$: a constant fraction of elements in $\mathbb{Z}_{2^n-1}$ change from low to high Hamming weight under translation by $\frac{1}{3}$.

Definition 1.1. Let $M = \{x \in \mathbb{Z}_{2^n-1} \mid wt(x) \leq \frac{n}{2}\}$, where $wt(x)$ is the Hamming weight of x. We say that $\alpha \in \mathbb{Z}_{2^n-1}$ has the *ϵ-shifting property* if $M \cup (\alpha + M) \geq \left(\frac{1}{2} + \epsilon\right) 2^n$.

We say that any binary string in M is *light*, and any binary string not in M is *heavy*. Then α has the ϵ-shifting property if translating $\mathbb{Z}_{2^n-1}$ by α takes a constant fraction of light strings to heavy strings.

Kopparty proved that powering by any α with the ϵ-shifting property requires exponential circuit size in $\mathrm{AC}^0(\oplus)$ [7]. We use the term *block* to denote a maximal length sequence of consecutive 0's or consecutive 1's in a binary string. Our main result is that any α with many blocks of 0's and 1's in its binary representation has the ϵ-shifting property, proving a conjecture of Kopparty.

Theorem 1.2. *Let $\sigma \in \{0,1\}^n$ be a bit-string of the form $\sigma = \sigma_1\sigma_2 \cdots \sigma_m$, where $m \geq cn$ for some constant $c > 0$, σ_i is either 0^{L_i} or 1^{L_i}, and each L_i is chosen to be maximal. Let $\alpha \in \mathbb{Z}_{2^n-1}$ have base 2 representation given by σ. Then there exists $\epsilon > 0$ such that α has the ϵ-shifting property.*

Note that the theorem still applies even in the setting of integer addition, not just when doing addition modulo $2^n - 1$. Our result states that α with $\Theta(n)$ blocks has the ϵ-shifting property. It is not difficult to show that α with $o(\sqrt{n})$ blocks does not have the ϵ-shifting property. First, observe that $o(\sqrt{n})$-sparse α (i.e. α with Hamming weight $\leq o(\sqrt{n})$) does not have the ϵ-shifting property because addition by α can only increase the weight by $o(\sqrt{n})$. Since there are $O(\frac{2^n}{\sqrt{n}})$ light binary strings of a fixed weight, we get $o(2^n)$ light strings changing to heavy strings under translation by α.

Next, observe that any α with $o(\sqrt{n})$ blocks can be written as a difference of two $o(\sqrt{n})$-sparse strings: $\alpha = \beta - \gamma$. Since translating by α is equivalent to first translating by β and then by $-\gamma$, we find that α with $o(\sqrt{n})$ blocks do not have the ϵ-shifting property. Thus, at least qualitatively, we see a strong connection between the ϵ-shifting property and the number of blocks. Establishing a full characterization of the ϵ-shifting property remains an interesting open question.

1.1. Related Work

Kopparty gave a different condition for when α has the ϵ-shifting property: its binary representation consists mostly of a repeating constant-length string that does not consist of all zeros or ones [7]. Note that any integer expressible as $\frac{a \cdot 2^n + b}{q}$, where $a, b, q \in \mathbb{Z}$, $q > 1$ is odd, and $0 < |a|, |b| < q$, has binary representation of this form. As a consequence, taking q-th roots and computing q-th residue symbols cannot be done with polynomial-size $\mathrm{AC}^0(\oplus)$ circuits. Our main result generalizes Kopparty's condition, as the periodic strings form a small subset of the strings with $\Theta(n)$ blocks.

Beck and Li showed that the q-th residue map is hard to compute in $\mathrm{AC}^0(\oplus)$ by using the concept of algebraic immunity [1]. It is worth noting that their method does not say anything about the complexity of the q-th root map in $\mathrm{AC}^0(\oplus)$. So in this regard, there is something to be gained by analyzing the ϵ-shifting property condition. A more detailed history of the complexity of arithmetic operations using low-depth circuits can be found in [7].

2. Application

It is known that powering by sparse α has polynomial-size circuits in $\mathrm{AC}^0(\oplus)$. Kopparty's work shows that powering by α with the ϵ-shifting property requires exponential-size circuits in $\mathrm{AC}^0(\oplus)$. We will use this result, along with our new generalized criterion for when α has the ϵ-shifting property, to expand the class of powering maps that are difficult to compute in $\mathrm{AC}^0(\oplus)$.

The proof resembles the method of Razborov and Smolensky for showing that Majority is not in $\mathrm{AC}^0(\oplus)$ [8, 10, 11]. We can show that for α with the ϵ-shifting property, if powering by α is computable by an $\mathrm{AC}^0(\oplus)$ circuit, then every function $f : F_{2^n} \to F_{2^n}$ is well-approximated by the sum of a low-degree polynomial with a function that sits in a low-dimensional space. The fact that there are not enough such functions provides the desired contradiction. In this way, we show certain powering maps require exponential-size circuits in $\mathrm{AC}^0(\oplus)$.

As a consequence of Theorem 1.2 and the above Razborov-Smolensky method, we get that the powering by any α with $\Theta(n)$ maximal uniform blocks requires an exponential-size $\mathrm{AC}^0(\oplus)$ circuit, thus greatly expanding the class of powers that are hard to compute in $\mathrm{AC}^0(\oplus)$.

Theorem 2.1. *Let $\alpha \in \mathbb{Z}_{2^n-1}$ have base 2 representation in the form given by Theorem 1.2. Define $\Lambda : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ by $\Lambda(x) = x^\alpha$. Then for every $AC^0(\oplus)$ circuit $C : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ of depth d and size $M \leq 2^{n^{\frac{1}{5}d}}$, for sufficiently large n we have:*

$$Pr[C(x) = \Lambda(x)] \leq 1 - \epsilon_0,$$

where $\epsilon_0 > 0$ depends only on c and d.

3. The Proof of the Main Result

3.1. Outline of Proof

Suppose we have a bit-string of length n. The bit-string is called *light* if its Hamming weight is at most $\frac{n}{2}$. The bit-string is called *heavy* otherwise. It is enough to show that translation by α in $\mathbb{Z}_{2^n-1}$ transforms some positive constant fraction of the light bit-strings into heavy bit-strings.

We choose a binary string of length n uniformly at random, translate it by α, and look at the joint distribution of its initial weight X and final weight Y. Let $(\overline{X}, \overline{Y}) = (X - \mathbb{E}[X], Y - \mathbb{E}[Y])$, so that when plotted, the plane is split into four quadrants. The fraction of strings that shift weight from light to heavy is the proportion of the distribution in the second quadrant. By symmetry, the same proportion of the distribution should lie in the fourth quadrant. We will prove that some constant fraction of the distribution lies in the second or fourth quadrant.

