Journal of Computational and Applied Mathematics 41 (1992) 215–235
North-Holland

CAM 1174

An affine walk on the hypercube

Persi Diaconis
Department of Mathematics, Harvard University, Cambridge, MA 02138, United States

Ron Graham
AT&T Bell Laboratories, Murray Hill, NJ 07974, United States

Received 7 June 1991
Revised 30 July 1991

Abstract

Diaconis, P. and R. Graham, An affine walk on the hypercube, Journal of Computational and Applied Mathematics 41 (1992) 215–235.

Let $\mathbb{Z}_2^d$ be the group of binary d-tuples. We study the process $X_n = AX_{n-1} + \epsilon_n$ with $X_i \in \mathbb{Z}_2^d$, A fixed in $\mathrm{GL}_d(\mathbb{Z}_2)$ and ϵ_n a random vector of disturbance terms. This models algorithms in the presence of a "bad bit". For a class of situations we show that the distribution of X_n tends to the uniform distribution on $\mathbb{Z}_2^d$. We determine sharp rates of convergence and demonstrate the existence of cutoff phenomena. The analysis depends on understanding codes made from the binomial coefficients (mod 2). It leads to a novel type of oscillating behavior for the location of the cutoff and for the error terms.

Keywords: Markov chain; uniform distribution; cutoff phenomena; Fourier analysis; code.

1. Introduction

Let $\mathbb{Z}_2^d$ be the group of binary d-tuples under coordinatewise addition. Let A be the $d \times d$ lower triangular matrix with ones on the diagonal and subdiagonal and zero elsewhere. Thus if $d = 4$, A appears as

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}. \tag{1.1}$$

This paper analyzes Markov chains of the form

$$X_n = AX_{n-1} + \epsilon_n, \tag{1.2}$$

Correspondence to: Dr. R.L. Graham, Research, Information Sciences Division, AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974, United States.

with $X_0 = x \in \mathbb{Z}_2^d$, and ϵ_n independent and identically distributed vectors having common distribution

$$P\{\epsilon_n = 0\} = 1 - \theta, \qquad P(\epsilon_n = e_1) = \theta, \tag{1.3}$$

with $0 < \theta < 1$ fixed and e_1 the vector with a one in the first coordinate and zeros elsewhere.

As motivation, take $d = 4$, and consider the result of applying successive powers of A without error (all calculations are mod 2):

$$A\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_0 + x_1 \\ x_1 + x_2 \\ x_2 + x_3 \end{pmatrix}, \qquad A^2\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_1 \\ x_0 + x_2 \\ x_1 + x_3 \end{pmatrix}, \qquad A^3\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_0 \\ x_0 + x_1 \\ x_0 + x_1 + x_2 \\ x_0 + x_1 + x_2 + x_3 \end{pmatrix}.$$

Finally $A^4 = I$. Thus repeatedly applying A gives the partial sums of the 4 coordinates. If $d = 2^r$, A^{d-1} gives all partial sums as above. This is a typical repetitive process. The model (1.2) allows the first bit to err with probability θ. As shown below, this source of randomness eventually corrupts the entire vector and for any starting state and any $y \in \mathbb{Z}_2^d$,

$$\lim_{n \to \infty} P\{X_n = y\} = \frac{1}{2^d}. \tag{1.4}$$

We determine sharp rates of convergence for this limiting behavior. Note that if $\theta = \frac{1}{2}$, then X_n becomes uniform by the time $n = d$. We henceforth assume $\theta \neq \frac{1}{2}$. Let $U(y) = 1/2^d$ denote the uniform distribution and let $Q_n(y) = P\{X_n = y\}$. Let

$$\| Q_n - U \| = \max_{B \subset \mathbb{Z}_2^d} | P\{X_n \in B\} - U(B) | \tag{1.5}$$

denote the total variation distance. Our main result shows that n of order $d \log d$ steps are necessary and suffice to make the total variation distance small. An exact statement is complicated by an oscillating lead term.

Theorem 1.1. *Let d be a positive integer with $2^{r-1} < d < 2^r$ for some integer r. Suppose that the binary expansion of d begins with s consecutive ones, $1 \leqslant s \leqslant r$, (so if $d = 25 = 11001_{(2)}$, $s = s(d) = 2$). If*

$$n = \alpha(d) d (\log d + c), \tag{1.6}$$

with

$$\alpha(d) = \frac{(1 + s/r)}{4 |\log(1 - 2\theta)|} \frac{2^r}{d},$$

then, for the random walk defined by (1.2), (1.3) *there are universal constants a, b such that*

$$\| Q_n - U \| \leqslant a \, \mathrm{e}^{-bc}. \tag{1.7}$$

This result is sharp in the following sense. For n of the form (1.6) *with $c < 0$,*

$$f(c) \leqslant \| Q_n - U \|, \tag{1.8}$$

with $0 < f(c) \uparrow 1$ as $c \to -\infty$.

Theorem 1.1 shows that the variation distance is essentially 1 for n substantially smaller than $\alpha(d) d \log d$ and tends to zero exponentially fast for n substantially larger than this cutoff

value. Similar behavior holds for a variety of other chains; see, e.g., [1,2,4,7]. Explaining these cutoffs is a major unsolved problem in this area.

The lead term $\alpha(d)$ is a bounded oscillating function of d. Suppose d is large. Then $\alpha(d)$ lies between

$$\frac{1}{4|\log(1-2\theta)|} \quad \text{and} \quad \frac{1}{2|\log(1-2\theta)|},$$

as d varies in $2^{r-1} < d < 2^r$. This is the first example we know of this type of oscillating behavior in Markov chain theory.

The upper bound (1.7) is proved in Section 2 by Fourier analysis. The lower bound (1.8) is proved in Section 3. Both arguments depend on a careful analysis of the weight enumerator of the code generated by d vectors W_i, $0 \leqslant i \leqslant d-1$, where $W_i \in \mathbb{Z}_2^{2^r}$ has $W_i(j) = \binom{j}{i} \pmod 2$. This code is treated in the Appendix.

Theorem 1.1 omits powers of 2. In this case, the exact asymptotics of the variation distance can be determined. The following result is proved in Section 4 without Fourier analysis.

Theorem 1.2. *Let $d = 2^r$. For the random walk defined by* (1.2), (1.3) *and*

$$n = \frac{d(\log d + c)}{2|\log(1-2\theta)|},$$

we have

$$\| Q_n - U \| = 1 - 2\Phi\left(-\tfrac{1}{2}\, e^{-c/2} b\left(\frac{n}{d}\right)\right) + O(d^{-1/2}),$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

and $b(n/d)$ is the bounded oscillating function given by

$$b\left(\frac{n}{d}\right) = (1-2\theta)^{-\{n/d\}} \left(1 - 4\left\{\frac{n}{d}\right\}\theta(1-\theta)\right)^{1/2},$$

with $\{x\}$ denoting the fractional part of x.

Theorem 1.2 determines the shape of the cutoff function. The usual bounds on the error function show that the variation distance tends to zero like $e^{-c/2} b(n/d)/\sqrt{2\pi}$ for c large. The variation distance tends to 1 doubly exponentially in $c < 0$. The oscillating form of the error is a novelty to us.

Theorems 1.1 and 1.2 hold as stated for any matrix A having a single Jordan block with eigenvalue 1, for example, for any lower triangular A having ones on the diagonal and subdiagonal and arbitrary entries elsewhere below. Also, Theorems 1.1 and 1.2 hold as stated for ϵ_n taking values 0 and ν with probability $1-\theta$ and θ with ν any fixed vector having a nonzero first component. These and related variations are discussed in Section 5.

The basic chain (1.2) with a general matrix A includes the problem of running a binary recurrence (or pseudo-random number generator)

$$Y_n = a_1 Y_{n-1} + \cdots + a_d Y_{n-d} + \epsilon_n.$$

If the generating function of the coefficients is an irreducible polynomial, a rather different analysis obtains. For example, the order of such a recurrence can be $2^d - 1$. We treat such problems in [6].

More generally, if G is a group and X is a semigroup on which G acts, one may define a process of the form $X_n = a_n X_{n+1} + \epsilon_n$. Chung et al. [3] took $X = G = \mathbb{Z}_p$, the integers mod p. There, a was taken as 2 and ϵ_n took values 0, ± 1 with probability $\frac{1}{3}$. Hildebrand [12] showed that their main findings essentially extend to any fixed a and independent and identically distributed law for ϵ_n. Hildebrand also developed a theory when a is allowed to be random (e.g., $a = 2$ or $\frac{1}{2}$ with probability $\frac{1}{2}$). The results are counter-intuitive and hard to make sense of. Diaconis [5] develops some general theory but much lies in the future.

