

Simplified Proof of the Fourier Sampling Theorem

Peter Høyer *

BRICS †

May 29, 2000

Abstract

We give a short and simple proof of Hales and Hallgren's Fourier Sampling Theorem [“Quantum Fourier Sampling Simplified”, *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, ACM Press, May 1999]. The transparency of our proof-technique allows us to generalize and tighten their result.

1 Introduction

In the recent years, the theory of quantum computing has been greatly developed and expanded. Two of the most striking results in the area are Grover's algorithm for searching [3] and Shor's algorithms for factoring and finding discrete logarithms [6]. For an excellent introduction to quantum computing, see for example [2].

Any quantum algorithm works on a finite Hilbert space \mathcal{H} . Two types of operations are allowed, the first is unitary operators on \mathcal{H} , the second is measurements of the whole or parts of the system. Since we are interested in

* BRICS, Department of Computer Science, University of Aarhus, DK-8000 Århus C, Denmark. email: hoyer@brics.dk.

† Basic Research in Computer Science, Centre of the Danish National Research Foundation.

the computational complexity of the algorithms, we restrict the operations allowed to only those that can be implemented efficiently.

One of the primary operators used in the quantum algorithms developed so far, is the quantum Fourier transform. Two of its main uses are to set up a quantum system in an initial state and to perform quantum Fourier sampling. The quantum Fourier transform is actually not a single operator, but a family of operators. One can define Fourier transforms for any finite group G . If the group G is Abelian, then there exists exactly one Fourier transform for G , and if G is non-commutative, then there are infinitely many Fourier transforms for G .

For every integer $n \geq 1$, the quantum Fourier transform over the cyclic group \mathbb{Z}_n is defined by

$$\mathbf{F}_n = \frac{1}{\sqrt{n}} \sum_{i,j=0}^{n-1} \omega_n^{ij} |i\rangle \langle j|, \quad (1)$$

where $\omega_n = \exp(2\pi\sqrt{-1}/n)$ denotes the n th principal root of unity. *Quantum Fourier sampling over \mathbb{Z}_n* is, given a superposition $|u\rangle = \sum_{i=0}^{n-1} u_i |i\rangle$, apply the Fourier transform \mathbf{F}_n and measure the resulting superposition [1]. The measurement induces a discrete probability distribution \mathcal{D} over the possible outcomes $\{0, 1, \dots, n-1\}$ where the probability for outcome i is $|\langle i | \mathbf{F}_n | u \rangle|^2$.

We have here adapted the Dirac notation that is commonly used in quantum algorithms. The “ket” notation $| \rangle$ is used to easily identify vectors from the Hilbert space, and the “bra” notation $\langle |$ is similarly used to easily identify functionals from the dual space. For the purpose of this paper, one may think of these objects in terms of matrices. Then the ket $|i\rangle$ can be identified with the $n \times 1$ column vector of all zeroes but a one at the i th entry. Similarly, the bra $\langle j|$ can be identified with the $1 \times n$ row vector of all zeroes but a one at the j th entry. Please see [2] for further information.

In many quantum algorithms, including Shor’s celebrated algorithms for factoring and discrete logarithms [6], quantum Fourier sampling is an essential ingredient. But unfortunately, often quantum Fourier sampling involves at least one of two difficulties: either we do not know the order n of the group over which we would like to perform the sampling, or the order n is known, but it has large prime factors, complicating efficient implementations of the Fourier transform [5].

Overcoming these two difficulties has in earlier work been based on an intriguing idea of Shor [6]: Instead of performing quantum Fourier sampling

over \mathbb{Z}_n , perform quantum Fourier sampling over \mathbb{Z}_m for some m sufficiently large compared to n . This idea, however, adds complications to the analysis of the modified algorithm; for example, one then has to show that the relevant data is still attainable via sampling from the modified distribution \mathcal{D}' .

Recently, Hales and Hallgren [4] proposed a general technique for circumventing such complications. They showed that, for any input state $|u\rangle$, the original distribution \mathcal{D} is contained in the modified distribution \mathcal{D}' by restriction. This allows us to sample from \mathcal{D} via sampling from \mathcal{D}' . We first explain the notation involved and then we state their theorem.

Let $1 < N < M$ be integers. For any integer $0 \leq i < N$, let $i' = \lfloor iM/N + 1/2 \rfloor$ denote a closest integer to iM/N , and set $\delta_i = i' - iM/N$. Note that $|\delta_i| \leq 1/2$. Given an input state $|u\rangle = \sum_{i=0}^{N-1} u_i |i\rangle$, set $|v\rangle = \mathbf{F}_N |u\rangle$ and $|w\rangle = \mathbf{F}_M |u\rangle$.