To get a handle on the distribution, we break up α into its m uniform blocks of 0's and 1's, and consider addition on each block separately. Addition on each block is performed from right to left, and if the leftmost sum is at least 2, then we say the addition produces a carry and the *carry bit* is 1. Otherwise, the carry bit is 0. The joint distribution of the initial weight and final weight of any block is determined by the carry bit from the addition on the previous block and the carry bit going into the next block. Thus, if the carry bits are given, then the weight distributions on the blocks are now independent. Although we will not be able to specify the distribution of the carry bits, we will show that with probability $\frac{1}{6}$, the carry bits have a certain property, and whenever they have this property, then the conditional distribution of (X, Y) has a positive constant fraction of its mass in the second or fourth quadrants.

3.2. Notation and Overview

First, observe that it suffices to prove the main result for M as viewed as a subset of $\mathbb{Z}_{2^n}$ instead of $\mathbb{Z}_{2^n-1}$. Note that only one element, $1^n \in \mathbb{Z}_{2^n}$, is not an element of $\mathbb{Z}_{2^n-1}$. Also, when translating by α, the resulting bit-string in $\mathbb{Z}_{2^n-1}$ is either the same or one more than the resulting bit-string in $\mathbb{Z}_{2^n}$. We know that $o(2^n)$ of the light bit-strings of $\mathbb{Z}_{2^n}$ transform into heavy bit-strings under translation by 1, because the only bit-strings that change from light to heavy have weight $n/2$ and end in 0. By symmetry, $o(2^n)$ of the heavy bit-strings transform into light bit-strings under translation by 1, so if $\Theta(2^n)$ light bit-strings become heavy under translation by α in $\mathbb{Z}_{2^n}$, then at least $\Theta(2^n) - o(2^n) = \Theta(2^n)$ light bit-strings become heavy under translation by α in $\mathbb{Z}_{2^n-1}$. This shows that we can work in the symmetric environment of all bit-strings of length n, $\mathbb{Z}_{2^n}$, and still achieve the result we want.

Let $S \in \mathbb{Z}_{2^n}$ be chosen uniformly at random. Let $T = \alpha + S$. Let $X = wt(S)$

and $Y = wt(T)$.

Write $\alpha = \alpha_1\alpha_2 \cdots \alpha_m$, where each α_i is a maximal length block of 0's or 1's of length L_i. Write $S = S_1S_2 \cdots S_m$, and $T = T_1T_2 \cdots T_m$, where each of the i-th parts have length L_i. Let $X_i = wt(S_i)$ and $Y_i = wt(T_i)$. Then $(X, Y) = \left(\sum_{i=1}^{m} X_i, \sum_{i=1}^{m} Y_i\right)$. Let $(\overline{X}, \overline{Y}) = (X - \mathbb{E}[X], Y - \mathbb{E}[Y])$. Then the part of the distribution of $(\overline{X}, \overline{Y})$ in the second quadrant corresponds to light bit-strings translating to heavy bit-strings. Similarly, the fourth quadrant corresponds to heavy to light bit-string translation. To avoid having to pass to analogously defined $(\overline{X_i}, \overline{Y_i})$ all the time, any reference to the second or fourth quadrant will be understood to be relative to $\left(\frac{L_i}{2}, \frac{L_i}{2}\right)$, the mean of (X_i, Y_i). We want to show that a positive constant fraction of the distribution lies in the second or fourth quadrants.

The random variables in the sum $\left(\sum_{i=1}^{m} X_i, \sum_{i=1}^{m} Y_i\right)$ are highly dependent. To get around this, we will condition on the fixing of the carry bits. Once the carry bits are fixed, the terms in the sum are independent. We will show that with probability at least $\frac{1}{6}$, we can find $\Theta(n)$ terms with identical distribution. Since the terms are independent, we will use the multidimensional Central Limit Theorem to prove that these identical distributions sum to a Gaussian distribution with dimensions of size $\Theta(\sqrt{n})$.

The remaining $O(n)$ terms can be divided into two categories. Either the term has non-zero covariance matrix or it is a translation along the line $y = -x$ relative to the mean, $\left(\frac{L_i}{2}, \frac{L_i}{2}\right)$. By applying the 2-dimensional Chebyshev Inequality to the terms with non-zero covariance matrix, we show that at least half of the distribution lies in a square with dimensions $O(\sqrt{n})$ centered at the origin. Any Gaussian with dimensions $\Theta(\sqrt{n})$ centered in the square of dimensions $O(\sqrt{n})$ will have a fixed positive proportion p of its distribution in the second quadrant and p of its distribution in the fourth quadrant. Finally, a translation of any magnitude along the line $y = -x$ still gives at least p of the distribution in the second or fourth quadrant (although we don't know which one!). However, as the addition map is a bijection from $\mathbb{Z}_{2^n}$ to itself, we get that the number of strings that go from light to heavy equals the number of strings that go from heavy to light. So we conclude that at least p of the distribution lies in the second quadrant and at least p of the distribution lies in the fourth quadrant.

3.3. Computing the Distribution

We first compute the 2-dimensional distribution of the initial and final weights of the i-th block conditioned on the carry bit from the $(i+1)$-th block. If the carry bit from the i-th block is denoted by c_i, then we want to understand the distribution of (X_i, Y_i) given the carry bit c_{i+1}. Suppose that $\alpha_i = 1^{L_i}$. The case where $\alpha_i = 0^{L_i}$

is similar.

Lemma 3.1. *Suppose that* $\alpha_i = 1^{L_i}$. *The joint distribution of* (X_i, Y_i) *conditioned on the carry bit* c_{i+1} *is given by:*

$$p_i(x, y \mid c_{i+1} = 1) = \begin{cases} \frac{1}{2^{L_i}}\binom{L_i}{x} & \textit{if } x = y \, (\textit{then } c_i = 1) \\ 0 & \textit{else} \end{cases};$$

$$p_i(x, y \mid c_{i+1} = 0) = \begin{cases} \frac{1}{2^{L_i}} & \textit{if } (x, y) = (0, L_i) \, (\textit{then } c_i = 0) \\ \frac{1}{2^{L_i}}\binom{L_i - y + x - 2}{x - 1} & \textit{if } L_i - 1 \geq y \geq x - 1 \geq 0 \, (\textit{then } c_i = 1) \end{cases}.$$

If $c_{i+1} = 1$, then $X_i = Y_i$ and $c_i = 1$. Hence, the probability mass function for (X_i, Y_i) given $c_{i+1} = 1$ is given by

$$p_i(x, y \mid c_{i+1} = 1) = \begin{cases} \frac{1}{2^{L_i}}\binom{L_i}{x} & \text{if } x = y \\ 0 & \text{else} \end{cases}.$$

If $c_{i+1} = 0$, then the distribution of (X_i, Y_i) depends solely on the number of trailing zeros, Z_i, of S_i:

$$Y_i = \begin{cases} X_i & \text{if } X_{i+1} - Y_{i+1} = L_{i+1} \\ X_i + Z_i - 1 & \text{if } Z_i < L_i \\ L_i & \text{if } Z_i = L_i \end{cases}.$$

