On Golden Gates and Discrepancy: Examining the Efficiency of Universal Gate Sets

Brent A.W. Mode*

Department of Physics and Astronomy, University of Lousiville, Louisville, KY

Advisor: Dr. Steven Damelin

Abstract

Quantum computation is of great interest to physicists both as a technical challenge and for its potential use in the simulation of quantum systems. Further, implementation of Shor's algorithm for factoring large numbers in polynomial time will render many of the most common encryption methods obsolete. While experimentalists struggle to create quantum computers with more than a handful of qubits, the problem of choosing efficient sets of quantum logic gates to implement remains to be fully understood. The most direct way to consider the problem is in terms of sets of gates considered as unitary matrices in $SU(2)$. However, with regards to determining the efficiency of gate sets, it is more useful to consider the problem of discrepancy by mapping gates in $SU(2)$ to points on S^3 and considering covering radius. With support from the REU program at the University of Michigan, I investigated this problem in a variety of settings and coauthored an article on the subject [2].

1 Background Mathematics

In the interest of completely approachable elucidation of the research herein, this section is an optional introduction to most of the mathematical machinery needed to study the problem at hand. On the other hand, many important sets and pieces of notation are introduced here, so it might be wise to quickly review this section. Largely the information in this section is adapted from [2].

Definition. A quantum bit or qubit is a linear combination of the states $|0\rangle$ and $|1\rangle$ that is a unit vector in $\mathbb{C}^2$. In terms of an arbitrary state, we have

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, |\alpha|^2 + |\beta|^2 = 1$$

*bamode01@louisville.edu

Definition. A 1-qubit quantum gate acts linearly on a qubit to make it a new qubit. To note, there are several quantum gates that can be used to reconstruct classical logic universally. So,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \to |\psi'\rangle = \alpha'|0\rangle + \beta'|1\rangle$$

This is the same action as a matrix on a vector, so let $X \in GL_2(\mathbb{C})$ and then an equivalent statement would be $X|\psi\rangle = |\psi'\rangle$ where $|\psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$. More specifically, since $|\psi'\rangle$ mut still be a unit vector, we require that $X \in U(2)$.

The unitary group is defined as $U(n) = \{X \in GL_n(\mathbb{C}) \mid X^\dagger X = I\}$. In addition, we can consider the special unitary group to be an equally appropriate collection of all possible quantum gates where $SU(n) = \{X \in U(n) \mid \det X = 1\}$. Finally, we are also sometimes interested in the projective special unitary group, which is not isomorphic to $SU(2)$ but can also serve as the collection of arbitrary 1-qubit quantum gates, $PSU(n) \equiv SU(n)/Z(SU(n))$, where for an arbitrary group G, Z(G) is the center of G.

So, the set of quantum gates is most simply $U(2)$. Up to a constant, we can restrict ourselves to $SU(2)$ by taking elements $X \in U(2)$ and considering them equivalent to $\frac{X}{\sqrt{\det X}} \in SU(2)$. Similarly, we could further restrict ourselves to $PSU(2)$. This restriction can simplify a question of efficiency, though for producing quantum circuits, it is more common to consider the gates as $SU(2)$. It is also of note, and should be intuitively clear, that $PU(2) \equiv U(2)/Z(U(2)) \approx PSU(2)$. Again, it is also the case that $PSU(2)$ and $SU(2)$ are not isomporphic.

The problem of constructing an efficienct universal subset of quantum gates can be reduced to finding an efficient universal subst of $PSU(2)$.

1.1 Topological Structure of $PSU(2)$

Definition. Let X be a set and τ be a collection of subsets of X. A topological space follows the following axioms:

1. $\emptyset, X \in \tau$
2. Any union of elements of τ are in τ.
3. The intersection of a finite number of elements of τ is in τ

where τ is called a topology.

Definition. Let S be a subset of a topological space X. A point $x \in X$ is a limit point of S if every neighborhood of x contains at least one point in S.

Definition. A subset A of a topological space X is called dense if $\forall x \in X$:

1. $x \in A$ or

2. x is a limit point of A.

Definition. Let G be a group, and $S \leq G$, such that $S = \{s_1, s_2, \cdots, s_n\}$. Let $\Gamma = \langle S \rangle$. S is called universal if Γ is dense in G.

Definition. A nonempty subset $S \leq G$ is symmetric if $S = S^{-1}$ where $S^{-1} = \{s^{-1} \mid s \in S\}$.

