The Quantum Fourier Transform Has Small Entanglement

Jielun (Chris) Chen, E.M. Stoudenmire, and Steven R. White

arXiv:2210.08468, 2022

Quantum Colloquium, Simons Institute
March 14, 2023


Overview

The **Quantum Fourier Transform** (QFT) is one of the most important ingredients in quantum algorithms.

However, we will show that the **core part** of the QFT can only generate **small entanglement** on qubits.

We can then simulate the QFT classically using tensor networks, which can be **faster** than fast Fourier transform (FFT) in many practical cases!

But this is not a bad news for quantum computers, because it doesn't mean the QFT can be simulated classically for arbitrary inputs (e.g. states in Shor's algorithm).

Outline

- Introduction
- Main result
- Implications
- Proof outline
- Simulating the QFT
- Outlook

Outline

- Introduction
- Main result
- Implications
- Proof outline
- Simulating the QFT
- Outlook

Introduction to the QFT

The QFT has been widely used in quantum algorithms for provable speed-up over their classical counterparts. Examples are:

- Shor's algorithm (factoring integers)
- Hidden subgroup problems
- Phase estimation
- Simulating quantum dynamics
- Quantum arithmetics
- Solving linear systems of equations
- …

The QFT's power stems from its ability to find periodic structures of quantum states, which many quantum algorithms rely on to gain a provable speed-up.

Introduction to the QFT

The QFT (over cyclic group $\mathbb{Z}_{2^n}$) = the discrete Fourier transform (DFT) over Hilbert space spanned by n qubits. In terms of a matrix:

$$F_n = \frac{1}{\sqrt{2^n}} \sum_{q=0}^{2^n-1} \sum_{q'=0}^{2^n-1} \omega^{qq'} |q\rangle\langle q'|$$

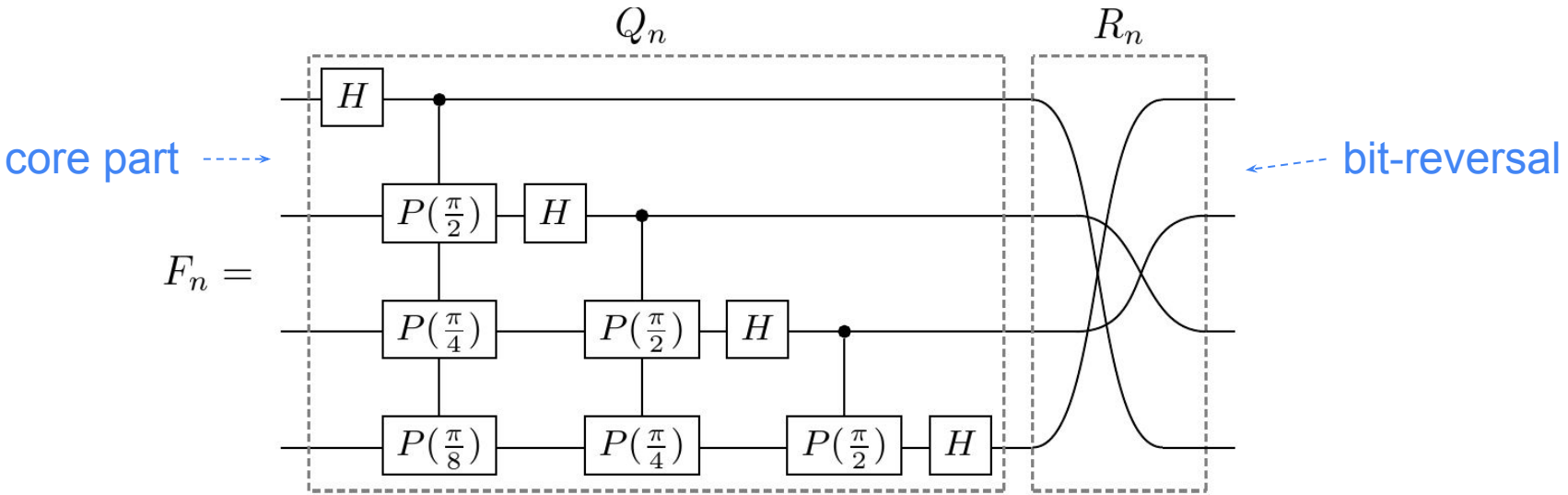
$$q = q_1 q_2 \ldots q_n$$

$$\omega = \exp(i2\pi/2^n)$$

$$= \frac{1}{\sqrt{2^n}} \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & \omega & \omega^2 & \ldots & \omega^{2^n-1} \\ 1 & \omega^2 & \omega^4 & \ldots & \omega^{2(2^n-1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \omega^{2^n-1} & \omega^{2(2^n-1)} & \ldots & \omega^{(2^n-1)(2^n-1)} \end{pmatrix}$$

Introduction to the QFT

The quantum circuit of F_n can be decomposed into two parts, a core part Q_n and the bit-reversal part R_n. E.g. for 4-qubits:



$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \qquad P(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$

Introduction to the QFT

From now on, I will refer Q_n as the QFT, and F_n as the DFT

Outline

- Introduction
- Main result
- Implications
- Proof outline
- Simulating the QFT
- Outlook

Main result (informal)

Informal statement of our result:

The Schmidt coefficients (to be defined) of Q_n and F_n differ significantly:

- Q_n has exponentially decaying Schmidt coefficients for any partition of the operator, suggesting small entanglement

- F_n has uniform Schmidt coefficients for any partition of the operator [1, 2], suggesting maximal entanglement

That is, most entanglement in F_n comes from the bit reversal R_n.

[1] Tyson, Jon. “Operator-Schmidt decomposition of the quantum Fourier transform on C^N1 ⊗ C^N2.” *Journal of Physics A* 36 (2003)
[2] Nielsen, Michael A. et al. “Quantum dynamics as a physical resource.” *Physical Review A* 67 (2003)

Why do we consider Schmidt coefficients?

There are many variants of an operator's entanglement measure, e.g. see [1]. We specifically consider operator Schmidt decomposition of the QFT, because the Schmidt coefficients:

- Lower-bound non-local communications required to implement the operator.
- Indicate representability as an efficient tensor network.
- Often capture the maximal entanglement the operator can generate on a bipartite (mixed) quantum state.

R_n is very non-local and can generate large entanglement on subsystems.

[1] Nielsen, Michael A. et al. "Quantum dynamics as a physical resource." *Physical Review A* 67 (2003)

Operator Schmidt decomposition: definition

An Operator Schmidt Decomposition of a unitary matrix acting on a bipartite systems $\mathcal{A} \otimes \mathcal{B}$ is a decomposition of the form:

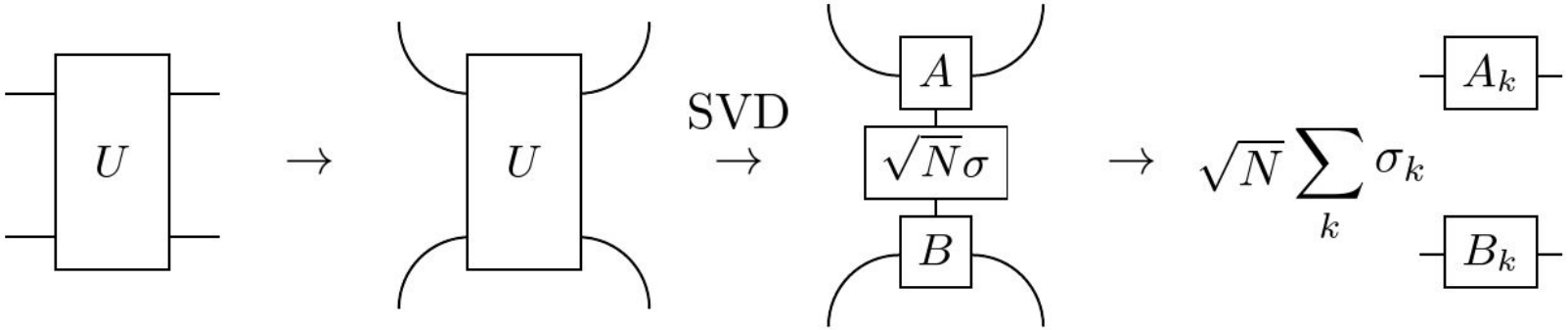
$$U = \sqrt{N} \sum_{k=0}^{\chi-1} \sigma_k A_k \otimes B_k$$

where:

