Letter to Scott Aaronson and Andy Pollington on the Solovay–Kitaev Theorem and Golden Gates

Peter Sarnak

February 2015

Typeset by Zachary Stier, June 2022

Dear Scott and Andy,

Thanks for pointing me to the papers [KMM'12, RS'14]. I was not aware of these interesting developments in connection with the Solovay–Kitaev theorem and the design of universal 1-qubit quantum gates. These papers refer to some others which are also very interesting such as [BGS'13]. All of these results can be understood in a unified way in terms of the arithmetic of quaternion algebras. Doing so clarifies (at least for me) the constructions and it also allows one to prove some fundamental properties for these gates as well as to relate them to some older and more recent developments. I explicate these points below, for myself as much as for the reader.

The problem is one of constructing an efficient universal set of quantum gates. That such exist is the Solovay–Kitaev theorem [NC'00]. The basic state in quantum computation is a single qubit, which is a unit vector in $\mathbb{C}^2$. The construction of universal quantum gates reduces to finding efficient topological generators of $G = \mathrm{PSU}(2)$ (or $\mathrm{SU}(2)$) [NC'00]. G comes with a left and right invariant metric:

$$\begin{aligned} d_G^2(x,y) &= 1 - \frac{|\mathrm{trace}(x^*y)|}{2} && (1)\\ &= d_G^2(hx,hy) = d_G^2(xh,yh). && (2)\end{aligned}$$

A set $S = \{s_1, \dots, s_\nu\}$ of elements of G (the gates) are universal if Γ, the group generated by the s_j's, is topologically dense in G. Different gates might have different costs $w(s_j) \geqslant 0$ to be implemented. So define the height $h(\gamma)$ of $\gamma \in \Gamma$ to be

$$h(\gamma) = \min\left\{\sum_{k=1}^{\ell} w(s_{j_k}) : \gamma = s_{j_1} \cdots s_{j_\ell}\right\}. \tag{3}$$

The efficiency of the gate set S is measured by their ability to approximate any $x \in G$ by γ's of small height (this corresponds to the size of the corresponding quantum circuit). Given $\varepsilon > 0$ let t_ε be the least t for which

$$G \subset \bigcup_{\gamma \in V(t)} B_G(\gamma, \varepsilon) \tag{4}$$

where

$$V(t) = \{\gamma \in \Gamma : h(\gamma) \leqslant t\} \tag{5}$$

and $B_G(x,r)$ is a ball centered at x with radius r in G.

Clearly

$$|V(t_\varepsilon)|\mu(B_G(\varepsilon)) \geqslant 1 \tag{6}$$

where μ is the normalized Haar measure on G and $B_G(\varepsilon)$ any ball of radius ε (they all have the same measure which is $\sim c\varepsilon^3$ as $\varepsilon \to 0$). To measure the covering efficiency of the points $V(t)$, we define the covering exponent $K(S)$ by

$$K(S) := \limsup_{\varepsilon \to 0} \frac{\log|V(t_\varepsilon)|}{\log \frac{1}{\mu(B_G(\varepsilon))}}. \tag{7}$$

Clearly $K(S) \geqslant 1$ and if $K(S) = 1$ then the generating gates are optimally asymptotically efficient. For example if the points in $V(t)$ were behaving like a random set of $|V(t)|$ points, then we would have $K(S) = 1$.

The mathematical problems are*:

(A) How small can we make $K(S)$ by choosing the gates in S suitably?

(B) To give a poly $\log(1/\varepsilon)$ time algorithm to find good approximations to these minimal cost circuits.

The Solovay–Kitaev theorem asserts that for any universal gate set S, given $\varepsilon > 0$ and $x \in G$, an element $\gamma \in \Gamma$ can be found in $O((\log(1/\varepsilon))^{2.71})$ steps for which $d_G(x, \gamma)$ is at most ε and the height of γ is $O((\log(1/\varepsilon))^{3.97})$; these explicit powers of $\log(1/\varepsilon)$ being due to [DN'06].

This is satisfactory from a theoretical point of view but in practice one would like to do much better. Indeed from the above it does not even follow that $K(S) < \infty$. The last is in fact true, at least if the entries in each s_j are algebraic numbers. This follows from the spectral gap theorem [BG'08] which asserts that the self adjoint operator (assuming S is symmetric, $s_j \in S \iff s_j^{-1} \in S$ and $w(s_j) = w(s_j^{-1})$) $T_S : L^2(G) \longrightarrow L^2(G)$, given by

$$T_S f(x) = \sum_{j=1}^{\nu} w(s_j) f(s_j x), \tag{8}$$

has a gap in its spectrum below its top eigenvalue, which is $\sum\limits_{j=1}^{\nu} w(s_j)$.

The proof of the spectral gap does not yield any feasible value of the gap and hence for $K(S)$. Moreover being a counting argument it offers nothing on problem (B). Still that $K(S) < \infty$ is a good step and an indication of what one might achieve by choosing the gates carefully.

All the known good gates S come from arithmetic and number theory and the proof that they are efficient uses the theory of automorphic forms on associated groups. We review the constructions, they all have the same pros and cons.

*The super-efficient continued fraction algorithm approximates a real number by rationals using $x \mapsto x + n$ and $x \mapsto 1/x$, and one is seeking something like it for PU(2).

(i) $\underline{p = 5}$ ([LPS, BGS'13]):

$S = \left\{ s_1, s_1^{-1}, s_2, s_2^{-1}, s_3, s_3^{-1} \right\}$ where

$$s_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1+2i & \\ & 1-2i \end{pmatrix}, \qquad s_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2i \\ 2i & 1 \end{pmatrix}, \qquad s_3 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}.$$

The weights $w(s_j)$ are all taken to be 1, Γ is a free group on s_1, s_2, s_3 and $h(\gamma)$ is the reduced word length of γ. $|V(t)| = 6 \cdot 5^{t-1}$, for $t \geqslant 1$.

(a)
$$\frac{4}{3} \leqslant K(S) \leqslant 2.$$

The last asserts (essentially) that for $\varepsilon > 0$ and $x \in G$ there is a $\gamma \in G$ with $h(\gamma) \leqslant 6 \log_5({}^{1}\!/\!_{\varepsilon})$ and $d_G(x, \gamma) < \varepsilon$, while there are y's in G for which $d_G(y, \gamma) < \varepsilon \implies h(\gamma) \geqslant 4 \log_5({}^{1}\!/\!_{\varepsilon})$.