2. Proof of Theorem 1.1, upper bound

Let A be a $d \times d$ binary matrix of the form (1.1). Here $2^{r-1} < d \leqslant 2^r$. The first order of business is to determine the order of A. The argument develops a representation used crucially in what follows.

Lemma 2.1. *For $2^{r-1} < d \leqslant 2^r$, the order of a $d \times d$ matrix of the form* (1.1) *is 2^r.*

Proof. Let $B = A^t$. By an easy induction, B^k has first row $\binom{k}{0} \cdots \binom{k}{k} \pmod 2$, second row $0\binom{k}{0} \cdots \binom{k}{k} \pmod 2$ and successive rows successive shifts of the first row. Terms which "drop off the end" are omitted. More formally, if rows and columns are numbered starting at 0, then B^k is an upper triangular matrix with (i, j) entry:

$$B^k_{i,j} = \binom{k}{j-i} \pmod 2. \tag{2.1}$$

Now Kummer's lemma (see, e.g., [13, p.68]) says that for integer a, b, the highest power of 2 dividing $\binom{a+b}{a}$ equals the number of "carries" if the integers a and b are added in binary. For example, $31 = 11111_{(2)}$ and $1 = 00001_{(2)}$ have 5 carries and 2^5 is the highest power of 2 dividing $\binom{31+1}{1}$. In particular, $\binom{a+b}{a}$ is odd if and only if there are no carries. Clearly, if $k = 2^r$, $2^r - (j-i)$ and $(j-i)$ have a carry for $j > i$ and no carry when $j = i$. Thus $A^{2^r} = I$. A similar argument shows that 2^r is the exact order. □

Remark 2.2. The group $\mathrm{GL}_d(\mathbb{Z}_2)$ has order $2^{d(d-1)/2}\prod_{i=1}^{d}(2^i - 1)$. The lower triangular matrices form a Sylow subgroup U_d of order $2^{d(d-1)/2}$. Thus a priori the order of A must be a power of 2. We have also proved the curious result that *every* element in U_d has order at most 2^r.

Without loss of generality, the process starts at $X_0 = 0$. Iterating the basic recurrence (1.2):

$$X_0 = 0, \qquad X_1 = AX_0 + \epsilon_1 = \epsilon_1, \qquad X_2 = A\epsilon_1 + \epsilon_2, \ldots,$$
$$X_n = A^{n-1}\epsilon_1 + A^{n-2}\epsilon_2 + \cdots + \epsilon_n. \tag{2.2}$$

Suppose until further notice that

$$n = m2^r, \quad \text{for integer } m. \tag{2.3}$$

Then the terms in (2.2) can be grouped by powers of A:

$$X_n = A^{2^r-1}\tau_1 + A^{2^r-2}\tau_2 + \cdots + \tau_{2^r}, \tag{2.4}$$

where τ_i are independent and identically distributed, each τ having the distribution of a sum of m independent vectors distributed as ϵ_1. This representation will be useful in computing the Fourier transform of the distribution of X_n to which we now turn.

2.1. Fourier analysis on $\mathbb{Z}_2^d$

Let Q be a probability on $\mathbb{Z}_2^d$. For $y \in \mathbb{Z}_2^d$, the Fourier transform of Q at y is defined by

$$\hat{Q}(y) = \sum_{x \in \mathbb{Z}_2^d} (-1)^{y^t x} Q(x). \tag{2.5}$$

In (2.5) (and throughout) x and y are column vectors and the dot product is taken mod 2. If P is a second probability on $\mathbb{Z}_2^d$, define convolution as

$$P * Q(x) = \sum_{z \in \mathbb{Z}_2^d} P(z)Q(x-z).$$

Fourier analysis turns convolution into product since $\widehat{P * Q}(y) = \hat{P}(y)\hat{Q}(y)$. For this and other basic properties of Fourier analysis on finite groups see [4, Chapters 2 and 3] or [14]. The uniform distribution has Fourier transform $\hat{U}(y) = 0$ for $y \neq 0$ with $\hat{U}(0) = 1$. The proof of Theorem 1.1 proceeds by calculating the Fourier transform of the distribution of X_n and showing that it is close to $\hat{U}$ for n suitably large. The connection to variation distance comes through the following upper bound lemma.

Lemma 2.3. *If Q is a probability on $\mathbb{Z}_2^d$, then*

$$\| Q - U \|^2 \leqslant \tfrac{1}{4} \sum_{y \neq 0} \hat{Q}^2(y).$$

Proof. The variation distance defined in (1.5) can be written

$$\| Q - U \| = \tfrac{1}{2} \sum_{x \in \mathbb{Z}_2^d} |Q(x) - U(x)|.$$

From this,

$$4 \| Q - U \|^2 = \Big(\sum_x |Q(x) - U(x)| \Big)^2 \leqslant 2^d \sum_x |Q(x) - U(x)|^2 = \sum_{y \neq 0} |\hat{Q}(y)|^2.$$

The inequality is Cauchy–Schwarz. Then, the Plancherel theorem and $\hat{U}(y) = 0$ for $y \neq 0$, $\hat{Q}(0) = \hat{U}(0) = 1$ were used. In [8] further details are given. □

The next lemma computes the Fourier transform of the walk at time n.

Lemma 2.4. *Let $Q_n(x) = P(X_n = x)$ for X_n defined in (1.2) with $n = m2^r$. Then*

$$\hat{Q}_n(y) = (1-2\theta)^{mM_n(y)}, \quad \text{with } M_n(y) = \sum_{j=0}^{2^r-1} \delta_1 \Big(\sum_{k=0}^{d-1} \binom{j}{k} y_k \Big),$$

where $\delta_1(\omega)$ is 1 or 0 (in $\mathbb{R}$) as $\omega = 1$ or 0 (mod 2).

Proof. Using the representations (2.2), (2.4) in the definition (2.5),

$$\hat{Q}_n(y) = E\left[(-1)^{y^t X_n}\right] = \prod_{j=0}^{n-1} E\left[(-1)^{y^t A^{n-j}\epsilon_j}\right] = \left\{\prod_{j=0}^{2^r-1} E\left[(-1)^{y^t A^j \epsilon_j}\right]\right\}^m. \tag{2.6}$$

Here

$$E\left[(-1)^{y^t A^j \epsilon_j}\right] = \begin{cases} 1-2\theta, & \text{if } (y^t A^j)_0 = 1, \\ 1, & \text{if } (y^t A^j)_0 = 0. \end{cases}$$

Thus the term in braces in (2.6) is $(1-2\theta)^{m(y)}$ with $m(y) = \sum_{j=0}^{2^r-1} \delta_1((y^t A^j)_0)$. Let $B = A^t$ and recall the explicit description of B^j from (2.1). We see

$$(y^t A^j)_0 = (B^j y)_0 = \sum_{k=0}^{d-1} \binom{j}{k} y_k. \tag{2.7}$$

Inside δ_1 the sum is taken mod 2. □

2.2. A binary code

The next step is to interpret $M_n(y)$ in Lemma 2.4 in terms of the weight enumerator of a code. Consider the $d \times 2^r$ array V_d with (i, j) entry $\binom{j}{i}$ (mod 2). This appears (with all entries modulo 2):

	0	1	2	⋯	j	⋯	2^r-1
0	1	1	1	⋯	$\binom{j}{0}$	⋯	1
1	0	1	0	⋯	$\binom{j}{1}$	⋯	$\binom{2^r-1}{1}$
⋮					⋮		⋮
i	0	0	0	⋯	$\binom{j}{i}$	⋯	$\binom{2^r-1}{i}$
⋮					⋮		⋮
$d-1$	0	0	0	⋯	$\binom{j}{d-1}$	⋯	$\binom{2^r-1}{d-1}$

$= V_d.$ (2.8)

The rows can be taken as vectors in $\mathbb{Z}_2^{2^r}$. The vector space spanned by the rows makes up a code which is studied in the Appendix. The array has full rank because the left-hand $d \times d$ block is upper triangular. The sum inside δ_1 in (2.7) is $(y^t V_d)_j$. Thus $M_n(y)$ is the weight (number of nonzero terms) of the code word $y^t V_d$. Let N_j be the number of y such that $y^t V_d$ has weight j. Define the *weight enumerator* $D_d(t)$ as

$$D_d(t) = \sum_{j=0}^{2^r} N_j t^j. \tag{2.9}$$

Using the upper bound of Lemma 2.3, the definitions above give the next result.