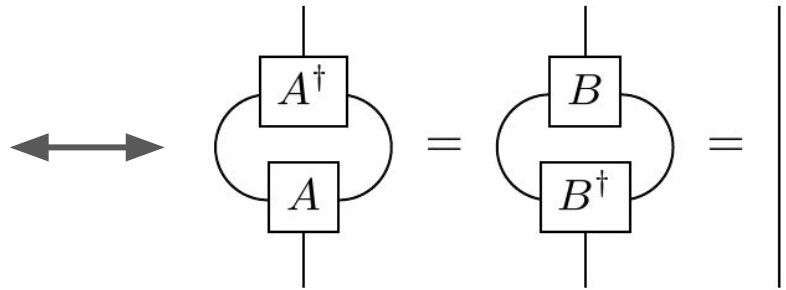
- χ is the Schmidt rank, $\{\sigma_k\}$ are the Schmidt coefficients
- $\mathrm{Tr}(A_k^\dagger A_{k'}) = \delta_{k,k'}, \mathrm{Tr}(B_k^\dagger B_{k'}) = \delta_{k,k'}$
- $N = dim(\mathcal{A} \otimes \mathcal{B}) = dim(\mathcal{A}) \times dim(\mathcal{B})$ (only apply to unitary)
- $\sum_k \sigma_k^2 = 1$ and $\sigma_0 \geq \sigma_1 \geq ...\sigma_{\chi-1}$

Operator Schmidt decomposition: definition

This is equivalent as the singular value decomposition (SVD). We first reshape the unitary into a matrix in the vertical direction, and apply SVD on it.



$$\mathrm{Tr}(A_k^\dagger A_{k'}) = \delta_{k,k'}$$
$$\mathrm{Tr}(B_k^\dagger B_{k'}) = \delta_{k,k'}$$



Operator Schmidt decomposition example: SWAP

The operator Schmidt decomposition of a SWAP gate is the sum of paulis.

It has maximal Schmidt rank and uniform Schmidt coefficients!

$$\mathrm{SWAP} = \frac{1}{2} I \otimes I + \frac{1}{2} X \otimes X + \frac{1}{2} Y \otimes Y + \frac{1}{2} Z \otimes Z$$

This explains why the bit-reversal operator has high operator entanglement!

Main result (formal)

The main theorem of our paper:

Consider the QFT operator Q_n. By partitioning the Hilbert space into $\mathcal{A}$ with qubits 1 to j, and $\mathcal{B}$ with qubits $j+1$ to n, Q_n has the operator Schmidt decomposition:

Exponentially large Schmidt-rank

$$Q_n = \sqrt{2^n} \sum_{k=0}^{\min(2^j, 2^{n-j})-1} \sigma_{n,j}^k A_{n,j}^k \otimes B_{n,j}^k$$

It then follows that for $k \geq 2$, the Schmidt coefficients $\{\sigma_{n,j}^k\}$ satisfy:

$$\sigma_{n,j}^k \leq \frac{1}{\sqrt{k}} \exp\left(-\frac{2k+1}{2} \log\left(\frac{4k+4}{e\pi}\right)\right)$$

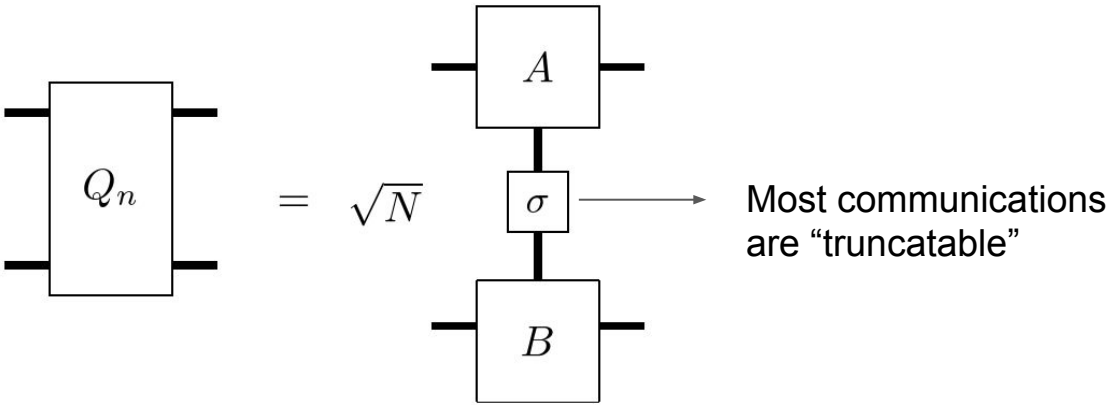
Independent of the number of qubits (n) & partition of the system (j)!

Outline

- Introduction
- Main result
- Implications
- Proof outline
- Simulating the QFT
- Outlook

A rough picture

- The QFT has high Schmidt-rank but limited “correlation length”.
- Require only few bipartite communications to accurately approximate the operator.



Implications

Does the result pose any complexity hierarchy issues?

No.

- Classical stimulability requires both operator and states to have small entanglement. E.g. Shor's algorithm has exponentially expensive state before QFT.

- Even without knowing the QFT's low entanglement, early results already showed (approximate) QFT is efficiently simulable if input circuit has limited interaction range and log depth [1] or log bubble width [2]. Semi-classical QFT is also efficiently simulable [3].

[1] Nadav Yoran, Anthony J. Short, "Efficient classical simulation of the approximate quantum Fourier transform" *Phys. Rev. A 76, 042321 (2007)*
[2] Dorit Aharonov, Zeph Landau, Johann Makowsky, "The quantum FFT can be classically simulated" *arxiv-0611241*
[3] Daniel E. Browne, "Efficient classical simulation of the semi-classical Quantum Fourier Transform", *New J. Phys. 9 146 (2007).*

Implications

Does it tell anything new on implementing the QFT on a quantum computer?

Perhaps?

- Generally low entanglement implies fewer resources & more robust to noise.
- However, some qubit architectures may not favor input & output to have different qubit ordering.
- Not clear whether it implies lower circuit depth since any QFT with approximation error below 1/10 requires at least log(n) depth [1].
- May imply an underlying exponentially-decaying Hamiltonian, but hard to find (to be explained in details later)

[1] Richard Cleve, John Watrous, “Fast parallel circuits for the quantum Fourier transform”, arXiv:quant-ph/0006004

Implications

Does it tell anything new on the classical Fourier transform?

Yes!

- With the low-entanglement nature, simulating the QFT can gain speed-up over FFT in many classical algorithms!

- If input is certain types of tensor networks, almost always favor over FFT.

- If input is a vector, may still gain advantage if it's compressible into certain tensor network.

(More details later!)

Outline

- Introduction
- Main result
- Implications
- Proof methods
- QFT tensor network
- Outlook

Proof methods: overview

The underlying mathematical reason for the exponential decay of the QFT's Schmidt coefficients can be understood through 3 steps:

1. Q_n has a generalized recursive circuit, from which we can see that only phase gates are relevant to Schmidt coefficients.

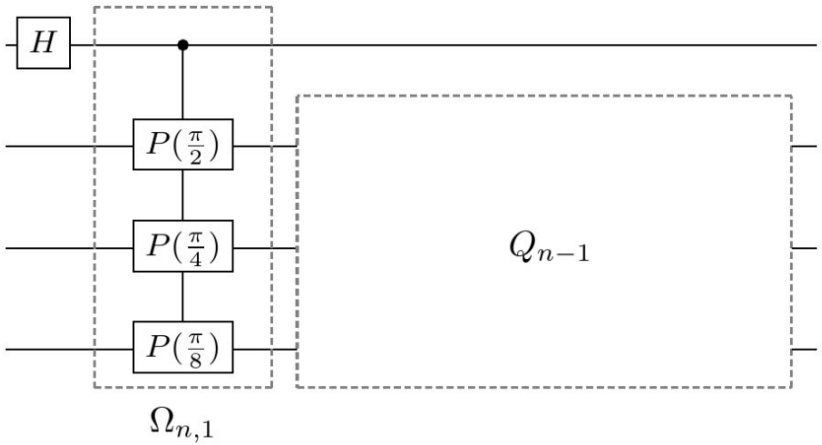
2. From the phase gates we find that the Schmidt coefficients of Q_n are equivalent to singular values of the top-left submatrices of the DFT matrix (i.e. F_n)

3. The submatrices of the DFT matrix are known be approximately low-rank. They are extensively studied in signal processing as the spectral concentration problem.