(b) For $\varepsilon > 0$ there is t_ε with

$$\frac{\log |V_{t_\varepsilon}|}{\log \varepsilon^{-3}} \to 1 \text{ as } \varepsilon \to 0,$$

such that for most points $y \in G$ w.r.t. Haar measure, there is an ε-approximation of y of height t_ε (so optimal for most y's).

(c) There is a poly $\log({}^{1}\!/\!_{\varepsilon})$ time algorithm which assuming some reasonable conjectures about the distribution of primes produces for any $x \in G$ a $\gamma \in \Gamma$ with $h(\gamma) \leqslant 12 \log_5({}^{1}\!/\!_{\varepsilon})$ and $d_G(x, \gamma) < \varepsilon$. Moreover if x is diagonal then γ has $h(\gamma) \leqslant 4 \log_5({}^{1}\!/\!_{\varepsilon})$.

(d) There is a probabilistic algorithm with expected running time poly $\log({}^{1}\!/\!_{\varepsilon})$ which for x diagonal produces (if it stops) the γ of smallest height in $B_G(x, \varepsilon)$ (assuming that one has a polynomial time factoring algorithm) [RS'14].

(ii) $\underline{p = 2}$ ([Ch'92]):

$S = \{s_0, s_1, s_1^{-1}\}$, where

$$s_0 = \begin{pmatrix} i & \\ & -i \end{pmatrix}, \qquad\qquad s_1 = \frac{1}{\sqrt{32}} \begin{pmatrix} 2 + i\sqrt{2} & \sqrt{26}i \\ \sqrt{26}i & 2 - i\sqrt{2} \end{pmatrix}.$$

The weights $w(s_j)$ are all chosen to be 1. $\Gamma \cong \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}$ and $h(\gamma)$ is reduced word length in s_0 ($= s_0^{-1}$) and s_1 (in PSU(2)). $|V(t)| = 3 \cdot 2^{t-1}$.

(a)

$$\frac{4}{3} \leqslant K(S) \leqslant 2.$$

The last asserts (essentially) that for $\varepsilon > 0$ and $x \in G$ there is a $\gamma \in G$ with $h(\gamma) \leqslant 6 \log_5(1/\varepsilon)$ and $d_G(x, \gamma) < \varepsilon$, while there are y's in G for which $d_G(y, \gamma) < \varepsilon \implies h(\gamma) \geqslant 4 \log_5(1/\varepsilon)$.

(b) For $\varepsilon > 0$ there is t_ε with

$$\frac{\log|V_{t_\varepsilon}|}{\log \varepsilon^{-3}} \to 1 \text{ as } \varepsilon \to 0,$$

such that for most points $y \in G$ w.r.t. Haar measure, there is an ε-approximation of y of height t_ε (so optimal for most y's).

(c) There is a $\operatorname{poly}\log(1/\varepsilon)$ time algorithm which assuming some reasonable conjectures about the distribution of primes produces for any $x \in G$ a $\gamma \in \Gamma$ with $h(\gamma) \leqslant 12 \log_5(1/\varepsilon)$ and $d_G(x, \gamma) < \varepsilon$. Moreover if x is diagonal then γ has $h(\gamma) \leqslant 4 \log_5(1/\varepsilon)$.

(d) There is a probabilistic algorithm with expected running time $\operatorname{poly}\log(1/\varepsilon)$ which for x diagonal produces (if it stops) the γ of smallest height in $B_G(x, \varepsilon)$ (assuming that one has a polynomial time factoring algorithm).

(iii) H–T gates ([AM'08, BS'12, KMM'12, RS'14]):

$S = \{H, T\}$

$H = \frac{i}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ the "Hadamard gate"

$T = \begin{pmatrix} e^{i\pi/8} & \\ & e^{-i\pi/8} \end{pmatrix}$ the "$\pi/8$ gate."

Gates such as H which lie in the finite Clifford group C ($C = \langle H, T^2 \rangle \cong S_4$ is a quotient of C_1 the usual Clifford group of order 192, since we work here in $\mathrm{PSU}(2)$) are apparently much easier to prepare. To obtain a universal gate set one has to add an element that is expensive. The T-gate which is of order 8 is a popular one to add and whose implementation has been studied. Set $w(c) = 0$ for $c \in C$ and $w(T) = 1$ (the weights can be chosen as in [BS'12]), so that $h(\gamma)$ only counts applications of the T gates. $|V_t| \sim c2^t$.

(a)

$$\frac{4}{3} \leqslant K(S) \leqslant 2.$$

The last asserts (essentially) that for $\varepsilon > 0$ and $x \in G$ there is a $\gamma \in G$ with $h(\gamma) \leqslant 6 \log_5(1/\varepsilon)$ and $d_G(x, \gamma) < \varepsilon$, while there are y's in G for which $d_G(y, \gamma) < \varepsilon \implies h(\gamma) \geqslant 4 \log_5(1/\varepsilon)$.

(b) For $\varepsilon > 0$ there is t_ε with

$$\frac{\log|V_{t_\varepsilon}|}{\log \varepsilon^{-3}} \to 1 \text{ as } \varepsilon \to 0,$$

such that for most points $y \in G$ w.r.t. Haar measure, there is an ε-approximation of y of height t_ε (so optimal for most y's).

(c) There is a $\operatorname{poly}\log(1/\varepsilon)$ time algorithm which assuming some reasonable conjectures about the distribution of primes produces for any $x \in G$ a $\gamma \in \Gamma$ with $h(\gamma) \leqslant 12\log_5(1/\varepsilon)$ and $d_G(x,\gamma) < \varepsilon$. Moreover if x is diagonal then γ has $h(\gamma) \leqslant 4\log_5(1/\varepsilon)$.

(d) There is a probabilistic algorithm with expected running time $\operatorname{poly}\log(1/\varepsilon)$ which for x diagonal produces (if it stops) the γ of smallest height in $B_G(x,\varepsilon)$ (assuming that one has a polynomial time factoring algorithm).