We therefore first compute the distribution of Z_i conditioned on X_i and use that to compute the joint distribution of (X_i, Y_i). The distribution of $Z_i \mid X_i$ is given by

$$p_{Z_i}(z \mid x) = \begin{cases} 1 & \text{if } (x, z) = (0, L_i) \\ \frac{\binom{L_i - z - 1}{x - 1}}{\binom{L_i}{x}} & \text{if } L_i - z \geq x \\ 0 & \text{else} \end{cases}.$$

Since $p_i(x, y) = p_{X_i}(x) p_{Y_i}(y \mid x)$, we compute $p_{X_i}(x)$ and $p_{Y_i}(y \mid x)$. As X_i is a binomial random variable having L_i trials with success probability $\frac{1}{2}$,

$$p_{X_i}(x) = \frac{1}{2^{L_i}}\binom{L_i}{x} \text{for } x = 0, 1, \cdots, L_i.$$

We can also write the distribution of $Y_i \mid X_i$ in terms of the distribution of $Z_i \mid X_i$:

$$p_{Y_i}(y_i \mid x_i) = \begin{cases} p_{Z_i}(y_i - x_i + 1 \mid x_i) & \text{if } 0 \leq y_i - x_i + 1 < L_i \\ p_{Z_i}(y_i - x_i \mid x_i) & \text{if } (x_i, y_i) = (0, L_i) \end{cases}.$$

Hence, we have that the joint distribution of (X_i, Y_i) is

$$p_i(x, y \mid c_{i+1} = 0) = \begin{cases} \frac{1}{2^{L_i}} & \text{if } (x, y) = (0, L_i) \, (\text{then } c_i = 0) \\ \frac{1}{2^{L_i}} \binom{L_i - y + x - 2}{x - 1} & \text{if } L_i - 1 \geq y \geq x - 1 \geq 0 \, (\text{then } c_i = 1) \end{cases}.$$

Similarly, if $\alpha_i = 0^{L_i}$, then the distribution of (X_i, Y_i) is as follows: if $c_{i+1} = 0$, then $X_i = Y_i$, $c_i = 0$ and

$$p_i(x, y \mid c_{i+1} = 0) = \begin{cases} \frac{1}{2^{L_i}} \binom{L_i}{x} & \text{if } x = y \, (\text{then } c_i = 0) \\ 0 & \text{else} \end{cases}.$$

When the carry bit makes the addition trivial, we call the resulting distribution the trivial distribution. Otherwise, the carry bit $c_{i+1} = 1$. In this case, the distribution of (X_i, Y_i) turns out to be symmetric with the case where $\alpha_i = 1^{L_i}$ and $c_{i+1} = 0$:

$$p_i(x, y \mid c_{i+1} = 1) = \begin{cases} \frac{1}{2^{L_i}} & \text{if } (x, y) = (L_i, 0) \, (\text{then } c_i = 1) \\ \frac{1}{2^{L_i}} \binom{L_i - x + y - 2}{y - 1} & \text{if } L_i - 1 \geq x \geq y - 1 \geq 0 \, (\text{then } c_i = 0) \end{cases}.$$

When the carry bit makes the addition nontrivial, as in this case, we call the resulting distribution the nontrivial distribution.

Fixing the carry bits leads to four types of distributions for the blocks, which are based on the carry bit coming in from the previous block addition and the resulting carry bit from the current block addition.

1. The block distribution is trivial and produces a carry bit that makes the subsequent block distribution non-trivial (Trivial to non-trivial).

2. Non-trivial to trivial.

3. Non-trivial to non-trivial (block length $L = 1$).

4. Non-trivial to non-trivial (block length $L \geq 2$).

We make the distinction between block lengths 1 and 2 for non-trivial to non-trivial distributions, as the latter is the only distribution with invertible covariance matrix. Ideally, we will find many identical distributions of type 4, which will sum to a Gaussian with large enough dimensions. This will not be possible when most of the blocks have length 1, which we deal with separately.

Knowing the weight distribution of a block given the previous carry, it is straightforward to write down the distribution given both the previous carry and the produced carry. Again, we assume the block $\alpha_i = 1^{L_i}$.

As a trivial distribution always produces a non-trivial carry, we get the trivial to non-trivial distribution is the same is the trivial distribution:

$$p_i(x, y \mid c_{i+1} = 1, c_i = 1) = \begin{cases} \frac{1}{2^{L_i}} \binom{L_i}{x} & \text{if } x = y \\ 0 & \text{else} \end{cases}.$$

A non-trivial distribution that produces a trivial carry must have $(X_i, Y_i) = (0, L_i)$. Also, a non-trivial distribution of block length 1 that produces a non-trivial carry must have $(X_i, Y_i) = (1, 0)$.

Finally, a non-trivial distribution of block length greater than 1 that produces a non-trivial carry has distribution:

$$p_i(x, y \mid c_{i+1} = 0, c_i = 1) = \begin{cases} \frac{1}{2^{L_i}-1} \binom{L_i - y + x - 2}{x - 1} & \text{if } L_i - 1 \geq y \geq x - 1 \geq 0 \\ 0 & \text{else} \end{cases}.$$

We summarize these distributions in the next lemma:

Lemma 3.2. *Suppose that $\alpha_i = 1^{L_i}$. The joint distribution of (X_i, Y_i) conditioned on the carry bits c_{i+1} and c_i is given by:*

$$p_i(x, y \mid c_{i+1} = 1, c_i = 1) = \begin{cases} \frac{1}{2^{L_i}} \binom{L_i}{x} & \textit{if } x = y \\ 0 & \textit{else} \end{cases};$$

$$p_i(x, y \mid c_{i+1} = 0, c_i = 0) = \begin{cases} 1 & \textit{if } (x, y) = (0, L_i) \\ 0 & \textit{else} \end{cases};$$

$$p_i(x, y \mid c_{i+1} = 0, c_i = 1) = \begin{cases} \frac{1}{2^{L_i}-1} \binom{L_i - y + x - 2}{x - 1} & \textit{if } L_i - 1 \geq y \geq x - 1 \geq 0 \\ 0 & \textit{else} \end{cases}.$$

Observe that the last non-trivial to non-trivial probability distribution works for all lengths $L_i \geq 1$. However, when $L_i = 1$, $(x, y) = (0, 1)$ with probability 1. We will still consider this as a separate type of distribution as its covariance matrix is all zeros, and consequently not invertible, which will be important in our analysis.

