PROPERTIES OF RANDOM MATRICES AND APPLICATIONS

IAN F. BLAKE AND CHRIS STUDHOLME

ABSTRACT. This report surveys certain results on random matrices over finite fields and their applications, especially to coding theory. Extensive experimental work on such matrices is reported on and resulting conjectures are noted.

December 15, 2006

1. INTRODUCTION

The study of random matrices over a finite field arises most naturally in a variety of contexts covered by the term "probabilistic combinatorics". Perhaps the prime example of this area is the study of graphical evolution, and in particular the study of threshold phenomena on graphs as more edges are added in a prescribed random manner to a set of graph vertices. However, many other aspects, such as the study of random permutations, random equations over finite fields, and many others are also of importance. The particular application of interest in this report is to the study of rank properties of rectangular matrices over finite fields, and their use in coding theory. The intent is a compilation and survey of relevant results of interest. It is by no means encyclopdaeic. The only contribution of the report is in the experimental results given in section 6.

Our main interest will be the study of rank properties of random $k \times (k+m)$ matrices where $m > -k$, over $\mathbb{F}_q$ which will be designated $M_{k,k+m}(q)$. The q will be omitted if it is understood. When interest is restricted to square matrices over $\mathbb{F}_q$ of size $n \times n$ we will use the notation $M_n(q)$, to emphasize the difference, and again omit the q when it is understood. In either the square or rectangular case, we say the matrix is of full rank if it has rank $\min(k, k+m)$ (or n, respectively). Where possible, we adapt results from the literature to this notation and note where this has not been done. Later, the notation will be modified to accommodate the probability with which each element of the finite field is chosen.

The reader is reminded of the standard algorithmic complexity notation [13] for a function of an integer N, $f(N)$: i) $g(N) = O(f(N))$ iff $\mid g(N)/f(N) \mid$ is bounded from above as $N \to \infty$ ii) $g(N) = o(F(N))$ iff $g(N)/f(N) \to 0$ as $N \to \infty$ iii) $g(N) = \Omega(N)$ iff $\mid g(N)/f(N) \mid$ is bounded from above by a strictly positive number as iv) $g(N) = \Theta(N)$ iff $\mid g(N)/f(N) \mid$ is bounded both from above and below by a strictly positive number as $N \to \infty$.

Date: March, 2006.

1991 *Mathematics Subject Classification.* Primary 54C40, 14E20; Secondary 46E25, 20C20.

Key words and phrases. random matrices, bipartite graphs, coding theory.

The first author was supported in part by NSERC Grant A632.

The outline of the report is as follows: The next section gives the well known enumeration of certain subspaces of vectors spaces over finite fields. The following two sections discuss random matrices, especially their rank properties, over $\mathbb{F}_2$ and $\mathbb{F}_q,\ q > 2$. Section 5 considers other aspects of of both random and nonrandom matrices that are relevant to our interests. These include the properties of windowed random matrices, and the algorithmic construction of rectangular matrices with the property that each column has a maximum weight and any k columns are independent. Some questions relating to the eigenvalues of random matrices and certain randomness questions of some matrix groups are noted.

Section 6 reports extensively on experiments with certain rank properties of random matrices. This leads to certain conjectures believed to be of interest. This is followed by a section on codes for the erasure channel that are derived from the windowed matrices previously considered. These codes, while having a slightly increased decoding complexity, have a very low overhead and high probability of decoding completion.

It is empasized again that, apart from the experimental data generated, the report is a compilation of certain of the approaches to these matrices found in the literature. It is intended only as a summary of certain of the approaches of interest.

2. A preliminary result

Let $V_n(q)$ denote the vector space of dimension n over the finite field $\mathbb{F}_q$ The number of ways of choosing a basis for $V_n(q)$ is easily determined as

$$(q^n-1)(q^n-q)\cdots(q^n-q^{n-1}) = q^{n(n-1)/2}\prod_{i=1}^{n}(q^i-1).$$

Likewise the number of ways of choosing a basis of a k dimensional subspace of $V_n(q)$ is

$$(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})$$

and for each such subspace there are

$$(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})$$

ways of choosing a basis. Thus the number of distinct k-dimensional subspaces of $V_n(q)$, a quantity we denote by $\begin{bmatrix} n \\ k \end{bmatrix}_q$, is

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1}\frac{(q^{n-i}-1)}{(q^{k-i}-1)}.$$

The quantities $\begin{bmatrix} n \\ k \end{bmatrix}_q$ are referred to as Gaussian coefficients and enjoy many properties that are similar to those of ordinary binomial coefficients (e.g. [4]):

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k \end{bmatrix}_q \text{ and } \begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)}{(q^k-1)}\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q.$$

It is easy to see that one can view $\begin{bmatrix} n \\ k \end{bmatrix}_q$ as a polynomial (not a rational function) in q and, if one replaces q with 1 in this polynomial, one obtains $\binom{n}{k}$ ([29]).

One has the following theorem ([29]):

Theorem 2.1 ([29], page 303)**.** *The number of surjective linear transformations from an n-dimensional vector space $V_n(q)$ to an m-dimensional vector space over $\mathbb{F}_q$ is*

$$\sum_{k=0}^{m}(-1)^{m-k}\begin{bmatrix} m \\ k \end{bmatrix}_q q^{nk+\binom{m-k}{2}}. \tag{2.1}$$

The proof (omitted here) uses Möbius inversion on the lattice of subspaces of the vector space $V_n(q)$. An important corollary of this theorem for our purposes is ([29]), recast with the notation we will use in the sequel, is:

Corollary 2.2. *The number of $k \times n$ matrices over $\mathbb{F}_q$ that have rank r is*

$$N_q(k,n,r) = \begin{bmatrix} n \\ r \end{bmatrix}_q \sum_{\ell=0}^{r}(-1)^{r-\ell}\begin{bmatrix} r \\ \ell \end{bmatrix}_q q^{k\ell+\binom{r-\ell}{2}}. \tag{2.2}$$

Here, the first Gaussian coefficient is the number of distinct subspaces of an n-dimensional space of dimension r and the summation is the number of surjective maps from a k-dimensional space to an r-dimensional one and the result follows.

Our interest in subsequent sections will be in $k\times(k+m)$ matrices wth $m > -k$, $(m+k > 0)$.

3. Random matrices over $\mathbb{F}_2$

The problem of determining the probability that a set of m randomly chosen k-tuples over a finite field $\mathbb{F}_2$ are linearly independent (over a $\mathbb{F}_2$) is quite old. As noted earlier, we denote by $M_{k,k+m}(q)$ a $k\times(k+m)$ random matrix, over the finite field $\mathbb{F}_q$. We will normally be interested in the case where $m \geq 0$ but will also consider the case for $-k \leq m \leq 0$. When the finite field is understood we delete the q in the matrix notaton. Denote by $\rho(M_{k,k+m})$ the rank of the matrix ($\leq \min(k, k+m)$). If this probability is denoted $Q_{k,k+m}$ then Berlekamp [2] (adapted to our notation) gives the argument (credited to Lansdurg [17]) that for $m \leq 0$

$$P(\rho(M_{k,k+m}) = k+m) = \prod_{j=0}^{k+m-1}\big(1-\frac{1}{2^{k-j}}\big),\ -k \leq m \leq 0.$$

For $k \to \infty$ for fixed $m \geq 0$, then we will see that

$$P\rho(M_{k,k+m}) = k) = Q_m = \prod_{i=-m+1}^{\infty}\big(1-\frac{1}{2^i}\big), m \geq 0.$$

A useful rapidly converging form ([2]) of this equation is

$$\log Q_m = \sum_{i=m+1}^{\infty}\log(1-2^{-i}) = \sum_{j=1}^{\infty}\frac{-2^{-sj}}{j(2^j-1)},\ s \geq 0.$$

In particular it is noted that

$$Q_0 = \prod_{j=1}^{\infty}(1-\frac{1}{2^j}) = 0.2887880951\cdots$$

i.e. the probability a square random binary matrix is nonsingular as its dimension tends to infinity. As will be noted later, this expression is very accurate even for k as low as 10.

For future reference we give an interesting approximation of Brent et al [7]. Denote by

$$\eta(n,x) = (1-x)(1-x^2)\cdots(1-x^n) \tag{3.1}$$

and from the above it is clear we will have interest in expressions of the form $\eta(n,1/q)$. A very simple and useful argument gives the result ($m \geq 0$)

$$\prod_{i=1}^{k}\left(1-\frac{1}{q^{m+i}}\right) = \begin{cases} 0.288 & q=2\ m=n, \\ 1-\frac{1}{q^m(q-1)} & \text{otherwise .} \end{cases}$$

To see this, note that for $m \geq n$ we have

$$\begin{aligned} \tfrac{\eta(m,x)}{\eta(m-n,x)} &= (1-x^{m-n+1})(1-x^{m-n+2})\cdots(1-x^m) \\ &\geq 1-(x^{m-n+q}+x^{m-n+2}+\cdots+x^m) \\ &\geq 1-x^{m-n+1}(1+x+x^2+\cdots+x^{n-1}+\cdots) \\ &\geq 1-x^{m-n+1}\cdot\tfrac{1}{1-x} \end{aligned}$$

from which the second result follows. When $q=2$ and $m=n$ the bound is zero and for this case we compute:

$$\begin{aligned} &\left(1-\tfrac{1}{2}\right)\left(1-\tfrac{1}{2^2}\right)\cdots\left(1-\tfrac{1}{2^m}\right) \\ &> \textstyle\prod_{i=1}^{\infty}\left(1-\frac{1}{2^i}\right) \\ &> \left(1-\tfrac{1}{2}\right)\left(1-\tfrac{1}{2^2}\right)\left(1-\tfrac{1}{2^3}\right)\left(1-\tfrac{1}{2^4}\right)\left(1-\left(\tfrac{1}{2^5}+\cdots+\tfrac{1}{2^m}+\cdots\right)\right) \\ &= \left(1-\tfrac{1}{2}\right)\left(1-\tfrac{1}{2^2}\right)\left(1-\tfrac{1}{2^3}\right)\left(1-\tfrac{1}{2^4}\right)\left(1-\left(\tfrac{1}{2^4}\right)\right. \\ &> 0.288. \end{aligned}$$

The problems of random matrices over finite fields (especially binary) and their applications to graph theory, owe much to researchers in the former Soviet union. We begin by considering the work of Kolchin ([16]) and transfer his results to the notation of interest here. Namely, we are interested in the probabilty a $k \times (k+m)$ binary (i.e. over $\mathbb{F}_2$, although there are many results for arbitrary finite fields $\mathbb{F}_q$) matrix having rank $r = k-s$ for $s \geq 0,\ m+s \geq 0$. Clearly, if the matrix entries are chosen independently and equally likely ($P(a_{ij}=0)=1/2$), the number of such matrices is $N(k,k+m,r)_2$ as in equation (2.1) and the matrices are equally likely and the result is

$$N(k,k+m,r)_2/2^{k(k+m)}.$$

Kolchin gives a different argument as follows:

Theorem 3.1 ([16], page 126, adapted)**.** *Let $M_{k,k+m}$ be a binary random $k\times(k+m)$, $m > -k$, matrix with entries chosen equally likely and $\rho(M_{k,k+m})$ its rank (over $\mathbb{F}_2$). It will be convenient to denote the rank by $\rho(M_{k,k+m}) = r = k-s,\ s \geq 0$ Then for $k-s \leq \min(k,k+m)$, as $k\to\infty$ we refer to s as the nullity or defect of the matrix and we have:*

$$P(\rho(M_{k,k+m}) = k-s) \to 2^{-(s)(m+s)} \prod_{i=s+1}^{\infty}\left(1-\frac{1}{2^i}\right)\prod_{i=1}^{m+s}\left(1-\frac{1}{2^i}\right)^{-1},$$

where the last product is 1 if $m+s=0$ (i.e. the matrix is of full rank).

Proof: (informal) The proof technique is interesting and an indication of it is given. Consider adding random binary columns, random k-tuples over $\mathbb{F}_2$, to the matrix, column by column. If the matrix is currently of size $k \times (k+m-1)$ and rank j (the $k+m-1$ columns span a j dimensional space) the probability the $(k+m)$-th column is in this space is $2^j/2^k$. Hence the probability the rank is increased by one with the added column is

$$1 - \frac{2^j}{2^k}.$$

There is an easy case to dispose of, that of full rank when $m+s=0$. In this case, to achieve full rank, the rank must increase by one each time a column is added and the probability of this is

$$P(\rho(M_{k,k-s}) = k-s) = \prod_{j=0}^{k-s-1} \big(1 - \frac{2^j}{2^k}\big) = \prod_{i=s+1}^{k} \big(1 - \frac{1}{2^i}\big),\ s \geq 0.$$

In the general case, we consider the rank as a discrete Markov chain as a function of the number of columns and for convenience denote $\rho(M_{k,k+m})$ by ρ_{k+m} as the rank of the $k \times (k+m)$ matrix. Denote by ξ_ℓ the random variable which takes on the value 1 if the ℓ-th column increases the rank (by one) and zero otherwise. We have the probabilities:

$$P(\xi_\ell = 0 \mid \rho_{\ell-1} = a) = 2^a/2^k, \text{ and } P(\xi_n = 1 \mid \rho_{\ell-1} = a) = 1 - 2^a/2^k.$$

The probability the rank of the $k \times (k+m)$ matrix is $k-s$ is then the probability exactly $k-s$ of the (independent) random variables $\{\xi_1, \xi_2, \cdots, \xi_{k+m}\}$ have the value 1. Thus, suppressing our previous notation using $M_{k,k+m}$, assuming the ranks remained the same when columns $\mathcal{R} = \{1 \leq i_1, i_2, \cdots, i_{m+s} \leq k+m\}$ were added, (note that $k+m-(k-s) = m+s$ and that $\mathcal{R}^c$ is the complement set in $[1, k+m]$) we have:

$$\begin{aligned} P(\rho_{k+m} = k-s) &= \textstyle\sum_{i_a \in \mathcal{R}, j_b \in \mathcal{R}^c} \prod_{a=1}^{m+s} P(\xi_{i_a} = 0) \prod_{b=1}^{k-s} P(\xi_{j_b} = 1) \\ &= \textstyle\sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_{k+m} \leq n} \big(1 - \frac{1}{2^k}\big)\big(1 - \frac{2}{2^k}\big) \cdots \big(1 - \frac{2^{k-s+1}}{2^k}\big) \\ &\quad \times \frac{2^{i_1 - 1 + i_2 - 2 + \cdots + i_{m+s} - (m+s)}}{2^{k(m+s)}}. \end{aligned}$$