Discussion

The source for (i) is the Hamilton quaternion algebra D (that is generated by $1, i, j, k$ with $i^2 = j^2 = -1$, $ij = -ji = k, \dots$) over $\mathbb{Q}$. D is split at $p = 5$ (one could use any other $p \neq 2$ just as well) and is ramified at the real place ∞ and at 2. So $D \otimes \mathbb{Q}_5 \cong \mathbb{Q}_5^{2\times 2}$ and if $D^*(\mathbb{Z}[1/5])$ consists of all elements in D whose entries and those of their inverses are in $\mathbb{Z}[1/5]$, then the diagonal embedding corresponding to $p = 5$ and ∞

$$D^*(\mathbb{Z}[1/5]) \hookrightarrow \mathrm{PGL}_2(\mathbb{Q}_5) \times (\mathrm{SU}(2) \times \mathbb{R}^\times)$$

has its image in $\mathrm{SU}(2)$ equal to Γ in (i). At the same time from the first factor $D^*(\mathbb{Z}[1/5])$ acts isometrically on the 6-regular tree $X = \mathrm{PGL}_2(\mathbb{Q}_5)/\mathrm{PGL}_2(\mathbb{Z}_5)$ [LPS]. In fact as shown in [LPS] it acts simply transitively on X and $h(\gamma)$ is simply the distance in X from $\gamma\xi$ to ξ, where $\xi = \mathrm{PGL}_2(\mathbb{Z}_5)$. This allows one to use automorphic forms to analyze the equidistribution in G of the points in V_t which correspond to a ball in X of radius t [LPS]. The emphasis there was to establish sharp equidistribution, but for the purpose of the covering number one can improve the exponent by a positivity argument. This was done in the thesis [Ch'95] in a somewhat different setting. Since this leads to the best bound that we have for $K(S)$ namely $K(S) \leqslant 2$, we give a variation of the argument in Appendix 1. The number 2 is a reflection of the square root cancellation coming from the Ramanujan conjectures (Deligne's theorem). That is we need $|V(t_\varepsilon)|$ to be the square of $\frac{1}{\mu(B_G(\varepsilon))}$ in order to ensure that V_{t_ε} meets every such ball. In Appendix 1 we also show that this square root feature leads to most points having optimally short circuit approximations.

That $K(S) \geqslant 4/3$ is more elementary and is a consequence of the points in V_t having big holes near the projections onto the unit sphere S^3 ($\cong \mathrm{SU}(2)$ metrically) of integer points in $\mathbb{Z}^4$. This big hole phenomenon was first observed in the context of approximating real 2×2 matrices of unit determinant by projections of ones of determinant p [Ha'90].

In our setting of (i) the points V_t correspond to integer solutions of

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 5^h, h \leqslant t. \tag{9}$$

Claim 10. If $0 \neq y \in \mathbb{Z}^4$ then the ball $B_{S^3}\left(\frac{y}{|y|}, \frac{1}{10(|y|5^{t/2})^{1/2}}\right)$ contains a ball in S^3 of radius $\frac{1}{20t(|y|5^{t/2})^{1/2}}$ which has no points of V_t.

We give a proof of (10) in Appendix 2. That $K(S) \geqslant 4/3$ follows immediately from (10).

The analysis of (ii) is similar. This time the gate set S comes from the quaternion algebra $D/\mathbb{Q}$ where $i^2 = -2$, $j^2 = -13$. D is ramified at ∞ and $p = 13$. So in this case one can localize at $p = 2$ and form the group $D^*(\mathbb{Z}[1/2])$. In [Ch'92] it is shown that $D^*(\mathbb{Z}[1/2])$ acts simply transitively on the 3-regular tree $\mathrm{PGL}_2(\mathbb{Q}_2)/\,\mathrm{PGL}_2(\mathbb{Z}_2)$. The generating set S corresponds again to the 3-neighbors of $\mathrm{PGL}_2(\mathbb{Z}_2)$. The rest of the analysis is the same as in example (i).

We turn to (iii) which has been investigated directly and quite intensively recently ([KMM'12, Se'12, RS'14], ...). These H–T gates also come from a quaternion algebra and this allows us, among other things, to conclude that $K(S) \leqslant 2$. The quaternion algebra is again the Hamilton quaternions D, but this time we consider it over the field $k = \mathbb{Q}(\sqrt{2})$. k has two archimedian places at both of which D is ramified, giving the algebra $H(\mathbb{R})$. We fix one of these real places. The prime 2 is ramified in k, $(2) = (\sqrt{2})^2 := P^2$. $k_P = \mathbb{Q}_2(\sqrt{5})$, and D is split over k_P ($\sqrt{-7} + 2i + \sqrt{2}j + k$ is a zero divisor in $D_P := D \otimes k_P$; note $\sqrt{7} \in \mathbb{Q}_2 \subset k_P$). In fact D/k is ramified only at its two infinite places. Let $\mathcal{O}$ be the ring of integers of k and O_P the integers in k_P. The key arithmetic group is $\Delta = D^*(O[1/2])$. Under the diagonal embedding

$$\Delta = D^*(O[1/2]) \hookrightarrow \mathrm{PGL}_2(k_P) \times (\mathrm{SU}(2) \times \mathbb{R}^\times) \tag{11}$$

(using our chosen real place), Δ is mapped to a subgroup of $\mathrm{SU}(2)$. We show that Δ is equal to $\Gamma = \langle H, T \rangle$.

Let

$$\tilde{H} = \frac{i+k}{\sqrt{2}} \text{ and } \tilde{T} = \frac{2+\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i. \tag{12}$$

Then

$$\mathrm{Norm}(\tilde{H}) = 1 \text{ and } \mathrm{Norm}(\tilde{T}) = \sqrt{2}(1+\sqrt{2}). \tag{13}$$

Hence $\tilde{H}$ and $\tilde{T}$ are in Δ.

Also $\tilde{T}^2 = (1+\sqrt{2})(1+i)$ so that under the usual identifications of $H(\mathbb{R})^\times/R^\times$ with $\mathrm{SU}(2)$ ($x_1 + x_2 i + x_3 j + x_4 k \longmapsto \begin{pmatrix} x_1 + ix_2 & x_3 + ix_4 \\ -x_3 + ix_4 & x_1 - ix_2 \end{pmatrix}$) $\tilde{H}$ corresponds to H and $\tilde{T}$ to T.