Step 1: Generalized recursive QFT circuit

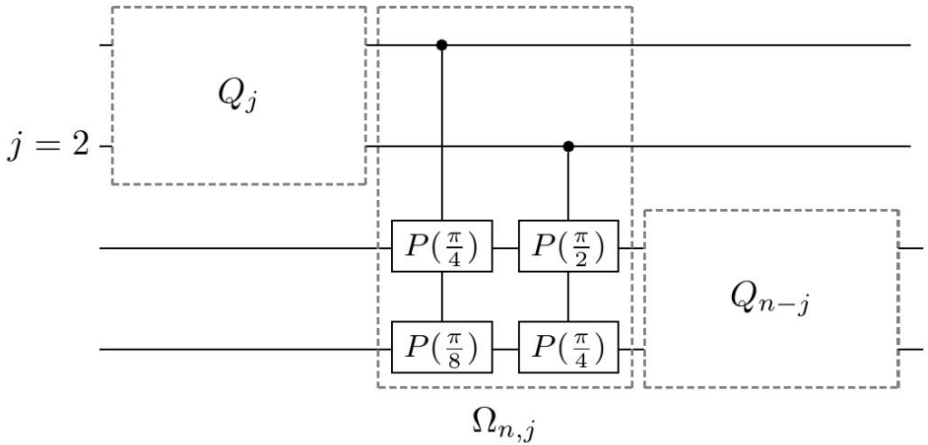
Standard recursive circuit

$$Q_n = (I_1 \otimes Q_{n-1})\Omega_{n,1}(H \otimes I_{n-1})$$



Generalized recursive circuit

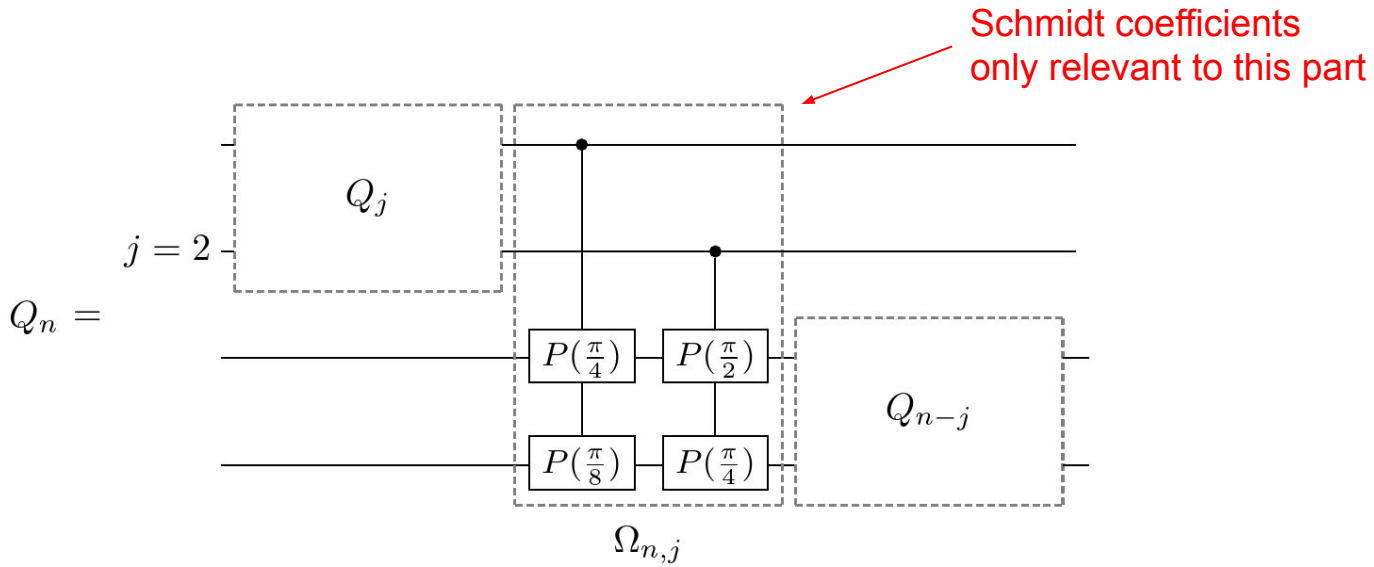
$$Q_n = (I_j \otimes Q_{n-j})\Omega_{n,j}(Q_j \otimes I_{n-j})$$



[1] Richard Cleve, John Watrous, “Fast parallel circuits for the quantum Fourier transform”, arXiv:quant-ph/0006004

Step 1: Generalized recursive QFT circuit

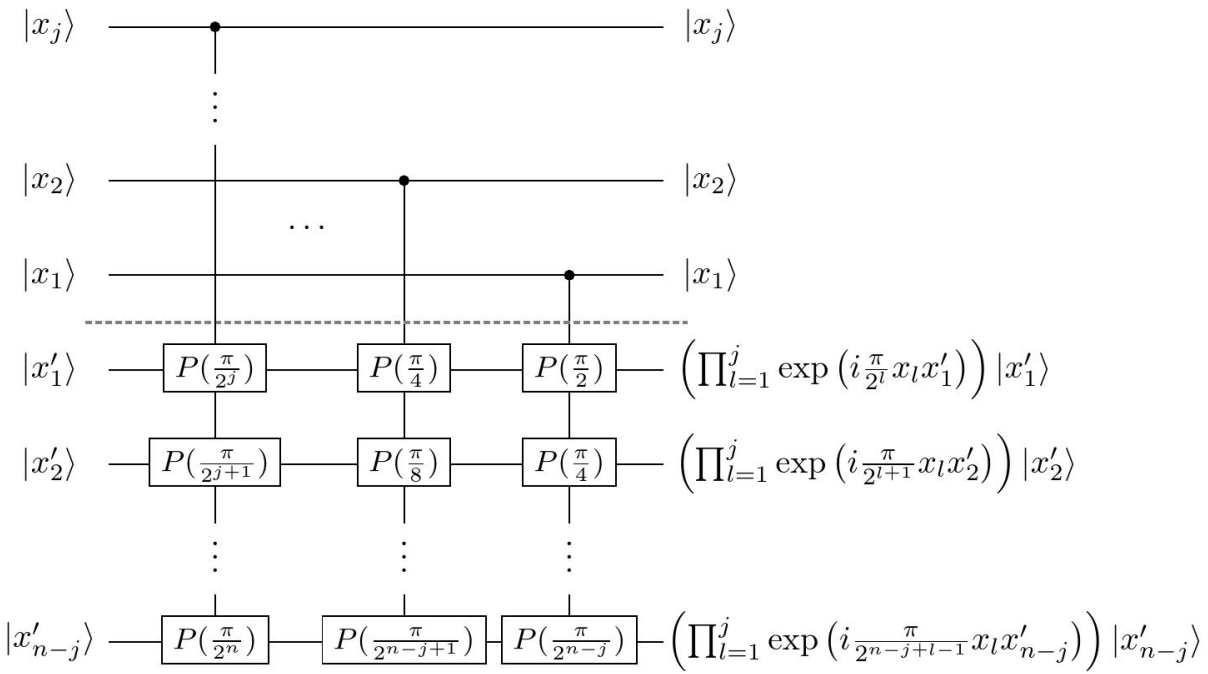
Because the two smaller QFT are local unitaries (with respect to the partition), only the central part $\Omega_{n,j}$ is relevant to the Schmidt coefficients.