Lemma 2.5. *For the random walk defined in* (1.2) *with* $2^{r-1} < d \leqslant 2^r$ *and* $n = m2^r$,

$$4\|Q_n - U\|^2 \leqslant D_d\left((1-2\theta)^{2m}\right) - 1. \tag{2.10}$$

As will emerge, the code generated by the rows of V_d has $N_0 = 1$, so the right-hand side of (2.10) tends to zero when m tends to infinity.

The weight enumerators $D_d(t)$ are studied in the Appendix. They satisfy the following recursive relations which determine them completely.

Theorem A.2. *The polynomials defined in* (2.9) *satisfy*

(1) $D_1(t) = 1 + t$,

(2) $D_{2m}(t) = D_m^2(t)$, $1 \leqslant m < \infty$,

(3) *For* $m \neq 2^s$,

$$D_{2m+1}(t) - D_{2m}(t) = \left(D_{m+1}(t) - D_m(t)\right)^2,$$

(4) $D_{2^s+1}(t) = (1+t^2)^{2^s} + (2t)^{2^s}$.

For example, $D_1(t) = (1+t)$, $D_2(t) = (1+t)^2$ and $D_3(t) = 1 + 6t^2 + t^4$. As a check, when $d = 3$, the matrix V_3 of (2.8) is

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

The linear combinations of the rows generate the code words

$$\{0000, 1111, 0101, 1010, 0011, 1100, 0110, 1001\}.$$

This has one word of weight zero, one word of weight four and six words of weight two.

In the remainder of this section, we assume $2^{r-1} < d < 2^r$ and the binary expansion of d begins with s ones. Theorem A.2 yields the following.

Lemma 2.6. For $2^{r-1} < d < 2^r - 1$, *the weight enumerator of* (2.9) *satisfies*

$$D_d(t) \leqslant \left[\tfrac{1}{2}\left\{(1+t)^{2^{s+1}} + (1-t)^{2^{s+1}}\right\}\right]^{2^{r-s-1}}, \quad \textit{for } 0 \leqslant t \leqslant 1,$$

while

$$D_{2^r-1}(t) = \tfrac{1}{2}\left\{(1+t)^{2^r} + (1-t)^{2^r}\right\}.$$

Proof. For $2^{r-1} < d \leqslant d' < 2^r$, the code generated by the rows of the $d \times 2^r$ matrix V_d of (2.8) is contained in the code generated by the $d' \times 2^r$ matrix $V_{d'}$. It follows that $D_d(t) \leqslant D_{d'}(t)$ for $0 \leqslant t \leqslant 1$. Take d' with binary expansion starting with $s+1$ ones and the other bits 0. Thus if $d = 9 = 1001_{(2)}$, $d' = 12 = 1100_{(2)}$. Here $d' = 2^{r-s-1}(2^{s+1} - 1)$. The matrix $V_{2^{s+1}-1}$ generates the code of all even weight words of length 2^{s+1}. To see this, observe that the rows of $V_{2^{s+1}-1}$ all have even weight. The left-hand $2^{s+1} - 1$ by $2^{s+1} - 1$ block of $V_{2^{s+1}-1}$ is upper triangular with ones on the diagonal so the rows of $V_{2^{s+1}-1}$ are linearly independent and so generate all even codewords of length 2^{s+1}. Thus we have

$$D_{2^{s+1}-1}(t) = \tfrac{1}{2}\left\{(1+t)^{2^{s+1}} + (1-t)^{2^{s+1}}\right\}. \tag{2.11}$$

The result follows from this and (2) in Theorem A.2. □

Proof of the upper bound (1.7). Consider $D_d(t)$ of Lemma 2.6 for $2^{r-1} < d < 2^r$ and $t = (1-2\theta)^{2m}$ with

$$m = \frac{(1+s/r)(\log d + c)}{4|\log(1-2\theta)|},$$

so that $t = f(d, c)/d^{(1+s/r)/2}$ with

$$f(d, c) = e^{-(1+s/r)c/2}. \tag{2.12}$$

Note that $1 < (1+s/r) < 2$ for $2^{r-1} < d < 2^r$.

The argument will show

$$D_d(t) \leqslant e^{8f^2 + O(f^4/2^{r-s})}. \tag{2.13}$$

Observe first that $2^{s+1}/d^{(r+s)/2r} \leqslant 4/2^{(r-s)/2}$. Thus

$$(1+t)^{2^{s+1}} = 1 + 2^{s+1}t + \binom{2^{s+1}}{2}t^2 + \binom{2^{s+1}}{3}t^3 + O\left(\sum_{j=4}^{\infty} \frac{(2^{s+1}t)^j}{j!}\right)$$

$$= 1 + 2^{s+1}t + \binom{2^{s+1}}{2}t^2 + \binom{2^{s+1}}{3}t^3 + O\left(\left(\frac{f(d, c)}{2^{(r-s)/2}}\right)^4\right),$$

where the implied constant is uniform for $|f(d, c)/2^{(r-s)/2}| \leqslant \frac{1}{2}$. Using a similar bound for $(1-t)^{2^{s+1}}$ gives

$$\tfrac{1}{2}\{(1+t)^{2^{s+1}} + (1-t)^{2^{s+1}}\} = 1 + \binom{2^{s+1}}{2}t^2 + O\left(\left(\frac{f(d, c)}{2^{(r-s)/2}}\right)^4\right).$$

Now, using $\log(1+x) = x + O(x^2)$,

$$D_d(t) = \exp\left\{\frac{2^r}{2^{s+1}}\frac{\binom{2^{s+1}}{2}}{d^{(r+s)/r}}f^2 + O\left(\frac{2^r}{2^{s+1}}f^4\left(\frac{2^{4r}}{d^{2(r+s)/r}} + \frac{1}{2^{2(r-s)}}\right)\right)\right\}.$$

The term in front of f^2 is bounded above by 8 while the error term is $O(f^4/2^{r-s})$.

The argument of this section has been developed under the restriction that $n = m2^r$ for integer m (see (2.3)). For general n, the total variation distance is monotone in n. Let m^* be such that $m^*2^r \leqslant n < (m^*+1)2^r$. Changing m by 1 amounts to changing c by a fixed amount (depending on θ which we have as fixed). This only changes ae^{bc} to $a'e^{bc}$.

The stated bound (1.7) follows easily from these considerations and (2.13). □

Remark 2.7. (i) As shown in Section 4, the correct asymptotic approximation for the total variation when $d = 2^r$ has an oscillatory quality. We presume that this is also the case for other values of d.

(ii) For fixed finite d it is a straightforward matter to evaluate the explicit bound in Lemma 2.5 or 2.6. These give satisfactory results for any "real" problem. The asymptotics give a useful rule of thumb for where the cutoff occurs.

3. Proof of Theorem 1.1, lower bound

Let Q_n be the probability on $\mathbb{Z}_2^d$ corresponding to the distribution of the Markov chain defined in (1.2). This section proves the lower bound (1.8) on the distance $\| Q_n - U \|$. From the definition (1.5), any set $B \subseteq \mathbb{Z}_2^d$ gives a lower bound by $|Q_n(B) - U(B)| \leq \| Q_n - U \|$. A suitable set B will be chosen by looking at the slow term in the Fourier analytic upper bound. Throughout this section $2^{r-1} < d < 2^r$ and $n = m2^r = \alpha(d)d(\log d + c)$ with $\alpha(d) = \{(1 + s/r)2^r/d\}/(4|\log(1-2\theta)|)$, so that

$$m = \left(1 + \frac{s}{r}\right) \frac{\log d + c}{4|\log(1-2\theta)|}. \tag{3.1}$$

Let $S_2 = \{y \in \mathbb{Z}_2^d: |y^t V_d| = 2\}$ with V_d defined in (2.8) and $|z|$ the number of ones (or weight) of the binary 2^r-tuple z. The cardinality $|S_2| = N_2$ is thus the number of codewords of weight 2. Define $g: \mathbb{Z}_2^d \to \mathbb{R}$ by

$$g(x) = \frac{1}{\sqrt{N_2}} \sum_{y \in S_2} (-1)^{y^t x}. \tag{3.2}$$

This g will be used as a test function and B will be chosen as the set of x where $|g(x)|$ is large. Of course $(-1)^{y^t x}$ is the yth character of $\mathbb{Z}_2^d$.