It follows that $\langle \tilde{H}, \tilde{T} \rangle \subset \Delta$. One way of seeing that these groups are equal is to consider their image under the usual double cover $\pi : D^* \longrightarrow \mathrm{SO}_f$ (over k) where $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ over k. Now $\pi(\Delta) \subset \mathrm{SO}_f(\mathcal{O}[1/2])$ and as shown in [Ser'11]* the last group is the quaquaversal group $G(4,8)$ generated by rotations R_4 and R_8 of orders 4 and 8 about orthogonal axes (and also $G(4,8) \cong S_4 \underset{D_4}{*} D_8$). Moreover $\pi(\tilde{T}) = R_8$ and $\pi(\tilde{H}\tilde{T}^2\tilde{H}^{-1}) = R_4$. Hence $G(4,8) \subset \pi(\Delta)$ and so $\pi(\Delta) = \mathrm{SO}_f(\mathcal{O}[1/2])$ and $\Gamma = \Delta$. To exploit this realization of Γ as the arithmetic group $D^*(\mathcal{O}[1/2])$ we need to relate the height $h(\gamma)$ in terms of T counts to the action of γ on the 3-regular tree $X = \mathrm{PGL}_2(k_P) / \mathrm{PGL}_2(\mathcal{O}_P)$ under the embedding (11). The elements in D_P of the form

$$\mathcal{O}_P \frac{1+i+j+k}{2} + \mathcal{O}_P \frac{1+i}{2} + \mathcal{O}_P + \mathcal{O}_P i + \mathcal{O}_P j + \mathcal{O}_P k$$

form a maximal order M_P of $\mathcal{O}_P$. Note that $\tilde{H}$ and $\tilde{T}^2$ are units in M_P, so that the Clifford group $C = \langle \tilde{H}, \tilde{T}^2 \rangle$ is contained in the unit group of M_P. It follows that C stabilizes $\xi = \mathrm{PGL}_2(\mathcal{O}_P)$ in X. On the other hand since $\mathrm{Norm}(\tilde{T}) = (\sqrt{2})$ it follows that $\tilde{T}$ moves ξ to one of its neighbors in X and $\tilde{H}\tilde{T}$ and $\tilde{T}^{-1}$ move ξ to the other two nearest neighbors. Thus Δ acts transitively on X and the height of any $\gamma \in \Delta$ in terms of $\tilde{T}$ and $\tilde{T}^{-1}$ counts, is it most the distance from ξ to $\gamma\xi$ in X plus $O(1)$. One can also use T-counts alone as is done in [BS'12, RS'14], where one treats the gates in C as being free. Again the distance in X from ξ to $\gamma\xi$ is essentially this T-count.

With this analysis we are in the same position as in examples (i) and (ii) and everything there applies equally well for the H–T gates (the Ramanujan conjectures for definite division algebras over totally real number fields are known).

One can also study this example (iii) in terms of an arithmetic unitary group. This view point is perhaps even more natural when examining gates sets such as the cyclotomic-Clifford gates in [FGKM'15]. Let $n \geqslant 3$ and E the "CM-field" $\mathbb{Q}(\zeta_{2n})$ and F the corresponding totally real subfield $\mathbb{Q}(\zeta_{2n} + \zeta_{2n}^{-1})$, here ζ_{2n} is a primitive $2n$th root of 1. The unitary group U/F consists of all 2×2 matrices with entries in E, which preserve the hermitian form $h(u,w) = \overline{u_1}w_1 + \overline{u_2}w_2$, where $u = (u_1, u_2)$, $w = (w_1, w_2)$ and $\bar{\cdot}$ is Galois conjugation E/F. As a group over F one can localize U to places ν of F. Denote by F_ν the completion of F at ν and by U_ν the corresponding unitary group over F_ν (its elements are matrices in $E_\nu = E \otimes F_\nu$). If ν is split in E then U_ν is isomorphic to $\mathrm{GL}_2(F_\nu)$, while if ν is inert or ramified in E then U_ν is a genuine unitary group over F_ν with corresponding field extension E_ν. At the archimedean places σ of F (all such places are real) U_σ is a definite (compact) unitary group. The group $\mathrm{U}_2(R_n)$ in [FGKM'15] is the full S-arithmetic group $\mathrm{U}(\mathcal{O}[1/2])$ where $\mathcal{O}$ is the ring of integers of F. To see this as an S-arithmetic group, factor 2 in $\mathcal{O}$; $(2) = (P_1 \cdots P_g)^e$ where $efg = \deg[F : G]$, e is the ramification and f the degree of the residue fields of any F_{P_j}. In this way $\mathrm{U}_2(R_n)$ is the S-arithmetic group consisting of

*in answer to a question posed in [Ro'06].

all elements in $\mathrm{U}(F)$ which are integral outside of $S = \{P_1, \ldots, P_g\}$. The local groups U_{P_j}, $j = 1, \ldots, g$, are noncompact and isomorphic. The diagonal embedding

$$\mathrm{U}_2(R_n) \hookrightarrow \mathrm{U}_{P_1} \times \cdots \times \mathrm{U}_{P_g} \tag{$*$}$$

realizes $\mathrm{U}_2(R_n)$ as a cocompact lattice in the latter group. The gate group $\mathfrak{g}_n$ in [FGKM'15] is the group generated by the finite Clifford group C and the element $T_n = \begin{pmatrix} 1 & \\ & e^{i\pi/n} \end{pmatrix}$. Clearly $\mathfrak{g}_n \leqslant \mathrm{U}_2(R_n)$ and we want to know when these are equal or at least when does $\mathfrak{g}_n$ have finite index in $\mathrm{U}_2(R_n)$; that is whether $\mathfrak{g}_n$ is thin or not in the sense of [Sa'14].

The first thing to not is that if $g > 1$ then $\mathfrak{g}_n$ is thin. Indeed as explained in [FGKM'15], $\mathfrak{g}_n$ has a presentation as a simple amalgamated product. It follows from rigidity theorems for lattices in different groups, namely ones in a rank one type group (such as an amalgamated product as above) and ones in a higher rank gruop, cannot be isomorphic. So if $g > 1$ so that the product group in ($*$) is of higher rank, then $\mathfrak{g}_n$ cannot be a lattice and so it is thin.

One can analyze the factorization of 2 in $\mathcal{O}_F$ using Dirichlet characters and one finds that if $2n = 2^k s$ with s odd, then $g = 1$ iff $\{-1, 2\}$ generated $(\mathbb{Z}/s\mathbb{Z})^\times$ and $f = |(\mathbb{Z}/s\mathbb{Z})^\times/(\pm 1)|$. On the other hand [FGKM'15] show that the intersection of $\mathfrak{g}_n$ with the diagonal torus $A = \left\{ \begin{pmatrix} \star & \\ & \star \end{pmatrix} \in \mathrm{U} \right\}$ is finite, while that of $\mathrm{U}_2(R_n)$ with A is finite iff -1 is in the group generated by 2 in $(\mathbb{Z}/s\mathbb{Z})^\times$. Combining these we have that if $\mathfrak{g}_n$ is arithmetic then 2 must generate $(\mathbb{Z}/s\mathbb{Z})^\times$.

