A Sharp Diameter Bound
For an Upper Triangular Matrix Group

Jordan S. Ellenberg

Harvard University

Abstract: The group of upper triangular $n+1 \times n+1$ matrices over $\mathbf{Z} / p\mathbf{Z}$ with 1's on the diagonal, endowed with a natural set of n generators, has diameter bounded above and below by constant multiples of $np + n^2 \log p$.

Introduction

In combinatorial group theory it is common to encounter problems of the following sort: given a group G and a symmetric set E of generators, determine the diameter of G in E, where the diameter is defined to be the smallest k such that any element of G can be written as a product of some sequence of k or fewer elements of E.

Let $U_n(p)$ be the group of upper triangular matrices over the field of p elements with 1's along the diagonals. We define $e_{i,j}$ to be the element of $U_n(p)$ which has 1's along the diagonal and at coordinate (i,j) and zeroes elsewhere. Then the set $\{e_{i,i+1}^{\pm 1}\} \cup id$, as i ranges from 1 to n, is a generating set for $U_n(p)$ with cardinality $2n+1$.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Fig. 1. The generating set for $\mathbf{U}_2(\mathbf{p})$.

Let $d(n, p)$ be the diameter of $U_n(p)$ in the generators described above. The main result of this paper is the following:

MAIN THEOREM. *Let $f(n,p) = np + n^2 \log p$; then there exist constants c_1, c_2 such that*

$$c_1 f(n,p) \leq d(n,p) \leq c_2 f(n,p).$$

In the body of the paper, we show that $c_1 = 1/8$, and $c_2 = 216$ suffice if $n > 100$ or $p > 10{,}000$.

One standard approach to general diameter problems involves bounding the second-largest eigenvalue of the adjacency matrix A of the Cayley graph. (The Cayley graph has vertices corresponding to the elements of G, with two vertices x and y adjacent if $xy^{-1} \in E$.) The (1,g) entry in A^m counts the number of ways to express g as a product of m elements of E. Thus, the diameter of the group is the smallest m for which A^m has no zero entries. Chung has shown, using the spectral decomposition of A, that the diameter of a k-regular graph is at most $\log(|G|-1)/\log(k/\lambda)$ **[Ch]**, where $|G|$ is the number of vertices (for our purposes, the order of the group), k is the largest eigenvalue and the degree of each vertex (respectively, the number of generators), and λ is the second-largest eigenvalue of A. Neither this method nor any of its refinements, however, do a very good job of bounding $d(n, p)$, since the adjacency matrix of the Cayley graph of $U_n(p)$ has eigenvalues which are very close to n.

Specifically, let $v: U_n(p) \to \mathbf{C}$ map an element u of $U_n(p)$ to $e^{ik\pi/p}$, where k is the entry in the (1,2)-coordinate of u. It is straightforward that this is a homomorphism; it can also be viewed as a vector in $\mathbf{C}^{|U_n(p)|}$. Now left multiplication by the adjacency matrix of the Cayley graph of $U_n(p)$ takes $v(u)$ to

$$v(e_{1,2}u) + v(e_{1,2}^{-1}u) + \ldots + v(e_{n,n+1}u) + v(e_{n,n+1}^{-1}u)$$
$$= v(e_{1,2})v(u) + v(e_{1,2}^{-1})v(u) + \ldots + v(e_{n,n+1})v(u) + v(e_{n,n+1}^{-1})v(u)$$

$$= [n - 2 + 2\cos(\pi/p)]v(u).$$

Therefore, v is an eigenvector for $U_n(p)$ with eigenvalue $n - 2 + 2\cos(\pi/p)$. So

$$\lambda \geq n - 2 + 2\cos(\pi/p) \geq n - c/p^2.$$

This bound is due to Stong **[St]**. Substituting this value of λ into Chung's formula yields an upper bound of order at least $n^3 p^2 \log p$ for the diameter.

(A forthcoming paper by Chung et. al. **[ChFM]** improves the bound in **[Ch]** to $\cosh^{-1}(|G|-1)/\cosh^{-1}(k/\lambda)$. However, this result still yields an upper bound for the diameter of order at least $n^{5/2} p \log p$.)

