Title

Approximate quantum circuit synthesis using block encodings

Journal

Physical Review A, 102(5)

ISSN

2469-9926

Authors

Camps, Daan

Van Beeumen, Roel

Publication Date

2020-11-01

DOI

10.1103/physreva.102.052411

Peer reviewed

Approximate Quantum Circuit Synthesis using Block-Encodings

Daan Camps[1,*] and Roel Van Beeumen[1,†]

[1] *Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

One of the challenges in quantum computing is the synthesis of unitary operators into quantum circuits with polylogarithmic gate complexity. Exact synthesis of generic unitaries requires an exponential number of gates in general. We propose a novel approximate quantum circuit synthesis technique by relaxing the unitary constraints and interchanging them for ancilla qubits via block-encodings. This approach combines smaller block-encodings, which are easier to synthesize, into quantum circuits for larger operators. Due to the use of block-encodings, our technique is not limited to unitary operators and can also be applied for the synthesis of arbitrary operators. We show that operators which can be approximated by a canonical polyadic expression can be synthesized with polylogarithmic gate complexity under certain assumptions.

I. INTRODUCTION

Quantum computing holds the promise of speeding up computations in a wide variety of fields [1]. After early breakthroughs such as Shor's algorithm [2] for factoring and Grover's algorithm [3] for searching, there have been substantial developments in various quantum algorithms over the past two decades. Noteworthy are the quantum walk algorithm of Szegedy [4, 5], and the quantum linear systems algorithm by Harrow, Hassidim, and Loyd [6]. These developments have lead to quantum linear systems [7] and Hamiltonian simulation [8] algorithms inspired by quantum walks. A unifying framework called the quantum singular value transformation, which combines the notion of qubitization [9] and quantum signal processing [10] by Low and Chuang, was recently proposed by Gilyén et al. [11, 12]. The quantum singular value transformation can describe all aforementioned quantum algorithms except factoring. Besides that, it has sparked an interest in the use of block-encodings since they can directly be used as input for a quantum singular value transformation. A block-encoding is the embedding of a –not necessarily unitary– operator as the leading principal block in a larger unitary

$$U = \begin{bmatrix} A/\alpha & * \\ * & * \end{bmatrix} \quad \Longleftrightarrow \quad A = \alpha \left(\langle 0| \otimes I\right) U \left(|0\rangle \otimes I\right).$$

In this paper, we propose the use of block-encodings, not as a building block for quantum algorithms, but as a technique for *approximate* quantum circuit synthesis and, more generally, the synthesis of arbitrary operators into quantum circuits. One of the major challenges on noisy intermediate-scale quantum (NISQ) devices is the limited circuit depth [13]. In general, exact synthesis of generic unitary operators requires exponentially many quantum gates [14–16]. The noise in NISQ devices limits the circuit depth but also relaxes the need for exact synthesis. In other words, we only need to approximate the action of some n-qubit operator up to an error proportional to the noise level. A polynomial dependence of the circuit depth on n is necessary to obtain efficient quantum circuits. Examples of other approximate synthesis approaches have been proposed in [17–19].

We show that, under certain assumptions, an efficient quantum circuit can be devised if the operator can be ϵ-approximated by a canonical polyadic (CP) expression [20, 21] with a number of terms that depends polylogarithmically on the operator dimension. CP decompositions have found applications in many scientific disciplines because they can often be computed approximately using optimization algorithms. However, their calculation is an NP-hard problem in general. We also demonstrate that the class of operators that we can efficiently synthesize is a linear combination of terms with Kronecker product structure, which is more general than standard CP decompositions. We call these expressions *CP-like* decompositions.

The proposed technique uses two operations to efficiently combine block-encodings: the Kronecker product of block-encodings and a linear combination of block-encodings. This allows us to combine block-encodings of small matrices into quantum circuits for larger operators. We show that in practice the scheme requires at most a logarithmic number of ancilla qubits and study the relation between the errors on the individual encodings and the overall circuit. Finally, we provide an example of a class of non-unitary operators that naturally have a CP-like structure and can efficiently be encoded using the proposed technique.