3.4. Computing the Covariance Matrix

Lemma 3.3. *The covariance matrix M of the trivial to non-trivial distribution of the random vector (X_i, Y_i) is given by*

$$M = \begin{pmatrix} \frac{L_i}{4} & \frac{L_i}{4} \\ \frac{L_i}{4} & \frac{L_i}{4} \end{pmatrix}.$$

The covariance matrix M of the non-trivial to non-trivial distribution of the random vector (X_i, Y_i) is given by

$$M = \begin{pmatrix} c & d \\ d & c \end{pmatrix},$$

where $c = \frac{L_i}{4}\left(1+\frac{1}{2^{L_i}-1}\right) - \frac{L_i^2}{4}\left(1+\frac{1}{2^{L_i}-1}\right)\frac{1}{2^{L_i}-1}$, *and* $d = \frac{L_i}{4}\left(1+\frac{1}{2^{L_i}-1}\right) + \frac{L_i^2}{4}\left(1+\frac{1}{2^{L_i}-1}\right)\frac{1}{2^{L_i}-1} - 1$.

We will need the following binomial coefficient identities:

Lemma 3.4. *The following identities are true for every positive integer L:*

$$\begin{aligned}\sum_{n=1}^{L}\binom{L}{n}n^2 &= \frac{L(L+1)}{4}2^L;\\ \sum_{x=1}^{L-1}x(x+1)\binom{L}{x} &= 2^{L-2}L(L+3)-(L^2+L);\\ \sum_{x=1}^{L-1}x\binom{L}{x-1} &= 2^{L-1}(L+2)-(L^2+L+1).\end{aligned}$$

Proof. The proofs of these identities are standard and are included for completeness.

We use repeated differentiation of the binomial theorem to compute $\sum_{n=1}^{L}\binom{L}{n}n^2$.

$$\sum_{n=1}^{L}\binom{L}{n}x^n = (x+1)^L - 1.$$

Differentiating with respect to x yields:

$$\sum_{n=1}^{L}\binom{L}{n}nx^{n-1} = L(x+1)^{L-1} \text{ so that } \sum_{n=1}^{L}\binom{L}{n}nx^n = L(x+1)^{L-1}x.$$

Differentiating a second time with respect to x gives:

$$\sum_{n=1}^{L}\binom{L}{n}n^2x^{n-1} = L(L-1)(x+1)^{L-2}x + L(x+1)^{L-1}.$$

Plugging in $x = 1$ gives us the sum we want:

$$\begin{aligned}\sum_{n=1}^{L}\binom{L}{n}n^2 = L(L-1)2^{L-2} + L2^{L-1} &= L2^L\left(\frac{L-1}{4}+\frac{1}{2}\right)\\ &= \frac{L(L+1)}{4}2^L.\end{aligned}$$

Let $B = \sum_{x=1}^{L-1}x(x+1)\binom{L}{x}$, and let $C = \sum_{x=1}^{L-1}x\binom{L}{x-1}$. We can simplify B and C by starting with the binomial theorem and applying standard generating function methods:

$$\sum_{k=0}^{L}\binom{L}{k}x^k = (1+x)^L \text{ so that } \sum_{k=0}^{L}\binom{L}{k}x^{k+1} = x(1+x)^L.$$

Differentiating both sides with respect to x gives:

$$\sum_{k=0}^{L}(k+1)\binom{L}{k}x^k = (1+x)^{L-1}(1+(L+1)x). \tag{1}$$

Substituting $x = 1$ in Equation (1) gives:

$$\begin{aligned}\sum_{k=0}^{L}(k+1)\binom{L}{k} &= (1+x)^{L-1}(L+2)\\ C + L\binom{L}{L-1} + (L+1)\binom{L}{L} &= 2^{L-1}(L+2)\\ C &= 2^{L-1}(L+2) - (L^2+L+1).\end{aligned}$$

To get B, we differentiate Equation (1) with respect to x once more:

$$\sum_{k=1}^{L}k(k+1)\binom{L}{k}x^{k-1} = L(1+x)^{L-2}(2+(L+1)x). \tag{2}$$

Substituting $x = 1$ in Equation (2) gives:

$$\begin{aligned}\sum_{k=1}^{L}k(k+1)\binom{L}{k} &= L2^{L-2}(L+3)\\ B + L(L+1)\binom{L}{L} &= 2^{L-2}L(L+3)\\ B &= 2^{L-2}L(L+3) - (L^2+L).\end{aligned}$$

□

Proof of Lemma 3.3 To simplify notation, let $(X(L), Y(L))$ denote some (X_i, Y_i) with $L_i = L$. We begin with the trivial to non-trivial distribution. Since $X(L)$ is a binomial random variable having L trials with success probability $\frac{1}{2}$, $Var(X(L)) = \frac{L}{4}$. Since the bit string corresponding to $Y(L)$ can be viewed as a translation of the bit string corresponding to $X(L)$ in $\mathbb{Z}_{2^L}$, the distribution of $Y(L)$ is the same as the distribution of $X(L)$. Hence, $Var(Y(L)) = \frac{L}{4}$. It remains to compute $Cov(X(L), Y(L))$. In the case of the trivial distribution, $X(L) = Y(L)$. So $Cov(X(L), Y(L)) = Var(X(L)) = \frac{L}{4}$.

The case of the non-trivial to non-trivial distribution requires more work. For our computation, we assume $\alpha_i = 1^{L_i}$. As the nontrivial distributions are symmetric in

x and y, the covariances will be the same. We begin by evaluating $Var(X(L)) = \mathbb{E}[X(L)^2] - \mathbb{E}[X(L)]^2$.

As the block with weight $X(L)$ is chosen uniformly at random among all non-zero strings of length L, we have

$$\mathbb{E}[X(L)] = \frac{L}{2}\left(1 + \frac{1}{2^L - 1}\right) \text{ and } \mathbb{E}[X(L)^2] = \frac{1}{2^L - 1}\sum_{n=1}^{L}\binom{L}{n}n^2.$$

Using Lemma 3.4, the variance of $X(L)$ is given by:

$$\begin{aligned} Var(X(L)) &= \frac{L(L+1)}{4}\frac{2^L}{2^L - 1} - \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)^2 \\ &= \frac{L^2 + L}{4}\left(1 + \frac{1}{2^L - 1}\right) - \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right) - \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1} \\ &= \frac{L}{4}\left(1 + \frac{1}{2^L - 1}\right) - \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1}. \end{aligned}$$

Observe that Y is the weight of a block of length L chosen uniformly at random from all strings except 1^L. So by symmetry, $Var(Y) = Var(X)$. We now compute $Cov(X, Y) = \mathbb{E}[X(L)Y(L)] - \mathbb{E}[X(L)]\mathbb{E}[Y(L)]$:

$$\begin{aligned} \mathbb{E}[X(L)Y(L)] &= \frac{1}{2^L - 1}\sum_{1 \le y \le x+1 \le L} xy\binom{L - x + y - 2}{y - 1} \\ &= \frac{1}{2^L - 1}\sum_{x=0}^{L-1} x \sum_{y=1}^{x+1} y\binom{L - x + y - 2}{y - 1} \\ &= \frac{1}{2^L - 1}\sum_{x=0}^{L-1} x \sum_{y=0}^{x} (y+1)\binom{L - x + y - 1}{y}. \end{aligned}$$