Let $\mathcal{D}_v : \{0, \dots, N-1\} \rightarrow [0, 1]$ denote the probability distribution induced by measuring $|v\rangle$, that is, $\mathcal{D}_v(i) = |\langle i|v\rangle|^2$. Define probability distribution $\mathcal{D}_w : \{0, \dots, M-1\} \rightarrow [0, 1]$ similarly. Let $\mathcal{D}_{w'} : \{0, \dots, N-1\} \rightarrow [0, 1]$ denote the probability distribution defined by $\mathcal{D}_{w'}(i) = c \cdot \mathcal{D}_w(i')$, where $c = (\sum_{i=0}^{N-1} \mathcal{D}_w(i'))^{-1}$ is the normalization factor. Thus, we obtain distribution $\mathcal{D}_{w'}$ by restricting \mathcal{D}_w to outcomes j for which $j = i'$ for some $0 \leq i < N$, and then relabeling i' by i . Finally, for any two probability distributions \mathcal{D} and \mathcal{D}' over $\{0, \dots, N-1\}$, let $|\mathcal{D} - \mathcal{D}'| = \sum_{i=0}^{N-1} |\mathcal{D}(i) - \mathcal{D}'(i)|$ denote their total variation distance.

Theorem 1 (Hales and Hallgren) *For any polynomial $s(n)$, there exists a polynomial $t(n)$ such that for all integers $N \leq 2^n$ and $M \geq t(n)N$, and all input states $|u\rangle = \sum_{i=0}^{N-1} u_i |i\rangle$, we have*

$$|\mathcal{D}_v - \mathcal{D}_{w'}| \leq \frac{1}{s(n)},$$

where $|v\rangle = \mathbf{F}_N |u\rangle$ and $|w\rangle = \mathbf{F}_M |u\rangle$.

Hales and Hallgren's theorem says that we, for all $0 \leq i < N$, can match the probability of measuring $|v\rangle$ yields i with the probability of measuring $|w\rangle$ yields i' , up to a global normalization factor. In the next section, we give a short proof of their theorem, or rather, we give a short proof of a generalization of their theorem. We improve upon their result in two ways. Firstly, we show that we can also match the *amplitudes*, not only the probabilities,

and secondly, we show that in Theorem 1, it suffices to pick $t(n)$ to be on the order of $s(n)n$.

The applications of Hales and Hallgren's theorem are many. For instance, it allows a simplified proof of Shor's theorem for factoring (see Section 3 of [4]). When applying their theorem, we would use the Fourier transform \mathbf{F}_M instead of \mathbf{F}_N . We set up the input state $|u\rangle$, apply \mathbf{F}_M and then measure the system. We repeat this experiment until the measurement produces an outcome j such that $j = i'$ for some $0 \leq i < N$. When that happens, we output i and stop. By Theorem 2 below, the expected number of repetitions is on the order of $\frac{M}{N}$. Furthermore, by Theorem 2 below, it suffices to pick M to be on the order of $N \log_2(N)$, in which case the expected number of repetitions is on the order of $\log_2(N)$.

2 A Simple Proof

The key object in our proof is the operator

$$\mathbf{A} = \sqrt{\frac{M}{N}} \mathbf{R} \mathbf{F}_M \mathbf{F}_N^{-1}, \quad (2)$$

where \mathbf{R} denotes the combined projection and permutation defined by

$$\mathbf{R} = \sum_{i=0}^{N-1} |i\rangle\langle i'|.$$

We use operator \mathbf{R} to rephrase Hales and Hallgren's theorem in terms of operators, and then using operator \mathbf{A} , we give a simple proof.

Theorem 2 *Let $N \geq 16$ and $s \geq 1$ be given. Then the following holds for all integers $M \geq s \cdot (12N \log_2(N))$. Let $|u\rangle = \sum_{i=0}^{N-1} u_i |i\rangle$ be any normalized state. Let $|v\rangle = \mathbf{F}_N |u\rangle$ and $|w'\rangle = c \mathbf{R} \mathbf{F}_M |u\rangle$ where $c > 0$ is the normalization factor such that $|w'\rangle$ has unit norm. Then*

$$\begin{aligned} \||v\rangle - |w'\rangle\| &\leq \frac{1}{s} \\ |\mathcal{D}_v - \mathcal{D}_{w'}| &\leq 4 \frac{1}{s}, \end{aligned}$$

and $c = \sqrt{\frac{M}{N}} (1 + O(\frac{1}{s}))$.