Restricting to the cases where $\mathfrak{g}_n$ might be arithmetic, let p be the unique prime in F above 2. U_p is a split 2-dimensional unitary group over F_p (since we are not assuming that $\mathfrak{g}_n$ is infinite, U_p cannot be definite). The projective group PU_p is isomorphic to $\mathrm{PGL}_2(F_p)$. Hence $\mathrm{PU}_2(R_n)$ is realized as a lattice in $\mathrm{PGL}_2(F_p)$ and it acts discontinuously and isometrically on the $2^T + 1$ regular tree $X = \mathrm{PGL}_2(F_p) / \mathrm{PGL}_2(\mathcal{O}_p)$. In this setting deciding anything about $\mathfrak{g}_n$ such as a presentation, or whether it acts with a compact quotient (i.e. it is arithmetic since $\mathrm{U}_2(R_n)$ does so) can be done in any given instance using the techniques in [Ser'77]. In fact using these he shows in [Ser'11] that in the case $s = 1$, for $n = 2, 4, 8$ $\mathfrak{g}_n$ is arithmetic (his generators are H and T_n but since H and T_2 generate C, his group is the same as $\mathfrak{g}_n$) and in fact $\mathfrak{g}_n = \mathrm{U}_2(R_n)$ while if $n = 2^\nu$ with $\nu \geqslant 4$ then $\mathfrak{g}_n$ is thin (for the latter he compares Euler characteristics and estimates these using Tamagawa numbers, i.e. arithmetic formulae for the volumes of $\mathrm{U}_2(R_n) \backslash X$). Presumably one can proceed similarly in the remaining cases when $s \neq 1$. However given the normal forms for members of $\mathfrak{g}_n$ in [FGKM'15] it seems to me that one can complete the analysis by simply counting elements in $\mathfrak{g}_n$. Note that if $\mathfrak{g}_n$ were arithmetic then $|\{\gamma \in \mathfrak{g}_n : d_X(\gamma e, e) \leqslant t\}| \sim c_{\mathfrak{g}_n} 2^{ft}$ as $t \to \infty$, with $c_{\mathfrak{g}_n} > 0$. This follows from a lattice point count in the tree X. However according to the normal form [FGKM'15, Cor 4.2] and relating the T-count of $\gamma \in \mathfrak{g}_n$ to the distance it moves e in X (I didn't check that the relation is similar to the $n = 4$ case) one only has $O(2^t)$ such elements in $\mathfrak{g}_n$. So if $\mathfrak{g}_n$ is to be arithmetic then we must have

$f = 1$ and hence $s = 1$ or 3. If so the only possible new arithemtic cases are $n = 3 \cdot 2^\nu$ with $\nu \geqslant 1$. According to [FGKM'15] $\nu = 1$ and 2 are in fact arithmetic (and again $\mathfrak{g}_n = \mathrm{U}_2(R_n)$ in these cases). For ν larger I expect that again by counting the number of elements in $\mathfrak{g}_n$ with $d_X(\gamma e, e) \leqslant t$, the count will be too small (and certainly by estimating Tamagawa numbers).

In summary it is likely that the cases $n = 3, 4, 8, 12$ from [FGKM'15] are the full list of arithmetic Clifford-cyclotomic gates. As for the case $n = 4$ discussed in (iii) above, the Ramanujan conjectures are known for these unitary groups and (a), (b), (c), (d) holds for them.

The basic algorithm

We review the algorithm ([KMM'12, BGS'13, Se'12, RS'14]) for finding good approximations of small height. Consider the case (i) (the same ideas work in (ii) and (iii)). The elements in V_t are solutions to (9). A key point is that if we have an integral solution x to (9) which approximates some $y \in G$ suitably, then finding the short circuit for x in terms of the gates S can be done quickly (i.e. polynomial in $t = O(\log(1/\varepsilon))$). The action of Γ on the 6-regular tree X tells us how to move γ of height h to the identity coset ξ using the gates; one simply navigates along the geodesics in X. An equivalent way of achieving this pointed out in [BGS'13] is to use the left and right unique factorization theory in $D(\mathbb{Z})$. Thus the problem of finding a short circuit is reduced to finding solutions to (9) with $x/5^{h/2}$ doing the required approximation. The idea is that for a diagonal matrix $Z = \begin{pmatrix} \alpha & \\ & \overline{\alpha} \end{pmatrix}$, $|\alpha| = 1$ one can find a good approximation as follows:

We seek x of height t such that x satisfies (9) and

$$|x_1 - \xi_1| \leqslant 5^{t/2}\varepsilon, \qquad |x_2 - \xi_2| \leqslant 5^{t/2}\varepsilon, \qquad |x_3| \leqslant 5^{t/2}\varepsilon, \qquad |x_4| \leqslant 5^{t/2}\varepsilon \tag{14'}$$

where

$$\xi_1^2 + \xi_2^2 = 5^t \qquad \left(\alpha = \frac{\xi_1 + i\xi_2}{5^{t/2}}\right). \tag{14}$$

Choose

$$x_1 = \lfloor \xi_1 \rfloor - k_1, \quad x_2 = \lfloor \xi_2 \rfloor - k_2, \quad \text{with } 0 \leqslant k_1, k_2 \leqslant t^A \text{ for some fixed } A > 1. \tag{15}$$

Then

$$0 < 5^t - (x_1^2 + x_2^2) \leqslant 2t^A 5^{t/2}. \tag{16}$$

For each choice of k_1, k_2 as above we check if $5^t - x_1^2 - x_2^2$ is a prime $p \equiv 1 \pmod 4$. This can be done in poly t steps [AKS'04] and if this happens then one can find x_3, x_4 in poly t steps [Sc'85], satisfying:

$$x_3^2 + x_4^2 = 5^t - x_1^2 - x_2^2. \tag{17}$$

In this case we arrive at a solution to (14) with $\varepsilon = 2t^{A/2}/5^{t/4}$. That is we arrive at a point of height (essentially) $4\log_5({}^1/_\varepsilon)$. If our choice of x_1, x_2 fails to produce such a p we repeat with different choices of k_1, k_2 and do so t^{2A} (poly t) times. Given the density of the distribution of primes we expect that if A is big enough (but fixed) that one will always arrive at such a prime p. So assuming this heuristic about the distributino of primes of this form, one will find (in poly t steps) an ε approximation to z of height at most $4\log_5 {}^1/_\varepsilon$. To approximate the general $y \in G$ one factors y as $r_1r_2r_3$ with the r's diagonal about different orthogonal axes and approximates each r_j by the above algorithm. This leads to the $12\log_5 {}^1/_\varepsilon$ length circuit approximating y.