One application of the main theorem occurs in the study of random walks. A random walk on a group endowed with a set of generators is defined as the random walk on the corresponding Cayley graph. Diaconis and Saloff-Coste have shown that the length of time necessary for a random group walk to converge to a nearly uniform distribution, given a bounded number of generators and a bounded nilpotency class, is proportional to the square of the diameter **[DS-C]**. We have shown in the current paper that $d(n, p)$ grows with order p as p grows large and n is held constant; combined with the results of Diaconis and Saloff-Coste, this implies that the number of steps necessary for convergence grows with order at most p^2. This bound is an improvement over Stong's result **[St]**, which tells us only that order $p^2 \log p$ steps will do. In fact, since the time for a random walk on $\mathbf{Z}/p\mathbf{Z}$ to converge grows with order p^2, this is a lower bound as well as an upper bound.

We will make great use of the following easily verified identity:

$$[e_{i,j}, e_{k,l}] = \begin{cases} id & j \neq k \\ e_{i,l} & j = k \end{cases}$$

where $[u, v]$ signifies the commutator $uvu^{-1}v^{-1}$.

The paper is structured as follows. In section 1, we prove two lower bounds for $d(n,p)$, and show that their maximum is greater than a constant multiple of $f(n,p)$. The first of these bounds comes from a straightforward counting argument. The second relies on the commutativity relations between the generators, and generalizes to a possibly novel lower bound on the diameter of Cayley graphs in general. In section 2, we provide an upper bound by presenting an explicit procedure for producing any element of $U_n(p)$ using a word of length $c_2 f(n,p)$ or smaller. In the Addendum, we present a sharper bound for the case n=2 (the Heisenberg group) and propose some questions for further research.

1. The Lower Bound For $d(n,p)$

In this section we will demonstrate that $d(n,p)$ is greater than or equal to $\frac{1}{8}(np + n^2 \log p)$ whenever $n > 100$ or $p > 10{,}000$. We start with the following result:

Theorem 1.1. $d(n,p) \geq \frac{1}{2}n(p-1)$.

Proof. The map $h_i\colon U_n(p) \to \mathbb{Z}/p\mathbb{Z}$ taking a matrix u to the entry in coordinate $(i,i+1)$ of u is a homomorphism; this fact is immediate from the multiplication. Note that h_i maps $e_{i,i+1}^{\pm 1}$ to ± 1, and all other generators to 0. Thus, a word which evaluates to a matrix with some coefficient r in the $(i,i+1)$ coordinate must contain $r \pmod p$ occurences of the generator $e_{i,i+1}$. A matrix with $\frac{1}{2}(p-1)$ (or 1, if $p=2$) in each coordinate just above the diagonal is therefore inexpressible by any word of length less than $\frac{1}{2}n(p-1)$. ($\therefore$ 1.1)

Theorem 1.2. $d(n,p) \geq n^2 \log p\left(\frac{1}{6\log 2} - \frac{1}{6n\log 2} - \frac{n+2}{3n^2 \log p}\right)$.

Proof. The proof will rest on a novel application of a result of Cartier and Foata from [CF]. We will start by raising a general combinatorial question. Let S be some set of k letters. Let W be the set of all formal words generated by the letters in S. Let T be some set of pairs of letters from S. We'll say that $s_i, s_j \in S$ *commute* if $\{s_i, s_j\}$ is in T.

We define an equivalence relation $\sim_T$ on W by $w_1 \sim_T w_2$ if and only if there is some sequence of transpositions of adjacent commuting letters which carries w_1 into w_2. Let E be the set of equivalence classes of W under $\sim_T$. For simplicity's sake, we will hereafter refer to $\sim_T$ simply as $\sim$. Since any two equivalent words have the same number of occurences of each letter in S, we can define $w_T(m_1, m_2, \ldots, m_k)$ to be the number of equivalence classes of words from S containing exactly m_i occurences of each s_i. Such a word will be called an **m**-word, where **m** is the vector $(m_1, m_2, \ldots, m_k)$.

Example. Suppose $S = \{s_1, s_2, s_3\}$ and T consists of the two pairs $\{\{s_1, s_2\}, \{s_2, s_3\}\}$. Then $w_T(1,1,1) = 2$, the relevant equivalence classes being $s_2 s_1 s_3 \sim s_1 s_2 s_3 \sim s_1 s_3 s_2$ and $s_2 s_3 s_1 \sim s_3 s_2 s_1 \sim s_3 s_1 s_2$.