Let $G = PSU(2)$ and $S = \{s_1, s_2, \cdots, s_n\}$ be a symmetric universal subset which generates the group Γ dense in G.

Definition. We define a weight function $w : S \to \mathbb{R}_{\geq 0}$. The function $w(s_i)$ is nonnegative on S and assigns a cost to each $s \in S$.

Definition. Next we define a height function to give the weight a reduced word $h : \Gamma \to \mathbb{R}_{\geq 0}$, defined as

$$h(\gamma) = \min\{\sum_{k=1}^{m} w(x_k) \mid \gamma = x_1 x_2 \cdots x_m, x_k \in S\}$$

For each $\gamma \in \Gamma$, we find that there are multiple ways of constructing γ from s_i, and we would like to find the minimum cost method of constructing each γ.

Example. Let $w(s_i) = 1$ for $1 \leq i \leq n, s_i \in S$. For $\gamma = s_1 s_2 \in \Gamma$, the cost is $w(s_1) + w(s_2) = 2$. But $\gamma = s_1 s_1^{-1} s_1 s_2$ with cost 4. We prefer the former construction and note that, assuming $s_2 \neq s_1^{-1}$, $h(\gamma) \leq 2$.

Using these concepts, we would now like to define a few important sets.

Definition. Using our height function, we can define the set with specific height

$$U(t) = \{\gamma \in \Gamma \mid h(\gamma) = t\}.$$

This is the subset of Γ containing every element of height exactly t.

Definition. We can also define the more important set with bounded height

$$V(t) = \{\gamma \in \Gamma \mid h(\gamma) \leq t\} = \bigcup_{i=0}^{t} (U(i)).$$

This is the subset of Γ containing every element of height less than or equal to t. We will use these kinds of subsets to approximate any $g \in G$.

1.2 Metric Structure of $PSU(2)$

We also wish to make use of the metric properties of $G = PSU(2)$.

Definition. A metric or distance function on a set X is defined as $d : X \times X \to \mathbb{R}_{\geq 0}$ satisfying $\forall x, y, z \in X$:

1. $d(x,y) \geq 0$
2. $d(x,y) = 0 \Leftrightarrow x = y$
3. $d(x,y) = d(y,x)$
4. $d(x,z) \leq d(x,y) + d(y,z)$.

Definition. We define the following metric on $G = PSU(2)$ by $d : G \times G \to \mathbb{R}_{\geq 0}$

$$d_G(X,Y) = \sqrt{1 - \frac{|Tr(X^\dagger Y)|}{2}}, \forall X, Y \in G.$$

Lemma. The metric $d(x,y)$ is an invariant metric in G, i.e. $\forall h \in G$ and $\forall x, y \in G$ held fixed, $d(hx, hy) = d(xh, yh) = d(x,y)$.

Our goal is to use the elements of $V(t) \subseteq G$ to approximate each point $x \in G$. Further we want that our approximation not have error exceeding some given value $\varepsilon > 0$. That is, $\forall x \in G, \exists \gamma \in V(t) \ni d(x,\gamma) < \varepsilon$ and we use γ to approximate x. Alternatively, $\forall x \in G, \exists \gamma \in V(t) \ni x$ is within the ball centered at γ with radius ε, i.e. $x \in B(\gamma, \varepsilon)$.

Definition. A ball in a metric space is defined such that $B(\gamma, \varepsilon) = \{x \in G \mid d(x,y) < \varepsilon\}$.

For a given $\varepsilon > 0$ let $t = t_\varepsilon > 0$ be the smallest t such that $V(t)$ can be used to approximate all of $G = PSU(2)$. That is, G is covered by balls such that

$$G = \bigcup_{\gamma \in V(t_\varepsilon)} B(\gamma, \varepsilon).$$

Theorem (Solovay-Kitaev)**.** Let an $\varepsilon > 0$ and a symmetric, universal gate set of $SU(2)$ be given. Then there exists a constant c such that $\forall X \in SU(2), \exists \gamma$, a finite sequence of gates from the universal gates set of length $O\left(\log^c \frac{1}{\varepsilon}\right) \ni d(X,\gamma) < \varepsilon$.

However, we do not know how efficient this approximation is. The Solovay-Kitaev theorem guarantees the existence of an approximation of all points but does not guarantee any particular efficiency. Further, the constant c gives you a computation time, but does not function well as a measure of the efficiency of any particular universal gate set. We find that we would like to define a measure of efficiency that is more robust. A recent examination of the theorem can be found in [4]. A discussion of the algorithm used by the theorem is found in [3].