Let $A(x) = \sum_{y=0}^{x}(y+1)\binom{L - x + y - 1}{y}$ be the inner summation. Then by repeated application of the hockey stick identity,

$$\begin{aligned} A(x) &= \sum_{y=0}^{x}(x+1)\binom{L - x + y - 1}{y} - \sum_{y=0}^{x-1}(x - y)\binom{L - x + y - 1}{y} \\ &= (x+1)\binom{L}{x} - \sum_{y=0}^{x-1}\sum_{j=0}^{y}\binom{L - x + j - 1}{j} \\ &= (x+1)\binom{L}{x} - \sum_{y=0}^{x-1}\binom{L - x + y}{y} \\ &= (x+1)\binom{L}{x} - \binom{L}{x - 1}. \end{aligned}$$

Substituting $A(x)$ back into our expression for $\mathbb{E}[X(L)Y(L)]$ yields:

$$\mathbb{E}[X(L)Y(L)] = \frac{1}{2^L-1}\left[\sum_{x=1}^{L-1} x(x+1)\binom{L}{x} - \sum_{x=1}^{L-1} x\binom{L}{x-1}\right].$$

Using Lemma 3.4, we get:

$$\begin{aligned}
\mathbb{E}[X(L)Y(L)] &= \frac{1}{2^L-1}(2^{L-2}L(L+3)-(L^2+L)-2^{L-1}(L+2)+(L^2+L+1)) \\
&= \frac{2^L}{2^L-1}\left(\frac{L(L+1)}{4}-\frac{L(L+1)}{2^L}-\frac{L+2}{2}+\frac{L^2+L+1}{2^L}\right) \\
&= \left(1+\frac{1}{2^L-1}\right)\left(\frac{L^2+L-4}{4}+\frac{1}{2^L}\right) \\
&= \left(1+\frac{1}{2^L-1}\right)\left(\frac{L^2+L}{4}-\left(1-\frac{1}{2^L}\right)\right) \\
&= \frac{L^2+L}{4}\left(1+\frac{1}{2^L-1}\right)-1.
\end{aligned}$$

Hence the covariance of X(L) and Y(L) is

$Cov(X(L), Y(L))$

$$\begin{aligned}
&= \frac{L^2+L}{4}\left(1+\frac{1}{2^L-1}\right)-1-\frac{L}{2}\left(1+\frac{1}{2^L-1}\right)\frac{L}{2}\left(1-\frac{1}{2^L-1}\right) \\
&= \frac{L^2+L}{4}\left(1+\frac{1}{2^L-1}\right)-1-\frac{L^2}{4}\left(1+\frac{1}{2^L-1}\right)+\frac{L^2}{4}\left(1+\frac{1}{2^L-1}\right)\frac{1}{2^L-1} \\
&= \frac{L}{4}\left(1+\frac{1}{2^L-1}\right)+\frac{L^2}{4}\left(1+\frac{1}{2^L-1}\right)\frac{1}{2^L-1}-1.
\end{aligned}$$

□

3.5. Breaking Up the Weight Distribution

Recall that the joint distribution of initial and final weights is the sum of the joint distribution of initial and final weights for $m = \Theta(n)$ smaller parts: $(X, Y) = \sum_{i=1}^{m}(X_i, Y_i) := \left(\sum_{i=1}^{m} X_i, \sum_{i=1}^{m} Y_i\right)$. However, the terms of this sum are dependent. We can remove the dependence by first sampling the carry bits according to their distribution. Given the carry bits, all of the terms in the sum are independent and have one of the four types of distributions given by Lemma 3.2. This gives us access to the Central Limit Theorem and the fact that covariance matrices add, both of which will be used in the proof.

We will break up (X, Y) into a sum of a Gaussian (X_G, Y_G), a translation (X_T, Y_T), and some remainder (X_R, Y_R), which we show is well-behaved. As the non-trivial to non-trivial distribution with block length at least 2 (type 4) is the only type with invertible covariance matrix, our goal will be to find many identical distributions of this type. By the Central Limit Theorem, these sum to a two-dimensional Gaussian (X_G, Y_G) of dimensions $\Theta(\sqrt{n})$. This will be the main part of the sum that pushes the distribution into the second and fourth quadrants.

It is not always possible to find many identical distributions of type 4. If there are $o(n)$ blocks of length at least 2, then it is trivially impossible. We deal with this case separately with a slightly modified argument. Otherwise, there are $\Theta(n)$ blocks of length at least 2. We will show that with probability at least $\frac{1}{6}$, the carry bits arrange themselves in such a way so that there are $\Theta(n)$ distributions of type 4. This is enough to find many identical distributions of type 4.

We then consider the sum of the remainder of the type 4 distributions along with the trivial to non-trivial type 1 distributions, (X_R, Y_R), and show that it is well-behaved. As the covariance matrices add, we will be able to apply the two-dimensional Chebyshev inequality to guarantee that half of the distribution lies inside an ellipse of dimensions $O(\sqrt{n})$. This will be enough to guarantee some constant proportion p of the distribution of $(X_G, Y_G) + (X_R, Y_R)$ in the second quadrant, and the same proportion p in the fourth quadrant. The rest of the distribution of (X, Y) is a translation (X_T, Y_T) along the line $y = -x$ relative to the mean. After translation, we still have proportion p of the distribution in either the second or fourth quadrant.

We first consider the case where there are $m' = \Theta(n)$ blocks of length at least 2. The following lemma says that with probability at least $\frac{1}{6}$, we get many distributions of type 4.

Lemma 3.5. *Suppose there are $m' = \Theta(n)$ blocks of length at least 2. Let X be the number of non-trivial to non-trivial distributions with block length at least 2. Then*

$$\mathbb{P}\left(X > \frac{m'}{4}\right) > \frac{1}{6}.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{m'} \mathbb{P}(\text{Block } i \text{ is non-trivial to non-trivial}) \\
&= \sum_{i=1}^{m'} \mathbb{P}(\text{Non-trivial carry out} \mid \text{non-trivial carry in}) \cdot \mathbb{P}(\text{Non-trivial carry in}) \\
&\geq \sum_{i=1}^{m'} \frac{3}{4} \cdot \frac{1}{2} \\
&= \frac{3}{8} m'.
\end{aligned}$$

Let Y be the number of blocks of length at least 2 that are not non-trivial to non-trivial. Then $\mathbb{E}[Y] \leq \frac{5}{8}m'$. By Markov's inequality,

$$\mathbb{P}(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t} \leq \frac{5}{8} \cdot \frac{m'}{t}.$$

Taking $t = \frac{3}{4}m'$ yields

$$\mathbb{P}(Y \geq \frac{3}{4}m') \leq \frac{5}{6}, \quad \mathbb{P}(Y \leq \frac{3}{4}m') \geq \frac{1}{6}, \quad \text{and} \quad \mathbb{P}(X \geq \frac{1}{4}m') \geq \frac{1}{6}. \qquad \square$$

We now show that having many distributions of type 4 implies having many identical distributions of type 4.