In the calculations to come, we use many inequalities and bounds. Several of these bounds are not tight as our primary aim is to give a simple and basic proof.

Operator \mathbf{A} is not necessarily unitary, but it is linear and can be written as a sum $\mathbf{A} = \sum_{i,j=0}^{N-1} a_{ij} |i\rangle\langle j|$ where $a_{ij} \in \mathbb{C}$ is given by

$$a_{ij} = \frac{1}{N} \sum_{k=0}^{N-1} \omega_N^{k(i-j)} \omega_M^{k\delta_i} = \frac{1}{N} \sum_{k=0}^{N-1} \omega_N^{k(i-j+\delta_i N/M)}. \quad (3)$$

Note that $|a_{ij}| \leq 1$ for all $0 \leq i, j < N$, that is, every coefficient has absolute value at most 1. The next lemma expresses that every diagonal element a_{ii} is close to 1, whereas every off-diagonal element a_{ij} ($i \neq j$) has small absolute value. The lemma is a variant of Claim 1 in [4].

Lemma 3 For operator $\mathbf{A} = \sum_{i,j=0}^{N-1} a_{ij} |i\rangle\langle j|$ given by Equation 2,

$$\begin{aligned} \operatorname{Re}(a_{ii}) &\geq 1 - 5\left(\frac{N}{M}\right)^2 \\ |a_{ij}| &\leq \frac{2}{|i-j|_N} \frac{N}{M} \quad (i \neq j), \end{aligned}$$

where

$$|x|_N = \begin{cases} x \bmod N & \text{if } (x \bmod N) \leq N/2 \\ (-x) \bmod N & \text{if } (x \bmod N) > N/2. \end{cases}$$

Proof First consider the diagonal element a_{ii} for some $0 \leq i < N$. Since $|\delta_i| \leq \frac{1}{2}$ then, for all $0 \leq k < N$, we have $\operatorname{Re}(\omega_M^{k\delta_i}) \geq \cos(\pi N/M) \geq 1 - 5\left(\frac{N}{M}\right)^2$. Thus, by Equation 3, it follows that $\operatorname{Re}(a_{ii}) \geq 1 - 5\left(\frac{N}{M}\right)^2$.

Now consider the off-diagonal element a_{ij} for some $0 \leq i, j < N$ with $i \neq j$. If $\delta_i = 0$ then $a_{ij} = 0$, so suppose otherwise. Using that the rightmost sum in Equation 3 is a geometric series, rewrite

$$a_{ij} = \frac{1}{N} \sum_{k=0}^{N-1} \omega_N^{k(i-j+\delta_i N/M)} = \frac{1}{N} \frac{1 - \omega_M^{\delta_i N}}{1 - \omega_N^{i-j+\delta_i N/M}}.$$

We upper bound the absolute value of the numerator in the above expression on the right, $|1 - \omega_M^{\delta_i N}| \leq \pi \frac{N}{M}$. To lower bound the absolute value of the

denominator, write $|1 - \omega_N^{i-j+\delta_i N/M}| = |\sin(\pi \frac{i-j}{N} + \delta_i \frac{\pi}{M})| \geq |\sin(\pi \frac{i-j}{N})| - \frac{\pi}{2} \frac{1}{M}$. For any real number x , we have that $\sin(\pi x) \geq 2|x|$ if $0 \leq |x| \leq \frac{1}{2}$ and, by symmetry, that $\sin(\pi x) \geq 2(1 - |x|)$ if $\frac{1}{2} \leq |x| \leq 1$. It follows that the absolute value of the denominator is lower bounded by $\frac{2}{N} |i - j|_N - \frac{\pi}{2} \frac{1}{M}$, allowing us to conclude that $|a_{ij}| \leq \frac{2}{|i-j|_N} \frac{N}{M}$ provided $M \geq 8N$. \square

Lemma 3 tells us that operator \mathbf{A} acts as the identity $\mathbf{I} = \sum_{i=0}^{N-1} |i\rangle\langle i|$, modulo some error terms. To analyze the “damage” caused by those error terms, write

$$\mathbf{A} = \mathbf{I} + \mathbf{E}.$$

Let $\mathbf{E} = \sum_{i,j=0}^{N-1} e_{ij} |i\rangle\langle j|$. Then $|e_{ii}|^2 = |\operatorname{Re}(a_{ii}) - 1|^2 + |\operatorname{Im}(a_{ii})|^2 = 1 + |a_{ii}|^2 - 2\operatorname{Re}(a_{ii}) \leq 10(\frac{N}{M})^2$, since $|a_{ii}| \leq 1$ by Equation 3. Hence $|e_{ii}| \leq \sqrt{10} \frac{N}{M}$ and $|e_{ij}| \leq \frac{2}{|i-j|_N} \frac{N}{M}$ for $i \neq j$.