1.3 Measure Structure of $PSU(2)$

Definition. Let X be a set and $\mathcal{P}(X)$ be the power set of X. Then $\Sigma \subseteq \mathcal{P}(X)$ is called a σ-algebra if it satisfies the following:

1. $X \in \Sigma$
2. $\forall A \in \Sigma, X - A \in \Sigma$
3. $\forall A_1, A_2, \cdots \in \Sigma, A_1 \cup A_2 \cup \cdots \in \Sigma$

The elements of a σ-algebra are called measurable sets.

Definition. In a topological space X, a Borel set is any set that can be formed from open sets using countable unions, countable intersections, and relative complements. The collection of all Borel sets on X forms a σ-algebra called the Borel algebra. Further, the Borel algebra is the smallest algebra containing all open sets.

Definition. In a metric space (X, d), compactness is equivalent to the statement that every infinite subset of X has at least one limit point in X. Similarly, a compact group is a group whose topology is compact.

$PSU(2)$ is a compact group.

Definition. Let G be a compact group. A normalized Haar measure $\mu : \Sigma \to \mathbb{R}_{\geq 0}$ on G where Σ is the Borel algebra of G satisfies:

1. $\mu(G) = 1$
2. $\forall x \in G$ and $S \in \Sigma, \mu(xS) = \mu(S)$

Let μ be a normalized Haar measure on $G = PSU(2)$. For an $\varepsilon > 0$, let $t_\varepsilon > 0$ be the smallest t such that

$$G \subset \bigcup_{\gamma \in V(t_\varepsilon)} B(\gamma, \varepsilon).$$

Hence, $\mu\left(\bigcup_{\gamma \in V(t_\varepsilon)} B(\gamma, \varepsilon)\right) = \mu(G) = 1$.

Lemma. Let $t_\varepsilon > 0$ be the smallest t such that $G \subset \bigcup_{\gamma \in V(t_\varepsilon)} B(\gamma, \varepsilon)$. Then, $|V(t_\varepsilon)| \cdot \mu(B(\varepsilon)) \geq 1$.

It would also be beneficial to know what the value of $\mu(B(\varepsilon))$ is. For the Haar measure on $G = SU(2)$, we find that it is much better to consider a measure that it is rotation-invariant on the three-sphere, S^3. To that end, we define μ as

$$\mu(\theta, \phi, \varphi) = 2 \sin \theta \mathrm{d}\theta \mathrm{d}\phi \mathrm{d}\varphi.$$

Hence, $\mu(B(\varepsilon)) \sim c\varepsilon^3$, when ε is small.

From the previous result, we have that $|V(t_\varepsilon)| \geqslant \dfrac{1}{\mu(B(\varepsilon))}$ for any ε and universal subset S. This leads us to three equivalent statements:

1. $|V(t)|$ and $\dfrac{1}{\mu(B(\varepsilon))}$ become closer,

2. The intersection between balls decreases,

3. The approximation is more efficient.

We express the efficiency of an approximation with universal subset S by

$$K(S) \equiv \limsup_{\varepsilon \to 0} \frac{\log|V(t_\varepsilon)|}{\log\left(\frac{1}{\mu(B(\varepsilon))}\right)} \sim \limsup_{\varepsilon \to 0} \frac{\log|V(t_\varepsilon)|}{\log\left(\frac{1}{\varepsilon^3}\right)}.$$

Since $|V(t_\varepsilon)| \geqslant \frac{1}{\mu(B(\varepsilon))}$, $K(S) \geqslant 1$. If $K(S) = 1$, the universal gate set S optimally approximates all of G. We would like to find S with as small a $K(S)$ as possible.

2 The Universal Gate Set T

My work has focused specifically on examining the universal gate set $T = \{s_1, s_2, s_3, s_1^{-1}, s_2^{-1}, s_3^{-1}, I, iX, iY, iZ\}$, where $s_1 = \frac{1}{\sqrt{5}}(I + 2iX), s_2 = \frac{1}{\sqrt{5}}(I + 2iY), s_3 = \frac{1}{\sqrt{5}}(I + 2iZ)$, and X, Y, and Z are the Pauli matrices. This set was recently studied in [1]. The elements iX, iY, and iZ have weight 0, while all other elements have weight 1. It is proven by Sarnak in [5] that the following lower and upper bounds exist on efficiency for T:

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

More specifically, Sarnak utilizes the set $S = T - \{iX, iY, iZ\}$ [5]. One interesting result of my research is that Sarnak's proof of $K(S) \geqslant \frac{4}{3}$ contains an important mistake that necessitates the use of T. This will be presented shortly.