Step 1: Analyzing the central part

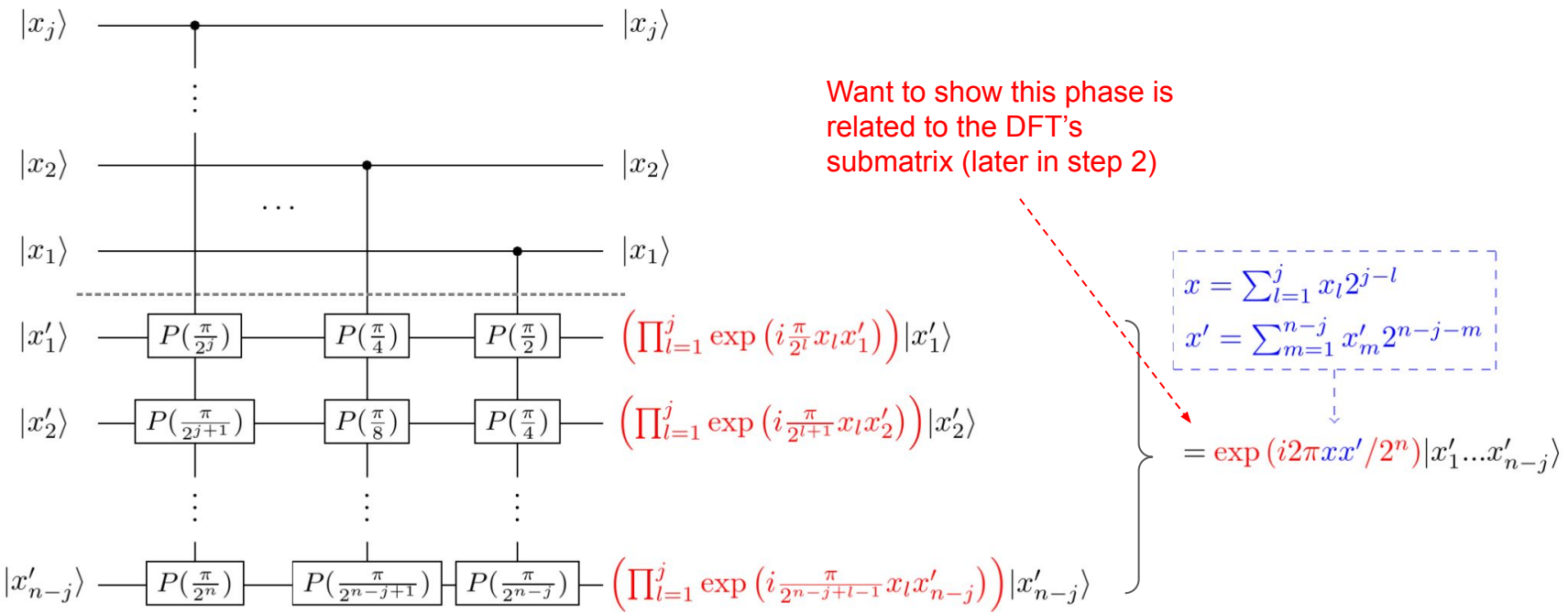
$\Omega_{n,j}$ is diagonal in the computational basis, and its matrix elements are product of phases conditioned on bit values of two registers x, x':

Notice two registers are labeled in the opposite order



Step 1: Analyzing the central part

The overall phase can be simplified to only dependent on the decimal values of two registers:



Want to show this phase is related to the DFT's submatrix (later in step 2)

$x = \sum_{l=1}^{j} x_l 2^{j-l}$

$x' = \sum_{m=1}^{n-j} x'_m 2^{n-j-m}$

$= \exp(i2\pi x x'/2^n)|x'_1 ... x'_{n-j}\rangle$

Step 1: Summary

To summarize, for step 1 we have:

1. The QFT operator Q_n can be decomposed as follows, where $|x^r\rangle = |x_j \ldots x_1\rangle, |x'\rangle = |x'_1 \ldots x'_{n-j}\rangle$

$$Q_n = (I_j \otimes Q_{n-j}) \underbrace{\left(\sum_{x,x'} \exp(i2\pi x x'/2^n) |x^r\rangle\langle x^r| \otimes |x'\rangle\langle x'| \right)}_{\Omega_{n,j}} (Q_j \otimes I_{n-j})$$

2. The Schmidt coefficients of Q_n at cut *j* is the same as those of $\Omega_{n,j}$ at cut *j*

$$\mathrm{Sch}_j(Q_n) = \mathrm{Sch}_j(\Omega_{n,j})$$

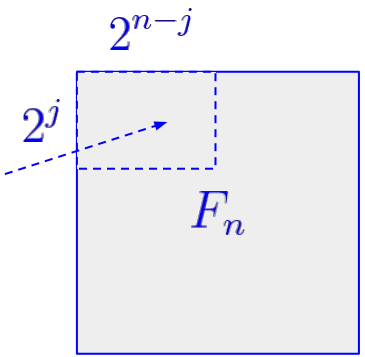
Step 2: The DFT submatrix

The non-zero Schmidt coefficients of $\Omega_{n,j}$ are equivalent to the singular values of a submatrix of the DFT:

$$\mathrm{Sch}_j(\Omega_{n,j}) = \mathrm{Sch}_j\left(\sum_{x=0}^{2^j-1}\sum_{x'=0}^{2^{n-j}-1} \exp(i2\pi xx'/2^n)|x^r\rangle\langle x^r| \otimes |x'\rangle\langle x'|\right)$$

$$= \sigma\left(\frac{1}{\sqrt{2^n}}\sum_{x=0}^{2^j-1}\sum_{x'=0}^{2^{n-j}-1} \exp(i2\pi xx'/2^n)|x^r x^r\rangle\langle x'x'|\right)$$

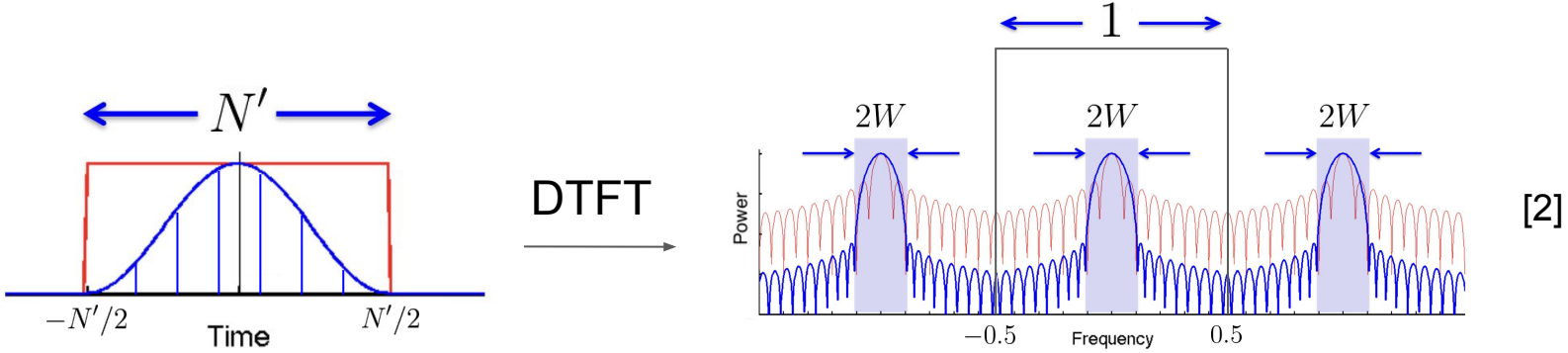
$$= \sigma\left(\frac{1}{\sqrt{2^n}}\sum_{x=0}^{2^j-1}\sum_{x'=0}^{2^{n-j}-1} \exp(i2\pi xx'/2^n)|x\rangle\langle x'|\right)$$



Step 3: The spectral concentration problem

The submatrix of the DFT matrix are known to be approximately low-rank. They relate to the spectral concentration problem [1], which has been extensively studied in signal processing. The original problem considers the discrete-time Fourier transform (DTFT), where in our case we use the DFT.

Spectral concentration problem: Find a strictly localized sequence [-N'/2, N'/2] whose discrete-time Fourier Transform is maximally localized within the a finite window [-W,W] relative to the frequency periodic window [-1/2, 1/2]. The solution sequences are known as discrete prolate spheroidal sequences (DPSSs).



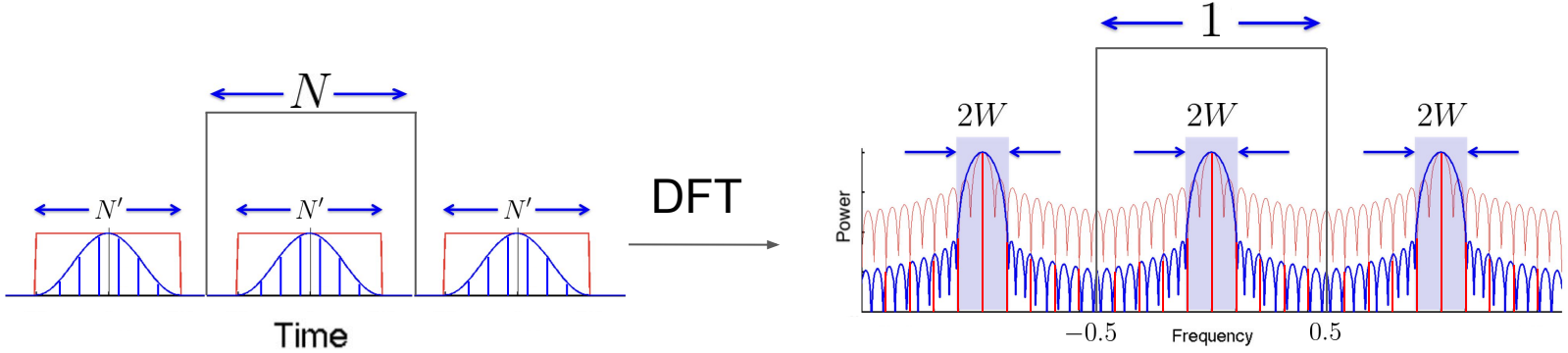
[2]

[1] D. Slepian et.al., Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty — I - V
[2] Credit: MIT OpenCourseWare, Introduction To Neural Computation, Lecture 13

Step 3: The spectral concentration problem

In the DFT version of the spectral concentration problem, instead of constrained to a strictly localized time sequence, we constrained it to be localized relative to a periodic window. This can be viewed as discretizing the frequency domain.