Notice that the product term here is independent of the particular instances of when the rank increases ($\xi_j = 1$) and so this last equation can be written as

$$\begin{aligned} P(\rho_{k+m} = k-s) &= 2^{-k(m+s)} \textstyle\prod_{j=0}^{k-s-1} \big(1 - \frac{2^j}{2^k}\big) \times \\ &\quad \textstyle\sum_{1 \leq i_1 \leq i_2 \leq \cdots \leq i_{m+s} \leq n} 2^{i_1 - 1 + i_2 - 2 + \cdots + i_{m+s} - (m+s)}. \end{aligned} \tag{3.2}$$

Notice that $1 \leq i_1 < i_2 < \cdots < i_{m+s} \leq k+m$ and so

$$0 \leq i_1 - 1 \leq i_2 - 2 \leq \cdots \leq i_{m+s} - (m+s) \leq k-s.$$

If we define the variables $j_\ell = k - s - (i_\ell - \ell),\ \ell = 1, \cdots, m+s$ then we have

$$0 \leq j_{m+s} \leq \cdots \leq j_1 \leq k-s.$$

Considering equation 3.2, noting that $k(m+s) = (k-s)(m+s) + s(m+s)$ to account for the subtraction of the rank, $k-s$, from each index in the exponent, and reversing the

product variable (replace i with $k-i$) we can rewrite the equation as

$$P(\rho_{k+m}=k-s) \quad \begin{array}{l} = 2^{-s(m+s)}\prod_{i=s+1}^{k}\left(1-\frac{1}{2^i}\right)\times \\ \sum_{0\le j_{m+s}\le j_{m+s-1}\le\cdots\le j_1\le k-s} 2^{-j_{m+s}-j_{m+s-1}+\cdots-j_1}. \end{array} \tag{3.3}$$

We examine the equation as $k\to\infty$ and note the product term tends to

$$\prod_{i=s+1}^{\infty}\left(1-\frac{1}{2^i}\right).$$

To evaluate the sum term in 3.3 (over all possible values of the variables j_ℓ as $k\to\infty$ we have:

$$\begin{aligned} &\sum_{0\le j_{m+s}\le j_{m+s-1}\le\cdots\le j_1} 2^{-j_{m+s}-j_{m+s-1}+\cdots-j_1} \\ &= \textstyle\sum_{0\le j_{m+s}\le j_{m+s-1}\le\cdots\le j_2} 2^{-j_{m+s}-j_{m+s-1}+\cdots-j_2}\sum_{j_2\le j_1}2^{-j_1} \\ &= \textstyle\sum_{0\le j_{m+s}\le j_{m+s-1}\le\cdots\le j_2} 2^{-j_{m+s}-j_{m+s-1}+\cdots-j_2}2^{-j_2}(1-\frac{1}{2})^{-1} \\ &= \textstyle\sum_{0\le j_{m+s}\le j_{m+s-1}\le\cdots\le j_2} 2^{-j_{m+s}-j_{m+s-1}+\cdots-j_3}\sum_{j_3\le j_2}2^{-2j_2}(1-\frac{1}{2})^{-1} \\ &= \textstyle\sum_{0\le j_{m+s}\le j_{m+s-1}\le\cdots\le j_2} 2^{-j_{m+s}-j_{m+s-1}+\cdots-j_3}(1-\frac{1}{2^2})^{-1}(1-\frac{1}{2})^{-1} \end{aligned}$$

and so on. The final expression is then as stated in the theorem. □

Notice that the probability a random $k\times k$ binary matrix M_k is of full rank k as k becomes large tends to the constant

$$P(\rho(M_k)=k)\to\prod_{i=1}^{\infty}\left(1-\frac{1}{2^i}\right)=0.2887880951\cdots\ .$$

More generally, the probability a random $k\times(k+m)$ binary matrix $M_{k,(k+m)}$ is of full rank k for $m\ge 0$, for large k is

$$Q_m=\prod_{i=m+1}^{\infty}\left(1-\frac{1}{2^i}\right),m=0,1,\cdots. \tag{3.4}$$

If we let $P_m,\ m\ge 0$, denote the probability that exactly m columns beyond k are needed to obtain rank k for the $k\times(k+m)$ matrix, then

$$P_m=Q_m-Q_{m-1}$$

and the average number of extra columns needed for full rank is

$$\bar{m}=\sum_{m=0}^{\infty}mP_m=\sum_{i=0}^{\infty}(1-Q_i)=1.60669515\cdots\ .$$

The simple bounds established by Brent atl [7] noted earlier can be useful for establishing bounds when working with such expressions.

It is interesting to observe that for such a matrix only two extra columns, on average, beyond the k, are required to achieve a full rank matrix and this, asymptotically, is independent of k. With 7 or 8 extra columns, the probability of achieving full rank is very close to 1 for even very small. While the above expressions are for large k, it has been observed experimentally that the expressions are remarkably accurate for k as small as 10.

The above result depends on the fact that, as a column (a randomly chosen binary k-tuple) is added to the matrix, the probability the rank of the matrix is increased depends

only on the current rank of the previous columns of the matrix. This assumes that the probability the added column is in the column space of the previously chosen columns is proportional to the size of the current columns space i.e. $2^\ell/2^k$ if the current rank is ℓ. Indeed, if the new column being added is not sufficiently random in being chosen, this observation may not hold.

It seems somewhat surprising then that the previous results do not depend on the matrix elements being chosen equally likely. In fact, as will be shown later in the work of Kolchin and Cooper (due in some part to several other Russian mathematicians whose work is in Russian and not used in this survey), as long as the probabilities tend to zero in a carefully prescribed manner, the same results will hold. The following result of Kolchin is one form of this kind of result.

Theorem 3.2 ([16], page 131, adapted)**.** *Let the elements of a random binary $k \times (k+m)$ matrix be independently chosen and suppose there is a constant δ such that the probabilities $p_{ij}^{(k)}$ of elements a_{ij} being 1 satisfy the inequalities*

$$\delta \leq p_{ij}^{(k)} \leq 1-\delta,\ i = 1, 2, \cdots, k,\ j = 1, 2, \cdots k+m,$$

hold. Let $s \geq 0$, and m be fixed integers, $m+s \geq 0$. Then as $k \to \infty$ we have

$$P(\rho_k(M_{k,k+m}) = k-s) \to 2^{-s(m+s)} \prod_{i=s+1}^{\infty} \left(1-\tfrac{1}{2^i}\right) \prod_{i=1}^{m+s} \left(1-\tfrac{1}{2^i}\right)^{-1},$$

where the last product is 1 if $m+s=0$.

The proof of the theorem is not given in [16]. However a variety of other results in the same direction, namely on the insensitivity of the results to variations in the matrix element probabilities, are available and some of these are noted here.

Let A be an $k \times (k+m)$ binary random matrix, then under fairly broad conditions on the probability distribution of the matrix elements, it is shown that the mean number of nonzero solutions to the matrix equation

$$XA = 0$$

tends to 2^{-m} as $k \to \infty$, where m is allowed to be negative. For m negative, the statement is a reflection of the expected number of vectors in the null space of the matrix (and hence of the expected rank of the null space). For m positive it simply reflects the probability the null space is of full rank, in some sense. The proof of this result is given in [16]. Perhaps of more interest is the fact the result remains true if the matrix element probabilities satisfy the inequalities

$$\frac{\log k + x}{k} \leq p_{ij}^{(k)} \leq 1 - \frac{\log k + x}{k}$$

where x is a constant. This result is made sharper in the work of Cooper to be discussed in the next section.

We pursue the ideas a little further. Let $A = A_{k,k+m}$ be a random binary $k \times (k+m)$ matrix. If columns $i_1, i_2, \cdots, i_r$ sum to the zero k-tuple we call the set of indices $C = \{i_1, i_2, \cdots, i_r\}$ a *critical set*. Note that if $C_1 \neq C_2$ are critical sets then their symmetric difference $C_1 \Delta C_2$ is also a critical set. One can then naturally define independent critical

sets. Clearly the maximum number of independent critical sets is the dimension of the null space of the matrix A, $s(A)$, (and the sum of this and the rank of the matrix is $k+m$).

Suppose now that the elements of the matrix A are chosen independently according to the distribution

$$P[a_{ij}=1]=p_{ij}^{(k)}=\frac{\log k+x}{k},\ \text{and}\ P[a_{ij}=0]=1-\frac{\log k+x}{k} \tag{3.5}$$

where x is a constant. The limit distribution for the dimension of the null space of A, $s(A)$ can then be found. In particular we have

Theorem 3.3 ([16], p. 135, theorem 3.3.1)**.** *If $k, k+m \to \infty$ such that $(k+m)/k \to \alpha < 1$ and the condition (3.5) is true, then the random variable $s(A)$ (maximum number of independent critical sets or rank of the null space) converges to a Poisson distribution with parameter $\lambda = \alpha e^{-x}$.*

A restatement of the theorem is, as noted in [10] (attributed to Balakin [1] and discussed in [16] (Theorem 3.3.2, page 142, adapted), if $A_{k,k+m}$ is a random matrix over $\mathbb{F}_2$, $m \geq 0$ and $p=(\ln(k)+d)/k,\ a=(k/(k+m))$ then

$$P(\rho(A_{k,k+m}=k-s) \sim \frac{(ae^{-d})^k}{k!}e^{-ae^{-d}}. \tag{3.6}$$

Furthermore it is shown that under the above conditions, the number of all zero columns of the matrix A has a Poisson distribution with parameter $\lambda=\alpha e^{-d}$ if $(k+m)/k \to \alpha$ for $0<\alpha<\infty$, a simple approximation of the binomial distribution. Further, if, under the previous conditions, and $(k+m)/k \to \alpha < 1$ then indeed, with probability tending to 1, the critical sets of A consist only of zero columns.

Similar to the above theorem we have:

Theorem 3.4 ([16] p. 142, theorem 3.3.2)**.** *Under the conditions 3.5, if $k, (k+m) \to \infty$ such that $(k+m)/k \to \alpha > 1$ then the distribution of $s(A)$ converges to a Poisson distribution with parameter $\lambda = e^{-d}/\alpha$.*

We have been a little informal with stating the results. In [16] a more precise statement of the results states that for a $T \times n$ matrix over $\mathbb{F}_2$, if $n,\ T \to \infty$ in such a way that $T/n \to a$, constant, either $a>1$ or $a<1$, the results hold.

In essence it was shown that in the equiprobable case as $k \to \infty$ the probability the matrix $A_{k,k+m}$ is of full rank tends to 1, under the conditions stated. The results are nontrivial to prove.

Similar results can be obtained for nonhomogeneous equations $AX=B$ and many other aspects of the problem are treated in [16].

We note one further aspect of the conditions in equation 3.5. Suppose we let the probability of a 1 element in the matrix be

$$p=\frac{\log(k)+x}{k} \tag{3.7}$$

where we view x as a constant. Suppose further that $(k+m)/k \to \alpha =$ constant, $0 < \alpha < \infty$. Then the probability of generating an all zero column is

$$p_k = \big(1 - \frac{\log(k) + x}{k}\big)^k$$

and as $k \to \infty$ in such a way that $(k+m)/k$ is constant then the probability of an all zero column tends to $p_k = e^{-x}/k$. Thus, the 'threshold value' of $p = \ln(k)/k$ noted above (and later, in the work of Cooper [9, 10] to be discussed) is somewhat natural. Similarly, if we take $p = (c\ln(k) + x)/k$ for $c > 0$, then $p_k \to e^{-x}/k^c$ and as c decreases the probability increases. A similar argument applied to rows, is more persuasive as to this threshold, since such a matrix with an all-zero row cannot be nonsingular. The expected number of all zero columns in the matrix then is $(k+m)p_k \to \alpha e^{-x}$, a constant. For lower values of p one would thus expect the number of all zero columns to increase as k increases. Thus the probability of a 1 in the matrix given by 3.7 leads to an expected constant number of all zero columns which perhaps explains somewhat the threshold effect of the probability.

4. Random matrices over $\mathbb{F}_q$, $q > 2$

The threshold probability, mentioned in the previous section, of the probability of a one in the random binary $n \times n$ matrix being

$$p = \frac{\log(n) - c}{n}$$

and its relationship to the number of zero rows or columns is explored further in a series of papers ([5], [9]. [10]) by extending the work to matrices over $\mathbb{F}_q$ and obtaining sharper estimates of the bound on the threshold probabilities. They consider only a square $n \times n$ matrix M_n over $\mathbb{F}_q$, $q > 2$, where the probability of a zero element is given by $1 - p$ where $p = (\log(n) - c)/n$ for some constant c where the probability of each nonzero element is equally likely at $p/(q-1)$. They observe that if one desires an expected rank of $n - O(1)$ then this is the critical probability, as noted previously and commented on further below. It is also shown that the rank of such a random matrix is, with high probability, not much less than its expected rank. In fact, the main technical result of the paper [5] is that the number of linear dependencies of the rows (or columns) of the matrix is bounded by a constant *iff* $p \geq (\log(n) - c)/n$ where c is some fixed positive constant.

If we ask the further question as to how small p can be chosen so that a random matrix is nonsingular (asymptotically with n) with some constant probability, then it seems the techniques of this paper [5] are insufficient to answer this question. This question is considered in the work of Cooper ([9], [10]) to be considered later. Notice that it is already shown that this is the case for random binary matrices (the constant .288...). (The proof in [25] that this is not the case is erroneous). Also it would be of interest to show that p can be nonconstant through the matrix elements, while preserving the property that the random matrix is nonsingular (asymptotically) with some constant probability.

We informally discuss the results of the paper [5] and adopt and adapt their notation. As before denote by M_n an $n \times n$ random matrix over $\mathbb{F}_q$ whose elements are chosen independently with the probability of a zero element given by $1-p$ where $p = (\log(n)-c)/n$

for some constant c and the probabilities of nonzero elements equally likely at $p/(q-1)$. Denote by $\delta(M_n)$ the *defect* or *nullity*, (dimension of the null space) of the matrix M_n i.e. $\delta(M_n) = n - \rho(M_n)$, where, as before $\rho(M_n)$ is the rank. (Note: we have used $s = \delta$ here, but maintain the dual notation!)

We give the following sequence of theorems and corollaries of that paper with minor adaptation of notation.