In order to study the behavior of w_T, we define the generating function

$$W_T(x_1, x_2, \ldots, x_k) = \sum_{m_1, m_2, \ldots, m_k = 0}^{\infty} w_T(m_1, m_2, \ldots, m_k) x_1^{m_1} x_2^{m_2} \ldots x_k^{m_k}.$$

Lemma 1.2.1. $W_T(x_1, x_2, \ldots, x_k) = \left(\sum_{R \in \mathbf{M}} (-1)^{|R|} \prod_{s_i \in R} x_i \right)^{-1}$, *where the outer sum is taken over the collection* **M** *of subsets of S whose elements commute pairwise.*

Proof. This is the result shown by Cartier and Foata in **[CF]**.

Let $w_T(m)$ be the number of non-equivalent words of length m. The generating function $W_T(x) = \sum_{m=0}^{\infty} w_T(m) x^m$ can be calculated by substituting x for each x_i in the generating function of Lemma 1.2.1, yielding

Corollary 1.2.2. $W_T(x) = \left(\sum_{i=0}^{\infty}(-1)^i k_i x^i\right)^{-1}$ *, where k_i is the number of pairwise commutative subsets of S with cardinality i.*

We wish to know the number of m-letter words in the n generators of $U_n(p)$. These generators are very nearly mutually commutative; $e_{i,i+1}$ and $e_{j,j+1}$ commute unless $|i-j|=1$. Let $w_n(m)$ be the number of distinct m-letter words in these letters (where we consider two words equivalent only if we can get from one to the other through the commutativity relations; thus, $w_n(m)$ may be greater than the actual number of distinct elements of $U_n(p)$ representable by m-letter words.) Let $W_n(x) = \sum_{m=0}^{\infty} w_n(m)x^m$. By Corollary 1.2.2,

$$W_n(x) = \left[\sum_{i=0}^{\left\lceil \frac{n}{2} \right\rceil}(-1)^i \binom{n-i+1}{i} x^i\right]^{-1}.$$

Let $p_n(x) = [W_n(x)]^{-1}$. We will show that the roots of $p_n(x)$ have magnitude greater than $\frac{1}{4}$.

Applying Pascal's identity allows us to write the following recurrence for $p_n(x)$:

$$\begin{aligned}
&p_0(x) = 1;\\
&p_1(x) = 1 - x;\\
&p_n(x) = p_{n-1}(x) - xp_{n-2}(x) \quad (n \geq 2)
\end{aligned}$$

The solution to this type of recurrence is of the form $Ar_1^n + Br_2^n$, where r_1 and r_2 are the roots of the characteristic equation $z^2 - z + x$. Solving, we find

$$p_n(x) = \tfrac{1}{2}\left[\left(1+\frac{1-2x}{\sqrt{1-4x}}\right)\left(\frac{1+\sqrt{1-4x}}{2}\right)^n + \left(1-\frac{1-2x}{\sqrt{1-4x}}\right)\left(\frac{1-\sqrt{1-4x}}{2}\right)^n\right]$$

Substituting $y = \sqrt{1-4x}$ gives

$$\tfrac{1}{2}\left[\left(1+\frac{1+y^2}{2y}\right)\left(\frac{1+y}{2}\right)^n+\left(1-\frac{1+y^2}{2y}\right)\left(\frac{1-y}{2}\right)^n\right]$$
$$=\tfrac{1}{2^{n+2}y}\left[(1+y)^{n+2}-(1-y)^{n+2}\right].$$

The roots of this polynomial in y are $i\tan[k\pi/(n+2)]$, $0 < k < n+2$. So the roots of $p_n(x)$ are the finite values greater than $\frac{1}{4}$ taken on by $\frac{1}{4}\sec^2[k\pi/(n+2)]$ as k ranges over the integers.

Since all the poles of $W_n(x)$ are greater than $\frac{1}{4}$, the power series centered at 0 is convergent at $\frac{1}{4}$, and so $4^{-m}w_n(m)$ is bounded above by some constant c_n. To bound c_n, we need only note that

$$c_n = \max_m\left[4^{-m}w_n(m)\right] < \sum_{m=0}^{\infty} 4^{-m}w_n(m) = W_n(\tfrac{1}{4}).$$

From the recurrence for $p_n(x)$, we find that $W_n(\frac{1}{4}) = 2^{n+1}/(n+2)$.