The probabilistic algorithm searches (quickly) for x_1, x_2 in the region defined by (14′) and $5^t - (x_1^2 + x_2^2) \leqslant \varepsilon^2 5^t$. We won't describe it further except to point out that it is in solving (17) for x_3, x_4 in poly t steps that one needs to assume that the right hand side of (17) can be factored quickly. In the H–T gate case (iii) this involves a search for integers in k in certain regions [RS'14]. The probabilistic algorithm apparently works very well in practice as illustrated with some runs in [RS'14].

Intrinsic Diophantine Approximation

In these arithmetic cases the problem of finding a global solution to (9) which approximates a given $g \in G$ is an extension of the problem of approximating points on spheres by rational and 'S-integral' points of small height. These are problems of intrinsic diophantine approximation that have been studied recently.

(a) The classic case of rational numbers ${}^a/_q$ in $[0, 1]$ with the height of ${}^a/_q$ being q. V_t consists of the Farey fractions of denominator at most t. There are $\sim t^2$ such points and their spacing vary from ${}^1/_t$ to ${}^1/_{t^2}$. Hence the exponent K is

$$K([0, 1], \text{Farey}) = 2.$$

The fact that most points in $[0, 1]$ can be approximated to within ${}^1/_{t^2}$ by these Farey points while at the same time there are points that cannot be approximated so well (i.e. $K \neq 1$) is the source of the rich theory of diophantine approximation.

(b) Consider the rational points in $[0, 1]$ which are '2-integral' that is the points $\mathbb{Z}[{}^1/_2] \cap [0, 1]$.

This time the 2^t points of height at most t are exactly 2^{-t} apart in $[0, 1]$. Thus the exponent is

$$K([0, 1], \mathbb{Z}[{}^1/_2]) = 1.$$

The theory of diophantine approximation here is simply the truncation of the dyadic expansion of a real number. So the theory is very simple.

The following concern the standard unit sphere S^n in $\mathbb{R}^{n+1}$ ($n > 0$) and subsets of rational points on S^n which we denote by Q.

(c) $Q = \left\{ {}^{x}/_{q} : X \text{ is integral}, q \geqslant 1, |{}^{x}/_{q}|^2 = 1 \right\}$
$h({}^{x}/_{q}) = q$ if $(x_1, \ldots, x_{n+1}, q) = 1$. The recent work [KM'13] shows that

$$K(S^n, Q) \leqslant 2.$$

The repulsion property (10) for these points takes a simpler and transparent form

$$d_{S^n}\left(\frac{x}{q}, \frac{x'}{q'}\right) \geqslant \frac{1}{(qq')^{1/2}}. \tag{18}$$

From this it follows that

$$K(S^n, Q) = 2,$$

and here too there is a rich theory of diophantine approximation.

(d) Consider S^1 with rational points which are say 5-integral that is in $\mathbb{Z}[{}^{1}/_{5}]$. These correspond to solutions of

$$x^2 + y^2 = 5^{2h}, \qquad\qquad x, y \in \mathbb{Z}.$$

The question of what $K(S^1, \mathbb{Z}[{}^{1}/_{5}])$ is reduces to the diophantine properties of the generator of these rational points, namely the numbers in $m\alpha \pmod{2\pi\mathbb{Z}}$, $m \geqslant 1$, where

$$\alpha = \cos^{-1}({}^{3}/_{5}).$$

If as we expect this number is a typical irrational in terms of its diophantine properties then

$$K(S^1, \mathbb{Z}[{}^{1}/_{5}]) = 1.$$

On the other hand if α is not typical then $K > 1$. I am not sure what if anything is known about the type of the number ${}^{\alpha}/_{2\pi}$.

(e) S^2 with the points in $\mathbb{Z}[{}^{1}/_{5}]$. This is a very interesting case and corresponds to solutions to

$$x_1^2 + x_2^2 + x_3^2 = 5^{2t}. \tag{19}$$

Using automorphic forms (again the Ramanujan conjectures) exploiting that the right hand side of (19) is a square, one can show that

$$K(S^2, \mathbb{Z}[{}^{1}/_{5}]) \leqslant 2. \tag{20}$$

In the recent paper [BRS'12] the local statistics of the distribution on S^2 of sums of 3-squares is investigated. The thesis there is that these behave like a random set of points placed on S^2, at least if the r.h.s. of (19) is squarefree. I expect that this also holds for (19) and hence that

$$K(S^2, \mathbb{Z}[1/5]) = 1.$$

If true this would be striking.

(f) S^3 with points in $\mathbb{Z}[1/5]$. That is solutions to

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 5^{2h}, \qquad h \leqslant t. \tag{21}$$

This is very close to (i) except that the exponent of 5 in (21) is restricted to be even. As discussed in (i) we have

$$\frac{4}{3} \leqslant K(S^3, \mathbb{Z}[1/5]) \leqslant 2. \tag{22}$$

So this is the smallest dimension for S^n for which there is definitely a rich theory of intrinsic diophantine approximation for 'S-integers'. An in depth study of such problems in great generality has been carried out in [GNN'12]. In particular they define local covering exponents which in this case are given as follows:

For $x \in S^3$ and $\varepsilon > 0$ let $t_\varepsilon(x)$ be the least t such that $B_{S^3}(x, \varepsilon)$ contains an element in S^3 with coordinates in $\mathbb{Z}[1/5]$ and of height 5^h with $h \leqslant t$. Set

$$K(x, \mathbb{Z}[1/5]) := \limsup_{\varepsilon \longrightarrow 0} \frac{\log |V_{t_\varepsilon}(x)|}{\log(1/\varepsilon^3)}.$$

They show that for almost all x (w.r.t. Lebesgue measure on S^3)

$$K(x, \mathbb{Z}[1/5]) = 1. \tag{23}$$

This of course is optimal (again the Ramanujan conjectures are an ingredient).