Theorem 4.1 ([5], theorem 2.1)**.** *Let M_n be a random $n \times n$ matrix over the fixed finite field $\mathbb{F}_q$ (the probability of a nonzero element being $p/(q-1)$) with $p = (\log(n) - c)/n$ with $n \geq e^c$ for a fixed $c \geq 0$. Then the defect $\delta(M_n)$ satisfies*

$$E(q^{\delta(M_n)}) = O(1).$$

Moreover, if the expectation is considered as a function of p then it is monotonically decreasing as a function in the range $0 \leq p \leq (q-1)/q$.

(Note: The last sentence of this theorem seems to require some clarification in that it does not seem to rule out that, for a constant n and c, if $p < (\ln(n) - c)/n$ the size of the null space increases as p decreases.)

The theorem implies in particular that for p above the threshold value, $p = (\log(n) - c)/n$ for c a constant, (in fact for $(\ln(n) - c)/n < p < (q-1)/q$) the number of possible linear dependencies (rows or columns) is upper bounded by a constant. Values of p close to the threshold are of particluar interest.

The following theorem addresses the possibility of p being a function of n to see what can be said about the rank behavior of the matrix in that case.

Theorem 4.2 ([5], theorem 2.2)**.** *Let M_n be a random matrix over $\mathbb{F}_q$ as before with $p(n) = (\log(n) - c(n))/n$ for $0 \leq c(n) < \log(n) - \log(q-1)$. Then the defect of M_n satisfies*

$$E(q^{\delta(M_n)}) = \Omega(e^{[(q-1)/4]e^c}).$$

The expected number of all zero rows of M_n is $\Omega(e^c)$.

Corollary 4.3 ([5], Corollary 2.3)**.** *Let $c(n)$ be a function with $0 \leq c(n) < \log(n)$ for all n. Then $E(\delta(M_n)) = \mathcal{O}(1)$ for random $n \times n$ matrices over $\mathbb{F}_q$ with $p = (\log(n) - c(n))/n$ iff the function $c(n)$ is bounded.*

Corollary 4.4 ([5], Corollary 2.4)**.** *For every $c \geq 0$ there exists a constant A_c such that a random $n \times n$ matrix M_n, $n > e^c$, with $p = (\log(n) - c)/n$ satisfies*

$$P(\delta(M_n) \geq \ell) \leq \frac{A_c}{q^\ell}$$

for all positive integers ℓ.

Recall that a linear dependency among the rows (or columns) of the matrix M_n is a nontrivial linear sum of rows adding to the zero row (critical set of indices in the terminology of Kolchin). Let $\ell(M_n)$ be the number of such dependencies and note that $\ell(M_n) = q^{\delta(M_n)} - 1$. The proof of the first of these theorems depends on the following:

Theorem 4.5 ([5], theorem 3.3)**.** *Let M_n be a random $n \times n$ matrix over $\mathbb{F}_q$ for an arbitrary p, $0 < p < 1$. Then*

$$E(\ell(M_n)) = \sum_{j=1}^{n} \binom{n}{j} (q-1)^j P_j^n$$

where P_j is given by

$$P_j = \frac{q-1}{q}\big(1 - \frac{qp}{q-1}\big)^j + \frac{1}{q}.$$

The argument of the theorem is straight forward, outlined as follows. Consider a fixed nonzero vector $c = (c_1, c_2, \cdots, c_n)$, $c_i \in \mathbb{F}_q$ with exactly k nonzero elements, which are assumed to be the first k elements without loss of generality. Consider the sum $\sum_{i=1}^{k} c_i m_i$ for $m_i \in \mathbb{F}_q{}^*$ chosen according to probability of a zero of $1-p$ and nonzero with equally likely with probabilities $p/(q-1)$. Since the c_i are fixed we have the following simple recursions:

$$P_0 = 1,\ P_k = P_{k-1}(1-p) + (1 - P_{k-1})p/(q-1)$$

from which it follows that

$$P_k = \frac{q-1}{q}\big(1 - \frac{qp}{q-1}\big)^k + \frac{1}{q}. \tag{4.1}$$

The work of Blömer et al [5] raises many interesting questions, apart from establishing many technical inequalities used in the proofs. They show that if $p = (\log(n) - c(n))/n$ where $0 \leq c(n) \leq a\log(n)$ and $0 \leq a < 1$ is an arbitrary fixed constant then the defect satisfies

$$E(q^{\delta(M_n)}) \leq e^{e^{\mathcal{O}(c(n)+1)}}$$

and in fact the defect of M_n increases exponentially with $c(n)$, as $n \to \infty$. They conjecture that this result is true for an arbitrary function $0 \leq c(n) < \log(n)$.

They ask the question of random binary matrices: what is the smallest p such that the probability that M_n is nonsingular is bounded below by some constant c? The results discussed indicate that for fixed $p > 0$, the probability that M_n is nonsingular tends to $.288788\cdots$ since as $n \to \infty$, p becomes greater than $\ln(n)/n$. However, Corollary 4.4 shows that if $p = (\log(n) - c)/n$, the matrix has constant defect with high probability, for the appropriate condition on c.

The final theorem of [5] (Theorem 6.3) shows a generalization of the binary case to matrices over $\mathbb{F}_q$ by showing that if the matrix elements are chosen zero with probability $1-p$ (where p is the probability of a nonzero element) and nonzero elements chosen equally likely (with probability $p/(q-1)$), for some arbitrarily small but constant p, then then the probability an $n \times n$ matrix M_n is nonsingular is at least

$$P(\rho(M_n) = n) = \prod_{i=1}^{n}(1 - \eta^i)$$

where $\eta = \max(p/(q-1), 1-p)$. This equation is lower bounded by the product to infinity which converges to some positive value. They then raise as a main open problem:

Open problem: Is there a function $p(n)$ that tends to 0 as $n \to \infty$ and a constant $c > 0$ such that a random matrix over $\mathbb{F}_q$ (the paper [5] poses the question only for $q = 2$ but it seems a valid question for the more general case) is nonsingular with probability at least c

This question is taken up by the work of Cooper, described next.

In the first of his two papers on this subject, Cooper [9], considers $n \times n$ matrices over $\mathbb{F}_q$ and considers first the equally likely case i.e. the probability an element is chosen nonzero is $p = (q-1)/q$ and each element in $\mathbb{F}_q$, including the zero element, is chosen equally likely with probability $1/q$. To prepare to discuss the generalizations this work introduces, we introduce the notation that $M_{m,n}(p, q)$ be an $m \times n$ matrix over $\mathbb{F}_q$ where an element is chosen to be zero with probability $1 - p$ (which may in general depend on the dimension, which is then explicitly shown), and each nonzero element of $\mathbb{F}_q$ is chosen equally likely with probability $p/(q-1)$. The case of $p = (q-1)/q$ is the equally likely case. Where one or more of the matrix parameters are understood, they are omitted. In particular we denote a square $n \times n$ matrix as M_n.

Let $P(\rho(M_n(p = (q-1)/q, q)) = n-s)$ be the probability the random square $n \times n$ matrix over $\mathbb{F}_q$ with equally likely probabilities, for nonzero elements, has rank $n - s$. Recall the probability the matrix is nonsingular ($s = 0$) is easily calculated as (recall eqn. (3.1))

$$\eta(n, 1/q) = \prod_{i=1}^{n} \big(1 - \frac{1}{q^i}\big).$$

It will be convenient to introduce the function $\pi(k, q)$, to use the notation of Cooper.

Theorem 4.6 ([9], theorem 1)**.**

$$\lim_{n\to\infty} P(\rho(M_n(p, q) = n - s)) = \pi(s, q) = \begin{cases} \lim_{n\to\infty} \prod_{j=1}^{\infty} \big(1 - \frac{1}{q^j}\big), & s = 0, \\ \prod_{j=s+1}^{\infty}(1 - 1/q^j)/\prod_{j=1}^{s}(1 - 1/q^j)q^{s^2}, & s \geq 1 \ . \end{cases}$$

(The function π defined here is closely related to the function , η defined previously (see Eqn. (3.1)) but we make no attempt to reconcile them here as both are convenient.)

It is noted that since we have an enumeration of rectangular matrices with a given rank (see eqn. (2.2)), in the model where these matrices are equally likely, we have the probability as given in this theorem i.e. it can be shown (not entirely trivial) that

$$\pi(s, q) = \lim_{n\to\infty} N(n, n, n - s)_q / q^{n^2}$$

The following result is for binary random matrices. It is convenient to discuss it here in the context of the above. Denote by $c_2 = \pi(0, 2)$, the probability a random square binary matrix is asymptotically nonsingular. The main theorem of [9] then is a sharper result on $p(n)$ than was in the work of Blömer et al. for matrices over $\mathbb{F}_2$:

Theorem 4.7 ([9], theorem 2)**.** *Let $M_n(p, 2)$ be a random binary matrix (over $\mathbb{F}_2$). Then:*
(i) If $p(n) = (\log(n) + d(n))/n \leq 1/2$ then

$$\lim_{n\to\infty} P(M_n(p, 2) \textit{is nonsingular}) = \begin{cases} 0, & d(n) \to -\infty \ , \\ c_2 \exp(-2e^{-d}), & d(n) \to d = \ \textit{constant} \ , \\ c_2, & d(n) \to \infty \ , \end{cases}$$

(ii) If $p(n) = 1 - (\log(n) + d(n))/n \geq 1/2$ *then*

$$\lim_{n\to\infty} P(M_n(p,2) \textit{is nonsingular}) = \begin{cases} 0, & d(n) \to -\infty \ , \\ c_2 \exp(-2e^{-d})(1+e^{-d})^2, & d(n) \to d = \ \textit{constant} \ , \\ c_2, & d(n) \to \infty \ , \end{cases}$$

(iii) Let $\mathcal{F}$ *be the event that* $M_n(p(n),2)$ *has no zero rows or columns and at most one row and column of all ones. If*

$$(\log(n) - \omega(n))/n \leq p(n) \leq 1 - (\log(n) - \omega(n))/n$$

where $\omega(n) = o(\log\log n)$ *then for any nonnegative integer* k

$$\lim_{n\to\infty} P(M_n(p(n),2) \textit{ has rank } n-s|\mathcal{F}) = \pi(s,2)$$

and in particular

$$\lim_{n\to\infty} P(M_n(p(n),2) \textit{ is nonsingular}|\mathcal{F}) = c_2.$$

The proof of the theorem is intricate and uses an investigation of linear dependencies of columns of the matrix. In particular he computes the expected values of the number of independent linear column dependencies (he refers to these as "simple" to avoid confusion with other references to linear independence - we continue to use the terminology of Kolchin) of a given size and their higher moments. He generalizes somewhat the equation of Blömer et al 2.2 and shows, using recursion as before, that the probability that, for a fixed vector of elements, $d = (d_1, \cdots, d_m)$, $d_i \in \mathbb{F}_q$ the probability that, for randomly chosen $a_i \in \mathbb{F}_q$, according to p, the equation $\sum_i d_i a_i = \gamma \in \mathbb{F}_q$, $\gamma \neq 0$ is

$$\frac{1}{q}\big(1 + (-1)\big(1 - \frac{q}{q-1}p\big)^m\big), \ m \neq 0.$$

The previous expression held only for $\gamma = 0$.

The second paper of Cooper [10] considers the case of rectangular matrices over an arbitrary finite field $\mathbb{F}_q$. (Cooper's terminology is adapted to ours.)

Theorem 4.8 ([10] theorem 1)**.** *Let* $M_{k.k+m}(p,q)$ *where all elements of* $\mathbb{F}_q$ *are equally likely (each having probability* $1/q$*). Then for* $m \geq 0$

$$\lim_{k\to\infty} P(\rho(M_{k,k+m}(p,q)) = k-s) = \pi_m(s,q) = \begin{cases} \prod_{j=m+1}^{\infty}\big(1 - \frac{1}{q^j}\big), & s = 0 \ , \\ \prod_{j=s+m+1}^{\infty}(1-1/q^j)/\prod_{j=1}^{s}(1-1/q^j)q^{s(s+m)}, & s \geq 1 \ . \end{cases}$$

Theorem 4.9 ([10] theorem 2)**.** *For the finite field* $\mathbb{F}_q$ *let* $q \geq 3$ *and* $q = \mathcal{O}(\log\log n)$*. Let* m *be a nonnegative integer and* $M_{k,k+m}(p,q)$ *a* $k \times (k+m)$ *random matrix over* $\mathbb{F}_q$ *with entries independently and identically distributed (zero element has probability* $1-p$ *and nonzero elements equally distributed with probability* $p/(q-1)$*). Let* $c_q = \pi_m(0,q)$ *be the asymptotic probability, as* $k \to \infty$*, the matrix has full rank (*k*). Let* $p(k) = (\log(k) + d(k))/k$ *where* $d(k) \geq -\log(\log(k/9q))$*. Then:*

(i)

$$\lim_{n\to\infty} P(M_{k,k+m}((q-1)/q, q) \textit{ is nonsingular}) = \begin{cases} 0 & d(k) \to -\infty \ , \\ c_q e^{-2e^{-d}} & d \textit{ constant} \ , \\ c_q & d(k) \to \infty \ , \end{cases}$$

(ii) Let $\mathcal{F}$ be the event there are no zero rows or columns in the matrix. For any non-negative integer s

$$\lim_{k\to\infty} P(M_{k,k+m}(p=(q-1)/q,q) \text{ has rank} k-s \mid \mathcal{F}) = \pi_m(s,q),$$

and in particular

$$\lim_{n\to\infty} P(M_{k,k+m}(p=(q-1)/q,q) \text{ is nonsingular} \mid \mathcal{F}) = c_q.$$

There is an interesting comment in [10] to the effect that the moments of the random variable representing the number of solutions of a random homogeneous set of linear equations does not satisfy the Carleman conditions necesary for the probability distribution to be uniqulet determined by its moments. However, it was noted by Alekseychuk, using other methods, that the moments do indeed uniquely specify the distribution in this case, which opens up the possibility of simpler proofs (see [10], page 199).