Lemma 3.1. *Let X be uniformly distributed on $\mathbb{Z}_2^d$. For g defined in* (3.2),

$$E(g(X)) = 0, \qquad \mathrm{Var}(g(X)) = 1. \tag{3.3}$$

Proof.

$$E(g(X)) = \frac{1}{2^d} \sum_x g(x) = \frac{1}{\sqrt{N_2}} \sum_{y \in S_2} \hat{U}(y) = 0,$$

$$E(g^2(X)) = \frac{1}{N_2} \sum_{y, y' \in S_2} \hat{U}(y + y') = 1. \qquad \square$$

The next lemma is the heart of this section. It uses a careful description of the set of words of weight 2 in the code of Section 2. This description is developed in the Appendix.

Lemma 3.2. *Let X_n be defined by* (1.2). *For g defined by* (3.2) *with $n = m2^r$,*

$$E(g(X_n)) = \sqrt{N_2}(1-2\theta)^{2m}, \tag{3.4}$$

$$E(g^2(X_n)) = \frac{1}{N_2}\Big\{N_2 + (2^s - 1)(2^s - 2)2^r(1-2\theta)^{2m} + \left(N_2^2 - N_2 - (2^s - 1)(2^s - 2)2^r\right)(1-2\theta)^{4m}\Big\}, \tag{3.5}$$

where $N_2 = \binom{2^s}{2} 2^{r-s}$ is the number of codewords of weight 2, $2^{r-1} < d < 2^r$, and d begins with s ones.

Proof. If Q_n is the distribution of X_n,

$$E(g(X_n)) = \frac{1}{\sqrt{N_2}} \sum_{y \in S_2} \hat{Q}_n(y) = \sqrt{N_2}(1-2\theta)^{2m}.$$

The last equality uses the computation of the Fourier transform of Lemma 2.4 with $M_n(y) = 2$ for $y \in S_2$.

To compute the second moment, square g and take expectations. Arguing as above,

$$E(g^2(X_n)) = \frac{1}{N_2}\left\{N_2 + A(1-2\theta)^{2m} + (N_2^2 - A - N_2)(1-2\theta)^{4m}\right\},$$

where $A = |\{x, y \in S_2 : x+y \in S_2\}|$. Indeed, for $x, y \in S_2$, $|(x+y)V_d|$ can only take values 0, 2 or 4. The value 0 occurs if and only if $x = y$ (this happens N_2 times). The value 2 occurs A times and the value 4 occurs for remaining pairs $N_2^2 - A - N_2$ times.

In Theorem A.5 the following explicit description of the codewords in $\mathscr{C}_2 = \{yV_d : y \in S_2\}$ is proved. Of course, $\mathscr{C}_2 \subset \mathbb{Z}_2^{2^r}$. Consider the 2^r coordinates in 2^s consecutive blocks each of length 2^{r-s}. A vector in $\mathscr{C}_2$ can be specified by choosing a pair of blocks ($\binom{2^s}{2}$ choices) and an integer $1 \leqslant k \leqslant 2^{r-s}$. Put a one in the kth coordinate of each of the two chosen blocks and zeros elsewhere. Each such word has weight 2, and each occurs for a unique $y \in \mathbb{Z}_2^d$. This shows $N_2 = \binom{2^s}{2}2^{r-s}$ as claimed. Further, the description makes it easy to count pairs such that the sum has weight 2. For each of N_2 values x, y can be chosen to kill one coordinate of x and have the second coordinate of the sum $x+y$ in a different block in $2(2^s - 2)$ ways. Thus $A = 2N_2(2^s - 2) = 2^r(2^s - 1)(2^s - 2)$. □

From Lemma 3.1 and Chebyshev's inequality, under the uniform distribution $g(x)$ is close to zero with high probability. Under Q_n, g has large mean and, as will emerge, small variance.

Proof of the lower bound (1.8). From Lemma 3.2 and the definitions,

$$E(g(X_n)) = a(d)\, e^{-(1+s/r)c/2}, \tag{3.6}$$

$$\mathrm{Var}(g(X_n)) = 1 + \frac{b(d)\, e^{-(1+s/r)}}{2^r} + \frac{c(d)\, e^{-(1+s/r)c/2}}{2^{(r-s)/2}}, \tag{3.7}$$

where $a(d)$, $b(d)$, $c(d)$ are bounded in absolute value by 3 uniformly in d and c.

Let

$$B = \{x \in \mathbb{Z}_2^d : |g(x)| \geqslant e^{-(1+s/r)c/4}\}.$$

From Lemma 3.1, $U(B) \to 1$ for c large. From (3.6), (3.7), under Q_n, $g(x)$ has mean $a(d)\, e^{-(1+s/r)c/2} \gg e^{-(1+s/r)c/4}$ and standard deviation of order, at worst, the square root of the mean. □

4. Asymptotics when $d = 2^r$: Proof of Theorem 1.2.

This section presents a direct (non-Fourier) proof of Theorem 1.2 when $d = 2^r$. The argument mimics the Fourier proof in an interesting way. The main technical difference is that $n = m2^r$ is *not* assumed. This complicates things a bit and leads to oscillations in the error

term. The first stage of the analysis uses symmetry to reduce the problem to calculating the difference between binomial random variables.

Consider the process X_n of (1.2) with A of the form (1.1). Here

$$n = \frac{d(\log d + c)}{2|\log(1-2\theta)|} = md + \mu d, \tag{4.1}$$

with

$$m = \left[\frac{\log d + c}{2|\log(1-2\theta)|}\right] \quad \text{and} \quad \mu = \left\{\frac{\log d + c}{2|\log(1-2\theta)|}\right\}, \tag{4.2}$$

with the brackets denoting greatest integer and the braces denoting fractional part, respectively, so $0 \leqslant \mu < 1$. Let

$$p = \mu d, \quad \text{so } n = md + p. \tag{4.3}$$

As in (2.2), (2.4) the walk can be written

$$X_n = A^{d-1}\tau_1 + A^{d-2}\tau_2 + \cdots + \tau_d, \tag{4.4}$$

where $\tau_1, \tau_2, \ldots, \tau_d$ are independent random vectors satisfying:

$\tau_1, \ldots, \tau_{d-p}$ are identically distributed with common law P^{*m},

$\tau_{d-p+1}, \ldots, \tau_d$ are identically distributed with common law $P^{*(m+1)}$,

with P the common law of ϵ_i given in (1.3). It follows that

$$P\{\tau_i = 0\} = \tfrac{1}{2}\left(1 + (1-2\theta)^m\right), \qquad P\{\tau_i = e_1\} = \tfrac{1}{2}\left(1 - (1-2\theta)^m\right),$$
$$1 \leqslant i \leqslant d-p, \tag{4.5a}$$

$$P\{\tau_i = 0\} = \tfrac{1}{2}\left(1 + (1-2\theta)^{m+1}\right), \quad P\{\tau_i = e_1\} = \tfrac{1}{2}\left(1 - (1-2\theta)^{m+1}\right),$$
$$d-p+1 \leqslant i \leqslant d. \tag{4.5b}$$

In (4.4) the term $A^{d-i}\tau_i$ is 0 for the first column of A^{d-i} with probabilities given by (4.5). Thus, X_n has the same distribution as

$$V^{d-1}\beta_1 + \cdots + V^0\beta_d = V_\beta, \tag{4.6}$$

where V is a $d \times d$ matrix having columns $V^{d-1}, V^{d-2}, \ldots, V^1, V^0$ with V^{d-i} being the first column of A^{d-i}. From (2.1), $V_j^{d-i} \equiv \binom{d-i}{j} \pmod 2$, $0 \leqslant j \leqslant d-1$, $1 \leqslant i \leqslant d$. The column vector β has independent random entries β_i with β_i taking values 0 or 1 with the probabilities given by the right-hand sides of (4.5a,b).

The matrix V is easily seen to be invertible (in fact $V^2 = I$). Now the variation distance to uniform is invariant under 1–1 transformations, so $\| Q_n - U \| = \| \mathscr{L}(\beta) - U \|$ with $\mathscr{L}(\beta)$ denoting the law of β on $\mathbb{Z}_2^d$. The measures $\mathscr{L}(\beta)$ and U are invariant under permutations of the first $d-p$ coordinates and permutations of the last p coordinates. It follows that $\| \mathscr{L}(\beta) - U \|$ is equal to the difference between the laws of the number of ones in these coordinates. In [9, Lemma 6.1] details are given. To write this out, note from (4.1),

$$m = \frac{\log d + c}{2|\log(1-2\theta)|} - \mu.$$

So

$$(1-2\theta)^{m+1} = \frac{e^{-c/2}}{\sqrt{d}}(1-2\theta)^{1-\mu} \stackrel{d}{=} \frac{\alpha}{\sqrt{d}},$$
$$(1-2\theta)^{m} = \frac{e^{-c/2}}{\sqrt{d}}(1-2\theta)^{-\mu} \stackrel{d}{=} \frac{\beta}{\sqrt{d}}. \tag{4.7}$$

Let X_α and Y_β be independent random variables with

$$X_\alpha \sim \text{binomial}\left(p, \tfrac{1}{2}\left(1-\frac{\alpha}{\sqrt{d}}\right)\right), \qquad Y_\beta \sim \text{binomial}\left(d-p, \tfrac{1}{2}\left(1-\frac{\beta}{\sqrt{d}}\right)\right),$$

and let X_0 and Y_0 be independent random variables with α and β both zero. The considerations above can be summarized as follows.

