Simulating Quantum Circuits with Sparse Output Distributions

Martin Schwarz\textsuperscript{1}, Maarten Van den Nest\textsuperscript{2}

\textsuperscript{1}Dahlem Center for Complex Quantum Systems, Freie Universität Berlin
Arnimallee 14, D-14195 Berlin-Dahlem, Germany

\textsuperscript{2}Max Planck Institut für Quantenoptik,
Hans-Kopfermann-Str. 1, D-85748 Garching, Germany

Quant. Inf. Comp. (accepted), arXiv:1310.6749
Overview

• Classical simulation of quantum circuits
• Quantum circuits with sparse output distributions
• Sketch of the proof
Classical sim. of quantum circuits

Strong simulation: compute $O(poly(n))$ bits of output probabilities
Weak simulation: sampling output distribution (or compute $O(log(n))$ bits)

Families of simulable quantum circuits, often defined by

- restricted set of input states
- restricted gate set (non-universal)
- restricted measurements (few quibts)
Computationally Tractable (CT) states

**Definition**[^VdN11]: A state is called *computationally tractable (CT)*, if
(a) $p_x = |\langle x | \psi \rangle|^2$ can be sampled efficiently classically, and if
(b) $\langle x | \psi \rangle$ can be computed efficiently (polynomial in the bit size)

A whole theory around CT states and operations on them exists.[^VdN11]

**CT states** capture two key properties of several important families of simulable quantum states, such as
- states generated by poly-size Clifford circuits,
- nearest-neighbor matchgate circuits,
- matrix product states (MPS) with polynomial bond dimension,
- bounded tree-width circuits,
- normalizer circuits over finite Abelian groups (acting on coset states)
- etc.

New class of circuits we can simulate

reversible circuit (e.g. Toffoli, CNOTs)

approximately sparse output distribution

computationally tractable state = \textit{CT state}^{[14]}

Compare: Shor’s quantum algorithm

\[ |0\rangle \quad \mathcal{F} \quad \mathcal{F}^{-1} \]

\[ |0\rangle \quad U^x_a \]

modular exponentiation

\[ c-U^x_a : |x\rangle|y\rangle \mapsto |x\rangle|ya^x \mod N\rangle \]
Approximate sparseness

- **Def. $t$-sparse:** A quantum state $|\psi\rangle$ written in the computational basis $|x\rangle$ is called $t$-sparse, if at most $t$ of its amplitudes $\langle x|\psi \rangle$ are non-zero. The set of the respective basis states is also called the *support* of $|\psi\rangle$.

- **Def. $\varepsilon$-close:** Two quantum states $|\psi\rangle, |\varphi\rangle$ are called $\varepsilon$-close, if $\| |\psi\rangle - |\varphi\rangle \|_2 \leq \varepsilon$ in $\ell_2$ norm.

- **Def. $\varepsilon$-approximately $t$-sparse:** A quantum state $|\psi\rangle$ written in the computational basis $|x\rangle$ is called $t$-sparse, if there exists a $t$-sparse state $|\psi\rangle$ that is $\varepsilon$-close to $|\varphi\rangle$.

- **Def: additive approximation:** A function $f: x \mapsto p_x$ can be additively approx. With error $\varepsilon$ and probability $1-\delta$, if there is a randomized algorithm computing $q_x$, s.t. $|p_x - q_x| < \varepsilon$ in time $\text{poly}(n, 1/\varepsilon, \log(1/\delta))$.

  (Analogous definitions apply to probability distributions and $\ell_1$ norm.)
Approximate sparseness

\[ |\langle x|\psi|\rangle|^2 \]

| \psi \rangle

\[ O(2^n) \]
probabilities non-zero

\[ \Omega(2^{-n}) \]

all probabilities

\[ \text{not sparse} \]
probabilities too small to be estimated
Approximate sparseness

$|\langle x|\psi\rangle|^2$

$|\psi\rangle$

$|0\rangle|1\rangle \ldots$

$|2^n - 1\rangle|x\rangle$

$O(2^n/poly(n))$

probabilities non-zero

all probabilities $
\Omega(poly(n)/2^n)$

→ not sparse

Shor's algorithm: $\Omega(N/\log(N))$ amplitudes non-zero. non-zero elements cannot be identified
Approximate sparseness

\[ O(\text{poly}(n)) \]

probabilities non-zero

\[ \Omega(1/\text{poly}(n)) \]

all non-zero probabilities

\( \Rightarrow \) sparse!