5. Results on other aspects of matrices over $\mathbb{F}_q$

5.1. **Windowed random binary matrices.** For application to the construction of codes, to be discussed in section 7, we will be interested in *windowed binary matrices*, where the nonzero elements in the matrix are restricted to fall within a window of length w, beginning at a randomly chosen row. Specifically, to add a column to the matrix, choose a row number at random and fill in the w elements at random, beginning with that row, with the probability of a 1 being p. If the initial row is chosen within w of the bottom row, the column will wrap around to the top of the matrix. We consider only square $k\times k$ matrices in this although the extension of the arguments to rectangular matrices is immediate. Suppose we divide the matrix into top half rows, where the initial position is chosen among the top $k/2$ rows, and bottom half rows. Suppose the number of top half rows is m_0 and bottom half rows m_1 ($m_0+m_1=k$). It is clear that if either m_0 or m_1 exceeds $k/2+w$ the matrix cannot achieve full rank. Furhermore, if this happens for one, the other will be less than $k/2-w$, and a necessary condition for full rank is that

$$\frac{k}{2} - w \le m_0 \le \frac{k}{2} + w.$$

The random variable m_0 is binomially distributed with mean $k/2$ and variance $\sigma = \sqrt{k}/2$. For k large, we use the normal approximation to the binomial and let

$$z = \frac{m_0 - (k/2)}{\sqrt{k}/2}$$

and note that

$$k/2 - w \le m_0 \le k/2 + w \;\Rightarrow\; \frac{-2w}{\sqrt{k}} \le z \le \frac{2w}{\sqrt{k}}.$$

If we choose $w=\sqrt{k}$ the probability needed is the probability a zero mean unit variance normal variate falls in the interval $(-2,2)$ which is approximately .95. In this case the probability the windowed matrix achieves full rank is upper bounded by $.95\times .288788\cdots$. In general we have:

Theorem 5.1. *For sufficiently large k, the probability that a $k \times k$ random, windowed binary matrix with window length $w = \delta\sqrt{k}/2$ has rank k is at most $2\Phi(\delta)Q_0$, where $\Phi(z)$ is the normal distribution function and $Q_0 = .288788\cdots$ as given by equation 3.4.*

In the experiments described in the next section it is observed that in fact we need a slightly larger window - it appears that a matrix with a window size of $2\sqrt{k}$ has a rank behaviour indistinguishable from that of a (unwindowed) random matrix.

These matrices will be used in section 7 to construct a class of codes that are particularly efficient in terms of coding and decoding complexity.

5.2. **Random binary matrices with fixed weight columns.** Calkin [8] posed the following problem: let $S_{n,k} \subset {\mathbb{F}_2}^n$ denote the set of binary n-tuples of weight k (Note the dramatic change of terminology here - we made no attempt to reconcile it with the previous work since it is a very different problem). How many such n-tuples must be chosen uniformly (with replacement) from $S_{n,k}$ to obtain a dependent set (over $\mathbb{F}_2$) with probability 1 (i.e. almost surely)? He notes that for $k = 1$ this is just the birthday surprise problem. The case of $k = 2$ a dependent set corresponds to a cycle in a graph on n vertices which relates the work to the theory of random graphs ([6, 16]). The proof techniques of this work are interesting and we outline the process here.

Let $k \geq 3$ be a fixed integer and denote by $p_{n,k}(m)$ the probability that the n-tuples $\mathcal{U} = \{u_1, u_2, \cdots, u_m \epsilon_R S_{n,k}\}$ chosen uniformly at random (with replacement) from $S_{n,k}$ are linearly dependent. The following two results are established:

Theorem 5.2 ([8], Theorem 1.1)**.** *For each k there is a constant β_k such that if $\beta < \beta_k$ then*

$$\lim_{n\to\infty} p_{n,k}(\beta n) = 0.$$

Furthermore, $\beta \approx 1 - e^{-k}/\log 2$ as $k \to \infty$.

Denote by r the rank of the set of m binary n-tuples of weight $k, \mathcal{U}$ and by $s = m - r$ its nullity.

Theorem 5.3 ([8], Theorem 1.2)**.** *(a) If $\beta < \beta_k$ and $m = m(n) < \beta n$ then $E(2^s) \to \infty$. (b) If $\beta > \beta_k$ and $m = m(n) > \beta n$ then $E(2^s) \to \infty$ as $n \to \infty$.*

The theorems indicate that if fewer than $\beta_k n$ columns are chosen then with high probability they will be full rank and, asympotically, if more than $\beta_k n$ columns are chosen the null space becomes arbitrarily large (as $n \to \infty$).

To compare these results with those of Blömer et al [5] and Cooper [9, 10] we assume that $c(n)$ is a function that is unbounded as $n \to \infty$ and tht $p = (\log(n) - \mathcal{O}(1))/n$, the probability of a 1 in a random $n \times n$ matrix. Then the probability that fewer than $n - c(n)$ of the columns of this matrix are linearly independent tends to 0 as $n \to \infty$.

As mentioned, the proof technique of [8] is of interest. For a set of vectors $u_1, u_2, \cdots$ define the set of vectors

$$x_0 = 0 \in {\mathbb{F}_2}^n,\ x_i = x_{i-1} + u_i\ ,\ i = 1, 2, \cdots .$$

Associate with the above sequence the Markov process $y_i = \omega(x_i)$ defined on the states $\{0, 1, 2, \cdots, n\}$ i.e. at each instant the state is the Hamming weight of the binary n-tuple x_i. The transition probabilities $A = (a_{p,q})$ between the states p and q, are easily calculated as

$$a_{p,q} = \binom{q}{\frac{k-p+q}{2}}\binom{n-q}{\frac{k+p-q}{2}}/\binom{n}{k}.$$

where the binomial coefficients are to be taken as 0 if $k + p + q$ is odd. Furthermore the eigenvalues λ_i and eigenvectors e_i of the transition matrix A may all be explicitly calculated as

$$\lambda_i = \sum_{t=0}^{k}(-1)^t\binom{i}{t}\binom{n-i}{k-t}/\binom{n}{k},\ e_i[j] = \sum_{t=0}^{j}(-1)^t\binom{i}{t}\binom{n-i}{j-t}.$$

The expressions here are strongly related to the well known Krawtchouk polynomials of coding theory. Interestingly the eigenvectors are independent of k. It is shown that if U is the matrix whose columns are the $n + 1$ eigenvectors, and Λ the diagonal matrix of eigenvalues, then

$$U^2 = 2^n I \text{ and } A = \frac{1}{2^n}U\Lambda U.$$

Thus the eigenvectors are linearly independent. The eigenvalues are also shown to have certain properties e.g. $|\lambda_i| < 1$ and for $i > n/2$, $\lambda_i = (-1)^k\lambda_{n-i}$.

To relate these eigenvalue properties to the problem of rank of the set of vectors, note that the 00-th entry of the t-step transition matrix A^t is the probability that the sum of the t vectors $u_1, u_2, \cdots, u^t$ is

$$\sum_{i=0}^{n}\frac{1}{2^n}\lambda_i^t\binom{n}{i},\ U = 2^nU^{-1},\ A = U\Lambda U2^n$$

Thus the number of linear dependencies among the vectors $u_1, \cdots, u_m$ is just

$$E(2^s) = \sum_{t=0}^{m}\binom{m}{t}\sum_{i=0}^{n}\frac{1}{2^n}\lambda_i^t\binom{n}{i} = \sum_{i=0}^{n}\frac{1}{2^n}\binom{n}{i}(1+\lambda_i)^m.$$

The previous two theorems are then proven by obtaining sufficiently good approximations to the eigenvalues in the above relation. Notice that the threshold $\beta_k = 1 - e^{-k}/\log 2 \approx 1 - 1.446e^{-k}$ and hence is less than 1 (as expected).

5.3. **Algorithmic construction of binary matrices with specified properties.** The remaining problems considered in this section are not strictly random matrix problems. The first considers algorithms for the construction of matrices with m rows over $\mathbb{F}_q$ with the maximum number of columns possible, each column of weight at most r with the property that any k columns are linearly independent over $\mathbb{F}_q$ (we make no attempt to resolve the terminolgy here with presious usage). The maximum number of such columns is denoted $N_q(m, k, r)$ (in conflict with an earlier use) and this quantity has been considered in [26, 18, 19]. The connection to coding theory is immediate since such a set of columns as a parity check matrix, yields a code of length $N_q(m, k, r)$ of dimension $N_q - m$ and minimum distance $d = k+1$. However it also has strong connections to random graph theory and the presence

of cycles of length $k+1$. The subject is already too large to deal with here. We mention only some of the results. For $q=2$ [18] gives a probabilistic lower bound of

$$N_2(m,k,r)=\Omega(m^{kr/2(k-1}).$$

For $q=2$ the deterministic algorithm results are given [18, 19] as follows. Since it is known that $N_2(m,2k+1,r)\geq(1/2)N_2(m,2k,r)$ it is sufficient, asymptotically, to consider only even dependencies. When $r=2$ the ties to graphs with no cycles of length at most k, and hence construction of graphs with large girth, is immediate. For $r\geq 1,\ k\geq 4$ even, we have [18]

$$N_2(m,k,r)=\Omega(m^{kr/(2(k-1))}) \text{ and for } \ k=2^i,\ N_2(m,k,r)=\mathcal{O}(m^{\lceil kr/(k-1)\rceil/2}).$$

When $k=2^i$ and $\gcd(k-1,t)=k-1$ the lower and upper bounds match. When $\gcd(k-1,r)=1$ the lower bound was improved ([3]) to

$$N_2(m,k,r)=\Omega(m^{kr/(2(k-1))}\cdot(\ln(m))^{1/(k-1)}).$$

For an arbitrary finite field we have [19] the lower bounds

$$N_q(m,k,r)=\begin{cases}\Omega(m^{(kr/2k(k-1))}), & \text{for } k \text{ even},\\ \Omega(m^{(k-1)r/(2(k-2))}), & \text{for odd } k\geq 3,\\ \Theta(m^{kr/(2(k-1))}), & \gcd(k-1,r)=k-1,\\ \Omega(m^{kr/(2(k-1))}\cdot(\log(m))^{1/(k-1)}), & \gcd(k-1,r)=1,\ k\geq 4, k \text{ even}.\end{cases}$$

In addition there are polytime algorithms that can achieve the lower bounds.

In the case that $k>r$ we have the further bound [26] that

$$N_q(m,k,r)=\mathcal{O}\big(m^{\frac{r}{2}+\frac{4r}{3k}}\big).$$

5.4. **Eigenvalues of random matrices.** Other work ([23, 24]) considers eigenvalues of "random" matrices. The matrices are not really random - the randomness comes from choosing a property and considering the probability of obtaining that in a randomly chosen matrix from a class of matrices. To describe this work, let $M_n(q)$ denote the space of $n\times n$ matrices over $\mathbb{F}_q$ and $GL_n(q)$ the group of invertible matrices over $\mathbb{F}_q$. The work of [24] describes some eigenvalue problems for certain types of matrices using the cycle index for these matrices. We glean only a few of the many deep results from that work.

For $\alpha\in GL_n(q)$ and $a\in\mathbb{F}_q$ denote by X_a the random variable that is the dimension of the a-eigenspace of α i.e. $X_a(\alpha)=\dim\ker(\alpha-aI)$. Thus X_a counts multiplicities and we emphasize that interest is limited to $a\in\mathbb{F}_q$ (and not an extension field). It is noted that the most likely multiplicity is 1. The randomness stems only from the random choice of $\alpha\in M_n(q)$. Then we have:

Theorem 5.4 (Theorem 12, [24]). *For $k\geq 1$ as $n\to\infty$*

$$P_n(X_a=k)=\frac{q^k}{(q-1)^2\cdots(q^k-1)^2}\prod_{r\geq 1}\big(1-\frac{1}{q^r}\big).$$

For $k=0$

$$P_n(X_a=0)=\prod_{r\geq 1}\big(1-\frac{1}{q^r}\big).$$

Note that the asymptotic (as $n \to \infty$) probability that $a \in \mathbb{F}_q$ is *not* an eigenvalue is $\prod_{r\geq 1}(1-1/q^r)$.

Lemma 5.5 (Proposition 15, [24]). *The asymptotic (as $n \to \infty$) expected number of eigenvalues (in $\mathbb{F}_q$, with multiplicity) of a matrix over $\mathbb{F}_q$ is*

$$q\prod_{r\geq 1}(1-\frac{1}{q^r})\sum_{k\geq 1}\frac{kq^k}{(q-1)^2\cdots(q^k-1)^2}.$$

If we define $\rho_k = P(X_a = k)$, then this gives a discrete probability distribution on the positive integers. We have the following:

Theorem 5.6 ([24], Theorem 16). *The asymptotic (as $n \to \infty$) probability that a matrix over $\mathbb{F}_q$ has k eigenvalues (counting multiplicity) in $\mathbb{F}_q$ is given by the coefficient of t^k in the power series*

$$(\sum_{k=0}^{\infty}\rho_k t^k)^q.$$

i.e.

$$\sum_{k_1+k_2+\cdots+k_q=k}\rho_{k_1}\cdots\rho_{k_q}.$$

Notice from this that the probability of no eigenvalue in $\mathbb{F}_q$ is the constant term i.e.

$$\rho_0^q = \prod_{r\geq 1}(1-\frac{1}{q^r})^q$$

and the probability of exactly one eigenvalue isin $\mathbb{F}_q$ is the coefficient of t in the expansion i.e.

$$q\rho_0^{q-1}\rho_1 = \frac{1}{(1-\frac{1}{q})^2}\prod_{r\geq 1}(1-\frac{1}{q^r})^q.$$

It is also shown [24] that as $q \to \infty$ that the distribution of $X = \sum_{a\mathbb{F}_q} X_a$, the number of eigenvalues in the base field, approaches a Poisson distribution with a mean of 1. Define a *linear derangement* as an invertible linear map that has no nonzero fixed vectors (any such map always fixes the zero vector). Then it can be shown that the asymptotic probability that an invertible linear map is a linear derangement is

$$\prod_{r\geq 1}(1-\frac{1}{q^r})$$

i.e. the probability that no vectors are fixed is the probability that $1 \in \mathbb{F}_q$ is not an eigenvalue.

Consider $\alpha \in GL_n(q)$ and ${\mathbb{F}_q}^n$ as the projective space $\mathbf{P}^{n-1}(\mathbb{F}_q)$ and define a *projective derangement* in the natural manner. Then we can show that the asymptotic probability that $\alpha \in GL_n(q)$ is a projective derangement is

$$\prod_{r\geq 1}(1-\frac{1}{q^r})^{q-1}.$$

Notice that the limit of this expression as $q \to \infty$ is $1/e$, a familiar expression for ordinary permutations.