Lemma 4.1. *For Q_n, the law of X_n defined in* (1.2), *and α, β, X_α, X_β as in* (4.6), (4.7), *we have*

$$\|Q_n - U\| = \|\mathcal{L}(X_\alpha \times Y_\beta) - \mathcal{L}(X_0 \times Y_0)\|$$
$$= \sum_{\theta(i,j) \geq 1/2^{2d}} \binom{p}{i}\binom{d-p}{j}\left(\theta(i,j) - \frac{1}{2^{2d}}\right), \tag{4.8}$$

with

$$\theta(i,j) = \frac{1}{2^{2d}}\left(1-\frac{\alpha}{\sqrt{d}}\right)^{i}\left(1+\frac{\alpha}{\sqrt{d}}\right)^{p-i}\left(1-\frac{\beta}{\sqrt{d}}\right)^{j}\left(1+\frac{\beta}{\sqrt{d}}\right)^{d-p-j}. \tag{4.9}$$

Proof. The first equality was argued in the comments preceding the lemma. For the second equality, if P and Q are two measures on a finite set Y, the variation distance $\|P-Q\|$ defined as in (1.5) can also be written $\|P-Q\| = \Sigma(P(y)-Q(y))$ where the sum is over y such that $P(y) \geq Q(y)$. □

The variation distance can now be approximated by using the normal approximation to the binomial distribution. In the asymptotics in the remainder of this section, α and β are in a fixed compact subset of $\mathbb{R}^2$ and d tends to infinity. Since both the sample size and parameters depend on d, a uniform version of the central limit theorem such as the Berry–Esseen theorem [10, Chapter 16] is needed. The first step identifies the set where $\theta(i,j) \geq 1/2^{2d}$.

Lemma 4.2. *With notation as in Lemma* 2.3, *let*

$$\mathcal{B}(d) = \left\{(i,j): 2\alpha i + 2\beta j \leq d\left[\mu\alpha + \bar{\mu}\beta - \frac{\mu\alpha^2 + \bar{\mu}\beta^2}{2\sqrt{d}}\right]\right\},$$

with $0 \leq i, j \leq d$ and $\bar{\mu} = 1-\mu$. Then

$$\|Q_n - U\| = \mathcal{L}(X_\alpha \times Y_\beta)(\mathcal{B}(d)) - \mathcal{L}(X_0 \times Y_0)(\mathcal{B}(d)) + o(1).$$

Proof. The variation distance $\|\mathscr{L}(X_\alpha \times X_\beta) - \mathscr{L}(X_0 \times Y_0)\|$ is achieved at $\{(i, j): \theta(i, j) \geqslant 1/2^{2d}\}$. All such (i, j) satisfy

$$\left(\frac{1-\alpha/\sqrt{d}}{1+\alpha/\sqrt{d}}\right)^i \left(\frac{1-\beta/\sqrt{d}}{1+\beta/\sqrt{d}}\right)^j \geqslant \left(1+\frac{\alpha}{\sqrt{d}}\right)^{-p} \left(1+\frac{\beta}{\sqrt{d}}\right)^{-(d-p)},$$

with

$$p = \mu d, \qquad d-p = \bar{\mu} d, \qquad \mu = \bar{\mu} = 1.$$

Taking logarithms, this is

$$i \log\left(\frac{1-\alpha/\sqrt{d}}{1+\alpha/\sqrt{d}}\right) + j \log\left(\frac{1-\beta/\sqrt{d}}{1+\beta/\sqrt{d}}\right) \geqslant -d\mu \log(1+\alpha/\sqrt{d}) - d\bar{\mu} \log(1+\beta/\sqrt{d}). \tag{4.10}$$

Approximate the left-hand side of (4.10), using $\log(1-\epsilon) - \log(1+\epsilon) = -2\epsilon + O(\epsilon^3)$ and $i, j = O(d)$:

$$-\frac{2\alpha}{\sqrt{d}} i - \frac{2\beta}{\sqrt{d}} j + O\left(\frac{1}{\sqrt{d}}\right).$$

For the right-hand side of (4.10), using $\log(1+\epsilon) = \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3)$, we obtain

$$\frac{\mu\alpha + \bar{\mu}\beta}{\sqrt{d}} - \frac{\mu\alpha^2 + \bar{\mu}\beta^2}{2d} + O\left(\frac{1}{d^{3/2}}\right).$$

This shows that (4.10) is equivalent to the set of (i, j) satisfying

$$2\alpha i + 2\beta j \leqslant d\left[\mu\alpha + \bar{\mu}\beta - \frac{\mu\alpha^2 + \bar{\mu}\beta^2}{2\sqrt{d}}\right] + O(1).$$

Under either pair of binomial distributions (X_α, Y_α) or (X_0, Y_0), the $O(1)$ terms are negligible. □

Proof of Theorem 1.2. Using the Berry–Esseen theorem, under $\mathscr{L}(X_\alpha, X_\beta)$ the coordinates i and j have independent approximately normal distributions with means

$$\mu_\alpha = \tfrac{1}{2}\mu d\left(1 - \frac{\alpha}{\sqrt{d}}\right), \qquad \mu_\beta = \tfrac{1}{2}\bar{\mu} d\left(1 - \frac{\beta}{\sqrt{d}}\right),$$

and variances

$$\sigma_\alpha^2 = \tfrac{1}{4}\mu d\left(1 - \frac{\alpha^2}{d}\right), \qquad \sigma_\beta^2 = \tfrac{1}{4}\bar{\mu} d\left(1 - \frac{\beta^2}{d}\right).$$

Normalizing, the $\mathscr{L}(X_\alpha, X_\beta)$ measure of the region $\mathscr{B}(d)$ is

$$P\left\{\frac{\alpha(X_\alpha - \mu_\alpha)}{\sqrt{d}} + \frac{\beta(X_\beta - \mu_\beta)}{\sqrt{d}} \leqslant \tfrac{1}{4}(\alpha^2\mu + \beta^2\bar{\mu})\right\} = \Phi\left(\sqrt{\tfrac{1}{4}(\mu\alpha^2 + \bar{\mu}\beta^2)}\right) + o(1),$$

with $\Phi(x)$ the standard normal distribution function. Under $\mathscr{L}(X_0, Y_0)$ the measure of the region $\mathscr{B}(d)$ is

$$\Phi\Big(-\sqrt{\tfrac{1}{4}(\mu\alpha^2+\bar{\mu}\beta^2)}\Big)+o(1).$$

From these approximations,

$$\|Q_n-U\|=1-2\Phi\Big(-\sqrt{\tfrac{1}{4}(\mu\alpha^2+\bar{\mu}\beta^2)}\Big)+o(1).$$

Finally,

$$\tfrac{1}{4}(\mu\alpha^2+\bar{\mu}\beta^2)=\tfrac{1}{4}\,e^{-c}(1-2\theta)^{-2\mu}(1-4\mu\theta(1-\theta)). \qquad \square$$

The direct probability arguments introduced above can be adapted to cases when $d \neq 2^r$. For example, when $d=3\cdot 2^r$ and $n=m2^r$, a linear transformation M can be found so that $MX_n=\delta+\delta'$ where δ and δ' are $d\times 1$ random vectors having the following laws:

δ has independent identically distributed binary coordinates
with common distribution as in (4.5),

$$(\delta')^t=(\delta'_1,\ldots,\delta'_{d/3},\delta'_1,\ldots,\delta'_{d/3},\delta'_1,\ldots,\delta'_{d/3}),$$

with δ'_i independent of each other and δ with common distribution as in (4.5).