(g) For $n \geqslant 4$ the determination of K for S-arithmetic points on S^n becomes easier. As pointed out to me by Bourgain, an application of the circle method as done in [BR'12, Appendix] can be promoted to an asymptotic (and not just an upper bound) when one has this many variables, for counting these points in small caps. As a consequence we have together with the repulsion (18) that for $n \geqslant 4$,

$$K(S^n, \mathbb{Z}[1/5]) = 2 - \frac{2}{n}. \tag{24}$$

We end with some basic problems:

(1) For the arithmetic S's in (i), (ii) and (iii) to show that $K(S) < 2$ and perhaps even that $K(S) = 4/3$.

(2) Is there a set of gates with $K(S) = 1$? Perhaps with 'thin' subgroups (see [Sa'14]) where all we know is that $K(S) < \infty$, there may be some cases where $K(S) = 1$.

(3) For the arithmetic gates S we know that for most points $y \in G$ there are optimally short circuits approximating them. Find a probabilistic algorithm with expected running time poly $\log(1/\varepsilon)$, which finds such short circuits. So for the H–T gates this means an algorithm to find for most y's in G and $\varepsilon > 0$ a circuit of T-count at most $(1+\delta)3\log_2(1/\varepsilon)$ which ε-approximates y (δ arbitrarily small). In practice achieving optimality for most y's in G, while being only slightly worse for singular y's, would render these arithmetic S's to be Golden Gates.

Appendix 1

We work in the context of a Hecke orbit on S^2 coming from the points on a ball in the p-regular tree, as in [LPS]. The analysis for $\mathrm{SU}(2)$ or more generally the setting of Hecke orbits in cases where the Ramanujan conjectures are known, is similar.

Let $V_t = \{S \in O_f\}$ (with $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$) be a set of representatives for the Hecke points. One can find an orthonormal basis ϕ_j of $L^2(S^2)$ of Hecke eigenfunctions (each is also a spherical harmonic that is an eigenfunction of the Laplacian Δ_{S^2}). If the Hecke orbit in S^2 comes from V_t applied to a fixed point $x_0 \in S^2$ then

$$\sum_{s \in V_t} \phi_j(sx_0) = \lambda_j(t)\phi_j(x_0) \qquad \phi_0 = \frac{1}{\sqrt{4\pi}} \text{ and } \lambda_0(t) = |V_t|. \tag{25}$$

$|V_t| = 6 \cdot 5^{t-1}$ for case (i) and for simplicity we stick to this case. The Ramanujan conjectures imply that

$$\left|\lambda_j(t)\right| \leqslant t|V_t|^{1/2} \qquad \text{for } j \neq 0. \tag{26}$$

Let $k_\varepsilon(x, y)$ be a point-pair invariant on S^2, that is $k_\varepsilon(\sigma x, \sigma y) = k_\varepsilon(x, y)$ for $\sigma \in O_f$, which is nonnegative and an approximation to the identity under $*$ ($k_1 * k_2(x, y) = \int_{S^2} k_1(x, z)k_2(z, y)\, \mathrm{d}A(y)$). Specifically

$$\begin{aligned} &k_\varepsilon(x, y) \geqslant 0 \\ &\int_{S^2} k_\varepsilon(x, y)\, \mathrm{d}A(y) = 1 \\ &k_\varepsilon(x, y) = 0 \qquad\qquad \text{if } d_{S^2}(x, y) > \varepsilon. \end{aligned}$$

We can also choose k to be positive-definite (i.e. $h_k(t) \geqslant 0$ where h_k is the spherical transform of k at a nonnegative eigenvalue t of Δ_{S^2}), by taking $k = k_{\varepsilon/2} * k_{\varepsilon/2}$ for example. Moreover we can choose k_ε so that

$$k_\varepsilon(x, x) \leqslant \frac{c}{\varepsilon^2} \qquad \text{for a fixed constant } c. \tag{27}$$

Expanding $k_\varepsilon(x,y)$ in the basis ϕ_j yields [Se'56]:

$$k_\varepsilon(x,y) = \sum_{j=0}^{\infty} h_{k_\varepsilon}(t_j)\phi_j(x)\phi_j(y). \tag{28}$$

Applying (25) with $x = x_0$ gives

$$\sum_{s\in V_t} k_\varepsilon(sx_0,y) = \frac{|V_t|}{\sqrt{4\pi}} + \sum_{j=1}^{\infty} h(t_j)\lambda_j(t)\phi_j(x_0)\phi_j(y). \tag{29}$$

If for some $y \in S^2$ $\sum\limits_{s\in V_t} k_\varepsilon(sx_0,y) = 0$, then from (29) and (26)

$$\begin{aligned}\frac{|V_t|}{\sqrt{4\pi}} &\leqslant \sum_{j=1}^{\infty} h(t_j)|\lambda_j(t)||\phi_j(x_0)||\phi_j(y)| \\ &\leqslant |V_t|^{1/2} t \sum_{j=1}^{\infty} h(t_j)\frac{|\phi_j(x_0)|^2 + |\phi_j(y)|^2}{2} \\ &\leqslant |V_t|^{1/2} t k_\varepsilon(z,z) \\ &\leqslant |V_t|^{1/2} t \frac{c}{\varepsilon^2}\end{aligned} \tag{30}$$

using (27).

That is

$$|V_t| \leqslant \frac{4\pi c t^2}{\varepsilon^4}. \tag{31}$$

Hence if $|V_t| > \frac{4\pi ct}{\varepsilon^4}$ then for every y, $\sum\limits_{s\in V_t} k_\varepsilon(sx_0,y) > 0$ and in particular $d_{S^2}(sx_0,y) < \varepsilon$ for some $s \in V_t$. This proves that $K \leqslant 2$ in this case, and in fact this sharper and non-asymptotic form.

To see that most y's have optimally good approximation by members of V_t we compute the variance over y in (29)

$$\begin{aligned}\int_{S^2}\left|\sum_{s\in V_t} k_\varepsilon(sx_0,y) - \frac{|V_t|}{\sqrt{4\pi}}\right|^2 \mathrm{d}A(y) &= \sum_{j=1}^{\infty} h_\varepsilon(t_j)^2|\lambda_j(t)|^2|\phi_j(x_0)|^2 \\ &\leqslant t^2|V_t|\sum_{j=1}^{\infty} h_\varepsilon(t_j)^2|\phi_j(x_0)|^2 \\ &\leqslant t^2|V_t|\int_{S^2}|k_\varepsilon(x_0,y)|^2\,\mathrm{d}A(y)\end{aligned} \tag{32}$$

(using (28)).