Lemma 3.6. *Suppose we have $m = \Theta(n)$ bit-strings of total length at most n. Then there is some fixed positive length L such that $l = \Theta(n)$ bit-stings have length L.*

Proof. Let n_k be the number of blocks of length k. Then we have the following equations about the total number of blocks and the total length of all the blocks:

$$\begin{aligned} \sum_{k=1}^{n} n_k &= m; \\ \sum_{k=1}^{n} k \cdot n_k &\leq n. \end{aligned}$$

Since $m = \Theta(n)$, we have $\frac{n}{m} = C = \Theta(1)$. Multiplying the first equation by C and subtracting from the second inequality, we get:

$$-C\sum_{k=1}^{C+1} n_k + \sum_{k>C+1} n_k \leq \sum_{k=1} n(k-C)n_k \leq 0.$$

So we have:

$$\begin{aligned} \frac{n}{C} = m &= \sum_{k=1}^{n} n_k \\ &= \sum_{k=1}^{C+1} n_k + \sum_{k>C+1} n_k \\ &\leq (C+1)\sum_{k=1}^{C+1} n_k \\ &\leq (C+1)^2 \max_{k\in[C]} n_k. \end{aligned}$$

So $\max_{k\in[C]} n_k \geq \frac{1}{C(C+1)^2}n$. Taking $a = \frac{1}{C(C+1)^2}$, we get that there exists a fixed length $L \in [C]$ such that an bit-strings have length L, as desired. □

So with probability at least $\frac{1}{6}$, we can find $l = \Theta(n)$ identical type 4 distributions. Since these identical distributions are independent of each other, the Central Limit Theorem together with Lemma 3.3 tells us that the distribution of their sum is a Gaussian with covariance matrix given by

$$M_G = \begin{pmatrix} cl & dl \\ dl & cl \end{pmatrix},$$

where $L \geq 2$,

$$c = \frac{L}{4}\left(1 + \frac{1}{2^L - 1}\right) - \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1},$$

and

$$d = \frac{L}{4}\left(1 + \frac{1}{2^L - 1}\right) + \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1} - 1.$$

Lemma 3.7. *Suppose G is a 2-dimensional Gaussian distribution with covariance matrix given by M_G. Then a fixed positive proportion of the distribution of G lies inside (and outside) an ellipse centered at the mean with dimensions $\Theta(\sqrt{n})$. Furthermore, the probability density function f_G satisfies $f_G(x, y) \geq \frac{1}{\pi n} e^{-\frac{144n}{l}}$ inside a circle of radius $4\sqrt{n}$ centered at the mean of G.*

Proof. Observe first that $c - d = 1 - \frac{L^2}{2}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1} \geq \frac{1}{9}$ and $d \geq \frac{1}{9}$ for any value of $L \geq 2$. As $\det(M_G) = (c^2 - d^2)l^2 \neq 0$, M_G is invertible. So letting $\overline{G} = (\overline{X_G}, \overline{Y_G})$ denote the distribution obtained by translating G to its mean, we get that a fixed proportion of the distribution lies in the ellipse defined by:

$$\begin{pmatrix} \overline{X_G} & \overline{Y_G} \end{pmatrix} M_G^{-1} \begin{pmatrix} \overline{X_G} \\ \overline{Y_G} \end{pmatrix} = 2.$$

The inverse of M_G is given by:

$$M_G^{-1} = \frac{1}{(c^2 - d^2)l^2} \begin{pmatrix} cl & -dl \\ -dl & cl \end{pmatrix}.$$

Substituting M_G^{-1} back into the equation of the ellipse gives

$$\begin{aligned} \frac{1}{(c^2 - d^2)l}[c\overline{X_G}^2 - 2d\overline{X_G Y_G} + c\overline{Y_G}^2] &= 2 \\ \overline{X_G}^2 - \frac{2d}{c}\overline{X_G Y_G} + \overline{Y_G}^2 &= \frac{2(c^2 - d^2)}{c}l. \end{aligned}$$

We have an equation of the form $x^2 - 2axy + y^2 = b$, where $a = \frac{d}{c} < 1$. This describes an ellipse rotated by $\frac{\pi}{4}$ counterclockwise. By rotating the ellipse clockwise by $\frac{\pi}{4}$, we can find the dimensions of the ellipse. Making the substitution:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2}(x' + y') \\ \frac{\sqrt{2}}{2}(y' - x') \end{pmatrix},$$

we get that the equation of the rotated ellipse is

$$\frac{x^2}{\frac{b}{1+a}} + \frac{y^2}{\frac{b}{1-a}} = 1. \tag{3}$$

Taking $a = \frac{d}{c}$ and $b = \frac{2(c^2-d^2)}{c}l$ as in the ellipse for our Gaussian, we find that the squares of the dimensions of the ellipse are given by:

$$\begin{aligned}\frac{b}{1-a} &= \frac{2(c^2-d^2)}{c}l \cdot \frac{c}{c-d} = 2(c+d)l \geq \frac{2}{3}l;\\ \frac{b}{1+a} &= \frac{2(c^2-d^2)}{c}l \cdot \frac{c}{c+d} = 2(c-d)l \geq \frac{2}{9}l.\end{aligned}$$

Hence, both dimensions of the ellipse are $\Theta(\sqrt{n})$. In fact, both dimensions exceed $\frac{1}{3}\sqrt{l}$. So the circle of radius $\frac{1}{3}\sqrt{l}$ centered at the mean lies completely inside the ellipse. Scaling every dimension up by a factor of $\sqrt{\frac{144n}{l}}$ tells us that inside the circle of radius $4\sqrt{n}$ centered at the mean, we have:

$$\begin{pmatrix}\overline{X_G} & \overline{Y_G}\end{pmatrix} M_G^{-1} \begin{pmatrix}\overline{X_G} \\ \overline{Y_G}\end{pmatrix} \leq \frac{144n}{l}.$$

Therefore, we have the following lower bound on the probability distribution function inside the circle of radius $4\sqrt{n}$:

$$\begin{aligned}f_G(x,y) &\geq \frac{1}{2\pi\sqrt{\det M_G}}e^{-\frac{144n}{l}}\\ &= \frac{1}{2\pi\sqrt{(c^2-d^2)l^2}}e^{-\frac{144n}{l}}\\ &\geq \frac{1}{2\pi cl}e^{-\frac{144n}{l}}\\ &\geq \frac{1}{\pi n}e^{-\frac{144n}{l}}.\end{aligned}$$

□

The above sequence of lemmas can be used to show that if we start with many blocks of length at least 2, then with positive constant probability, we can find many identical distributions that sum to a Gaussian of dimensions $\Theta(\sqrt{n})$. Suppose now that the exponent α has a total of m blocks, but fewer than $0.01m$ blocks of length at least 2. Then at least 0.99 fraction of the blocks have length 1. Consider all consecutive block pairs. At most 0.01 fraction of these pairs have their first block with length 2, and at most 0.01 fraction have their second block with length at least 2. So at most 0.02 fraction have a block of length at least 2. Hence, 0.98 fraction of the pairs consists of two blocks of length 1. By the Pigeonhole Principle, at least