Consider the situation of eigenvalues in extension fields of the ground field $\mathbb{F}_q$. It is shown in [24] that the probability a given $a \in \mathbb{F}_{q^m}$ is not an eigenvalue of a square matrix (over $\mathbb{F}_q$) is

$$\prod_{r \geq 1} (1 - \frac{1}{q^{rm}}).$$

One can also show that the probability that a given monic irreducible polynomial of degree m over $\mathbb{F}_q$ is a factor of the characteristic polynomial of a square matrix is

$$1 - \prod_{r \geq 1} (1 - \frac{1}{q^{rm}}).$$

The paper [24] contains many more results of a similar nature.

The same author [23] considers other problems on the enumeration of certain matrices, namely those over $\mathbb{F}_2$ with no eigenvalues of 0 or 1. Notice that this is equivalent to matrices that define a projective derangement (no fixed points). Interestingly, if the number of such matrices over $\mathbb{F}_2$ is e_n then the generating function is given by

$$1 + \sum_{n \geq 1} \frac{e_n}{\gamma_n} u^n = \frac{1}{1-u} \prod_{r \geq 1} (1 - \frac{1}{q^r})$$

where γ_n is the number $\mid GL_n(q) \mid = \prod_{i=0}^{n-1} (q^n - q^i)$. More generally he shows that if d_n is the number of $n \times n$ matrices over $\mathbb{F}_q$ with no eigenvalues in $\mathbb{F}_q$ then

$$1 + \sum_{n \geq 1} \frac{d_n}{\gamma_n} u^n = \frac{1}{1-u} \prod_{r \geq 1} (1 - \frac{u}{q^r})^{q-1}.$$

5.5. **Group theoretic aspects of random matrices.** We mention two final papers of interest. The work of Fulman [12] investigates a variety of problems of random matrices over finite fields but is more concerned with group theoretic and reduction type problems for $GL_n(q)$. The type of questions of interest there include the number of Jordan blocks in the rational canonical form of a random matrix, the distribution of the order of a random matrix, the probability the characteristic polynomial of a random matrix has no repeated factors and the like. It makes heavy use of the cycle index technique.

Finally we note the work of Brent et al [7]. The item of concern in this work is the action of a matrix on a subspace. To introduce this notion, let T Denote a linear mapping on $V_n(q)$ and let $S \subseteq V_n(q)$ be a set of m vectors. The Krylov subspace generated by S under T is defined as:

$$\mathrm{Kry}(S,T) = \{ \sum_{i=1}^{m} f_i(T) v_i \mid f_i(x) \in \mathbb{F}_q[x],\ v_i \in S,\ 1 \leq i \leq m \}.$$

Thus Kry(T,S) is just the space formed by all powers of the matrix T acting on the subset of vectors of S. Define

$$\kappa_m(T) = \frac{1}{q^{mn}} \cdot \| \{ (v_1, v_2, \cdots, v_m) \in V_n(q)^m,\ \mathrm{Kry}(T, (v_1, \cdots v_m)) = V_n(q) \} \|.$$

The paper determines lower bounds on $\kappa_m(T)$ using the Frobenius index of T, the number of invariants in the Frobenius decomposition of $V_n(q)$ under T. While they are concerned with the this particular question, several interesting and useful bounds for the questions of interest in this note are also developed and some of these have been noted.

6. Experimental results and conjectures

This section reports on experimental results generated and resulting conjectures, based on the theory presented in the previous sections. We restrict attention here to binary matrices and suppose unit elements are chosen with probability p. We first modify the expression for a $k \times (k+m)$ matrix to be of full rank and have no all zero rows or columns. Recall the expression for an $k \times (k+m)$ matrix is of full rank for $m \leq 0$ is given by

$$\prod_{i=1}^{k+m} (1 - \frac{1}{2^{k-i+1}})$$

and when $m \geq 0$ this is approximated by

$$Q_m = \prod_{i=m+1}^{\infty} (1 - \frac{1}{2^i}),$$

where the last expression is an asymptotic result as $k \to \infty$.

For relatively large p all zero rows or columns are unlikely. As p decreases however, they become a significant factor.

We first enumerate full rank matrices on the number of zero columns - the probability a column is all zeros is $(1-p)^k$. Hence the probability a $k \times (k+m)$ matrix is of full rank is

$$(1 - (1-p)^{k+m})^k \sum_{j=0}^{m} \binom{k+m}{j} (1-p)^{kj} (1 - (1-p)^k)^{k+m-j} Q_{m-j} \ , \ m > 0 \tag{6.1}$$

where the first term is the fraction of all $k \times (k+m)$ matrices that have no all-zero rows, the term after the binomial coefficient is the probability of exactly j particular all-zero columns, $(1-(1-p)^k)^{k+m-j}$ is the fraction of $k \times k+m-j$ matrices with no all-zero columns and Q_{m-j} is the fraction of all $k \times (k+m-j)$ matrices that are nonsingular. The expression is an approximation since it assumes the probabilities of all-zero rows and columns are independent. In addition, the expression for Q_{m-j} does not exclude the possibility of zero columns - the above argument essentially argues that the probability of full rank given no all-zero columns is the same as probability of full rank. Nonetheless this expression has been shown to be remarkably accurate for values of p well below the $\ln)(k)/k$ threshold and we note it here. We need further work to justify this approximation. For values of p near $1/2$ the expression is very near the previous expression Equation 3.4. Equation 6.1 is a much more accurate expression, as p decreases, although it requires further theoretical justification.

We report rather extensively on the experimental work done, in the hope that it will prove useful in suggesting conjectures and suggest further problems.

To begin, we confirm various results noted in Section 3, perhaps especially Theorem 3.2. The following graphs consider the rank of (random binary) matrices where each element is chosen at random and independently with probability p, as shown.

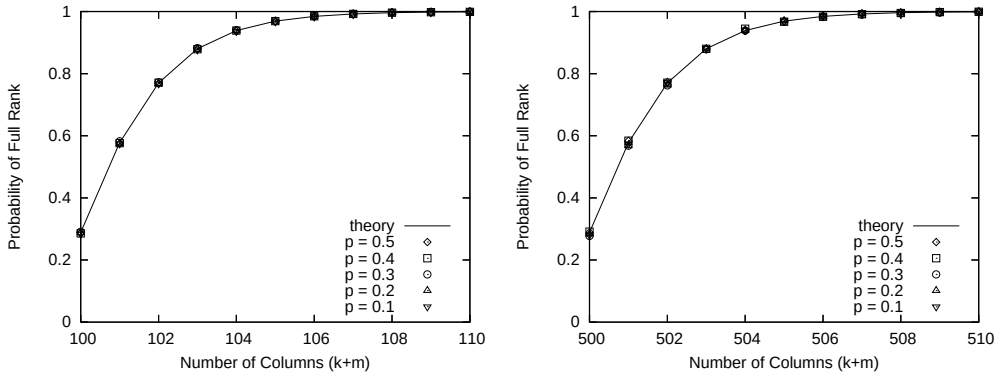


FIGURE 1. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 500$. In both graphs, the theoretical expression of Equation 3.4 is shown versus experimental results for the probability of a nonzero element $p = 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5$. Each point of the experimental results is for $50,000$ trials for $k = 100$ and $5,000$ trials for $k = 500$.

The agreement of experiment with theory, for p sufficiently large, in these graphs is quite surprising. Further experiments for k down to 10 showed similar agreement. Note that all values of p in these curves exceed the threshold value of $\ln(k)/k$.

It was noted in Section 3 that a critical value for the probability p for choosing the nonzero elements in a random matrix was $p = \ln k/k$ for a matrix with k rows. The following experimental results attempt to justify this, showing the behavior of the probability of full rank as p varies around $\log(k)/k$. Additionally we show that the expression for Q_m corrected for the probability of zero rows and columns (Equation 6.1)is accurate well below this value, although it becomes increasingly difficult to verify this experimentally.

Based on these experimental results and the expressions of the previous sections it can be justified that the rank properties of a random matrix over $\mathbb{F}_2$ where the matrix elements are chosen to be 1, independently and identically with probability p where $2\ln(k)/k < p < 1 - 2\ln(k)/k$, are indistinguishable from the purely random case where $p = 1/2$. For example the expression equation (3.6) can be used to show this [1]

Note that from the theorems of Section 3 the critical value of probability is $p = \ln(k)/k$ and for a value of p slightly larger, asymptotically as $k,\ m$ increase, the rank tends to full. size of p needed to behave as a "purely random" ($p = 1/2$) matrix is . Based on the work of Kolchin and Cooper, the correct lower bound for the value of p in order for the rank poprties of the matrix to be similar for the random ($p = 1/2$ case) is likely of the form $p = (\ln(k) + d(k))/k$ for a function $d(k)$ decreasing to 0 sufficiently slowly with k.

[1]The authors are grateful to Omran Ahmadi for pointing this out.

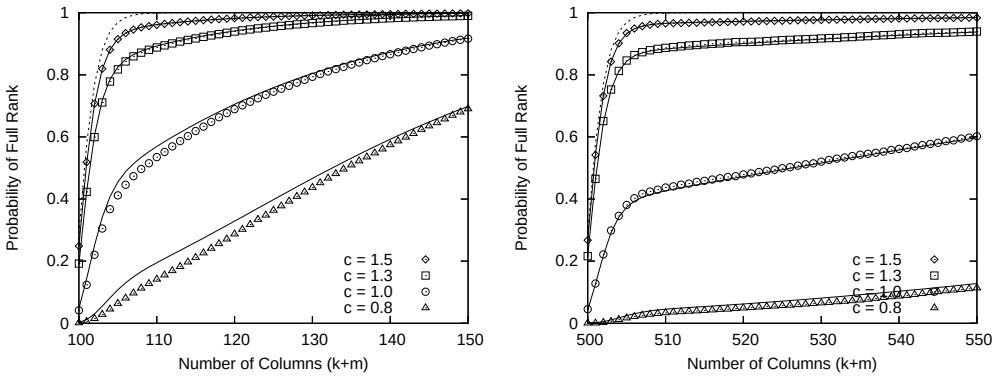


FIGURE 2. Probability of full rank for $p = c\ln(k)/k$ as the number of extra columns increases. Each data point is the result of $50,000$ trials (for $k = 100$) and $5,000$ trials (for $k = 500$). The solid line is computed using Equation 6.1 and the dotted line (purely random matrix with constant $p > 2\ln(k)/k$) for Equation 3.4.

It has been observed the the average number of columns, beyond k, for a purely random $k \times (k+m)$ to achieve full rank with high probability, is $1.60669515\cdots$. Based on the above observations, we suggest the following:

CONJECTURE 6.1. *If the elements of a random binary $k \times (k+m)$ matrix $M_{k,(k+m)}$ are chosen independently and identically at random to be 1 with probability p, then the expected number of columns beyond k required to achieve full rank (k), $\bar{m}$, is, asymptotically as $k \to \infty$*

i) $\bar{m} = \infty$ *for* $p < \ln(k)/k$
ii) $\bar{m} = 1.60669515\cdots$ *(as in section 3) for* $2\ln(k)/k < p < 1 - 2\ln(k)/k$

The behavior of $\bar{m}$ for $p = c\ln(k)/k$ for $1 < c < 2$ would be of interest.

So far we have only considered generating columns for the random matrices by choosing the individual column elements at random, identically and independently distributed. It is interesting to consider other mechanisms to generate the columns and we will have a use for one such mechanism when considering applications for the material to coding.

For the mechanism to be considered, assume we have a "degree distribution" available, $p(d),\ d = 1, 2, \cdots, k$ where $p(d)$ is the probability of choosing a degree of d. The terminology derives from coding theory where the binary matrix is viewed as a bipartite graph. The process will be, for each column to be added to the random matrix, d is chosen according to the distribution, and the column is formed by choosing a random d-tuple of integers from 1 to k to place the ones. For this we use the terms weight and degree interchangeably.

Several distributions are used.

Wedge distribution: To achieve a mean column weight of σ, which we assume is not of the form of an integer plus 0.5, we choose degrees, $d-1,\ d,\ d+1$ where d is σ rounded to the nearest integer. The probability assigned to d is 0.5 and the sum of probabilities for $d-1,\ d+1$ is 0.5, with mean degree σ, resulting in a unique distribution.

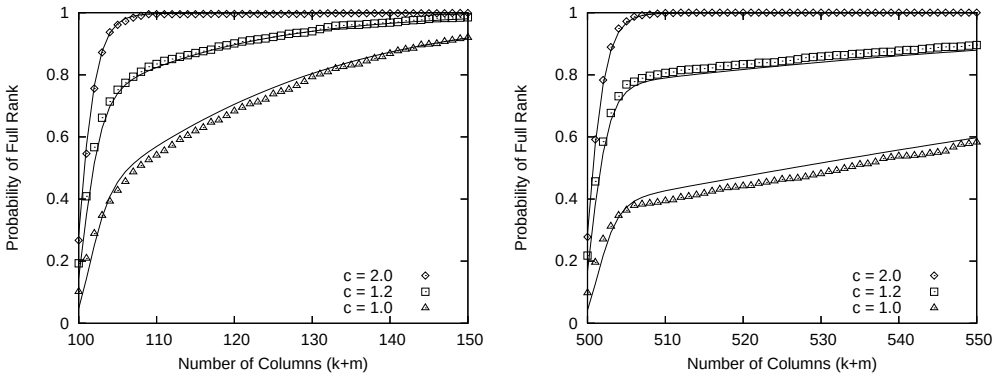


FIGURE 3. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 500$. The graphs use $p = c\ln(k)/k$ in Equation 6.1 and the data points are derived from the wedge distribution with the same mean.

Uniform distribution: To achieve a mean degree (weight) of σ we assign probabilities of approximately $1/2\sigma$ to degrees $1, 2, \cdots$, to approximately 2σ, adjusting the probability of the last degree to achieve the correct mean.

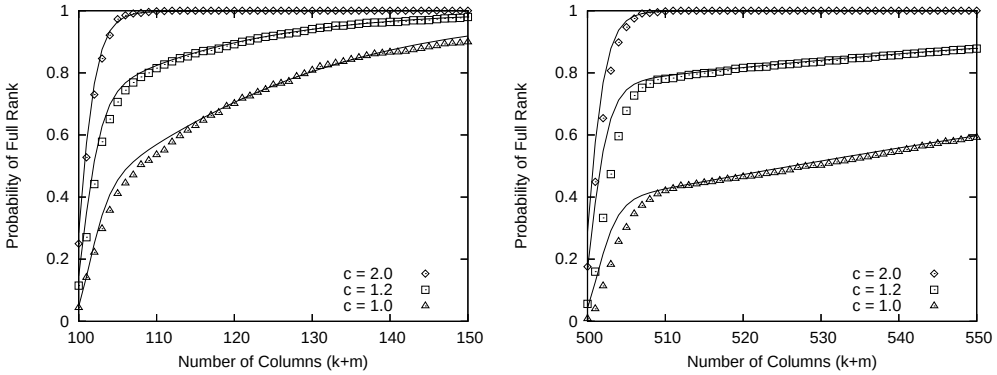


FIGURE 4. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 500$. The graphs use $p = c\ln(k)/k$ in Equation 6.1 and the data points are derived from the uniform distribution with the same mean.