elements can be identified!
probabilities can be estimated!
Approximate sparseness

\[ O(\text{poly}(n)) \]

probabilities non-zero and large

\[ |\langle x|\psi\rangle|^2 \]

\[ |\psi\rangle \]

\[ |0\rangle|1\rangle \ldots \]

\[ |2^n \rangle - 1\rangle \]

\[ |x\rangle \]

noise

→ approx. sparse
still works with noise
Main result

Theorem. Consider a unitary $n$-qubit quantum circuit composed of two blocks $C = U_2 U_1$ with input state $|\psi_{in}\rangle$. Suppose that the following conditions are fulfilled:
(a) the state $U_1 |\psi_{in}\rangle$ obtained after applying the first block is CT,
(b1) the second block $U_2$ is the QFT modulo $2^n$ or its inverse, or
(b2) the second block $U_2$ is a tensor product of unitaries $u_1 \otimes \cdots \otimes u_n$
(c) the final state $|\psi_{out}\rangle = C |\psi_{in}\rangle$ is promised to be $\sqrt{\varepsilon}$-approximately $t$-sparse for some $\varepsilon \leq 1/6$ and some $t$.

Then there exists a randomized classical algorithm with runtime $\text{poly}(n, t, 1/\varepsilon, \log 1/\delta)$ which outputs (by means of listing all nonzero amplitudes) an $s$-sparse state $|\psi\rangle$ which, with probability at least $1 - \delta$, is $O(\sqrt{\varepsilon})$-close to $|\psi_{out}\rangle$, where $s = O(t/\varepsilon)$.

(Theorem is stated for case of amplitudes and 2-norm.
Analogous theorem is true for probabilities and 1-norm.)
Simulating quantum circuits classically

\[ |\psi_{in}\rangle \xrightarrow{U_1} |\psi\rangle \xrightarrow{\mathcal{F}^{-1}} \text{approximately sparse output distribution} \]

computationally tractable state

= CT state\(^{[14]}\)

Simulating quantum circuits classically

\[ |\psi_{in}\rangle \xrightarrow{U_1} |\psi\rangle \]

A computationally tractable state = \textit{CT state}\textsuperscript{[14]}

\[ U_{2,1} \]
\[ U_{2,2} \]
\[ U_{2,3} \]


\textit{approximately sparse output distribution}
Shor’s algorithm (quantum part)

\[ |0\rangle \quad \mathcal{F} \quad \mathcal{F}^{-1} \]

\[ |1\rangle \quad U^x_a \]

modular exponentiation

\[ c-U^x_a : |x\rangle|y\rangle \rightarrow |x\rangle|ya^x \mod N\rangle \]

NO approximately sparse output distribution
Main result (again)

Theorem. Consider a unitary $n$-qubit quantum circuit composed of two blocks $C = U_2 U_1$ with input state $|\psi_{in}\rangle$. Suppose that the following conditions are fulfilled:

(a) the state $U_1 |\psi_{in}\rangle$ obtained after applying the first block is $CT$,
(b1) the second block $U_2$ is the QFT modulo $2^n$ or its inverse, or
(b2) the second block $U_2$ is a tensor product of unitaries $u_1 \otimes \cdots \otimes u_n$
(c) the final state $|\psi_{out}\rangle = C |\psi_{in}\rangle$ is promised to be $\sqrt{\varepsilon}$-approximately $t$-sparse for some $\varepsilon \leq 1/6$ and some $t$.

Then there exists a randomized classical algorithm with runtime $\text{poly}(n, t, 1/\varepsilon, \log \frac{1}{\delta})$ which outputs (by means of listing all nonzero amplitudes) an $s$-sparse state $|\psi\rangle$ which, with probability at least $1 - \delta$, is $O(\sqrt{\varepsilon})$-close to $|\psi_{out}\rangle$, where $s = O(t/\varepsilon)$.

(Theorem is stated for case of amplitudes and 2-norm. Analogous theorem is true for probabilities and 1-norm.)
Proof sketch

Main theorem requires to approximate list of outcome probabilities such as

\[ p(y) = \langle CT | [\mathcal{F}^\dagger P(y) \mathcal{F}] \otimes I | CT \rangle \]

where \( |y_1 \cdots y_m \rangle \langle y_1 \cdots y_m | \otimes I \equiv P(y) \) is a projector on \( m \)-bit string \( y \), and \( p(y) \) big.