Spectral concentration problem (DFT ver.): Find a localized sequence [-N'/2,N'/2] relative to a periodic window [-N/2, N/2], whose discrete Fourier Transform is maximally localized within the a finite window [-W,W] relative to the frequency periodic window [-1/2, 1/2] with N sample points. The solution sequences are known as periodic discrete prolate spheroidal sequences (P-DPSSs).



Step 3: The spectral concentration problem

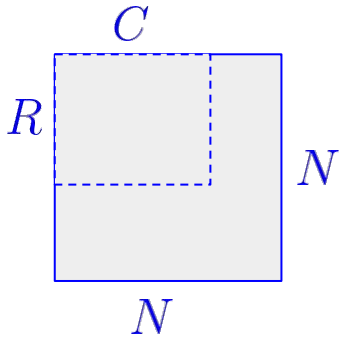
With some analysis, one can show that solving the spectral concentration problem corresponds to solving the singular value equation of the DFT's top-left submatrix.

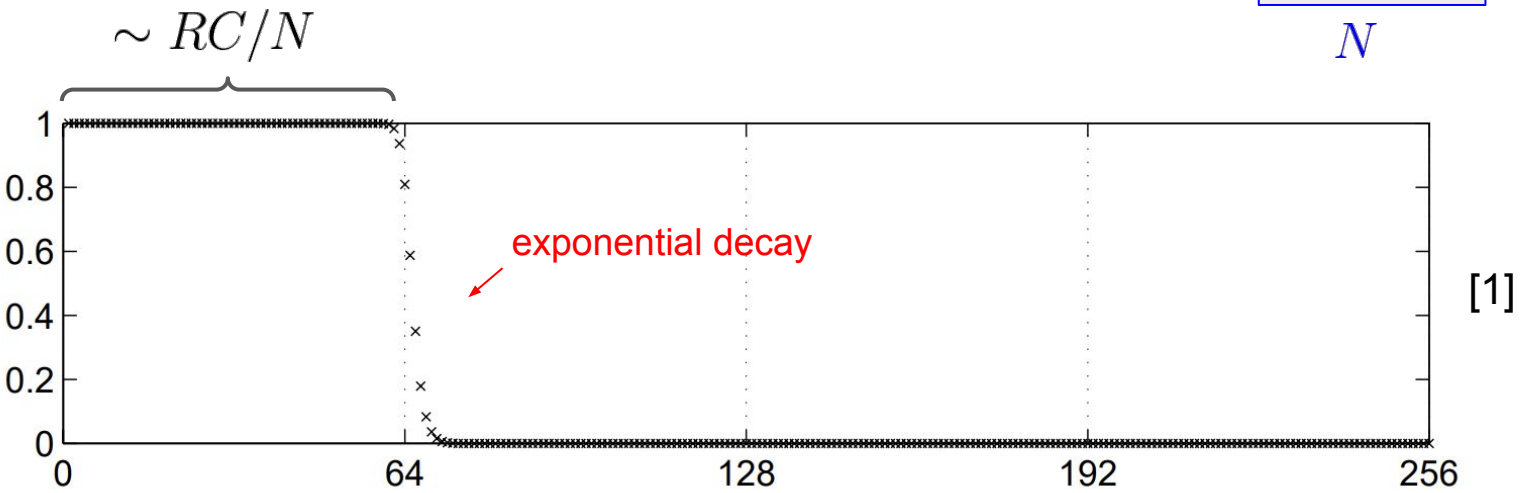
$2NW \times N'$ top-left submatrix of the DFT matrix

$$\max_{|v\rangle} \frac{\langle v|F_D^{-1}F_D|v\rangle}{\langle v|v\rangle} \Rightarrow F_D^{-1}F_D|v\rangle = \lambda|v\rangle$$

Step 3: The spectral concentration problem

For an $R \times C$ submatrix of the $N \times N$ DFT, the singular values have a clustering behavior [1]:





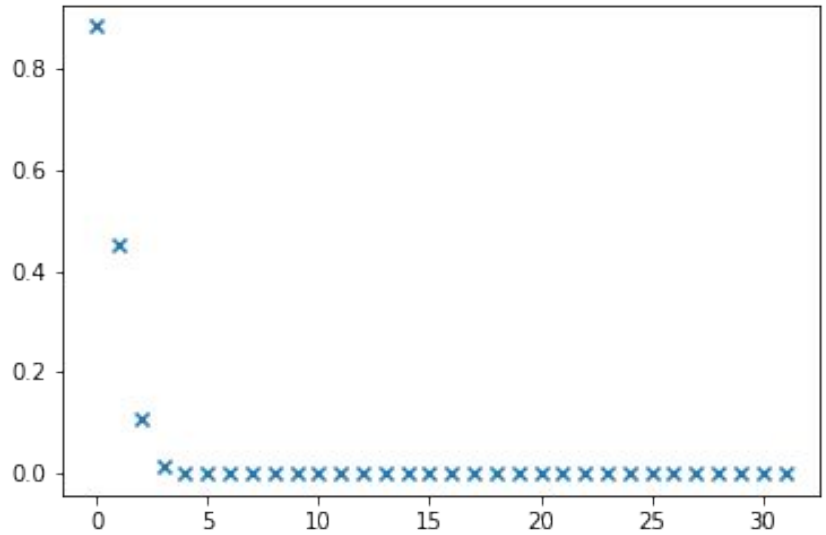
[1]

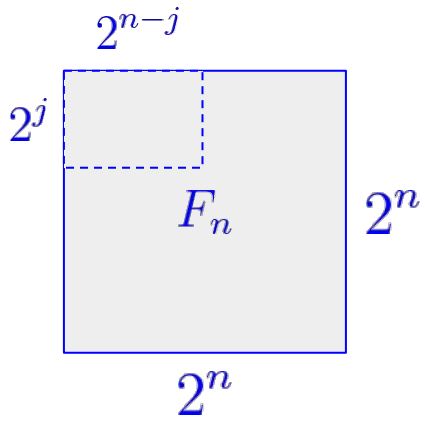
[1] Edelman, A et. al., The Future Fast Fourier Transform?

Step 3: The spectral concentration problem

Remember for the QFT's schmidt coefficients, they are singular values of the $2^j \times 2^{n-j}$ submatrix of the $2^n \times 2^n$ DFT

only one singular value is close to 1, independent of the choice of j, n!





Step 3: The spectral concentration problem

To prove rigorous exponential decaying bounds:

$$\sigma_{n,j}^{k} \leq \frac{1}{\sqrt{k}} \exp\left(-\frac{2k+1}{2} \log\left(\frac{4k+4}{e\pi}\right)\right)$$

We adopt techniques from [1, 2], which employs the decay property of half-order Bessel functions.

Audience interested in details shall refer to our paper!