Now, the law of MX_n can be analyzed by considering the law of the array (with all entries mod 2):

$$\begin{array}{ccc}
\delta'_1+\delta_1 & \delta'_1+\delta_{1+d/3} & \delta'_1+\delta_{1+2d/3}\\
\delta'_2+\delta_2 & \delta'_2+\delta_{2+d/3} & \delta'_2+\delta_{2+2d/3}\\
\vdots & \vdots & \vdots\\
\delta'_{d/3}+\delta_{d/3} & \delta'_{d/3}+\delta_{2d/3} & \delta'_{d/3}+\delta_d
\end{array}$$

and bounding its variation distance to the $\frac{1}{3}d\times 3$ array with independent fair coin tossing coordinates. Symmetry considerations show this can be reduced to the distance between the induced laws of $N(x)$, the number of rows in the array having pattern $x\in\mathbb{Z}_2^3$. Under both measures these rows of the array are independent, so the central limit theory can be used as above. The calculation is tedious and we give no further details since the problem has been treated by Fourier analysis in Section 2.

5. Extension to other multipliers and perturbation laws

Theorems 1.1 and 1.2 were proved for multipliers of the form (1.1). The present section shows that they imply the same results for a wider class of multipliers and perturbation laws. Thus consider a process of the form

$$X_n=CX_{n-1}+\epsilon_n, \tag{5.1}$$

with $C\in \mathrm{GL}_d(\mathbb{Z}_2)$ and ϵ_n independent and identically distributed random vectors taking values in $\mathbb{Z}_2^d$. The natural generalization of the matrix of (1.1) allows C with minimal polynomial

$(1-x)^d$. Thus C has all its characteristic roots in $\mathbb{Z}_2$ and its Jordan canonical form has one Jordan block of the form (1.1) so that C is conjugate to A in $\mathrm{GL}_d(\mathbb{Z}_2)$, say with

$$C = D^{-1}AD. \tag{5.2}$$

Of course, any such C has order 2^r where $2^{r-1} < d \leqslant 2^r$ as before. Write the random walk as

$$X_n = C^n\epsilon_1 + C^{n-1}\epsilon_2 + \cdots + \epsilon_n = D^{-1}A^{n-1}D\epsilon_1 + D^{-1}A^{n-2}D\epsilon_2 + \cdots + \epsilon_n.$$

If $Y_n = DX_n$, it follows that

$$Y_n = A^{n-1}\epsilon_1' + A^{n-2}\epsilon_2' + \cdots + \epsilon_n',$$

where $\epsilon_i' = D\epsilon_i$. One-to-one transformations do not change variation distance giving a first reduction.

For the second reduction, observe that a matrix C conjugate to A has a unique nonzero left fixed vector $V_C^t = V_C^t C$. Suppose that

$$P(\epsilon_i = 0) = 1-\theta, \qquad P(\epsilon_i = x) = \theta, \quad \text{with } x^t V_C \neq 0 \pmod 2. \tag{5.3}$$

In case $C = A$, $V_a = e_1$ and the condition becomes $x_0 \neq 0$. With these definitions, the following result can be stated.

Theorem 5.1. *Let X_n be defined by* (5.1)–(5.3). *Then the conclusions of Theorems* 1.1 *and* 1.2 *hold with Q_n replaced by the law of X_n.*

Proof. Using the first reduction, it is enough to bound the rate of convergence of Y_n based on ϵ' taking values 0 and Dx with probability $1-\theta$ and θ, respectively. From $C = D^{-1}AD$, $V_C = D^t e_1$ and the condition in (5.3) becomes $(Dx)^t e_1 \neq 0 \pmod 2$. Let $N(A) = \{F \in \mathrm{GL}_d(\mathbb{Z}_2)\colon FA = AF\}$. By an elementary computation, $N(A)$ consists of lower triangular matrices with constant diagonal entries. Thus for $d = 4$, for example, F consists of matrices of the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ x & 1 & 0 & 0 \\ y & x & 1 & 0 \\ z & y & x & 1 \end{pmatrix},$$

with x, y, z arbitrary in $\mathbb{Z}_2$. It is clear that if $y \in \mathbb{Z}_2^d$ satisfies $y^t e_1 \neq 0$, there is a unique $F \in N(A)$ such that $Fy = e_1$. Since

$$Y_n = A^{n-1}\epsilon_1' + \cdots + \epsilon_n',$$

$\mathbb{Z}_n = FY_n$ is distributed as Q_n as analyzed in Theorems 1.1 and 1.2. □

Remark 5.2. (i) The group $N_d(A)$ which was used in the proof of Theorem 5.1 is an Abelian 2-group. It has

$$\left[\frac{d}{2^{t-1}}\right] - 2\left[\frac{d}{2^t}\right] + \left[\frac{d}{2^{t+1}}\right]$$

factors isomorphic to $\mathbb{Z}_{2^t}$. For example, $N_2(A) \cong \mathbb{Z}_2$, $N_3(A) \cong \mathbb{Z}_4$, $N_4(A) \cong \mathbb{Z}_2 \times \mathbb{Z}_4$. If the matrix C has a richer canonical form, more randomness is required in the law of ϵ to yield a

uniform distribution. As an example, suppose C is lower triangular with ones on the diagonal, a single zero on the next diagonal, and zeros elsewhere. Then ϵ must take nonzero values in both the first coordinate and the coordinate matching the second diagonal zero. This is not enough. For example, if ϵ is zero or x with probability $1-\theta$, where x is nonzero in the two appropriate coordinates and zero elsewhere, then the coordinates of X_n stay dependent and never converge to the uniform distribution.

One natural choice for the law of ϵ such that X_n of (5.1) converges to the uniform distribution for every choice of C has each coordinate 1 or 0 independently with probability θ and $1-\theta$. When $C=I$, the process becomes a version of random walk on the hypercube. This has an extensive literature reviewed in [4]. Preliminary computations indicate that A of the form (1.1) and this choice of ϵ speeds things up to order at least d^a with $a \leqslant \log 3/\log 2 \doteq 0.6309$. We hope to do a more complete analysis.

(ii) Results of [15] show that most matrices C in $\mathrm{GL}_d(\mathbb{Z}_2)$ do not have their roots in the field $\mathbb{Z}_2$. Further, most C have order log d blocks in their rational canonical form. See [11] and [15, Propositions 10 and 11]. The analysis of walks with such C has a different flavor which is developed in [6].

Appendix: The code of binomial coefficients (mod 2)

Let $2^{r-1} < d \leqslant 2^r$ be fixed. Consider the d vectors

$$W_i = (W_i(0), W_i(1), \ldots, W_i(2^r-1)), \quad 0 \leqslant i < d, \text{ where } W_i(j) \equiv \binom{j}{i} \pmod 2. \tag{A.1}$$

Let $\mathscr{C}_d$ denote the subspace of $\mathbb{Z}_2^{2^r}$ generated by the W_i. This $\mathscr{C}_d$ is also called a *code* and elements of $\mathscr{C}_d$ are called *codewords*. For $W \in \mathscr{C}_d$, let $|W|$, *the weight of* W, denote the number of nonzero entries. Let N_k be the number of words of weight k in $\mathscr{C}_d$. Let

$$D_d(t) = \sum_{k=0}^{d} N_k t^k \tag{A.2}$$

denote the *weight enumerator* of $\mathscr{C}_d$.

Example A.1. If $d=3$ and $r=2$, then $W_0=(1111)$, $W_1=(0101)$, $W_2=(0011)$.

$$\mathscr{C}_3 = \{0000, 1111, 0101, 1010, 0011, 1100, 0110, 1001\}, \qquad D_3(t) = 1+6t^2+t^4.$$

This section studies the code $\mathscr{C}_d$ and N_2, the number of words of weight 2. The first result is proved below as a series of lemmas.

Theorem A.2. *The weight enumerators for the binomial coefficient codes defined in* (A.2) *satisfy*

(1) $D_1(t) = 1+t$,
(2) $D_{2m}(t) = D_m^2(t)$, $1 \leqslant m < \infty$,
(3) $D_{2m+1}(t) - D_{2m}(t) = (D_{m+1}(t) - D_m(t))^2$, $m \neq 2^s$,
(4) $D_{2^s+1}(t) = (1+t^2)^{2^s} + (2t)^{2^s}$.

Remark A.3. $D_1(t) = 1+t$ by inspection, so $D_{2^s}(t) = (1+t)^{2^s}$ implying that C_{2^s} consists of all binary vectors of length 2^s. From (1), (2) and (4), $D_3(t) = (1+t^2)^2 + (2t)^2$ in agreement with

the example above. The relations (1)–(4) determine D_d for all d. The form of D_d is particularly simple if $d = j2^s$ with j fixed, e.g., $D_{3\cdot 2^{r-1}}(t) = (1 + 6t^2 + t^4)^r$. In Section 2, a direct proof of $D_{2^s-1}(t) = \frac{1}{2}\{(1+t)^{2^s} + (1-t)^{2^s}\}$ was given. It is an amusing exercise to derive this from Theorem A.2.