Lemma 4 *For all states $|v\rangle$ of unit norm,*

$$\|\mathbf{E}|v\rangle\| \leq 3 \frac{N}{M} (2 + \log_2(N)).$$

Proof We prove this lemma by rephrasing it in terms of matrices and vectors, and then introduce a vector norm and its induced matrix norm which we easily can bound. Hence, consider \mathbf{E} an $N \times N$ matrix $(e_{ij})_{i,j=0}^{N-1}$, and let $\operatorname{Norm}(\cdot)$ denote the matrix norm defined by

$$\operatorname{Norm}(\mathbf{B}) = \max \{ \|\mathbf{B}x\|_2 : \|x\| = 1 \}$$

where $\|x\|_2 = (x^* \cdot x)^{1/2}$ denotes the Euclidean norm of the $N \times 1$ column vector x , and where x^* denotes the Hermitian adjoint of x . Then, clearly, we have that $\|\mathbf{E}|v\rangle\| \leq \operatorname{Norm}(\mathbf{E})$, and thus it suffices to upper bound the matrix norm of \mathbf{E} .

For this, note that $\operatorname{Norm}(\mathbf{E}) \leq \operatorname{Norm}(|\mathbf{E}|)$ where $|\mathbf{E}|$ denotes the matrix obtained by replacing each entry of \mathbf{E} with its absolute value. Observe that, by Lemma 3, we have $\operatorname{Norm}(|\mathbf{E}|) \leq \operatorname{Norm}(\mathbf{C})$ where $\mathbf{C} = (c_{ij})_{i,j=0}^{N-1}$ with

$$c_{ij} = \begin{cases} 6 \frac{N}{M} & \text{if } i = j \\ \frac{2}{|i-j|_N} \frac{N}{M} & \text{otherwise.} \end{cases}$$

Matrix \mathbf{C} is circulant with positive real-valued entries and hence its norm is equal the sum of any row or any column, $\text{Norm}(\mathbf{C}) = \sum_{j=0}^{N-1} c_{1j} = \sum_{i=0}^{N-1} c_{i1}$. We upper bound the leftmost sum

$$\sum_{j=0}^{N-1} c_{1j} = \frac{N}{M} \left(6 + 2 \sum_{j=1}^{N-1} \frac{1}{|j|_N} \right) \leq \frac{N}{M} \left(6 + 4 \sum_{j=1}^{\lfloor N/2 \rfloor} \frac{1}{j} \right) \leq \frac{N}{M} (6 + 4 \ln(N)).$$

Since $4 \ln(N) \leq 3 \log_2(N)$, the lemma follows. \square

Lemma 4 quantifies that operator \mathbf{A} is close to the identity. That bound implies a bound on the distance of the two states before and after applying \mathbf{A} .

Lemma 5 *Let $|v\rangle = \sum_{i=0}^{N-1} v_i |i\rangle$ be any normalized state. Let $|w'\rangle = \frac{1}{\sqrt{b}} \mathbf{A}|v\rangle$ where $b = \|\mathbf{A}|v\rangle\|$. Assume $\|\mathbf{E}|v\rangle\| \leq \frac{1}{2}$. Then $|w'\rangle$ has unit norm and*

$$\| |v\rangle - |w'\rangle \| \leq \frac{5}{2} \|\mathbf{E}|v\rangle\|.$$

Furthermore, the normalization factor is bounded by

$$1 - \frac{1}{2} \|\mathbf{E}|v\rangle\| \leq \frac{1}{\sqrt{b}} \leq 1 + \|\mathbf{E}|v\rangle\|. \quad (4)$$

Proof By definition, $b = \|\mathbf{A}|v\rangle\| = \|(\mathbf{I} + \mathbf{E})|v\rangle\|$, so $1 - \|\mathbf{E}|v\rangle\| \leq b \leq 1 + \|\mathbf{E}|v\rangle\|$. Equation 4 follows since we assume $\|\mathbf{E}|v\rangle\| \leq 1/2$.