[1] Santhosh Karnik et. al., “Improved bounds for the eigenvalues of prolate spheroidal wave functions and discrete prolate spheroidal sequences”, arXiv:2006.00427
[2 M. Boulsane et. al., “Discrete Prolate Spheroidal Wave Functions: Further spectral analysis and some related applications”, arXiv:1905.08354

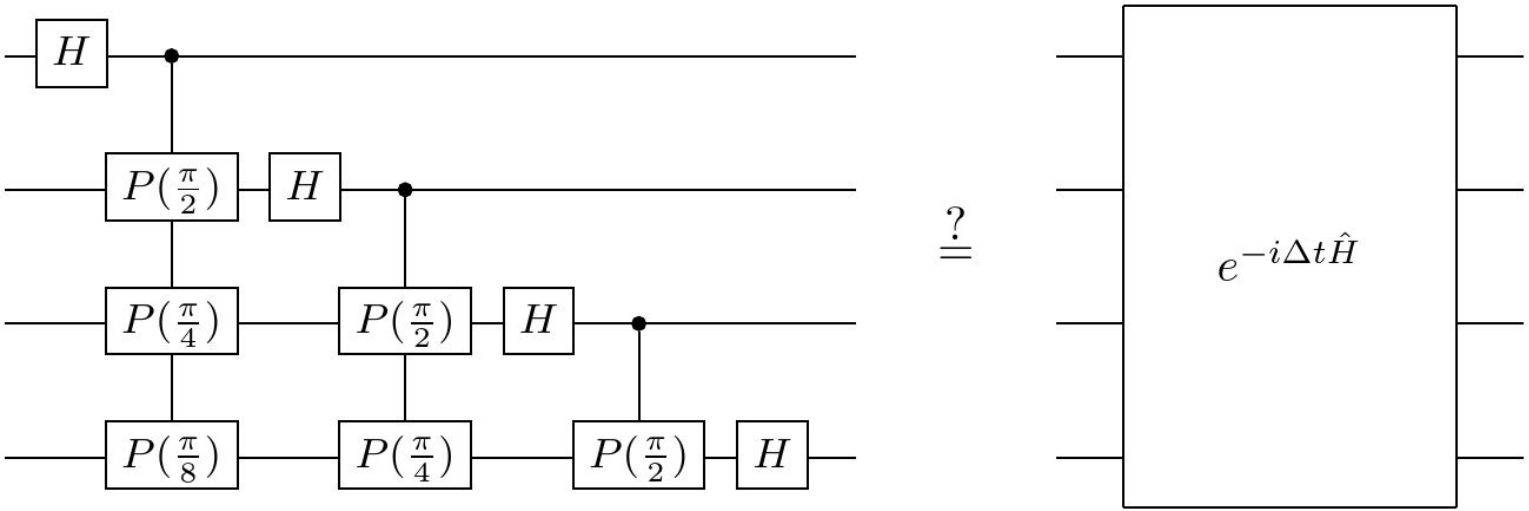
Proof summary

To summarize:

- The Schmidt coefficients of the QFT are singular values of the DFT's submatrix.
- The DFT's submatrix is related to the spectral concentration problem.
- Spectral concentration has decaying solutions (singular values), corresponding to the QFT's decaying Schmidt coefficients.
- The upper-bounds are independent of the number of qubits and the partition of the system.

Relation to the area law

It’s still not intuitive why the QFT circuit with long-range gates and non-constant depth can have small entanglement. Can we interpret the QFT as a physical process, i.e. evolving under some Hamiltonian with constant time, and understand the low-entanglement from there?



Relation to the area law

Area law for dynamics: A Hamiltonian on a one-dimensional lattice with interaction decaying faster than $1/r^2$ has a constant entangling rate [1,2].

In other words, when evolving an arbitrary 1D system with such Hamiltonian, the rate of change of the system's entanglement entropy does not increase with system size.

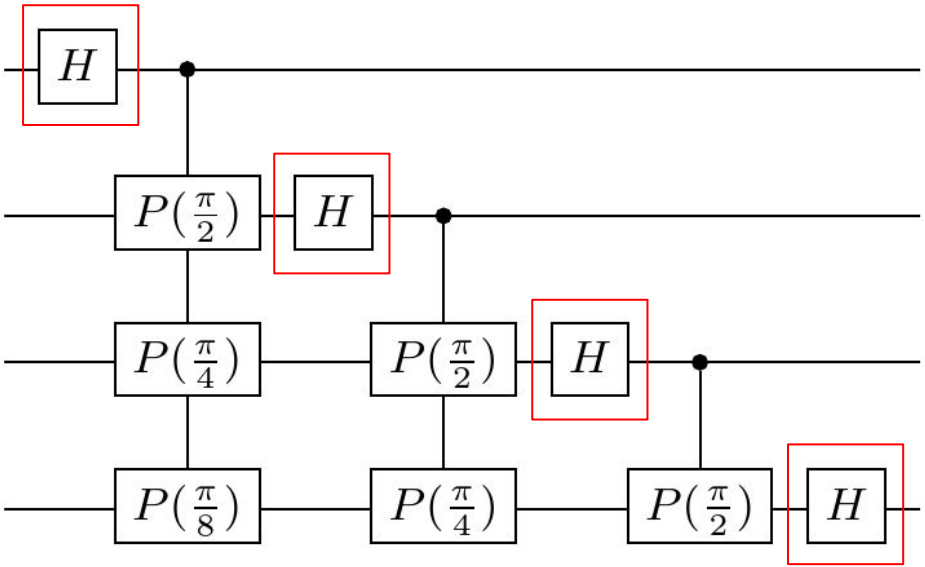
The exponential decay of the QFT's Schmidt coefficients suggests that it falls into this category, is it true?

[1] Van Acoleyen, Karel et al. "Entanglement rates and area laws." *Physical review letters* 111 17 (2013): 170501
[2] Gong, Zhe-Xuan et al. "Entanglement Area Laws for Long-Range Interacting Systems." *Physical review letters* 119 5 (2017): 050501

Relation to the area law

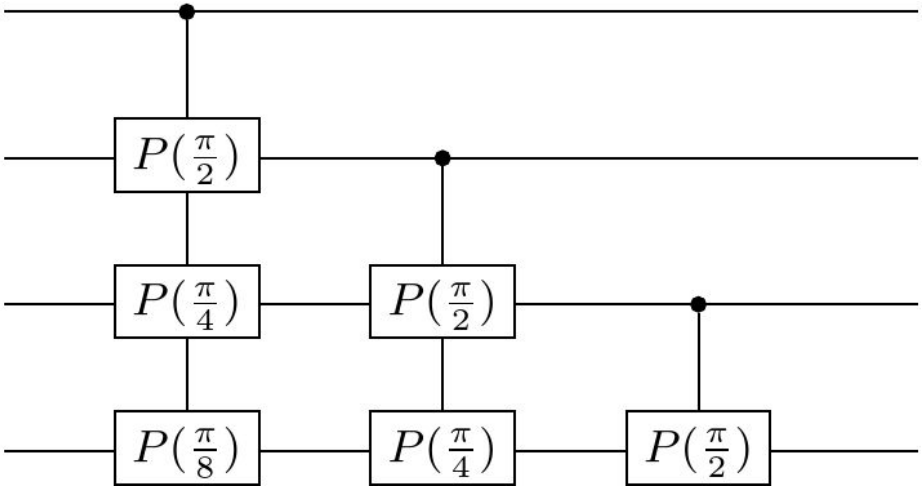
Explicitly solving the QFT's Hamiltonian is a challenging task, because the Hadamard gates and controlled phase gates do not commute.

Problematic!



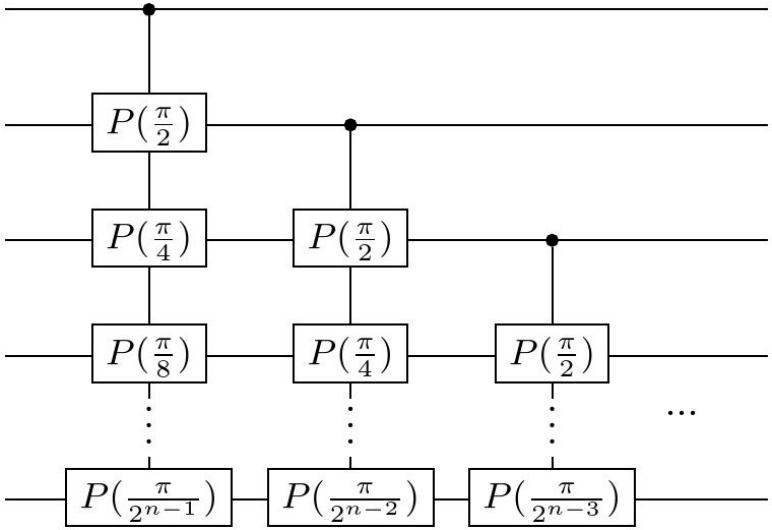
Relation to the area law

However, recall that the Schmidt coefficients only depend on phase gates. Therefore, if we remove all the H gates, the Schmidt coefficients will be invariant for any cut! In fact, we can show even more: any bipartite entanglement measure (under the LOCC principle) will remain invariant from removing H gates.