We now have an upper bound for $w_n(m)$, which is in turn an upper bound for the number of distinct elements of $U_n(p)$ expressible as an m-letter word from $\{e_{i,i+1}\}$. The inverses of these elements and the identity are also included in our set of generators. To account for the inverses, we need only allow signs on the letters, multiplying the number of distinct words by 2^m. Accounting for the identity just means counting all words of m letters *or fewer*-- since the ratio between consecutive terms is always at least 2 (once we've accounted for inverses), the number of words of m letters or fewer is no more than twice the number of words of exactly m letters. Thus, the number of elements of $U_n(p)$ expressible with m letters from $\{e_{i,i+1}^{\pm 1}\} \cup id$ is at most $(n+2)^{-1}2^{n+2}8^m$.

Therefore, we have

$$(n+2)^{-1}2^{n+2}8^{d(n,p)} \geq p^{\binom{n}{2}}$$

Dropping the factor of $(n+2)^{-1}$ and taking a logarithm, we have

$$\begin{aligned} d(n,p) &\geq \frac{1}{3\log 2}\binom{n}{2}\log p - \frac{n+2}{3} \\ &= n^2\log p\left(\frac{1}{6\log 2} - \frac{1}{6n\log 2} - \frac{n+2}{3n^2\log p}\right). \end{aligned}$$

$\therefore$ (1.2)

Now we are ready to demonstrate the promised lower bound for $d(n,p)$. It is clear from Theorem 1.1 that $d(n,p) \geq \frac{1}{3}np$. If $n > 100$, the lower bound of Theorem 1.2 is at least $\frac{1}{5}n^2\log p$. If $n \leq 100$ and $p > 10{,}000$, then $\frac{1}{3}np > \frac{1}{5}n^2\log p$. Therefore, if (n,p) lies outside the finite rectangle bounded by $n = 100$ and $p = 10{,}000$, we have

$$d(n,p) \geq \max(\tfrac{1}{3}np, \tfrac{1}{5}n^2\log p) \geq \tfrac{1}{8}(np + n^2\log p).$$

The technique used to prove Theorem 1.2 is applicable to groups in general; of course, it will be useful primarily when the chosen set of generators is highly commutative. We state the following general result:

Theorem 1.3. *Let G be a group and E a set of generators for G. Let k_i be the number of mutually commuting subsets of E with cardinality i. If c is a positive real number smaller than the magnitude of any root of the polynomial P(x) =* $\sum_{i=0}(-1)^i k_i x^i$,

$$\operatorname{diam}(G,E) \geq -\frac{\log|G| + \log P(c)}{\log c}.$$

Proof. The argument follows the course of Theorem 1.2 exactly.

Example. Consider the symmetric group on n elements, with the transpositions (12),(23),...,$(n\text{--}1\ n)$ as generators. By Stirling's formula, $\log|G|$ is of order $n\log n$. If we considered the set of all possible formal words without utilizing the commutativity relations, we could conclude only that the diameter of the group was of order at least n. In fact, the commutativity structure of these generators is exactly the same as the one we've discussed for $U_n(p)$. So we can substitute $c = \frac{1}{4}$ into Theorem 1.3, showing that the diameter is of order at least $n\log n$. (The actual diameter of S_n in these generators is of order n^2.)

2. THE UPPER BOUND FOR $d(n,p)$

In this section, we present an explicit algorithm for generating any element of $U_n(p)$ as a product of at most $216(np + n^2 \log p)$ generators.