0.49 fraction of the pairs are either all 01 or all 10. Without loss of generality, assume that 0.49 fraction of consecutive block pairs are 01. We now treat each block pair 01 as a single block of length 2. The initial and final weight distribution of this larger block, given there is no carry in and no carry out matches the type 4 distribution. We have proven the existence of a large number of modified blocks of length 2:

Lemma 3.8. *Suppose there are fewer than* 0.01 *fraction of the blocks have length at least* 2*. Then at least* 0.49 *fraction of consecutive block pairs are* 01 *or at least* 0.49 *fraction of consecutive block pairs are* 10*.*

Lemma 3.8 essentially reduces the case of having few blocks of length at least 2 to the case where there are many blocks of length at least 2 by consolidating many of the length 1 blocks. As there are $\Theta(n)$ such consolidated blocks, and each has type 4 distribution with probability at least $\frac{3}{8}$, we can again find $\Theta(n)$ identical type 4 distributions by Lemma 3.5. Lemma 3.7 then says that these identical distributions sum to a two-dimensional Gaussian with large x and y dimensions. So for any α, we can find many terms in the initial and final weight distribution summing to a large Gaussian.

3.6. Distribution of the Sum of the Remaining Terms

Consider the terms remaining in the distribution of $(X, Y) = \sum_{i=1}^{m} (X_i, Y_i)$, when the terms contributing to the Gaussian are removed. The terms with distribution types 2 or 3 are translations in the $< 1, -1 >$ direction relative to $\left(\frac{L}{2}, \frac{L}{2}\right)$. These will contribute to the translation part of the distribution (X_T, Y_T). The rest of the terms of type 4 along with the terms of type 1 sum to the remainder $R = (X_R, Y_R)$. There are $O(n)$ terms remaining. By Lemma 3.3, the covariance matrix of each of these terms is one of the following two forms:

$$\begin{pmatrix} \frac{L}{4} & \frac{L}{4} \\ \frac{L}{4} & \frac{L}{4} \end{pmatrix}; \qquad \begin{pmatrix} c & d \\ d & c \end{pmatrix},$$

where

$$c = \frac{L}{4}\left(1 + \frac{1}{2^L - 1}\right) - \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1},$$

and

$$d = \frac{L}{4}\left(1 + \frac{1}{2^L - 1}\right) + \frac{L^2}{4}\left(1 + \frac{1}{2^L - 1}\right)\frac{1}{2^L - 1} - 1.$$

As the terms are independent given a fixing of the carry bits, the covariance matrices sum. The total covariance matrix of the sum is:

$$M_R = \begin{pmatrix} C & D \\ D & C \end{pmatrix},$$

where $D \leq C \leq \frac{n}{3}$, with $D = C$ only when the remainder is a sum of type 1 distributions.

Lemma 3.9. *At least half of the distribution of the remainder lies in a circle of radius $\sqrt{2n}$ centered at the mean.*

Proof. When the remainder is a sum of type 1 distributions, then the remainder has the form (X_R, X_R), where X_R is a binomial distribution with $L_R \leq n$ trials and success probability $\frac{1}{2}$. So by Chebyshev's Inequality,

$$\mathbb{P}\left(|X_R - \frac{L_R}{2}| > \sqrt{\frac{L_R}{2}}\right) \leq \frac{1}{2}$$
$$\mathbb{P}\left(|X_R - \frac{L_R}{2}| > \sqrt{\frac{n}{2}}\right) \leq \frac{1}{2}$$
$$\mathbb{P}\left(|X_R - \frac{L_R}{2}| \leq \sqrt{\frac{n}{2}}\right) \geq \frac{1}{2}.$$

So in this case, at least half of the distribution of the remainder lies in a circle of radius $\sqrt{\frac{n}{2}}$. When the remainder contains some type 4 distributions, then $D < C \leq \frac{n}{3}$. Hence, the covariance matrix M_R is invertible. So we may apply the 2-dimensional Chebyshev's inequality to (X_R, Y_R) to get:

$$\mathbb{P}\left\{\overline{R} M_R^{-1} \overline{R}^T > t\right\} \leq \frac{2}{t^2}.$$

Taking $t = 2$ yields

$$\begin{aligned} Pr\left\{\overline{X_R}^2 + 2\frac{D}{C}\overline{X_R Y_R} + \overline{Y_R}^2 > \frac{2(C^2 - D^2)}{C}\right\} &\leq \frac{1}{2} \\ Pr\left\{\overline{X_R}^2 + 2\frac{D}{C}\overline{X_R Y_R} + \overline{Y_R}^2 \leq \frac{2(C^2 - D^2)}{C}\right\} &> \frac{1}{2}. \end{aligned}$$

Chebyshev's inequality tells us that at least half of the distribution lies in the ellipse centered at the origin defined by the inequality above. By a similar computation as with the Gaussian distribution, the squares of the dimensions of this ellipse are $2(C + D)$ and $2(C - D)$, both of which are less than $\frac{4n}{3}$. Hence, the ellipse lies inside a circle of radius $2\sqrt{\frac{n}{3}} < \sqrt{2n}$, and therefore over half of the distribution of the remainder must lie inside this circle. □

3.7. The Proof

We are ready to prove the main theorem.

Proof of Theorem 1.2. Suppose that $\alpha \in \mathbb{Z}_{2^n}$ has $m \geq cn$ blocks in its binary representation, where $0 < c < 1$ is constant. We know that either there are $0.01m$

blocks of length at least 2, or there are fewer than $0.01m$ blocks of length at least 2. In the second case, Lemma 3.8 tells us that we can find $0.49m$ identical pairs of consecutive blocks of length 1. Lemma 3.5 then says that with probability at least $\frac{1}{6}$, the carry bits arrange themselves in such a way that there are at least $\frac{0.49}{4}m \geq \frac{1}{9}m$ identical type 4 distributions.

If there are $m' > \frac{1}{100}m$ blocks of length at least 2, then Lemma 3.5 says that with probability at least $\frac{1}{6}$, the carry bits arrange themselves in such a way that there are at least $\frac{1}{4}m' > \frac{1}{400}m$ type 4 distributions. Since $\frac{1}{9} > \frac{1}{400}$, we conclude that for any α with m blocks, we can find $\frac{1}{400}m$ type 4 distributions with probability $\frac{1}{6}$.