Relation to the area law

This means the Hamiltonian for the QFT without H gates can be used to determine the entanglement structure of the QFT itself too. It turns out this Hamiltonian has exponentially decaying interaction, which is pseudo-local! Therefore, it will only generate constant amount of entanglement regardless of the system size, thus so do the QFT!



$$= \exp(i\pi \sum_{l=1}^{n-1} \sum_{m=l+1}^{n} \left(\frac{1}{2}\right)^{m-l} |1\rangle\langle 1|_l \otimes |1\rangle\langle 1|_m)$$

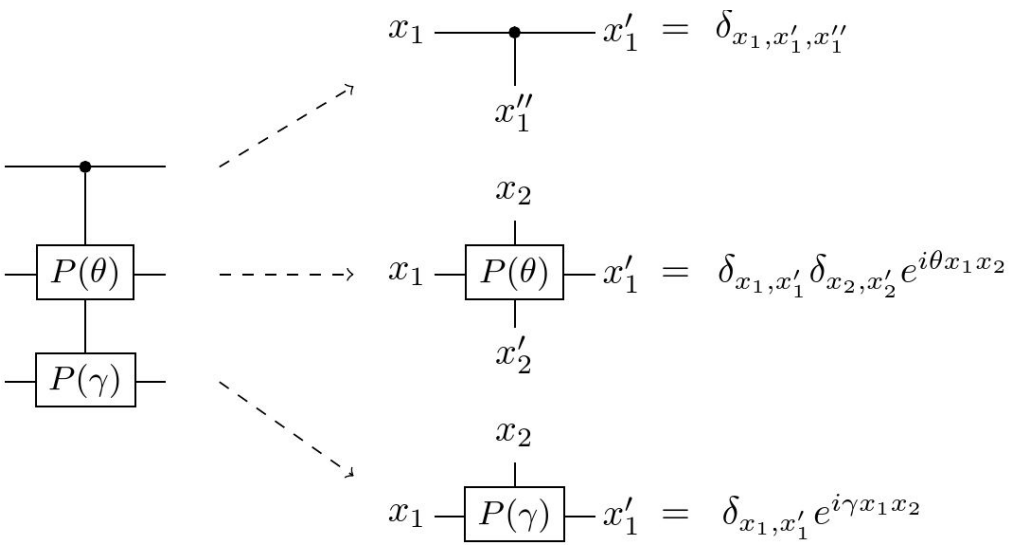
Hamiltonian with exponentially decaying interaction

Outline

- Introduction
- Main result
- Implications
- Proof outline
- Simulating the QFT
- Outlook

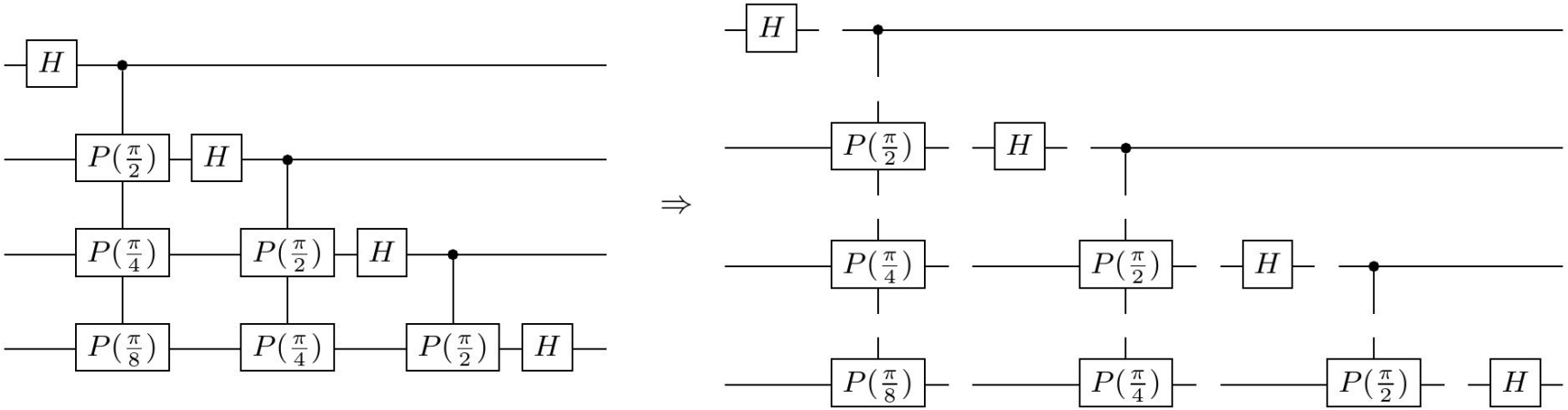
The QFT tensor network and matrix product operator

The QFT's small entanglement implies that it has an efficient tensor network representation, in particular an Matrix Product Operator (MPO) representation, i.e. a QFT-MPO. To find it, we first consider decomposing the controlled phase gates into low-rank local tensors:



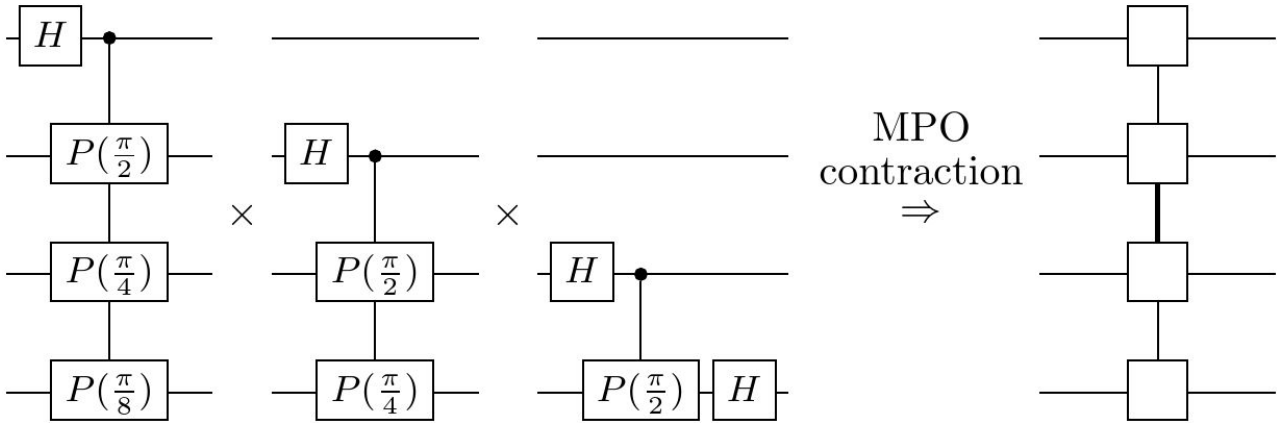
The QFT tensor network and matrix product operator

This allows the QFT circuit to be represented as a compact tensor network, for which we call the QFT-TN:



The QFT tensor network and matrix product operator

To efficiently construct the QFT-MPO, we can view the QFT-TN as a series of MPOs and apply any MPO x MPO algorithm with bond-dimension kept up to a constant. We use the zip-up algorithm [1] since it's both fast and stable in our case.



[1] E M Stoudenmire and Steven R White, "Minimally entangled typical thermal state algorithms, 2010 *New J. Phys.* 12 055026

Contraction time & error of the QFT-MPO

Assume we set the bond-dimension of the QFT-MPO to χ, the contraction algorithm will generally have:

- time complexity ~ $O(\chi^3 n^2)$ (around one second for 50 qubits)
- total truncation error ~ $O(ne^{-\chi \log(\chi/3)})$
- For fixed error, bond-dimension grows sub-logarithmically.

Less than machine precision; essentially exact!

Some numerics on the errors $\left\|Q_n - Q_n^{\mathrm{MPO}}\right\|_F^2 / 2^n$:

n \ χ	2	4	6	8
6	0.0237328	8.05738 e-7	1.58873 e-12	2.78383 e-16
7	0.0327034	2.33700 e-6	1.35423 e-11	3.69769 e-16
8	0.0416221	4.22659 e-6	4.64780 e-11	8.20772 e-17

Fourier transform through the QFT-MPO

Our QFT-MPO is efficient enough to be classically practical! i.e. a “superfast” classical Fourier transform

- If the input is an MPS with bond-dimension χ_m, applying the QFT-MPO has linear complexity $\sim O(n\chi\chi_m)$.