Note that the set of matrices with 1's along the main diagonal, entries in $\mathbf{Z}/p\mathbf{Z}$ elsewhere in the top row, and zeroes otherwise is an abelian subgroup of $U_n(p)$ isomorphic to $(\mathbf{Z}/p\mathbf{Z})^n$. Multiplication within this group corresponds to addition in $(\mathbf{Z}/p\mathbf{Z})^n$. Call this group $R_n(p)$ and let $r(n,p)$ be its diameter in the generators $\{e_{i,i+1}^{\pm 1}\} \cup id$. (This constitutes a slight abuse of the diameter definition, since $\{e_{i,i+1}^{\pm 1}\} \cup id$ are generators of $U_n(p)$, not $R_n(p)$. To be precise: $r(n,p)$ is the smallest k such that every element of $R_n(p)$ can be expressed as a product of k or fewer members of $\{e_{i,i+1}^{\pm 1}\} \cup id$.)

$$\begin{bmatrix} 1 & z_1 & & z_n \\ & 1 & & \\ & & & \\ & & & 1 \end{bmatrix}$$

Fig. 2. An element of $\mathbf{R}_n(\mathbf{p})$ $(z_i \in \mathbf{Z}/\mathbf{p}\mathbf{Z})$

The subgroup $V_n(p)$ of $U_n(p)$ generated by $\{e_{i,i+1} : i \geq 2\}$ is isomorphic to $U_{n-1}(p)$. Any element u of $U_n(p)$ can be expressed as vr, with v in $V_n(p)$ and r in $R_n(p)$. We conclude that

$$d(n,p) \leq d(n-1,p) + r(n,p)$$

and thus that

$$d(n,p) \leq \sum_{i=1}^{n} r(i,p).$$

Theorem 2.1.

$$r(n,p) \le \frac{1}{2}p + \frac{8}{\log^2 2}\sqrt{p}\log^2 p + \frac{8}{\log 2}\sqrt{p}\log p + \frac{8}{\log 2}n\log p + 8.$$

Proof. The proof will be split into two cases, $p > 2$ and $p = 2$.

Case i) $p > 2$. Let $k = \lfloor \log_2 p \rfloor$. We will consider $R_n(p)$ as the additive group of vectors in $(\mathbf{Z}/p\mathbf{Z})^n$. Then we can write $R_n(p) = R_1 \oplus R_2 \oplus R_3$, where R_1 is the subgroup supported on the first coordinate, R_2 is the subgroup supported on coordinates 2 through k, and R_3 is the subgroup supported on the remaining coordinates. (If n is small compared to p, R_3 may be trivial.) So $r(n,p)$ is bounded above by the sum of the diameters of these three subgroups. The diameter of R_1 is just $\frac{1}{2}(p-1)$.

Define a function (*not* a homomorphism) $q:(\mathbf{Z}/p\mathbf{Z})^n \to U_n(p)$, where

$$q(a_1, a_2, \ldots, a_n) = e_{n,n+1}^{-a_n} e_{n-1,n}^{-a_{n-1}} \ldots e_{2,3}^{-a_2} e_{1,2}^{a_1} e_{2,3}^{a_2} \ldots e_{n,n+1}^{a_n}.$$

It is straightforward that $q(a_1, a_2, \ldots, a_n)$ is the element of $R_n(p)$ whose top row is

$$[1 \quad a_1 \quad a_1 a_2 \quad \ldots \quad \prod a_i].$$

Notice that the definition of q gives us an expression for $q(a_1, a_2, \ldots, a_n)$ as a word of length $2\sum_{i=1}^{n} |a_i| - |a_1|$.

Lemma 2.1.1. *The diameter of R_2 is at most*

$$\frac{8}{\log^2 2}\sqrt{p}\log^2 p + \frac{8}{\log 2}\sqrt{p}\log p.$$

Proof. We start by claiming that an element of the form $e_{1,m+1}$ can be expressed by a word of fewer than $4m$ letters. For

$$e_{1,m+1} = q(1,1,\ldots,1,0,\ldots,0)q(-1,1,\ldots,1,0,\ldots,0),$$

the right-hand expression being a word of length $(2m - 1) + (2m - 3) = 4m - 4$.

Any element of R_2 can be written as $e_{1,3}^{s_2} e_{1,4}^{s_3} \ldots e_{1,k+1}^{s_k}$, with $0 \le s_i < p$. We will show that an element of the form $e_{i,i+m}^{s}$ can be expressed as a word of fewer than $16m\sqrt{s}$ letters whenever $m \ge 2$. Let $t = \lfloor \sqrt{s} \rfloor$. Then

$$e^{s}_{1,m+1} = e^{t^2}_{1,m+1}e^{s-t^2}_{1,m+1} = [e^{t}_{1,m}, e^{t}_{m,m+1}][e_{1,m}, e^{s-t^2}_{m,m+1}].$$