As $m \geq cn$, the number of type 4 distributions exceeds $\frac{1}{400}m \geq \frac{c}{400}n$. By Lemma 3.6, we can find $l \geq \left(\frac{c}{400}\right)^3 n$ identical type 4 distributions each with block length L. By Lemma 3.7, these sum to a Gaussian whose probability distribution function $f_G(x,y) \geq \frac{1}{\pi n}e^{-\frac{144n}{l}}$ inside a circle of radius $4\sqrt{n}$ centered at the mean of G. As each type 4 distribution has mean $\left(\frac{L}{2}, \frac{L}{2}\right) \pm \frac{L}{2} \cdot \frac{1}{2^L - 1}(1,-1)$, we decompose the Gaussian into a Gaussian $G = (X_G, Y_G)$ centered at $\left(\frac{Ll}{2}, \frac{Ll}{2}\right)$ and a translation in the $(1,-1)$ direction which contributes to the translation term (X_T, Y_T).

It is worth noting that every distribution type of block length L can be decomposed into the sum of a distribution centered at $\left(\frac{L}{2}, \frac{L}{2}\right)$ and a translation in the $(1,-1)$ direction. To see this, we will write the mean of each type of distribution as $\left(\frac{L}{2}, \frac{L}{2}\right) + k(1,-1)$, for some k depending on L.

Type 1 distributions have mean $\left(\frac{L}{2}, \frac{L}{2}\right)$. Type 2 distributions have mean $\left(\frac{L}{2}, \frac{L}{2}\right) \pm \frac{L}{2}(1,-1)$. Type 3 distributions have mean $\left(\frac{L}{2}, \frac{L}{2}\right) \pm \frac{L}{2}(1,-1)$. Type 4 distributions have mean $\left(\frac{L}{2}, \frac{L}{2}\right) \pm \frac{L}{2} \cdot \frac{1}{2^L - 1}(1,-1)$.

We extract the translation component from each term and call that sum (X_T, Y_T). Let $R = (X_R, Y_R) = (X,Y) - (X_G, Y_G) - (X_T, Y_T)$ be the remainder. If L_T denotes the total length of all blocks contributing to R, we have that $\left(\frac{L_T}{2}, \frac{L_T}{2}\right)$ is the mean of R. Let $\overline{R} = R - \left(\frac{L_T}{2}, \frac{L_T}{2}\right)$. By Lemma 3.9, at least half of the distribution of $\overline{R}$ lies in an circle centered at the origin with radius $\sqrt{2n}$. By taking the square W of side length $2\sqrt{2n}$ surrounding the circle, we see that at least half of the distribution of R lies in W.

For any point (p,q) in the square, consider the distribution of the Gaussian $\overline{G} + (p,q) = (\overline{X_G} + p, \overline{Y_G} + q)$, which is centered at (p,q). Lemma 3.7 guarantees that the probability distribution function exceeds $\frac{1}{\pi n}e^{-\frac{144n}{l}}$ inside a circle of radius $4\sqrt{n}$ centered at (p,q). Contained within this circle is a square with side length $\sqrt{2n}$ in the second quadrant. Hence, the probability that $\overline{G} + (p,q)$ lies in the second quadrant is at least $\frac{2}{\pi}e^{-\frac{144n}{l}}$.

Recall that $l \geq \frac{c^3}{400^3}n$, where $m \geq cn$. So we have:

$$\begin{aligned} \frac{2}{\pi}e^{-\frac{144n}{l}} &\geq \frac{2}{\pi}e^{-\frac{144 \cdot 64000000}{c^3}} \\ &= \frac{2}{\pi}e^{-\frac{9216000000}{c^2}}. \end{aligned}$$

Take $C = \frac{2}{\pi}e^{-\frac{9216000000}{c^2}}$. Then with probability $\frac{1}{6}$, at least C fraction of the distribution of $\overline{G} + \overline{R}$ conditioned on the carry bits lies in the second quadrant, where C is a constant depending only on c. By symmetry, the same fraction lies in the fourth quadrant. Finally, we must add the translation (X_T, Y_T) in the $(1, -1)$ direction. No matter the size of the translation, we are guaranteed C fraction in either the second or fourth quadrant. Hence, we have at least $\frac{C}{6}$ of the unconditioned distribution of $(X, Y) = (X_G, Y_G) + (X_R, Y_R) + (X_T, Y_T)$ lying in the second or fourth quadrants relative to the mean $\left(\frac{n}{2}, \frac{n}{2}\right)$, and so at least $\frac{C}{12}$ fraction lying in either the second or fourth quadrant. By symmetry, we get at least $\frac{C}{12}$ fraction lying in both the second and fourth quadrants. □

4. Heavily Shifting Numbers

We have shown that α with many uniform blocks of 0's and 1's has the ϵ-shifting property. An interesting related question is whether there is an α that is heavily shifting: that is, α shifts almost all of the light strings to heavy strings. More precisely, $o(1)$ fraction of light strings remain light under translation by α. We already know that when α has $o(\sqrt{n})$ blocks, α does not have the ϵ-shifting property, and can therefore not be heavily shifting.

Our current understanding of the joint initial and final weight distribution cannot quite show that α with $\Theta(n)$ blocks is not heavily shifting. The reason is that we have no handle on the size of the translation term (X_T, Y_T) in the $(1, -1)$ direction. It is possible that the translation is so large most of the time to make α heavily shifting, though we suspect this does not happen.

It is an open problem to figure out which α are heavily shifting.

Acknowledgements The author thanks Swastik Kopparty for many useful discussions. The author also would like to thank the referee for pointing out an elementary proof of Lemma 3.6.

References

[1] C. Beck and Y. Li, Represent MOD function by low degree polynomial with unbounded one-sided error, arXiv:1304.0713, (2013).

[2] W. Eberly, Very fast parallel polynomial arithmetic, *SIAM J. Comput.*, **18** (1989), 955-976.

[3] F. E. Fich and M. Tompa, The parallel complexity of exponentiating polynomials over finite fields, *J. AMC*, **35** (1988), 651-667.

[4] J. von zur Gathen, Efficient exponentiation in finite fields (extended abstract), *FOCS*, (1991), 384-391.

[5] A. Healy and E. Viola, Constant-depth circuits for arithmetic in finite fields of characteristic two, *STACS*, volume **3884** of *Lecture Notes in Computer Science*, Springer, New York, 2006.

[6] W. Hesse, E. Allender, and D. A. M. Barrington, Uniform constant-depth circuits for division and iterated multiplication, *J. Comput. Syst. Sci.*, **65** (2002), 695-716.

[7] S. Kopparty, On the complexity of powering in finite fields, *STOC* (2011).

[8] A. Razborov, Lower bounds on the size of bounded depth circuits over a complete basis with logical addition, *MATHNASUSSR: Mathematical Notes of the Academy of Sciences of the USSR*, **41**, (1987).

[9] I. E. Shparlinsky, Number Theoretic Methods in Cryptography: Complexity lower bounds, Volume **17** of *Progress in Computer Science and Applied Logic*, Birkhäuser Verlag, 1999.

[10] R. Smolensky, Algebraic methods in the theory of lower bounds for boolean circuit complexity, *STOC* (1987), 77-82.

[11] R. Smolensky, On representations by low-degree polynomials, *FOCS* (1993), 130-138.