The following preliminary lemma will be useful.

Lemma A.4. *For nonnegative integers i and j,*

$$\binom{2j}{2i} \equiv \binom{j}{i} \pmod 2, \qquad \binom{2j+1}{2i} \equiv \binom{j}{i} \pmod 2,$$

$$\binom{2j}{2i+1} \equiv 0 \pmod 2, \qquad \binom{2j+1}{2i+1} \equiv \binom{j}{i} \pmod 2.$$

Proof. These all follow by checking how the number of base 2 "carries" in $b + (a - b)$ occur as in [13, p.68]. For example, since $(2i + 1) + (2j - 2i - 1) = 2j$ always has at least one carry, $\binom{2j}{2i+1} \equiv 0 \pmod 2$. The other arguments are similar. □

The proofs of (2), (3) and (4) are recursive. To set up the basic recursion, let V be the $d \times 2^r$ array formed by the W_i. Let V_0 and V_1 be $d \times 2^{r-1}$ arrays formed from the even and odd columns of V, respectively. From the definition of V and Lemma A.4, V_0 and V_1 appear as (with entries mod 2):

$$V_0 = \begin{matrix}
\binom{0}{0} & \binom{1}{0} & \cdots & \binom{j}{0} & \cdots & \binom{2^{r-1}-1}{0} \\
0 & 0 & \cdots & 0 & \cdots & 0 \\
 & & \vdots & & & \\
\binom{0}{i} & \binom{1}{i} & \cdots & \binom{j}{i} & \cdots & \binom{2^{r-1}-1}{i} \\
0 & 0 & \cdots & 0 & \cdots & 0 \\
 & & \vdots & & &
\end{matrix}$$

$$V_1 = \begin{matrix}
\binom{0}{0} & \binom{1}{0} & \cdots & \binom{j}{0} & \cdots & \binom{2^{r-1}-1}{0} \\
\binom{0}{0} & \binom{1}{0} & \cdots & \binom{j}{0} & \cdots & \binom{2^{r-1}-1}{0} \\
 & & \vdots & & & \\
\binom{0}{i} & \binom{1}{i} & \cdots & \binom{j}{i} & \cdots & \binom{2^{r-1}-1}{i} \\
\binom{0}{i} & \binom{1}{i} & \cdots & \binom{j}{i} & \cdots & \binom{2^{r-1}-1}{i} \\
 & & \vdots & & &
\end{matrix}$$

The form of the last rows of V_0 and V_1 will depend on the parity of d and will determine the difference between parts (2), (3) and (4) of Theorem A.2.

Proof of Theorem A.2(2). Here $d = 2m$. The last row of V is W_{2m-1} and the last rows of V_0 and V_1 are (mod 2):

$$\text{for } V_0: \quad \binom{0}{m-1} \quad \binom{1}{m-1} \quad \cdots \quad \binom{j}{m-1} \quad \cdots \quad \binom{2^{r-1}-1}{m-1}$$
$$0 \qquad 0 \qquad \cdots \qquad 0 \qquad \cdots \qquad 0$$

$$\text{for } V_1: \quad \binom{0}{m-1} \quad \binom{1}{m-1} \quad \cdots \quad \binom{j}{m-1} \quad \cdots \quad \binom{2^{r-1}-1}{m-1}$$
$$\binom{0}{m-1} \quad \binom{1}{m-1} \quad \cdots \quad \binom{j}{m-1} \quad \cdots \quad \binom{2^{r-1}-1}{m-1}$$

The code $\mathscr{C}_{2m}$ is formed by adding all possible subsets of rows of V. For the rows W_{2i} and W_{2i+1} there are four possibilities: include neither, W_{2i} alone, W_{2i+1} alone, or both. Consider the effect of these choices on the corresponding pairs of rows in V_0 and V_1. In the first case (neither) $(0 \cdots 0)$ is added to both V_0 and V_1. In the second case (W_{2i} alone)

$$\left(\binom{0}{i} \quad \binom{1}{i} \quad \cdots \quad \binom{j}{i} \quad \cdots \quad \binom{2^{r-1}-1}{i}\right)$$

is added to both V_0 and V_1. In the third case (W_{2i+1} alone) $(0 \cdots 0)$ is added to V_0 and

$$\left(\binom{0}{i} \quad \cdots \quad \binom{2^{r-1}-1}{i}\right)$$

is added to V_1. Finally, in the last case (both W_{2i} and W_{2i+1}), it has the effect of adding

$$\left(\binom{0}{i} \quad \binom{1}{i} \quad \cdots \quad \binom{j}{i} \quad \cdots \quad \binom{2^{r-1}-1}{i}\right)$$

to V_0 and $(00 \cdots 0)$ to V_1. Thus, the four possibilities of adding or not adding the row

$$\left(\binom{0}{i} \quad \binom{1}{i} \quad \cdots \quad \binom{2^{r-1}-1}{i}\right)$$

to V_0 and V_1 each occur exactly once. Of course, this holds for each of the $\frac{1}{2}d = m$ pairs of rows in V. Hence in generating $\mathscr{C}_{2m}$ we are generating $\mathscr{C}_m$ independently in both V_0 and V_1. This immediately implies $D_{2m} = D_m^2$ which is (2). □

Proof of Theorem A.2(3). Here $d = 2m + 1$ with $m \neq 2^s$. The last three rows of V_0 and V_1 are (mod 2):

$$\text{for } V_0: \quad \binom{0}{m-1} \quad \binom{1}{m-1} \quad \cdots \quad \binom{j}{m-1} \quad \cdots \quad \binom{2^{r-1}-1}{m-1}$$
$$0 \qquad 0 \qquad \cdots \qquad 0 \qquad \cdots \qquad 0$$
$$\binom{0}{m} \quad \binom{1}{m} \quad \cdots \quad \binom{j}{m} \quad \cdots \quad \binom{2^{r-1}-1}{m}$$

$$\text{for } V_1: \quad \binom{0}{m-1} \quad \binom{1}{m-1} \quad \cdots \quad \binom{j}{m-1} \quad \cdots \quad \binom{2^{r-1}-1}{m-1}$$
$$\binom{0}{m-1} \quad \binom{1}{m-1} \quad \cdots \quad \binom{j}{m-1} \quad \cdots \quad \binom{2^{r-1}-1}{m-1}$$
$$\binom{0}{m} \quad \binom{1}{m} \quad \cdots \quad \binom{j}{m} \quad \cdots \quad \binom{2^{r-1}-1}{m}$$

In this case each of V_0 and V_1 have a single, repeated, unpaired row. In each of V_0 and V_1 the codes generated by the first $2m$ rows are $\mathscr{C}_m$ as before. The last row is what changes $\mathscr{C}_{2m}$ (and so D_{2m}) into $\mathscr{C}_{2m+1}$ (and so D_{2m+1}). Thus

$$D_{2m+1} = D_{2m} + (D_{m+1} - D_m)^2,$$

and this is (3). □

Proof of Theorem A.2(4). Here $d = 2^s + 1$. In going from 2^s to 2^{s+1}, the new row added has the form (mod 2)

$$W_{2^s} = \binom{0}{2^s} \quad \binom{1}{2^s} \quad \cdots \quad \binom{2^s-1}{2^s} \quad \binom{2^s}{2^s} \quad \cdots \quad \binom{2^{s+1}}{2^s}$$
$$= 0 \qquad 0 \qquad \cdots \qquad 0 \qquad 1 \qquad \cdots \qquad 1$$

The length of the codewords jumps from 2^s to 2^{s+1}. Thus the array V appears as

$$\begin{array}{ll} W_0 & W_0 \\ W_1 & W_1 \\ \vdots & \vdots \\ W_{2^s-1} & W_{2^s-1} \\ 0\cdots 0 & 1\cdots 1 \end{array}$$

with all vectors shown of length 2^s. Thus, a codeword consisting of any linear combination of the first 2^s rows has the form (Z, Z) where $Z \in \mathscr{C}_{2^s}$. Hence, the code generated by the first 2^s rows has weight enumerator $D_{2^s}(t^2)$. Adding the final row gives words of the form $(Z, \bar{Z})$ where $\bar{Z}$ is the coordinatewise complement of Z. Any such word has weight 2^s, and there are $|\mathscr{C}_{2^s}| = 2^{2^s}$ of them. Since $D_{2^s}(t) = (1+t)^{2^s}$, $D_{2^s+1}(t) = (1+t^2)^{2^s} + 2^{2^s}t^{2^s}$, which is (4). □

The final result of this Appendix is an exact description of the codewords of weight two.