Since $t \le \sqrt{s}$ and $s - t^2 \le [(t+1)^2 - 1] - t^2 \le 2t \le 2\sqrt{s}$, the product of commutators on the right has length at most $2(4m-4)\sqrt{s} + 2\sqrt{s} + 2(4m-4) + 2(2\sqrt{s}) \le 16m\sqrt{s}$. Since $s < p$, the diameter of R_2 is at most

$$\sum_{m=2}^{k} 16m\sqrt{p} \le 8(k^2+k)\sqrt{p} \le \frac{8}{\log^2 2}\sqrt{p}\log^2 p + \frac{8}{\log 2}\sqrt{p}\log p. \quad \therefore (2.1.1).$$

Lemma 2.1.2. *The diameter of* R_3 *is at most* $\dfrac{8}{\log 2} n \log p + 8$.

Proof. Let r be an element of R_3. Let $r(m), m > k$, be the mth coordinate of r. (Throughout this proof we will treat R_3 as a subgroup of $(\mathbf{Z}/p\mathbf{Z})^n$, with addition of vectors as the group law.) Consider $r(m)/2 \in \mathbf{Z}/p\mathbf{Z}$ as an integer; it has a binary expansion of the form $2^{d_1} + 2^{d_2} + \ldots + 2^{d_j}$, with $d_i < \log_2 p$. (The division by 2 in this step is the reason we need to consider the case $p = 2$ separately.) Let $r(m;i)$ be 2^{i+1} if 2^i appears in the binary expansion of $r(m)/2$, 0 otherwise. Finally, let r_i be the element of R_3 whose mth coordinate is $r(m;i)$. Clearly, $\sum_{i=0}^{k} r_i = r$. Since addition in R_3 is equivalent to multiplication in $U_n(p)$, the length of the shortest word in $\{e^{\pm 1}_{i,i+1}\} \cup id$ representing r is at most the sum of the lengths of the shortest words representing the r_i.

Each r_i is an element of R_3 whose entries are all either 0 or 2^{i+1}. We will show that such an element can be represented by a word of $8n$ letters or fewer.

Consider some r_i of the form described above. Let $I(m)$ be the function which is 1 when the mth coordinate of r_i is 0, and -1 otherwise. Now define two n-tuples of integers as follows:

$$a_m = b_m = 2 \quad (1 \le m < i);$$

$$a_m = 1 \quad (m \ge i);$$

$$b_m = I(m)I(m-1) \quad (m \ge i);$$

Now $q(\mathbf{a})$ is a vector whose entries increase by factors of 2 until reaching 2^i, and remain constant thereafter. Evidently, $q(\mathbf{b})$ agrees with the above vector for all coordinates up to and including place k; thereafter, its entries are 2^i wherever the corresponding entry in r_i is 0, and -2^i otherwise. Thus, the difference of the two vectors is r_i. Since $|a_m|,|b_m| \leq 2$ for all m, each of the two vectors can be produced by a word of length $4n$ or less. Thus, we can write r_i as a word of $8n$ letters.

Since $\sum_{i=0}^{k} r_i = r$, we can express r as a word with $8n(k+1) < \frac{8}{\log 2} n \log p + 8$ letters or fewer, as claimed. $\therefore$ (2.1.2)

Thus,

$$\begin{aligned} r(n,p) &\leq \operatorname{diam}(R_1) + \operatorname{diam}(R_2) + \operatorname{diam}(R_3) \\ &\leq \frac{1}{2}p + \frac{8}{\log^2 2}\sqrt{p}\log^2 p + \frac{8}{\log 2} n \log p + 8. \end{aligned}$$

This concludes case i).

Case ii) p=2. We define a new map $q' : (\mathbf{Z}/p\mathbf{Z})^n \to U_n(p)$ by

$$q'(\mathbf{a}) = e_{n,n+1}^{-a_n} e_{n,n-1}^{-a_{n-1}} \ldots e_{2,3}^{-a_2} t e_{1,2}^{a_1} e_{2,3}^{a_2} \ldots e_{n,n+1}^{a_n},$$

where t is the element of $R_n(p)$ with 1's along the top row. The range of q' lies within $R_n(p)$. If $q_i'(\mathbf{a})$ is the $(1, i+1)$ entry of $q'(\mathbf{a})$, then we have $q_{i+1}'(\mathbf{a}) = a_{i+1} q_i'(\mathbf{a}) + 1$.