To prove theorem, show that

1. \( \rightarrow \) function \( p(y) \) can be additively approximated for all marginals of \( y_i \)
2. \( \rightarrow \) list of \( y \) for all large \( p(y) \) can be approximated using marginals

(further: recover phases not just magnitudes, show how to sample from list)
CT states, Fourier basis, marginals

**Generalized Pauli operators**
- defined on d-level system with basis states $|x\rangle$, $x \in \mathbb{Z}_d$ as
  
  
  \[
  X_d|x\rangle = |x + 1\rangle \\
  Z_d|x\rangle = e^{\frac{2\pi i x}{d}}|x\rangle
  \]

  where $x+1$ is defined modulo $d$.

  - Note that the order of both $X$ and $Z$ is $d$.

**Fourier transforms over $\mathbb{Z}$**
- let $|\mathcal{F}_d\rangle$ be the Fourier transform over $\mathbb{Z}_d$, i.e. for $d=2^n$
  
  \[
  \mathcal{F}_{2^n} = \frac{1}{\sqrt{2^n}} \sum_{x,y \in \mathbb{Z}_{2^n}} \exp \left( \frac{2\pi i x y}{2^n} \right) |x\rangle\langle y|
  \]

  - then it follows that $\mathcal{F}_d^\dagger Z_d \mathcal{F}_d = X_d$. and $\mathcal{F}_d Z_d \mathcal{F}_d^\dagger = X_d^\dagger$. 

\[p(y) = \langle \text{CT} | [\mathcal{F}_d^\dagger P(y) \mathcal{F}_d] \otimes I \langle \text{CT} | \langle y_1 \cdots y_m | y_1 \cdots y_m | \otimes I \equiv P(y)\]
CT states, Fourier basis, marginals

**Proof idea:** Express $P(y)$ in terms of Pauli operators, with $d=2^k$, reduce to standard results on CT states.

(1) Indeed, $P(y)$ is projector onto the $1$-eigenspace of

$$M := \alpha y Z^{2^k-m}$$

where $\alpha y Z^{2^k-m} |\tilde{x}\rangle = |\tilde{x}\rangle$ with $\alpha := e^{-\frac{2\pi i}{2^m}}$, iff $\hat{x} \mod 2^m = \hat{y}$

where $\hat{y}$ is the integer value of bit string $y$.

(2) To arrive at the projector, we can average over the group generated by $M$

$$P(y) = \frac{1}{2^m} \sum_{u=0}^{2^m-1} M^u$$
(3) But now we can Fourier transform $P(y)$

\[
P(y) = \frac{1}{2^m} \sum_{u=0}^{2^m-1} M^u
\]

\[
\mathcal{F}^\dagger P(y) \mathcal{F} = \frac{1}{2^m} \sum_{u=0}^{2^m-1} N^u
\]

Thus we have switched $Z$ to $X$ operators, which are \textit{basis preserving}

\[
p(y_1 \cdots y_m) = \frac{1}{2^m} \sum_{u=0}^{2^m-1} \langle \text{CT} | N^u \otimes I | \text{CT} \rangle.
\]

where expectation values of basis preserving operators on CT states can be additively approximated by standard results on CT states\textsuperscript{14}, as well as the sum over $u$.

Proof sketch

Main theorem requires to approximate list of outcome probabilities such as

\[ p(y) = \langle \text{CT} \lvert \mathcal{F}^\dagger P(y) \mathcal{F} \rvert \otimes I \lvert \text{CT} \rangle \]

where \( |y_1 \cdots y_m\rangle \langle y_1 \cdots y_m| \otimes I \equiv P(y) \) is a projector on \( m \)-bit string \( y \), and \( p(y) \) big.