Horshoe distribution: Half of the columns have low degree (2 or 3) and half have high degree (approximately 2σ). We used two variants: in the degree 2 variant, we used half of the columns of weight 2 and the other half of weight d and $d+2$, d odd, choosing the probabilities to give the correct mean. In the degree 3 variant we choose half of the columns of weight 3 and the other half of degrees d and $d+1$ (note: a matrix with all columns of even weight cannot have full rank and these considerations are to avoid this possibility).

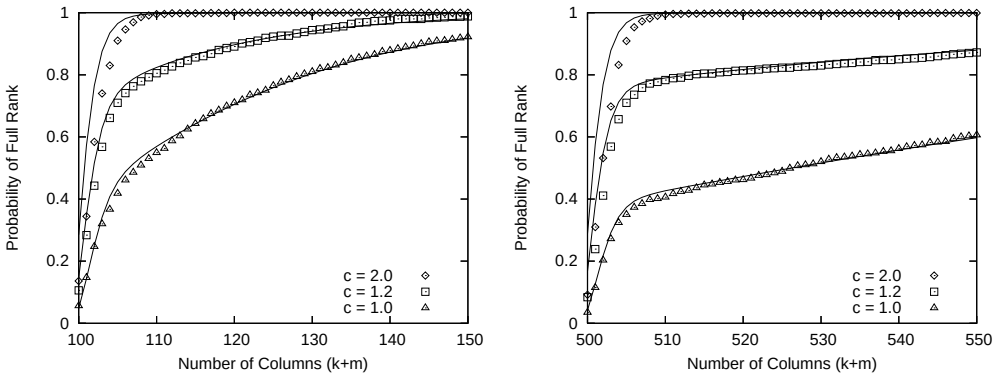


FIGURE 5. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 500$. The graphs use $p = c\ln(k)/k$ in Equation 6.1 and the data points are derived from the horshoe distribution (variant 2) with the same mean.

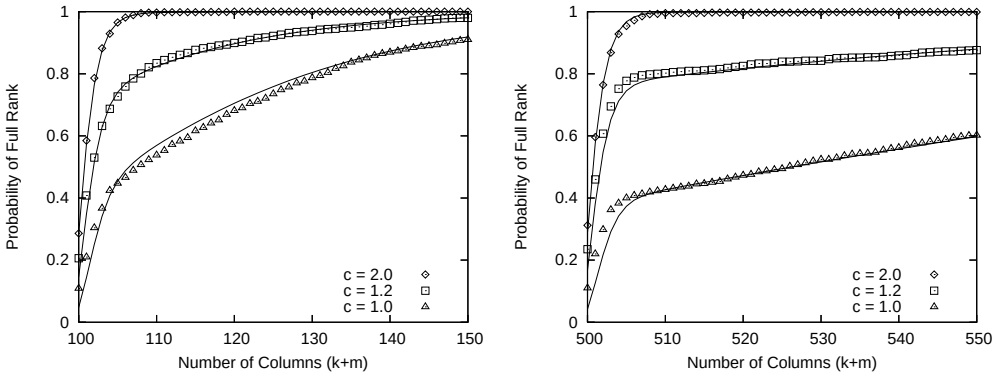


FIGURE 6. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 500$. The graphs use $p = c\ln(k)/k$ in Equation 6.1 and the data points are derived from the horshoe distribution (variant 3) with the same mean.

Soliton distribution: This distribution will be of interest in the coding section (Section 7). The distribution is given by

$$p(d) = \begin{cases} 1/k & i = 1 \\ 1/i(i-1) & 2 \le i \le k. \end{cases}$$

It is seen there is very little difference in the results of the wedge, uniform and horshoe distributions for the same mean column weight. As far as rank properties of the random matrices are concerned, as long as the mean column weights are the same, one can either choose the column elements at random with probability p or choose the column weight d from the distribution and then a random d-tuple from $[1, k]$.

However for the soliton distribution the results are significantly different (see Graph 7). No explanation was found for this.

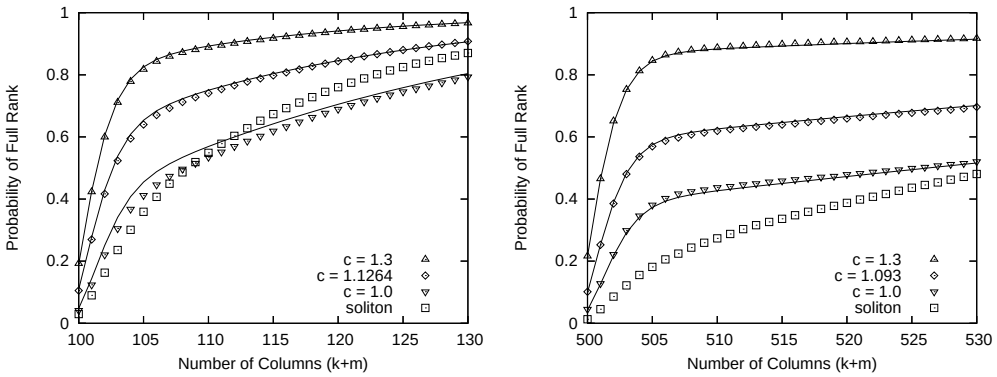


FIGURE 7. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 500$. The graphs use $p = c\ln(k)/k$ in Equation 6.1 and the data points are derived from the soliton distribution with similar mean. Note for $k = 100$ the mean of the soliton distribution is approximately $1.1264\ln(k)$ and for $k = 500$ it is approximately $1.093\ln(k)/k$.

Experiments for the windowed matrices described in Section 5.1 are considered.

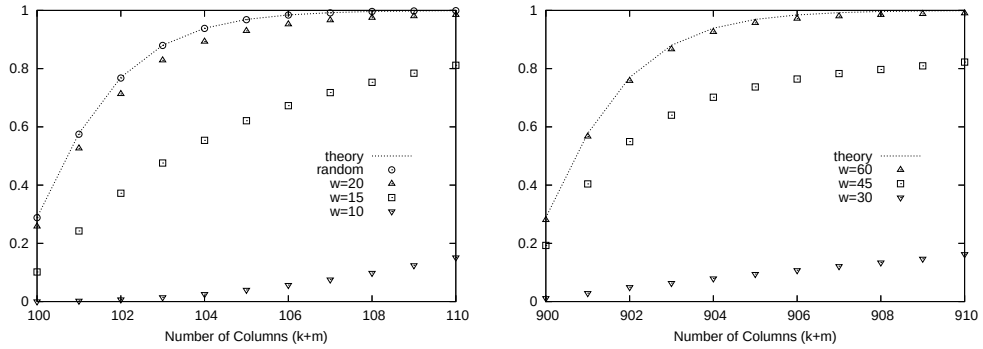


FIGURE 8. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 900$. In both graphs, the dotted line is Q_m (Equation 3.4), the open circles show results from random ($p = 1/2$) matrices, and the other symbols show results for windowed matrices (with the specified window size). For clarity, no open circles are shown on the right graph.

The curves of Graph 8 (and subsequent graphs) suggest the lower bound of $2\sqrt{k}$ for full rank falls slightly short for lower values of k (≤ 100) but appears quite accurate for larger values, in terms of obtaining rank behaviour of the windowed matrices equivalent to those of a random matrix.

The Graph 10 is similar to previous curves with slightly extended horizontal range.

The conclusions of this experimental evidence suggests that windowed matrices with window size at least $w \geq 2\sqrt{k}$ with mean column weight higher then $2\ln(k)$ give probabilities of full rank as those of purely random matrices, but that otherwise the effect of window size beyond $2\sqrt{k}$ has little effect.

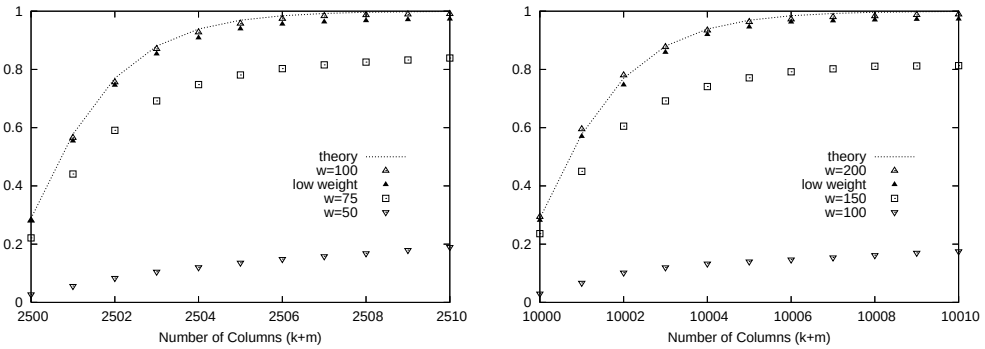


FIGURE 9. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 2500$, while the right is for $k = 10000$. In both graphs, the broken line is Q_m (Equation 3.4), the open triangles show results for random ($p = 1/2$) matrices (window size $2\sqrt{k}$, 100 and 200 respectively), and the other symbols show results for windowed matrices (with the specified window size). The *low weight* closed triangles are the results for a window size of $w = 2\sqrt{k}$ but with the mean column weight fixed at $2\log k$.

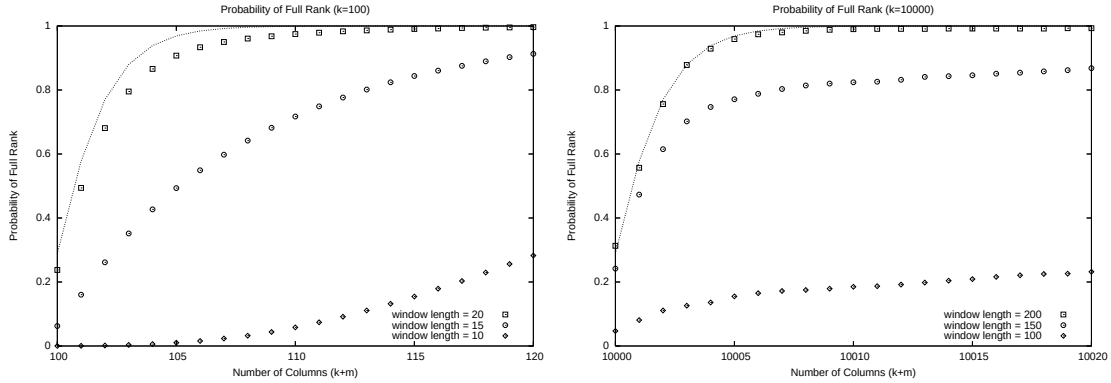


FIGURE 10. Probability of rank k for a $k \times (k+m)$ matrix. The left graph shows the results for $k = 100$, while the right is for $k = 10000$. In both graphs, the broken line is Q_m, and the other symbols show results for windowed matrices (with the specified window size), for $p = 1/2$ within a window.

The remaining graphs are given without caption and are self explanatory, where the term*forced start* refers to the first (topmost element) in the window is forced to be 1. Curves where the probability of a nonzero element is chosen to make the mean column weight $2\ln(k)$ are so marked. All other curves use $p = 1/2$. The dotted curve in each graph is Equation 3.4.

CONJECTURE 6.2. *The rank properties of a binary windowed $k \times (k+m)$ random matrix, as discussed in section 5.1, behave as a random binary matrix iff the window length is at least $2\sqrt{k}$ as long as the probability of an element being 1 is at least p, where p is chosen to give a mean column weight of $2\ln(k)$.*

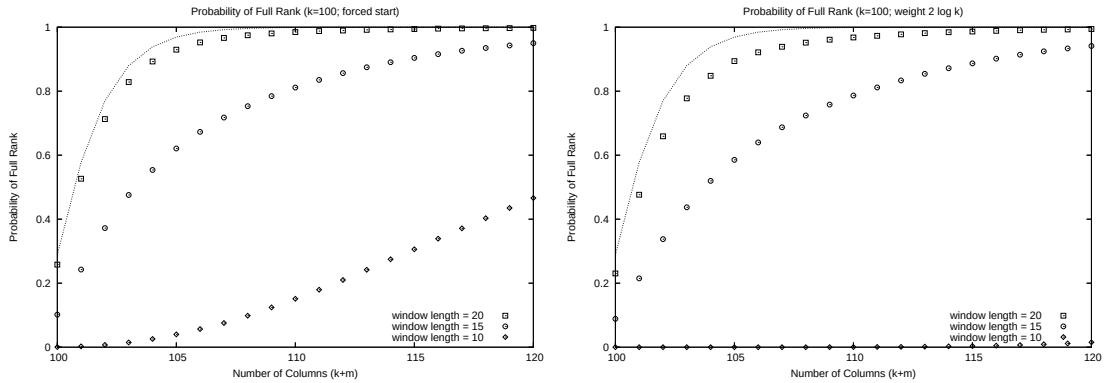


FIGURE 11. Probability of rank k for a $k \times (k+m)$ matrix for $k = 100$. In both graphs, the broken line is Q_m: the left graph shows results for random ($p = 1/2$) (windowed, forced start) matrices, and the other show results for low weight windowed matrices.Note low value of k here.