Theorem A.5. *Let $\mathscr{C}_d$ be the code generated by the vectors W_i of* (A.1), *and let S_2 be the set of weight-two words in $\mathscr{C}_d$. Suppose that d begins with s ones in its binary expansion and that $2^{r-1} < d < 2^r$. Then, the vectors in S_2 can be described as follows. Break up the 2^r coordinates into 2^s disjoint blocks each of length 2^{r-s}. A given codeword in S_2 can be uniquely specified by giving two blocks and an integer k, $0 \leqslant k < 2^{r-s}$. The word has a one in the kth position of each of these blocks and zeros elsewhere. In particular, $|S_2| = \binom{2^s}{2} 2^{r-s}$.*

Proof. Let $V = V(d)$ be the $d \times 2^r$ array with rows $W_0, W_1, \ldots, W_{d-1}$. This array has a recursive structure which we now describe. Let $r = s + t$ so that d begins with s ones, then a zero, then $t-1$ following binary digits. In particular, $d \leqslant 2^{r-1} + 2^{r-2} + \cdots + 2^t + 2^{t-1} - 1$. The array V can be pictured as in Fig. 1.

The lower line L which defines the lower boundary of the array is above row $W_{2^{r-1}+\cdots+2^t+2^{t-1}}$. Hence, any subset sum of rows between L and $L' = 2^{r-1} + 2^{r-2} + \cdots + 2^t$ has the form $0\cdots 0XX$ with X of length 2^{t-1} having even weight.

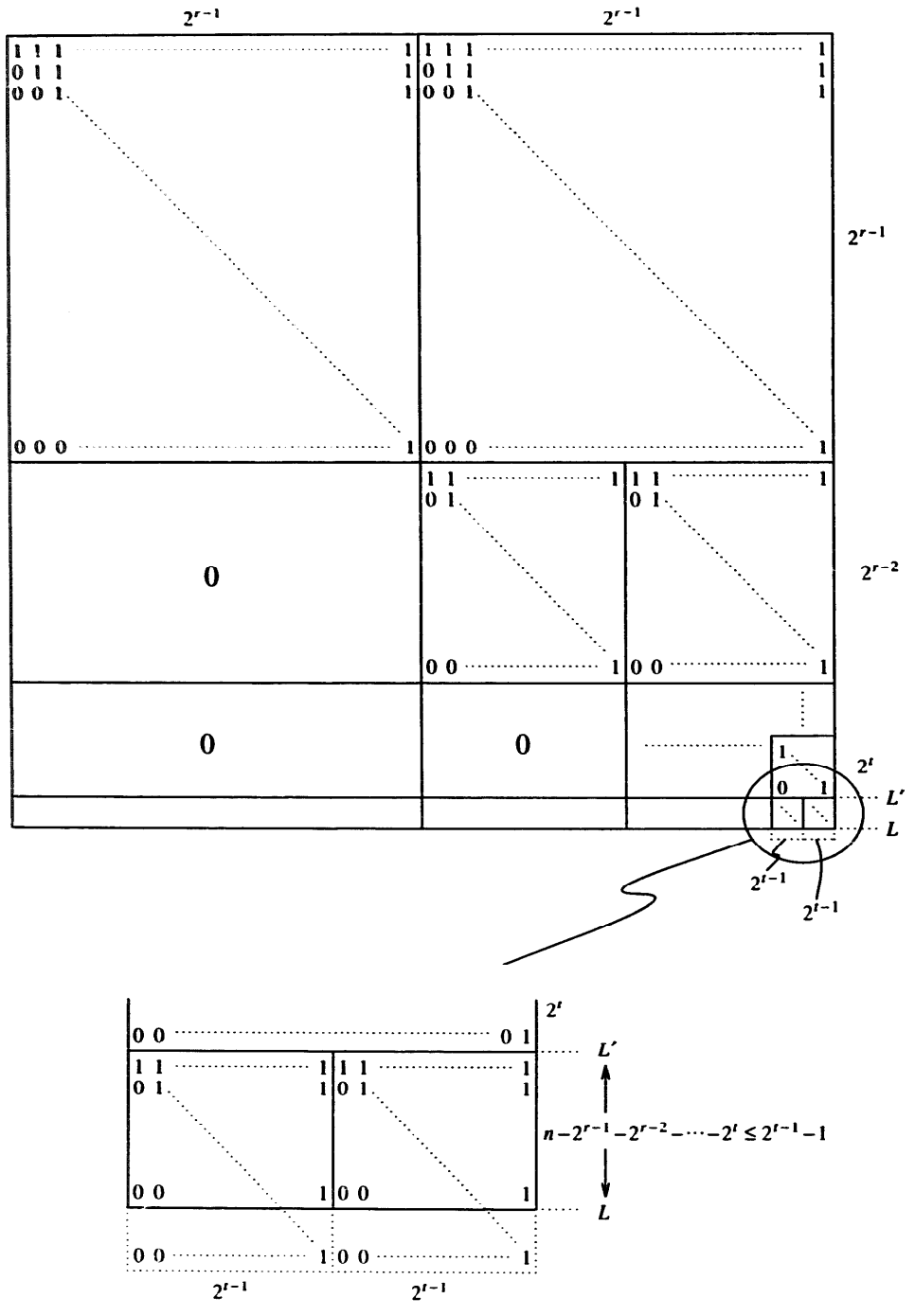

Fig. 1.

Now note the following.

(i) The codes generated by the rows above L' are exactly the codes in $\mathscr{C}_{2^r-2^t}$ (combining the proofs of Lemma 2.6 and Theorem A.2(2) or adapting the proof of Lemma 2.6). These are

exactly the codes with the following property: for every $k = 0, 1, \ldots, 2^t - 1$, the sum of the entries in spaces congruent to $k \pmod{2^t}$ is even.

(ii) The nonzero codes generated by rows below L' lie in the last block of spaces and have at least four ones in each codeword.

From these remarks Theorem A.5 is clear. □

Acknowledgement

The authors wish to acknowledge the useful technical comments of Richard Stong on an earlier draft of this paper.

References

[1] D. Aldous, Random walks on groups and rapidly mixing Markov chains, in: *Séminaire de Probabilités XVII*, Lecture Notes in Math. **986** (Springer, Berlin, 1983) 243–297.
[2] D. Aldous and P. Diaconis, Shuffling cards and stopping times, *Amer. Math. Monthly* **93** (5) (1986) 333–348.
[3] F. Chung, P. Diaconis and R.L. Graham, A random walk problem arising in random number generation, *Ann. Probab.* **15** (1987) 1148–1165.
[4] P. Diaconis, Group representations in probability and statistics, Inst. Math. Statist., Hayward, CA, 1988.
[5] P. Diaconis, Patterned matrices, in: C. Johnson, Ed., *Proc. Short Course on Matrix Theory and Its Applications* (Amer. Mathematical Soc., Providence, RI, 1990) 37–57.
[6] P. Diaconis and R.L. Graham, Running a recurrence with a bad bit, Unpublished manuscript, 1991.
[7] P. Diaconis, R.L. Graham and J. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, *Random Structures Algorithms* **1** (1989) 21–37.
[8] P. Diaconis and M. Shahshahani, Generating a random permutation with random transposition, *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **57** (1981) 159–179.
[9] P. Diaconis and S. Zabell, Updating subjective probability, *J. Amer. Statist. Assoc.* **77** (1982) 822–830.
[10] W. Feller, *An Introduction to Probability and Its Applications, Vol. II* (Wiley, New York, 2nd ed., 1971).
[11] W. Goh and E. Schmutz, A central limit theorem on $GL_n(\mathbb{F}_q)$, *Random Structures Algorithms* **2** (1) (1991) 47–53.
[12] M. Hildebrand, Analysis of random walk on the affine group (mod n), *Ann. Probab.*, to appear.
[13] D. Knuth, *The Art of Computer Programming, Vol. I* (Addison-Wesley, Menlo Park, CA, 2nd ed., 1973).
[14] J.P. Serre, *Linear Representations of Finite Groups* (Springer, New York, 1977).
[15] R. Stong, Some asymptotic results on finite vector spaces, *Adv. Appl. Math.* **9** (1988) 167–199.