Let $\Omega = \langle T \rangle$ be the group dense in $SU(2)$ generated by T and $V_T(t)$ be the set of all elements of Ω with height no more than t. By considering the nature of elements of $SU(2)$, it can be shown that points in $V_T(t)$ can be represented as integer solutions to the equation

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 5^l, l \leqslant t.$$

These integer solutions than can then be projected as points on the sphere S^3 [1]. Sarnak's proof that $K(S) \geqslant \frac{4}{3}$ relies on this correspondence. However, there are solutions to this equation that can be shown not to be in the group generated by S.

Counterexample. Consider the solution for $t = 1$ given by $(-2, 1, 0, 0)$. The corresponding matrix in $SU(2)$ is

$$\frac{1}{\sqrt{5}} \begin{pmatrix} -2+i & 0 \\ 0 & -2-i \end{pmatrix} = (iX)s_1.$$

Since $iX \notin S$, we have that there is no such bijection. i.e. $(-2, 1, 0, 0)$ does not correspond to an element of $V_S(t)$, although it does correspond to the given element of $V_T(t)$.

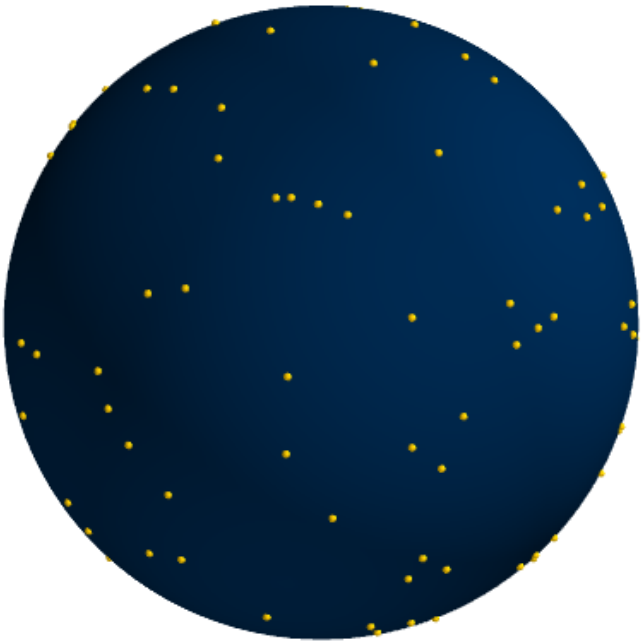

Figure 1: The points of $V_T(2)$, in Maize and Blue.

Also consider that the points of $V_T(t)$ must have a representation in $PSU(2)$ preserving the cardinality of the set, since both $SU(2)$ and $PSU(2)$ are valid ways to examine the points. Then note that $PSU(2) \approx SO(3)$, the rotation group on S^2. So under a proper homomorphism, the points of $V_T(t)$ can be plotted on the surface of S^2 as shown in Figures 1, 2, and 3.

3 A Conjecture on Mesh Norm

We conjecture a certain mesh norm on the points in $V_T(t_\varepsilon)$ [2]. Then we show that this is sufficient to prove that $K(T) = \frac{4}{3}$.

Definition. The mesh norm or covering radius of a point set with respect to S^d is given as

$$M(\mathcal{N}) \equiv \max_{y \in S^d} \min_{x \in \mathcal{N}} |x - y|$$

where $\mathcal{N}$ is the point set in question. Intuitively, the mesh norm is the radius that is required for balls centered at points of $\mathcal{N}$ to cover all of S^d.