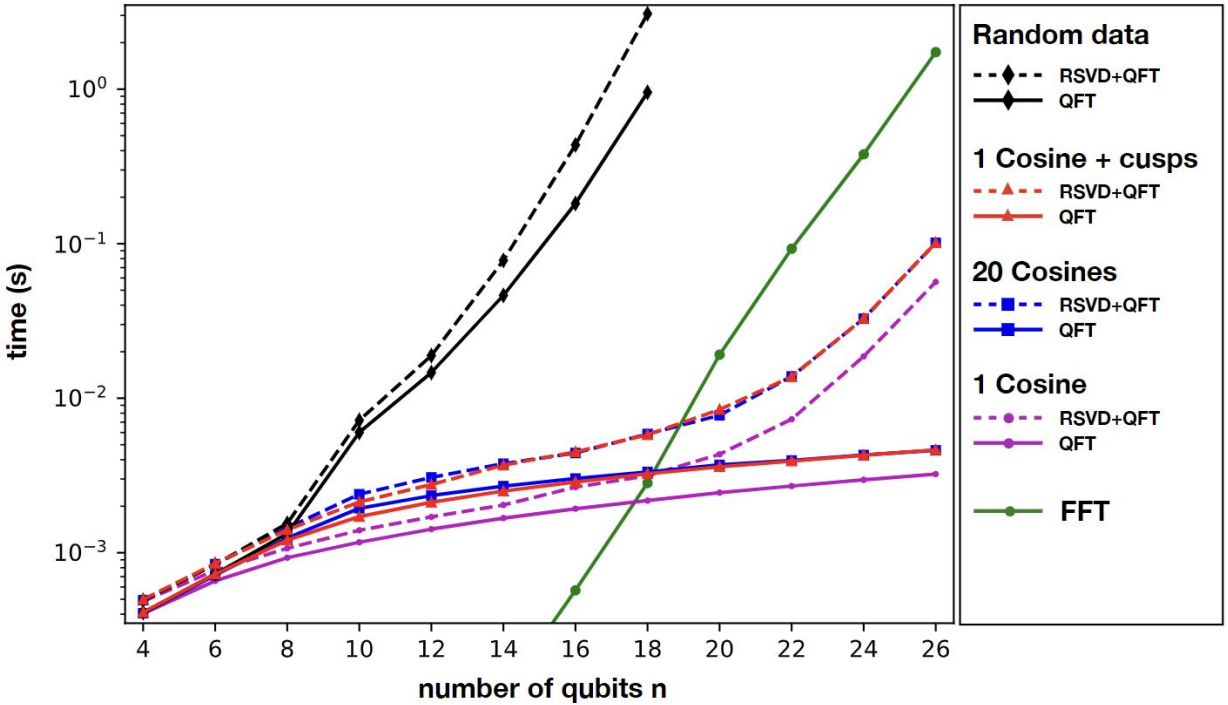
- If the input is a vector of length $N = 2^n$, converting it to an MPS using randomized SVD takes $\sim O(2^n\chi_m)$. Compared to FFT with time $O(N \log N) = O(2^n n)$, still has advantage if χ_m is sub-linear!

- Can be even faster using more efficient conversion methods, e.g. Tensor-train cross-interpolation [1].

[1] S. Dolgov, D. Savostyanov, Computer Physics Communications Volume 246, January 2020, 106869

Benchmark: QFT vs. FFT

Here are some benchmarks ran on a personal computer. For compressible data we can already see advantage at around 18 qubits!



Many classical applications!

Operations with FFT become cheap in MPS, can be applied to quantum chemistry, quantum dynamics, differential equations, etc. (work in progress)

- Fourier interpolation
- Convolution
- Laplacian
- Split-operator
- …

Many classical applications!

There's evidence that bitwise-encoding of functions through MPS/MPO is very powerful. Fourier transform is an addition to this family! Can combine all of them to do fast classical computation. Some examples:

- Many functions have explicit efficient construction as small-rank MPS [1], including: cos(x), sin(x), exp(x), finite-order polynomials, etc.

- Self-similar functions generally have small bond-dimension.

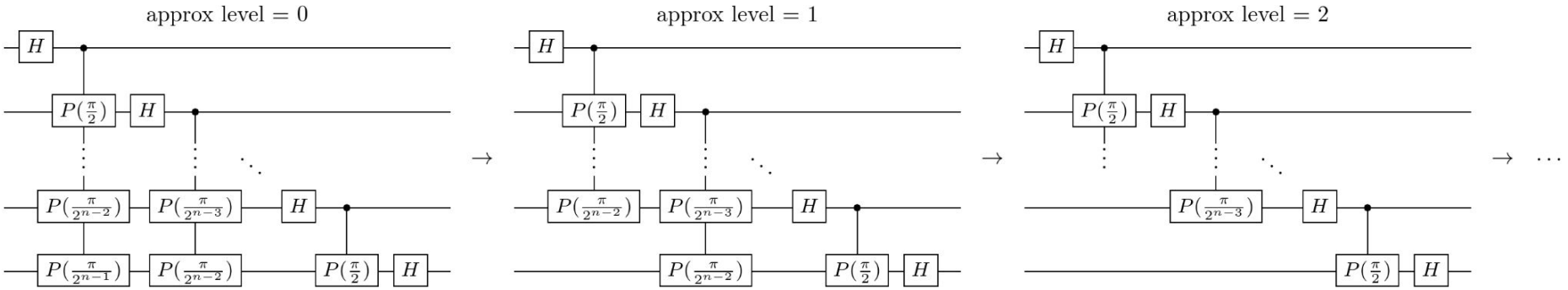
- There's evidence that all Lipschitz-continuous functions have small bond-dimension [2].

- Bitwise operations & finite state machine can be directly converted to efficient MPO, e.g. addition, shifting, etc.

[1] I. V. Oseledets, "Constructive Representation of Functions in Low-Rank Tensor Formats", Constructive Approximation volume 37, (2013)
[2] Griebel, M., Harbrecht, H. Analysis of Tensor Approximation Schemes for Continuous Functions. Found Comput Math 23, 219–240 (2023)

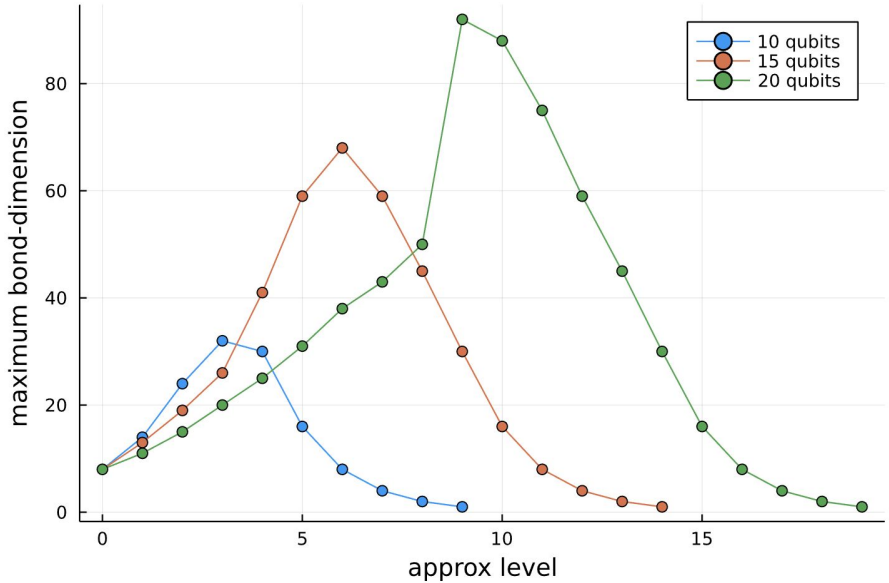
Side note: the approximate QFT

A natural question to ask is whether the the same low-rank structure applies to the approximate QFT (AQFT), which is defined by removing the long-range controlled-phase gates.



Side note: the approximate QFT

Numerical simulations suggest that MPOs for the AQFTs have higher bond-dimension [1]! The low-rank structure is a feature to the QFT's full circuit. No rigorous argument for this yet…



[1] Woolfe, Kieran J. and Hill, Charles D. and Hollenberg, Lloyd C. L., arXiv:1406.0931

Outline

- Introduction
- Main result
- Implications
- Proof outline
- Simulating the QFT
- Future outlook

Future outlook

1. Does this imply a better implementation of the QFT on a QC?
   - Converting the QFT-MPO to a quantum circuit
   - Time-evolving an exponentially-decaying Hamiltonian
   - …

2. Why does AQFT seem to have higher entanglement?

3. Can we generalize it to the QFT over more general groups?

4. Can we find more classical applications? Can we categorize a general class of problems for the QFT-MPO to be useful?

5. Can we connect the QFT-MPO to other “faster” Fourier transform, e.g. the sparse Fourier transform [1]?

[1] https://groups.csail.mit.edu/netmit/sFFT/index.html

Acknowledgement

1. Miles Stoudenmire, Steve White for the amazing collaborations.

2. Sandeep Sharma, Garnet Chan, and Chan group for insightful discussions on the classical applications of the result.

Thank you for your attention!