Which elements of $R_n(p)$ lie in the range of q'? Note that if some coordinate $q_i'(\mathbf{a})$ of $q'(\mathbf{a})$ is 0, the following coordinate is $0 \cdot a_{i+1} + 1 = 1$. However, if $q_i'(\mathbf{a}) = 1$, the following coordinate is just a_{i+1} + 1, which can be either 0 or 1 depending on our choice of a_{i+1}. Therefore, every element of $R_n(p)$ which does not contain two successive 0's can be expressed as $q'(\mathbf{a})$ for some $\mathbf{a} \in (\mathbf{Z}/2\mathbf{Z})^n$. Since $t = q(1,1,\ldots,1)$ can be expressed as a word of length $2n$ - 1, every $q'(\mathbf{a})$ can be expressed as a word with fewer than $4n$ letters.

We claim that every element of $(\mathbf{Z}/2\mathbf{Z})^n$ can be expressed as the sum of two other elements of $(\mathbf{Z}/2\mathbf{Z})^n$, neither of which contain two successive 0's. Let r be an arbitrary element of $(\mathbf{Z}/2\mathbf{Z})^n$; then we can separate it into contiguous blocks of 0's and 1's. We define r_1 to be the result of replacing each block of 0's in r by a block of 1's, and each block of 1's by a block of alternating 1's and 0's, starting with 1. We define r_2 similarly, except that the blocks of alternating 1's and 0's start with 0. Then r_1 and r_2 sum to r, and neither r_1 nor r_2 contains two successive 0's.

We conclude that we can express any element of $R_n(p)$ as a product $q'(\mathrm{a})q'(\mathrm{b})$, so $r(n,2) \le 8n \le \frac{8}{\log 2} n \log p$. Thus, Theorem 2.1 holds for $p = 2$. $\therefore$ (2.1)

As shown above, $d(n,p) \le \sum_{i=1}^{n} r(i,p)$. So we have

$$d(n,p) \le \frac{1}{2}np + \frac{8}{\log^2 2} n\sqrt{p}\log^2 p + \frac{8}{\log 2} n\sqrt{p}\log p + \frac{4}{\log 2} n^2 \log p + \frac{4}{\log 2} n \log p + 8n.$$

It's clear that the second, third, fifth and sixth summands on the right side are asymptotically small compared to the others. Let us formalize this notion. Suppose that either $n > 100$ or $p > 10{,}000$. Let $g(n,p) = \frac{1}{2}np + \frac{4}{\log 2} n^2 \log p$.

Consider the term $\frac{8}{\log^2 2} n\sqrt{p}\log^2 p$. If $n > 100$ and $p \le 10{,}000$, then the ratio between this term and the second summand in $g(n,p)$ is $\frac{2\sqrt{p}\log p}{n \log 2} \le \frac{800 \log 10}{100 \log 2} < 27$. If $p > 10{,}000$, the ratio between this term and the first summand is $\frac{16\log^2 p}{\sqrt{p}\log^2 2}$; it is easy to verify that this function is decreasing for $p > 10{,}000$, so it is at most $\frac{16\log^2 10{,}000}{100\log^2 2} < 29$. So in any case, this term is at most $29g(n,p)$. Similarly, the third term is at most $4g(n,p)$.

The fifth and sixth terms are less troublesome. The fifth term, $\frac{4}{\log 2} n \log p$, is evidently less than $g(n,p)$, and the sixth, $8n$, is less than $\frac{8}{\log 2} n^2 \log p < 2g(n,p)$. We conclude that $d(n,p) \le 36g(n,p)$. It is clear that $g(n,p) < 6(np + n^2 \log p)$. Therefore, setting $c_2 = 216$, the main result is true as claimed. Combining the results of sections 1 and 2, we have proven that $d(n,p)$ is asymptotically of order $np + n^2 \log p$.

ADDENDUM

A. The case n = 2

The group $U_2(p)$ is called the *Heisenberg group* over the field of p elements. On this small case, we can sharpen the bounds presented in the paper considerably. Specifically, we have

Theorem A.1. $p-1 \leq d(2,p) \leq p+2$.