Let $\nu(5^{t_\varepsilon})$ denote the set of integer solutions of the quadratic form: $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 5^{t_\varepsilon}$.Let $M \equiv M_{S^3}(\mathcal{N})$ denote the covering radius of the points $\mathcal{N} = \nu(5^{t_\varepsilon}) \cup \nu(5^{t_\varepsilon - 1})$ on the sphere S^3 in

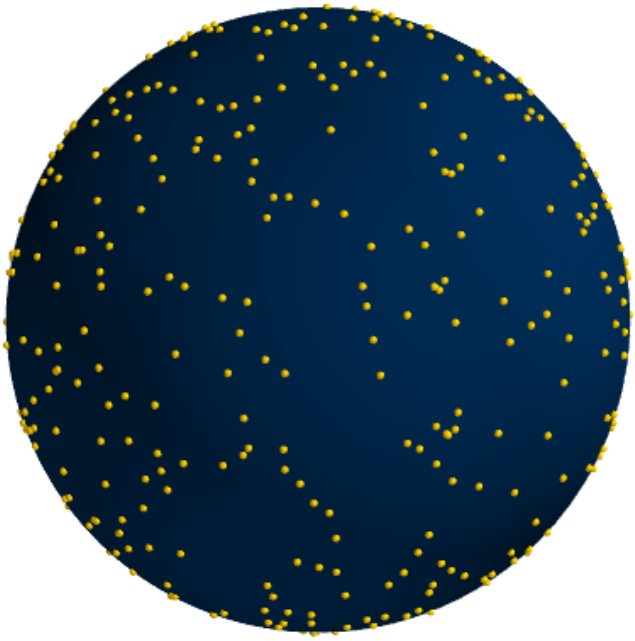

Figure 2: The points of $V_T(3)$.

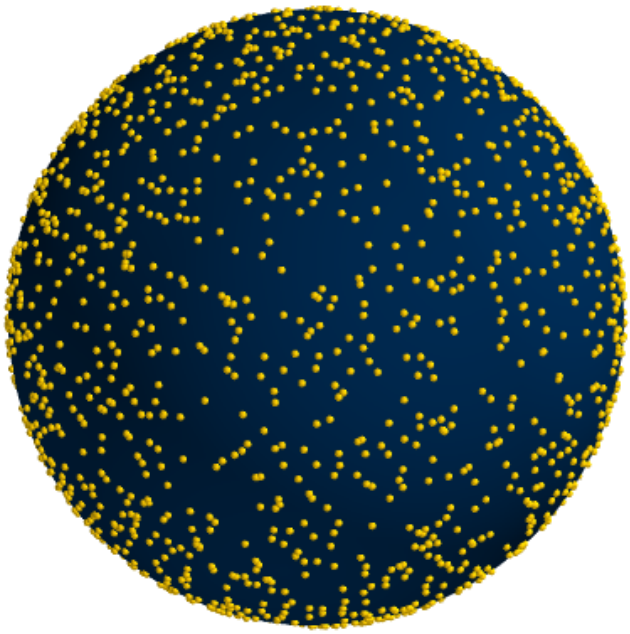

Figure 3: The points of $V_T(4)$.

$\mathbb{R}^4$.

Conjecture A. *We conjecture* $\varepsilon \leq f(t_\varepsilon)5^{-t_\varepsilon/4}$ *for a function* $f : (0,\infty) \to (1,\infty)$ *satisfying:*

$$\lim_{t_\varepsilon \to \infty} log(f(t_\varepsilon))/t_\varepsilon$$

exists with value 0. *Then* $M \sim f(\log N)N^{-1/4}$. *Here* $N \equiv N(\varepsilon) = 6 \cdot 5^{t_\varepsilon} - 2$.

Assuming this conjecture, we prove the following theorem. The proof is taken from [2].

Theorem 1. $K(T) \leqslant \dfrac{4}{3}$

Proof. Let $\varepsilon > 0$ and all log be $\log_5$, for convenience only. Then,

$$\begin{aligned}
K(T) &= \limsup_{\varepsilon \to 0} \frac{\log|V_T(t_\varepsilon)|}{\log\left(\dfrac{1}{\mu\left(B_G(\varepsilon)\right)}\right)} \\
&= \limsup_{\varepsilon \to 0} \frac{\log|V_T(t_\varepsilon)|}{\log\left(\dfrac{1}{\varepsilon^3}\right)} \\
&= \lim_{\varepsilon \to 0} \frac{\log(6 \cdot 5^{t_\varepsilon} - 2)}{3\log\left(\dfrac{1}{\varepsilon}\right)} \\
&\leqslant \lim_{\varepsilon \to 0} \frac{t_\varepsilon + \log\left(6 - \dfrac{2}{5^{t_\varepsilon}}\right)}{3\log\left(\dfrac{1}{f(t_\varepsilon)} \cdot 5^{\frac{t_\varepsilon}{4}}\right)} \\
&= \lim_{\varepsilon \to 0} \frac{t_\varepsilon + \log\left(6 - \dfrac{2}{5^{t_\varepsilon}}\right)}{\dfrac{3}{4}t_\varepsilon - 3\log(f(t_\varepsilon))} \\
&= \lim_{\varepsilon \to 0} \left[\frac{1}{\dfrac{3}{4} - \dfrac{3\log f(t_\varepsilon)}{t_\varepsilon}} + \frac{\log\left(6 - \dfrac{2}{5^{t_\varepsilon}}\right)}{\dfrac{3}{4}t_\varepsilon - 3\log(f(t_\varepsilon))} \right].
\end{aligned}$$