Let $|y\rangle = |v\rangle - |w'\rangle$. Then $|y\rangle = (\mathbf{I} - \frac{1}{\sqrt{b}}(\mathbf{I} + \mathbf{E}))|v\rangle = (1 - \frac{1}{\sqrt{b}})|v\rangle - \frac{1}{\sqrt{b}}\mathbf{E}|v\rangle$. Hence, $\| |y\rangle \| \leq |1 - \frac{1}{\sqrt{b}}| + \frac{1}{\sqrt{b}} \|\mathbf{E}|v\rangle\| \leq \frac{5}{2} \|\mathbf{E}|v\rangle\|$. \square

Finally, we require a fundamental result of Bernstein and Vazirani [1], saying that if the distance between any two states is small then their induced probability distributions are close.

Lemma 6 ([1, Lemma 3.6]) *Let $|v\rangle$ and $|w'\rangle$ be two normalized states with $\| |v\rangle - |w'\rangle \| \leq \epsilon$. Then the total variation distance between the probability distributions resulting from measurements of $|v\rangle$ and $|w'\rangle$ is at most 4ϵ ,*

$$|\mathcal{D}_v - \mathcal{D}_{w'}| \leq 4\epsilon.$$

This holds no matter what basis is used for the measurements.

Theorem 2 follows immediately by composing Lemmata 4, 5, and 6.

Proof of Proof of Theorem 2 Write $|w'\rangle = c\mathbf{R}\mathbf{F}_M|u\rangle = \frac{1}{\sqrt{b}}\mathbf{A}|v\rangle$ where $c = \frac{1}{\sqrt{b}}\sqrt{\frac{M}{N}}$. By Lemma 4, we have $d = \|\mathbf{E}|v\rangle\| \leq 3\frac{N}{M}(2 + \log_2(N)) \leq \frac{9}{2}\frac{N}{M}\log_2(N)$. Since we assume $M \geq s \cdot 12N\log_2(N)$, we have $d \leq \frac{2}{5}\frac{1}{s} \leq \frac{1}{2}$, so by Lemma 5, $\||v\rangle - |w'\rangle\| \leq \frac{5}{2}d \leq \frac{1}{s}$ and $|1 - \frac{1}{\sqrt{b}}| \leq d \leq \frac{1}{s}$, and, finally, by Lemma 6, $|\mathcal{D}_v - \mathcal{D}_{w'}| \leq 4\frac{1}{s}$. \square

3 Discussion

Our Theorem 2 generalizes Hales and Hallgren's theorem in two ways. Firstly, if we want to apply Fourier sampling over \mathbb{Z}_N , then, by Theorem 2, it suffices to be able to implement the Fourier transform \mathbf{F}_M for some M which is only a $\log(N)$ factor larger than N . Thus, for such an M , we only need $\log \log(N)$ additional qubits to implement \mathbf{F}_M instead of implementing \mathbf{F}_N .

Secondly, in Theorem 2, not only are the distributions \mathcal{D}_v and $\mathcal{D}_{w'}$ close, but so are the states $|v\rangle$ and $|w'\rangle$ just prior the final measurement. In the setup studied by Hales and Hallgren, we are given a state $|u\rangle$ on which we want to apply a Fourier transform \mathbf{F}_N which is immediately followed by a measurement of the system. Now, suppose we do not want to measure the state $\mathbf{F}_N|u\rangle$, but instead first apply some other operations on it and then measure it. In that case, we cannot apply Hales and Hallgren's theorem, but we can apply Theorem 2. The reason is that Hales and Hallgren's theorem says that the distributions are close, not that the states themselves are close, as stated in Theorem 2.

Acknowledgements

I am very grateful to Sean Hallgren for helpful discussions. I would like to thank an anonymous referee for a careful reading and for many constructive suggestions, and Gilles Brassard and Joan Boyar for comments.

References

- [1] E. Bernstein and U. Vazirani, Quantum complexity theory, *SIAM J. Comput.* 26 (1997) 1411–1473.

- [2] R. Cleve, An introduction to quantum complexity theory. To appear in *Collected Papers on Quantum Computation and Quantum Information Theory*, World Scientific, C. Macchiavello, G. M. Palma, and A. Zeilinger, editors, 2000. Also available at Los Alamos National Laboratory e-Print archive as <<http://arXiv.org/abs/quant-ph/9906111>>.
- [3] L. K. Grover, Quantum mechanics helps in searching for a needle in a haystack, *Phys. Rev. Lett.* 79 (1997) 325–328.
- [4] L. Hales and S. Hallgren, Quantum Fourier sampling simplified, *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, ACM Press (1999) 330–338.
- [5] A. Yu. Kitaev, *Quantum measurements and the Abelian stabilizer problem* (1995). Available at Los Alamos National Laboratory e-Print archive as <<http://arXiv.org/abs/quant-ph/9511026>>.
- [6] P. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM J. Comput.* 26 (1997) 1484–1509.