Proof. The lower bound follows directly from Theorem 1.1.

Let $a = e_{1,2}$ and $b = e_{2,3}$. Note that $U_2(p)$ is nilpotent of class 2; therefore, the commutator $[a,b]$ ($= e_{1,3}$) is in the center of the group, and $[a^m, b^n] = [a,b]^{mn}$.

For x in $(\mathbf{Z}/p\mathbf{Z})$, let $|x|$ be the distance between x and 0 when $(\mathbf{Z}/p\mathbf{Z})$ is considered as a cycle. Let u be an element of $U_2(p)$ of the form

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

with x and y nonzero.

Now a word of the form $b^{y-k}a^j b^k a^{x-j}$ is equal to u if $jk = z$. The length of this word is $|y-k| + |k| + |x-j| + |j|$. Suppose without loss of generality that $|y| \geq |x|$. It is straightforward that $|y-k| + |k| \leq p - |y|$. We can clearly choose j with $|j| = 1$ and $|x-j| = |x-1|$; just pick j on the short arc joining 0 and x in $(\mathbf{Z}/p\mathbf{Z})$. Now let $k = z/j$. Then we have a word equal to u whose total length is at most $p - |y| + |x| \leq p$.

In case one or both of x and y is 0, we can append an a on the right side of u and/or a b on the left, and thus make both entries nonzero. So any element of $U_2(p)$ can be expressed as the product of $p + 2$ or fewer letters. ($\therefore$ A.1)

B. Questions for Further Research

The results of this paper suggest many natural questions. I'll mention several:

- Which of the results here still hold if the prime parameter p is replaced by an arbitrary integer? The lower bounds of Chapter 1 are not affected. Lemma 2.1.2, however, will not work if p is replaced by an even integer.

- Is the equivalence relation $\sim_T$ the same as the equivalence relation $\sim_K$, defined by $w_1 \sim_K w_2$ if the two words, considered as elements of the free group on S, are in the same coset of the normal subgroup K generated by the commutators of the pairs in T? If so, Lemma 1.2.1.1 would become trivial. Moreover, Corollary 1.2.2 would imply that the group G/K has rational growth in the sense of Benson **[Be]**. Benson discusses only groups with nilpotent subgroups of finite index; the groups G/K will, in general, not be of this type.

- What if we consider subgroups of $U_n(p)$ generated by sets of elements $e_{i,j}$, where $j-i$ was not necessarily equal to 1? (These are the *closed subgroups* as discussed by G.D. James **[Ja]**.)

- What is the actual time necessary for a random walk on $U_n(p)$ to converge?

- How does the diameter change if we do not include the inverses of the $e_{i,i+1}$ in our generating set?

- To what extent can the techniques of Theorem 1.2 be applied to other diameter problems?

C. Acknowledgements

I would like to thank Persi Diaconis for suggesting the problem, and for providing unflagging support and unerring critique throughout the course of the work reported in this paper. Thanks also to Richard Stong for his comments on earlier drafts.

REFERENCES

[Be] M. Benson, *On the rational growth of virtually nilpotent groups*, in *Combinatorial Group Theory and Topology*, Annals of Math. Studies, vol. 111, Princeton University Press, Princeton, 1987.

[CF] P. Cartier and D. Foata, *Problèmes Combinatoires de Commutation et Réarrangements*, Lect. Notes in Math. vol. 85, Springer Verlag, New York, 1969.

[Ch] F.R.K.Chung, *Diameters and eigenvalues*, Journ. Amer. Math. Soc., vol.2, Amer, Math. Soc., Providence, R.I., 1989, pp.187-196

[ChFM] F.R.K.Chung, V.Faber, T.A.Manteuffel, *An upper bound on the diameter of a graph from eigenvalues associated with its Laplacian*, to appear in SIAM J. of Discrete Math.

[DS-C] P. Diaconis and L. Saloff-Coste, *Moderate growth and random walk in finite groups*, Geometric and Functional Analysis, vol.4, no.1, Birkhäuser Verlag, Basel, 1994.

[Ja] G.D.James, *Representations of General Linear Groups*, Lond. Math. Soc. Lect. Note Ser. 94, Cambridge University Press, London.

[St] R. Stong, *A random walk on the Heisenberg group*, preprint.