To prove theorem, show that

1. \( \rightarrow \) function \( p(y) \) can be additively approximated for all marginals of \( y_i \)
2. \( \rightarrow \) list of \( y \) for all large \( p(y) \) can be approximated using marginals

(further: recover phases not just magnitudes, show how to sample from list)
Finding strings with large probabilities

Theorem 10. Let \( \mathcal{P} = \{p_x : x \in B_k\} \) be a probability distribution. Let \( \mathcal{P}_m \) denote the marginal probability distribution of the first \( m \) bits, for every \( m \) ranging from 1 to \( k \) (with \( \mathcal{P}_k \equiv \mathcal{P} \)). Suppose that all distributions \( \mathcal{P}_m \) are additively approximable. Then the following holds: given \( \theta, \pi > 0 \), there exists a randomized classical algorithm with runtime \( \text{poly}(k, 1/\theta, \log(1/\pi)) \) which outputs a list \( L = \{x^1, \ldots, x^l\} \) where \( l \leq 2/\theta \) and where each \( x^i \) is an \( k \)-bit string such that, with probability at least \( 1 - \pi \):

(a) for all \( y \in L \), it holds that \( p(y) \geq \frac{\theta}{2} \);

(b) every \( k \)-bit string \( x \) satisfying \( p(x) \geq \theta \) belongs to the list \( L \);

The key idea of the proof is to

- perform a “binary search“ over all bit strings
- show that the search terminates in polynomial time
- argue that the strings found indeed satisfy (a)-(b)

This theorem is a generalization of a classical result from the learning theory of Boolean functions\(^{[KM91,GM89]}\).

Finding strings with large probabilities

“Binary search“ – use the fact, that \( p(0) \) and \( p(1) \) can be approximated, i.e. we compute a \( c(x) \), s.t. \( |p(x)-c(x)|<\theta/4 \), recurse if \( c(x) \) is larger than \( 3\theta/4 \).

\[
L_0 = \{\}
\]

\[
L_1 = \{0\} \cup \{1\}
\]

\[
L_2 = \{00\} \cup \{01\} \cup \{11\}
\]

\[
L_3 = \{000\} \cup \{011\} \cup \{110\} \cup \{111\}
\]

Terminate at level \( k \), if \( |L_k|>2/\theta \). Will never trigger, since \( |L_k| < 2/\theta \) due to norm=1!

(a) \( \leftarrow \) since for every \( x \) in the list \( c(x)>3\theta/4 \), and \( |p(x)-c(x)|<\theta/4 \rightarrow p(x)>\theta/2 \)

(b) \( \leftarrow \) show“if \( p(x_1...x_m)>2/\theta \), then \( x_1...x_m \in L_m “ \) Trivial for \( L_1 \). For any \( m \), suppose \( p(x_1...x_m)>\theta/2 \). Due to \( p(x_1...x_{m-1})>p(x_1...x_m) \rightarrow x_1...x_{m-1} \in L_{m-1} \rightarrow x_1...x_m \in L_m \)
Finding strings with large probabilities

(a) $\leftarrow$ since for every $x$ in the list, $c(x) > \frac{3\theta}{4}$ and $|p(x)-c(x)| < \frac{\theta}{4} \implies p(x) > \frac{\theta}{2}$

(b) $\leftarrow$ show “if $p(x_1...x_m) > \frac{2}{\theta}$, then $x_1...x_m \in L_m$” Trivial for $L_1$.

For the recursion, suppose for any $m$, $p(x_1...x_m) > \frac{\theta}{2}$. Then, due to $p(x_1...x_{m-1}) > p(x_1...x_m) \implies x_1...x_{m-1} \in L_{m-1} \implies$ def. of $L_m$ yields $x_1...x_m \in L_m$
Proof sketch

Main theorem requires to approximate list of outcome probabilities such as

\[ p(y) = \langle CT | [\mathcal{F}^\dagger P(y) \mathcal{F}] \otimes I | CT \rangle \]

where \( |y_1 \cdots y_m \rangle \langle y_1 \cdots y_m | \otimes I \equiv P(y) \) is a projector on \( m \)-bit string \( y \), and \( p(y) \) big.

To prove theorem, show that

1. \( \rightarrow \) function \( p(y) \) can be additively approximated for all marginals of \( y_i \)
2. \( \rightarrow \) list of \( y \) for all large \( p(y) \) can be approximated using marginals

(further: recover phases not just magnitudes, show how to sample from list)
Conclusions

• We have presented a new family of classically simulable quantum circuits, with a certain block structure and a novel assumption about the output probability distribution.

• The dense output distribution of Shor’s algorithm (or its generalizations) is a necessary feature for the (conjectured) exponential speed-up over classical computers.

• In order to extract meaningful information out of a dense superposition, additional structure (e.g. group structure) must be present, such that $O(poly(n))$ samples suffice to efficiently identify the structure.
Thank you!