Since $t_\varepsilon \to \infty$ as $\varepsilon \to 0$, we have that

$$\begin{aligned}
K(T) &\leqslant \lim_{t_\varepsilon \to \infty} \left[\frac{1}{\dfrac{3}{4} - \dfrac{3\log(f(t_\varepsilon))}{t_\varepsilon}} + \frac{\log\left(6 - \dfrac{2}{5^{t_\varepsilon}}\right)}{\dfrac{3}{4}t_\varepsilon - 3\log(f(t_\varepsilon))} \right] \\
&= \frac{1}{\dfrac{3}{4} - 0} + 0 \\
&= \frac{4}{3}.
\end{aligned}$$

Therefore,

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

□

We now provide two examples, one of such a function that disappears as required and satisfies the conjecture, and another that does not.

Example. With $t_\varepsilon \sim \log(N)$, consider:

$$f(t_\varepsilon) = t_\varepsilon^{(\log(t_\varepsilon^{\log(t_\varepsilon^{\cdot^{\cdot^{\cdot}}})}))}$$

where the term $\log(t_\varepsilon)$ is nested n times. Then easily we have

$$\log(f(t_\varepsilon))/t_\varepsilon \sim \frac{(\log(\log N))^{n+1}}{\log N}$$

which decays to 0 for large enough N.

Example. On the other hand for a function which grows faster, say

$$f(t_\varepsilon) = t_\varepsilon^{t_\varepsilon}$$

we easily have

$$\log(f(t_\varepsilon))/t_\varepsilon \sim (\log(\log N))$$

which diverges for large enough N.

4 Numerical Experiments on Mesh Norm

One of the my goals after drafting [2] was to calculate mesh norms for $V_T(t)$ for as high a t as computationally accessible. This turned out to not be a very large t, since N grows as 5^t, and the number of requisite computations with it. Given below in Table 1 are mesh norms for various t, the number of points N, and the mesh norm as an exponent, η, such that $M = N^{-\eta}$.

t	N	M	η
2	147	0.46	0.155
3	762	0.21	0.235
4	3857	0.117	0.26

Table 1: Numerical experiments on mesh norm.

The trend noticed here suggests that the some form of the previous conjecture might be true, though providing proof has shown to be a difficult problem.

5 Continued Research

While the contents of this paper describe the bulk of the work undertaken during the REU program, it is not a sole representation of the attempts of myself and Dr. Damelin to make progress. We have also considered attempting to redefine the covering exponent in terms of some kind of discrepancy as well as reformulating the problem as a conjecture on minimal energy points. We are discussing at least two more papers currently, one on minimal energy, and another presenting the problem to the physics community.

6 Acknowledgments

I would like to think Dr. Damelin for his collaboration and illuminating conversations. This work was supported by the NSF as part of an REU.

References

[1] A. Bocharov, Y. Gurevich, and K. Svore. "Efficient decomposition of single-qubit gates into V basis circuits," Phys. Rev. A 88.1 (2013): 012313.

[2] S.B. Damelin, Q. Liang, B.A.W. Mode, "On Golden Gates and Discrepancy," (2017) preprint arXiv:1506.05785, submitted to Journal of Complexity.

[3] C.M. Dawson, M.A. Nielsen, "The Solovay-Kitaev Algorithm," QIC, Vol 6, No 1 (2006), pp 081-095.

[4] N. Ross and P. Selinger, "Optimal ancilla-free Clifford+T approximation of z-rotations," QIC **16**(2016) (11-12), pp 901-953.

[5] P. Sarnak, "Letter to Scott Aaronson and Andy Pollington on the Solovay-Kitaev Theorem and Golden Gates," `http://publications.ias.edu/sarnak/paper/2637` (2015).