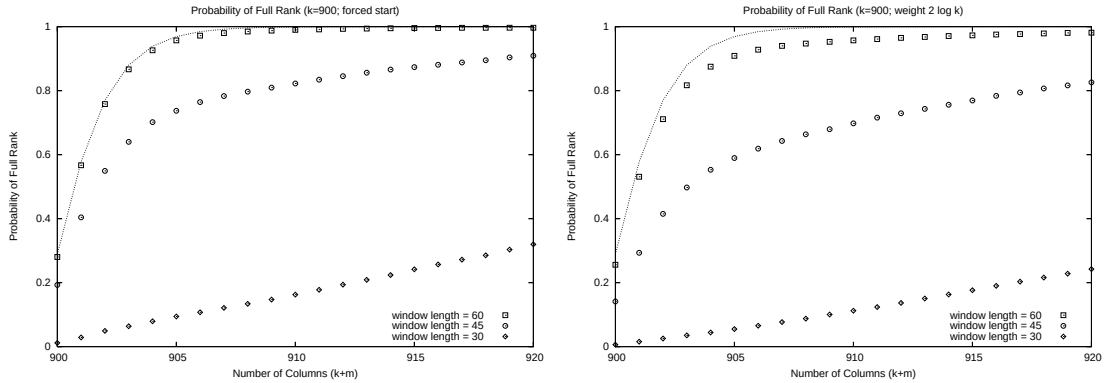


FIGURE 12. Probability of rank k for a $k \times (k+m)$ matrix. Similar to previous graphs only for $k = 900$. In both graphs, the broken line is Q_m. The left graph shows results for random ($p = 1/2$) and forced start matrices, and the right for low weight.

We conclude the section with some experimental results associated with the work of Calkin [8] on random matrices with constant weight columns and to verify the behaviour of the threshold of β_k mentioned there.

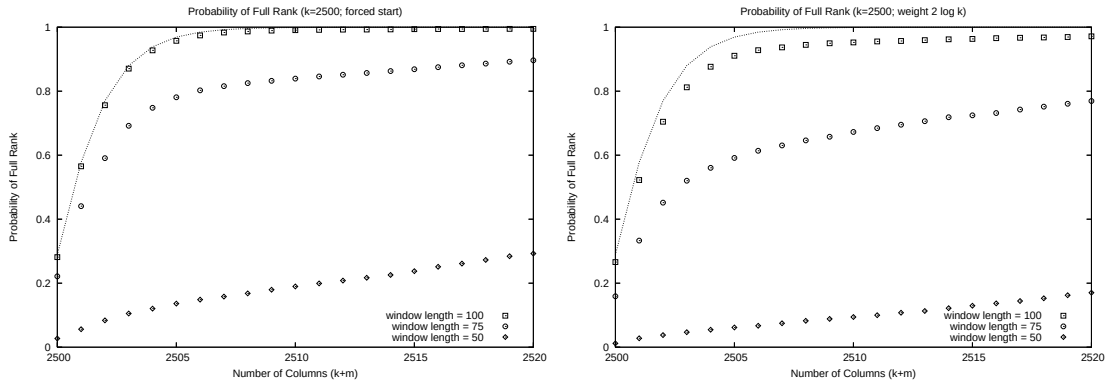


FIGURE 13. Probability of rank k for a $k \times (k+m)$ matrix. Similar to previous graphs only for $k = 2500$. In both graphs, the broken line is Q_m. The left graph shows results for random ($p = 1/2$) and forced start matrices, and the right for low weight.

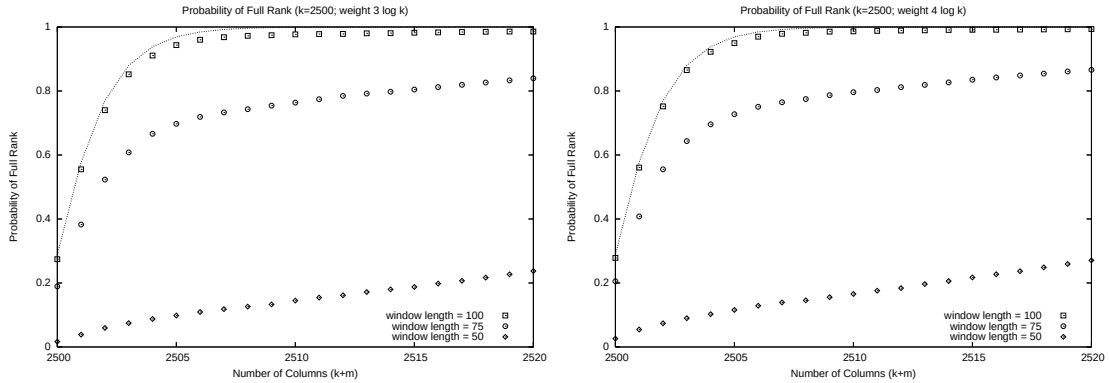


FIGURE 14. Probability of rank k for a $k \times (k+m)$ matrix. Similar to previous graphs with $k = 2500$ and higher mean column weight. In both graphs, the broken line is Q_m.

k	$n = 500$	$n = 1000$	$n = 2000$	$theory3$	$theory2$
2	$0.464 < b < 0.466$	$0.459 < b < 0.460$	$0.4563 < b < 0.4577$		
3	$0.914 < b < 0.916$	$0.916 < b < 0.917$	$0.9170 < b < 0.9175$	0.90912	0.92817
4	$0.972 < b < 0.974$	$0.974 < b < 0.975$	$0.9755 < b < 0.9760$	0.96909	0.97358
5	$0.990 < b < 0.992$	$0.991 < b < 0.992$	$0.9920 < b < 0.9925$	0.98935	0.99028
6	$0.992 < b < 0.994$	$0.995 < b < 0.996$	$0.9960 < b < 0.9965$	0.99625	0.99642
7	$0.996 < b < 0.998$	$0.998 < b < 0.999$	$0.9985 < b < 0.9990$	0.99865	0.99868
8			$0.9985 < b < 0.9990$	0.99951	0.99952
9			$0.9995 < b < 0.9999$	0.99982	0.99982

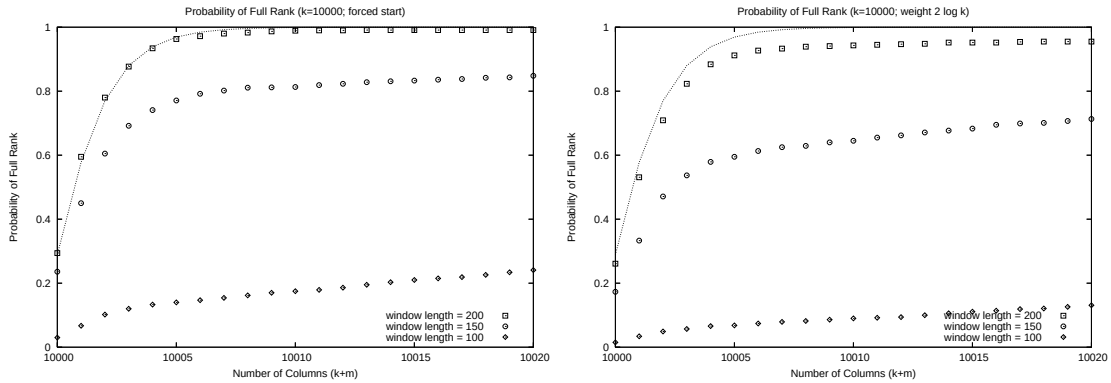


FIGURE 15. Probability of rank k for a $k \times (k+m)$ matrix. Similar to previous graphs with $k = 10000$ and mean column weight as indicated. In both graphs, the broken line is Q_m.

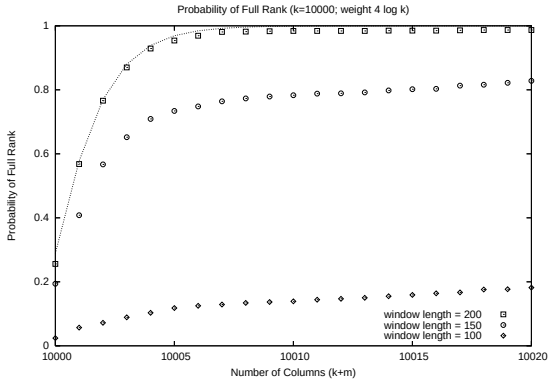


FIGURE 16. Probability of rank k for a $k \times (k+m)$ matrix. Similar to previous graphs with $k = 10000$ and higher mean column weight as indicated. The broken line is Q_m.

In this table, the column labelled 'theory3' is the 3 term expression for β_k mid way down on page 270 of [8] and 'theory2' is the approximation $\beta_k \sim 1 - exp(-k)/log(2)$. The ranges of values shown in columns 2 to 5 were extracted from the graph 17 below. It proved a very difficult problem to assess these values with accuracy.

For the probabilty of independence (full rank) in the square case:

k	n=500	n=1000	n=2000
3	0	0	0
5	0.007694	0.00012	0
7	0.183832	0.11501	0.0460
9	0.271318	0.25741	0.2316
11			0.2836

In both cases, tested 500,000 matrices were tested for $n = 500$, and $100,000$ for $n = 1000$ and $10,000$ for $n = 2000$. From previous results it seems that for a fixed n as k exceeds $2\ln(n)$ the values tend to the value $.288\cdots$ for random matrices.

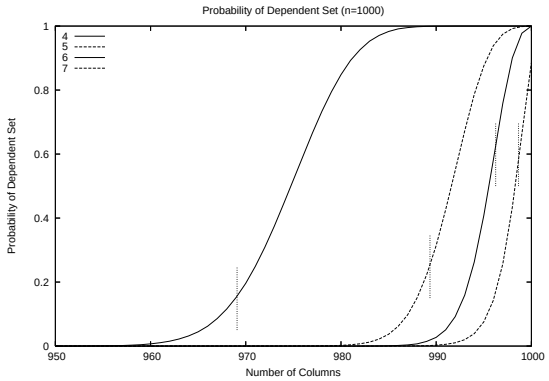


FIGURE 17. Probability of dependent set for constant weight matrices, $k = 4, 5, 6, 7$. for columns of size 1000

7. Applications to coding theory

7.1. Previous work. A very limited review of the remarkable recent work on erasure correcting codes is given with a view to take the results on windowed matrices discussed above to define a new class of codes for erasure correction. While work on low density parity check codes originates from the sixties [14], it was only more recent work that showed how such codes on the binary symmetric channel with errors (a transmitted binary 1 received as a zero or vice-versa) were capable of achieving the elusive goal of capacity. However, on the binary erasure channel a further step was taken to actually show how codes that achieve capacity can be constructed [20, 21, 22, 27]. Only this more limited case of coding for erasure channels is of interest here. Much of this work was concerned with designing (random) codes by constructing a bipartite graph whose incidence matrix is the parity check matrix of a linear code. A requirement of the construction [27] was to design two distributions on the degrees of the left and right vertices of this graph in such a way as to achieve capacity for a given decoding process. More recent work has not pursued this approach.

Subsequent to this, the work of Luby [20, 21] introduced the notion of a *rateless* code. Again, this can be thought of in terms of a bipartite graph where the left vertices are identified with, say, k data or input symbols (which we may take as binary symbols for simplicity although the extension to strings of binary symbols is trivial). The right vertices are thought of as parity checks on the input symbols or as coded symbols. These are generated as follows: for a given degree distribution on the integers $1, 2, \cdots, k$, $\rho(k)$, the distribution is sampled to give an integer d, $1 \leq d \leq k$. The corresponding code symbol is formed by choosing d input symbols at random. For decoding, if a sufficient number of coded symbols are obtained, the process starts by choosing a coded symbol of degree 1 i.e. a code symbol corresponding to a right vertex of degree 1. The value of the code symbol is

transferred to the corresponding input symbol, whose value is then transferred to all coded symbols containing it, and all the corresponding edges are removed from the graph. If a new right vertex is now of degree 1, the process continues.

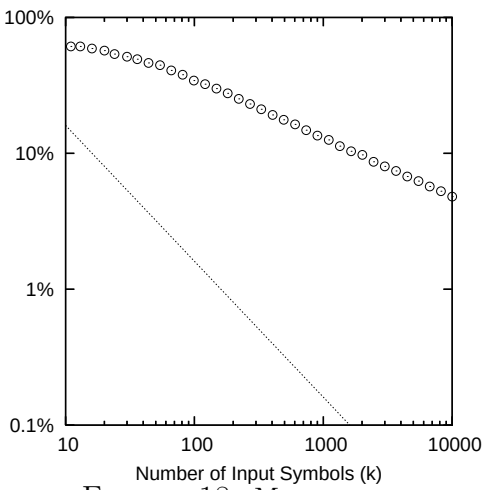
Clearly the decoding continues until it completes with all input symbols recovered or there are no right vertices of degree 1 at some stage, in which case the decoding fails. To minimize the probability of the latter, the distribution $\rho(i)$ is chosen carefully. Initially, it was suggested [21] to use the *soliton* distribution given by $\rho(i) = 1/i(i-1),\ i = 2, 3, \cdots, k,\ \rho(1) = 1/k$. The theoretical reasoning for this distribution is sound, but it turns out in practice that the probability, at each stage of the decoding, of having a vertex of degree 1 is too sensitive, giving too high a probability the decoding would not complete. To remedy the situation, a *robust* soliton distribution was suggested that proved more robust in practice. These codes are referred to as rateless since the decoder can continue to gather coded symbols from any source, on the same input data, until decoding completes – there is no fixed *a priori* rate associated with the code. In conventional coding, it would first be necessary to estimate the erasure probability to allow proper code design and this rateless feature is very useful.

For an original set of k data or input symbols, the number of extra (more than k) coded symbols required to give a high probability of decoding completion is referred to as the *overhead*. It is shown in [21] that if δ is the allowable decoding failure probability then by using the robust soliton distribution for the generation of the coded symbols, the overhead required to decode is $k + O(\sqrt{k}\ln^2(k/\delta))$ and the average degree of the right vertices is $D = \ln(k/\delta)$. Since the complexity of decoding is proportional to the number of edges in the graph, this last result shows the decoding complexity to be $O(k\ln(k/\delta))$.

To improve the situation further Shokrollahi [27] and independently Maymounkov [22] introduced the idea of *precoding*. Here one uses a good erasure correcting code to code the original input symbols resulting in a certain number of *auxiliary* symbols. These are then added to the input symbols (on the left hand side of the bipartite graph) and the combined set of left hand side symbols are coded using a truncated soliton-like encoder. This latter (outer) code has a mean right hand side degree that is independent of k, giving linear time encoding and decoding. In exchange for this complexity reduction, overhead is increased to $O(k)$ extra coded symbols. For convenience, we note Theorem 2 from [22], using our terminology. A similar result is found in [27, 11].