Now from (27) we get that

$$\int_{S^2} \left| \sum_{s \in V_t} k_\varepsilon(sx_0, y) - \frac{|V_t|}{\sqrt{4\pi}} \right|^2 \mathrm{d}A(y) \leqslant \frac{c^2 t^2 |V_t|}{\varepsilon^2}. \tag{33}$$

If $B = \left\{ y : \sum_{s \in V_t} k_\varepsilon(sx_0, y) = 0 \right\}$ then $B \supset \{y : d_{S^2}(sx_0, y) = 0 \text{ for all } s \in V_t\}$. According to (33)

$$\mu_{S^2}(B) \frac{|V_t|^2}{4\pi} \leqslant \frac{c^2 t^2 |V_t|}{\varepsilon^2}$$

or

$$\mu_{S^2}(B) \leqslant \frac{4\pi c^2 t^2}{\varepsilon^2 |V_t|}. \tag{34}$$

This is a quantitative and explicit form of the fact that B has a small measure if $|V_t|$ is somewhat larger than $\frac{1}{\mu_{S^2}(B)}$.

Appendix 2

Let $x, y \in \mathbb{Z}^4$ both not 0 and $\frac{x}{|x|} \neq \frac{y}{|y|}$. Set $|x|^2 = M$ and $|y|^2 = N$. Then

$$\begin{aligned} d^2_{S^3}\left(\frac{x}{|x|}, \frac{y}{|y|} \right) &= 2 - \frac{2\langle x, y \rangle}{\sqrt{MN}} \\ &= 2\left(\frac{\sqrt{MN} - \langle x, y \rangle}{\sqrt{MN}} \right). \end{aligned} \tag{34}$$

If $\frac{x}{|x|}$ and $\frac{y}{|y|}$ are close ethen $\langle x, y \rangle > 0$ and

$$\begin{aligned} d^2_{S^3}\left(\frac{x}{|x|}, \frac{y}{|y|} \right) &= \frac{2(MN - \langle x, y \rangle^2)}{\sqrt{MN}(\sqrt{MN} + \langle x, y \rangle)} \\ &= \frac{MN - \langle x, y \rangle^2}{MN}. \end{aligned} \tag{35}$$

In the interval $\left[MN - \sqrt{MN}, MN \right]$ there is at most one square number t_{MN}. It follows that if

$$\langle x, y \rangle^2 \neq t_{MN} \text{ then } d^2_{S^3}\left(\frac{x}{|x|}, \frac{y}{|y|} \right) \geqslant \frac{1}{\sqrt{MN}}. \tag{36}$$

So given a y as above then for x with $|x|^2 = M$, either $d\left(\frac{y}{|y|}, \frac{x}{|x|} \right) = \zeta_{MN}$ with

$$0 < \zeta_{MN} < \frac{1}{(MN)^{1/4}} \text{ or } d\left(\frac{x}{|x|}, \frac{y}{|y|} \right) \geqslant \frac{1}{(MN)^{1/4}}. \tag{37}$$

For $h \leqslant t$ and x varying over

$$|x|^2 = 5^h,$$

and our given y, we have either

$$d\left(\frac{x}{|x|}, \frac{y}{|y|}\right) = \xi_h, \qquad 0 < \xi_h < \frac{1}{(|y|^2 5^t)^{1/4}}, \qquad \text{or} \qquad d\left(\frac{x}{|x|}, \frac{y}{|y|}\right) \geqslant \frac{1}{(|y|^2 5^t)^{1/4}}.$$

Hence there is an annulus $A_{\alpha,\beta}$ about the point $\frac{y}{|y|}$ in S^3

$$A_{\alpha,\beta} = \left\{ z \in S^3 : \alpha \leqslant d\left(\frac{y}{|y|}, z\right) \leqslant \beta \right\}$$

with $\beta \leqslant \frac{1}{(|y|^2 5^t)^{1/4}}$ and $\beta - \alpha \geqslant \frac{1}{2t(|y|^2 5^t)^{1/4}}$, which is free of any $x \in V_t$. Choosing a maximal radius ball B in this annulus yields (10).

References

[AKS'04] Agrawal, Kayal, Saxena, *Annals of Mathematics*, 160 (2004) 781–793

[AM'08] Amano, Matsumoto, arXiv:0806.3834

[BG'08] Bourgain, Gamburd, *Inventiones Mathematicae* 171 (2008) 83–121

[BR'12] Bourgainm Rudnick, *Geometric and Functional Analysis* 22 (2012)

[BRS'12] Bourgain, Rudnick, Sarnak, arXiv:1204.0134

[BS'12] Bocharov, Svore, arXiv:1206.3223

[BGS'13] Bocharov, Gurevich, Svore, arXiv:1303.1411

[Ch'92] Chiu, *Combinatorica* 12(3) (1992) 275–285

[Ch'95] Chiu, *Journal of Number Theory* 53 (1995) 25–44

[DN'06] Dawson, Nielsen, *Quantum Information & Computation* 6 (2006) 81–95

[FGKM'15] Forest, Gosset, Kliuchnikov, McKinnon, arXiv:1501.04944

[GNN'12] Ghosh, Gorodnik, Nevo, arXiv:1205.4426

[KM'13] Kleinbock, Merrill, arXiv:1301.0989

[Ha'90] Harmon, *Journal of Number Theory* 34 (1990) 63–81

[KMM'12] Kliuchnikov, Maslov, Mosca, arXiv:1212.6964 and arXiv:1212.0822

[LPS] Lubotzky, Phillips, Sarnak, *Communications of Pure and Applied Math* XXXIX (1986) 149–186 and XL (1987) 401–420

[NC'00] Nielsen, Chuang, "Quantum computation and quantum information" *Cambridge University Press* (2000)

[Ro'06] Robinson, *Journal of Algebra* 306 (2006) 201–207

[RS'14] Ross, Selinger, arXiv:1403.2975

[Sa'14] Sarnak, "Notes on thin matrix groups" *MSRI Publications* 61 (2014) 343–362.

[Sc'85] Schoof, *Mathematics of Computation* 44 (1985) 483–494

[Se'12] Selinger, arXiv:1212.6253

[Se'56] Selberg, *Journal of the Indian Mathematical Society* 20 (1956) 47–87

[Ser'11] Serre, "Le groupe quaquaversal, ou comme group S-arithmetique" *Oberwolfach Report* 2011

[Ser'77] Serre, "Arbes, amalgamues, SL_2" *Asterique* 46 (1977)

Regards
Peter Sarnak