Theorem 7.1. *(Maymounkov): For any message of size k input symbols, and any parameter $\epsilon > 0$, there is a rateless locally encodable code (right distribution) that can recover the input symbols from any $(1+\epsilon)k$ coded symbols with high probability in time proportional to $k\log(1/\epsilon)$.*

In the presence of such a powerful result one might wonder why further work on this problem might be of interest. One aspect of the above is that they are essentially asymptotic results. In practice, with our simulations and others available in the references, a typical overhead for the



LT decoding might be as much as 15% for a number of input symbols on the order of a thousand and this drops to perhaps 3% to 4% for a much higher number of input symbols. The Online or Raptor codes can be constructed to have an overhead of 3% to 4%, of k for large k. For an idea of the improvement in overhead that is possible, see figure 18. The solid line in this figure illustrates the overhead achievable by our erasure code construction which we take as approximately 2 symbols, independent of k.

The class of codes presented here will have a decoding complexity of $O(k^{3/2})$ for k input symbols. Although significantly higher than the previous work, for low k the increase is not so large when all the constants in the estimates are taken into account. In return for this increased complexity, one obtains codes that are effective for values of k as low as 100, are easily generated requiring only a uniform distribution, and have a very high and quantifiable probability of decoding completion with an overhead that is constant, independent of k. It is likely there will be applications where such properties are of interest.

7.2. Erasure code construction. In this section we describe how one might use these windowed random matrices to construct an efficiently encodable and decodable erasure code.

7.2.1. Encoding. The encoder begins with the k data symbols (input blocks) that need to be transmitted. From these input blocks, the encoder generates, randomly and independently, columns of a windowed matrix along with output data blocks (coded symbols). There is one output block associated with each column.

Here are the steps the encoder takes to generate an output data block:

(1) Choose a starting row number (uniformly) within the range $1, \ldots, k$.
(2) For each of the w rows starting at this start row, decide whether a 1 or a 0 will be placed in that row. These two choices may be equally likely, or not (see below).
(3) For each 1 in the column generated in step 2, sum the corresponding input blocks using bitwise exclusive or.
(4) Send the column and output block (sum) to the decoder.

In this last step, sending the column need not require $O(k)$, or even $O(\sqrt{k})$ space. Instead, one might send a small seed for a pseudo-random number generator that is used by both parties to generate the column. This requires $O(\log k)$ space.

With $w = 2\sqrt{k}$, the maximum number of input blocks to be summed is w. For a densely packed window, the expected number is $w/2 = \sqrt{k}$; however, we can do better than this. In figure 11, right graph, the closed triangles give the probability of full rank for $w = 2\sqrt{k}$ but with a mean column weight of $2 \log k$. Note that these results are almost as good as the results for a densely packed window, and therefore, if one wishes to reduce the per output block encoding time to $O(\log k)$, one can modify step 2 to instead choose $2 \log k$ 1's within

the window chosen in step 1. Be careful, however, to ensure that it is not the case that every column has even weight. Such a matrix will be singular. Perhaps choose the next largest odd integer greater than $2 \log k$ and generate columns with that weight.

7.2.2. Avoiding Wrapping. In the next section, while discussing the decoding algorithm, we will notice that the decoder must treat columns that wrap from bottom to top as though they have length k. We will also see that these long length columns can increase the work to be done by the decoder by a factor of 2. Therefore, to avoid this extra work, we have developed a strategy to avoid columns that wrap; however, as we will see, this strategy may not reduce the work to be done by the decoder as much as we would like.

The strategy involves both a non-uniform distribution on the starting row selection (step 1 above) and a variable window length. The calculations required are all based on the binomial theorem and are similar to those of previous theorems. Unfortunately, we find that, to achieve a probability of full rank that is similar to what we see when wrapping is allowed, the required window length for columns with an initial 1 near the middle of the column is about $4\sqrt{k}$, and the mean window length is a little greater than $3\sqrt{k}$. This negates much of the benefit of avoiding wrapping. Regardless, we will give a complete description of the non-wrapping encoder and a detailed performance comparison in the full version of this paper.

7.3. Decoding. The decoding algorithm is simply *Gaussian Elimination*; however, to ensure our discussion of decoder complexity is precise, we will describe a specific decoding algorithm. Decoding has two phases:

(1) **Column collection.** During column collection, matrix columns, along with their associated data blocks, are inserted into a hash table and reduced as necessary to ensure they are independent.
(2) **Back filling.** At the conclusion of column collection, we will have a lower triangular matrix with 1's along the diagonal. This matrix is non-singular. With a series of data block sums, the matrix is essentially made diagonal and decoding is complete.

The hash table constructed during the first phase is to have exactly k bins. Each column received hashes to the bin whose index is the index of the first row containing a 1. When this table is full, we will have k columns, all of which start at a different row (ie. a lower triangular matrix with 1's along the diagonal). For the purposes of this algorithm, we do not consider any columns to wrap from bottom to top, and as a result, the first 1 in a column may not coincide with the start of the window.

During column collection, we will occasionally come across two columns whose initial 1's are in the same row. This is a hash collision and requires collision resolution. Such resolution is easy, simply add the two columns (and their associated data blocks) together. The resulting column is guaranteed not to have it's initial 1 in the same row. A subtle but important aspect of this algorithm is the choice of columns to keep after collision resolution. Obviously, the sum is to be kept. The other column to keep is the *shorter* of the two colliding columns. Here, the length of a column is the number of rows between the first 1 and the last 1 (inclusive). If the two columns are of equal length, either one may be kept.

It may also happen during collision resolution that the two colliding columns are identical and their sum is the all zero column. In this case, one of the two columns is simply discarded and an extra column must be collected.

When the hash table is full, back filling can begin. Back filling is done starting at the last row and working up through the matrix. First, the data block associated with the 1 on the diagonal is added to all of the data blocks associated with 1's in the last row. Then, the second to last row is processed in a similar manner. At the completion of back filling, the data blocks will be the original input blocks.

Theorem 7.2. *Worst case decoder complexity is $\bar{l}k$ data block additions, where $\bar{l}$ is the mean column length. Column length, l, as mentioned earlier in this section, is the number of rows between the first* 1 *and the last* 1*, inclusive.*

Proof. During the column collection phase, one data block addition is required each time there is a hash table collision. If two columns, one of length x and the other of length y, $x \leq y$, collide, their sum will be a column whose length is no greater than $y - 1$. Since $\bar{l}k$ is the sum of the length's of the columns and each collision reduces this total length by at least 1, there can be at most $\bar{l}k$ collisions.

During the back filling phase, the number of data block additions needed is exactly the weight of the matrix (after column collection) less k. Also, the weight of the matrix is no greater than the total length, and the total length after column collection no greater than the total length before column collection less the number of collisions. Therefore, the sum of the weight of the matrix after column collection and the number of collisions resolved during column collection is at most $\bar{l}k$. □

The average case complexity is $\bar{l}k/2$. This is easily seen by noting that when columns of length x and y, $x \leq y$, are added, the expected length of the resulting column is at most $y - 2$. Furthermore, the expected weight of the matrix after column collection is half the total length.

To see how the average case complexity may be calculated from w, first notice that for columns that do not wrap, $l \leq w$; however, for columns that do wrap, l may be as large as k. In the case where wrapping is allowed, we suggest $w = 2\sqrt{k}$ is sufficient to achieve low overhead. Note that the probability of generating a column that wraps is w/k. This means that, after k columns have been generated, we only expect w of them to wrap. The mean column length (as seen by the decoder) is thus $\bar{l} = wk + (k - w)w \approx 2wk$. This gives us an average case decoder complexity of $wk = 2k^{3/2}$.

It was suggested earlier that we may be able to avoid generating columns that wrap; however, in doing so columns with a window length as large as $4\sqrt{k}$ may be generated. This means $\bar{l}$ may be as large as $4\sqrt{k}$ and thus decoder complexity is still $2k^{3/2}$. Actually, in this no-wrapping case, $\bar{l}$ is closer to $3\sqrt{k}$ so there may yet be a benefit to avoiding wrapping.

7.4. Conclusions. An efficiently encodable and decodable class of erasure correcting codes, with decoding complexity $O(k^{3/2})$, based on the rank properties of windowed binary random matrices, has been formulated. They have the advantages over other classes of erasure

correcting codes of a low and fixed overhead and effectiveness at much lower block lengths. The necessity of the rank properties required for the windowed matrices are established here. While simulation and intuitive arguments show the conditions needed are clearly sufficient it is hoped to establish these properties analytically in future work.

8. A diversion

There are numerous applications of the matrices considered here as well as many other related problems, including combinatorics, coding, theoretical computer science etc.. We mention a problem that has been of interest in computer science, that of the singularity of random (equally likely) ± 1 matrices taken over the integers (or rationals). (Recall, the work of this report has been concerned exclusively with rank over the finite field of interest.) Let M_n be an $n \times n$ random ± 1 matrix. It long been conjectured that

$$P(\det(M_n) = 0) \geq (1 + o(1))n^2 2^{1-n},$$

a result of the fact that such a random matrix with two identical or opposite columns has a determinant of 0, which has also been conjectured as the main source of singularity. However the best actual result in this direction is:

Theorem 8.1.

$$P(\det(M_n) = 0) = \mathcal{O}((1 - \epsilon)^n)$$

where [15] $\epsilon = .001$ *improved to* $\epsilon = .0691\cdots$ [28].

(The constant $\epsilon = .0691\cdots$ in the theorem is the unique solution to the equation

$$h(\epsilon) + \epsilon/(\log_2(16/15)) = 1$$

in the interval $(0, 1/2)$, where $h(\cdot)$ is the binary entropy function.) It is also noted there ([28]) that from Hadamard's inequality we have that

$$\det(M_n) \leq n^{n/2}$$

with equality iff M_n is a Hadamard matrix. Furthermore it can be shown that

$$E((\det M_n)^2) = n!$$

and higher moments are also computable. A main result of [28] is that

$$P(\det(M_n) = \pm\sqrt{n!}\exp(-o(n^{1/2+\epsilon})) = \frac{1}{2} - o(1).$$

It is interesting, although not surprising, that the behavior of this case is so different than the binary $\{0, 1\}$ over $\mathbb{F}_2$.

9. Comments

A very brief overview of certain aspects of rank properties of random matrices over finite fields has been considered and certain questions of interest have been posed. It has been intended only as a document for possible use in pursuing further work in the area.

Acknowledgements: The authors would like to thank Igor Shparlinski for his careful reading of the report and suggestions for improvement.

References

1. G.V. Balakin, The disribution of the rank of random matrices over a finite field, *Theory Prob. Appl.,*, vol. 8, no. 4, pp. 594-605, 1968.
2. E.R. Berlekamp, The technology of error correcting codes, *Proc. IEEE*, vol. 68, pp. 564-593, 1980.
3. C. Bertram-Kretzberg, T. Hofmeister and H. Lefmann, Sparse $0-1$ matrices and forbidden hypergraphs, *Combinatorics, Probability and Computing*, vol. 8, pp. 417-427, 1999.
4. Ian F. Blake and Ronald C. Mullin, *The Mathematical Theory of Coding*, Academic Press, 1975.
5. Johannes Blömer and Richard Karp and Emo Welzl, The rank of sparse random matrices over finite fields, *Random Structures and algorithms*, vol. 10, pp 407–419, 1997.
6. Béla Bollobas, *Random Graphs*, 2nd edition, Cambridge, 2001.
7. R. Brent, S. Gao and A. Lauder, Random Krylov spaces over finite fields, *SIAM J. Discrete Math.*, vol. 16, pp. 276–287, 2003.
8. N. Calkin, Dependent sets of constant weight binary vectors, *Combinatorics, Probability and Computing*, vol. 6, pp. 263-271, 1997.
9. C. Cooper, On the rank of random matrices, *Random Structures and Algorithms*, vol. 16, no. 2, pp. 209-232, 2000.
10. C. Cooper, On the distribution of rank of a random matrix over a finite field, *Random Structures and Algorithms*, vol. 17, pp. 197-212, 2000.
11. O. Elesami and A. Shokrollahi, Raptor codes on binary memoryless symmetric channels, *IEEE Trans. Information Theory*, vol. 52, pp. 2033-2051, 2006.
12. J. Fulman, Random matrix theory over finite fields, *Bull. Amer. Math. Soc.*, vol. 39, no. 1, pp. 51-85, 2001.
13. R. Sedgewick and P. Flajolet, *Analysis of Algorithms*, Addison-Wesley, 1996.
14. R.G. Gallager, *Low Density Parity Check Codes*, MIT Press, 1963.
15. J. Kahn, J. Komlós and E. Szemerédi, On the probability that a random ± 1 matrix is singular, *J. American Math. Society*, vol. 8, no. 1, pp. 223-240, 1995.
16. V.F. Kolchin, *Random Graphs*, Cambridge, 1999.
17. G. Landsburg, Uber eine Anzahlbestimmiung und eine damit zusammenhangende Reihe, *J. Reine Angew. Math.* vol. 111, pp. 87-88, 1893.
18. H. Lefmann, P. Pudlick and P. Savicky, On sparse parity check matrices, *Designs, Codes and Cryptography*, vol. 12, pp. 107-130, 1997.
19. H. Lefmann, Sparse parity check matrices over finite fields, *COCOON 2003*, T. Warnow and B. Zhu eds., LNCS vol. 2697, pp. 112-121, 2003.
20. Michael Luby, Michael Mitzenmacher, M. Amin Shokrollahi, Daniel Speilman and Volker Stemann, Practical loss-resilient codes, *STOC*, pp. 150–159, 1997.
21. M. Luby, LT codes, 43rd *FOCS*, pp. 271-280, 2002.
22. P. Maymounkov, Online Codes, Technical Report TR2002-833, New York University, October 2002.
23. Kent Morrison, Eigenvalues of random matrices over finite fields, preprint, 1999.
24. Kent Morrison, Matrices over $\mathbb{F}_q$ with no eigenvalues of 0 or 1, preprint, 2004.
25. A. Mukhopadhyay, The probability that the determinant of an $n \times n$ matrix over a finite field vanishes, *Discrete Mathematics*, vol 51, pp. 311-315, 1995.
26. Assaf Naor and Jacques Verstraete, Improved bounds on the size of sparse parity check matrices, preprint, 2005.
27. A. Shokrollahi, Raptor codes, *IEEE Trans. Information Theory*, vol. 52, pp. 2551-2567, 2006.
28. T. Tao and Van Wu, On random ± 1 matrices: Singularity and determinant, *STOC 2005*, pp. 431-440. (Also, *Random Structures and Algorithms*, vol. 28, no. 4, pp. 1-23, 2005).
29. J.H. van Lint and R.M. Wilson, *A Course in Combinatorics*, Cambridge, 1992.

Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4
E-mail address: ifblake@comm.utoronto.ca

Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4
E-mail address: cvs@cs,utoronto.ca