II. BLOCK-ENCODINGS

Since an n-qubit quantum circuit performs a unitary operation, non-unitary operations cannot directly be handled by quantum computers. One way to overcome this limitation is by encoding the non-unitary matrix into a larger unitary one, so called *block-encoding* [11, 12]. We define an *approximate* block-encoding of an operator on s signal qubits, A_s, in a unitary U_n on n qubits as follows.

* DCamps@lbl.gov
† RVanBeeumen@lbl.gov

Definition 1 *Let $a, s, n \in \mathbb{N}$ such that $n = a + s$, and $\epsilon \in \mathbb{R}^+$. Then an n-qubit unitary U_n is an (α, a, ϵ)-block-encoding of an s-qubit operator A_s if*

$$\tilde{A}_s = \left(\langle 0|^{\otimes a} \otimes I_s\right) U_n \left(|0\rangle^{\otimes a} \otimes I_s\right),$$

and $\left\|A_s - \alpha\tilde{A}_s\right\|_2 \leq \epsilon$.

The parameters (α, a, ϵ) of the block-encoding are, respectively, the *subnormalization factor* to encode matrices of arbitrary norm, the number of *ancilla* qubits, and the *error* of the block-encoding. Since $\|U_n\|_2 = 1$, we have that $\|\tilde{A}_s\|_2 \leq 1$ and $\|A_s\|_2 \leq \alpha + \epsilon$. Note that every unitary U_s is already a $(1, 0, 0)$-block-encoding of itself and every non-unitary matrix A_s can be embedded in a $(\|A_s\|_2, 1, 0)$-block-encoding [22]. This does not guarantee the existence of an efficient quantum circuit.

An equivalent interpretation of Definition 1 is that $\tilde{A}_s$ is the partial trace of U_n over the zero state of the ancilla space. This naturally partitions the Hilbert space $\mathcal{H}_n$ into $\mathcal{H}_a \otimes \mathcal{H}_s$. Given an s qubit signal state, $|\psi_s\rangle \in \mathcal{H}_s$, the action of U_n on $|\psi_n\rangle = |0\rangle^{\otimes a} \otimes |\psi_s\rangle \in \mathcal{H}_n$ becomes

$$U_n |\psi_n\rangle = |0\rangle^{\otimes a} \otimes \tilde{A}_s |\psi_s\rangle + \sqrt{1 - \|\tilde{A}_s |\psi_s\rangle\|_2^2}\, |\phi_n^\perp\rangle,$$

with

$$\left(\langle 0|^{\otimes a} \otimes I_s\right) |\phi_n^\perp\rangle = 0, \qquad \left\||\phi_n^\perp\rangle\right\|_2 = 1,$$

and $|\phi_n^\perp\rangle$ the normalized state for which the ancilla register has a state orthogonal to $|0\rangle^{\otimes a}$. By construction, we see that a partial measurement of the ancilla register projects out $|\phi_n^\perp\rangle$ and results in $(|0\rangle^{\otimes a} \otimes \tilde{A}_s |\psi_s\rangle)/\|\tilde{A}_s |\psi_s\rangle\|_2$ with probability $\|\tilde{A}_s |\psi_s\rangle\|_2^2$. In this case, the ancilla register is measured in the zero state and the signal register is in the target state $\tilde{A}_s |\psi_s\rangle$, see Figure 1. An inadmissible state orthogonal to the desired outcome is obtained with probability $1 - \|\tilde{A}_s |\psi_s\rangle\|_2^2$.

Using amplitude amplification, the process must be repeated $1/\|\tilde{A}_s |\psi_s\rangle\|_2$ times for success on average. This makes our proposed synthesis technique probabilistic.

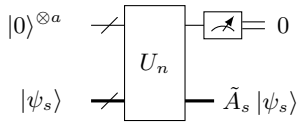


FIG. 1. Quantum circuit for U_n. The thick quantum wire carries the *signal* qubits, the other are the *ancilla* qubits. If the ancilla register is measured in the zero state, the signal register is in the desired state $\tilde{A}_s |\psi_s\rangle$.

III. COMBINING BLOCK-ENCODINGS

We introduce two operations on block-encodings that in combination allow us to build encodings of larger operators from encodings of small operators. The first operation creates a block-encoding of a Kronecker product of two matrices from the block-encodings of the individual matrices. We denote a SWAP-gate on the ith and jth qubits as SWAP_j^i.

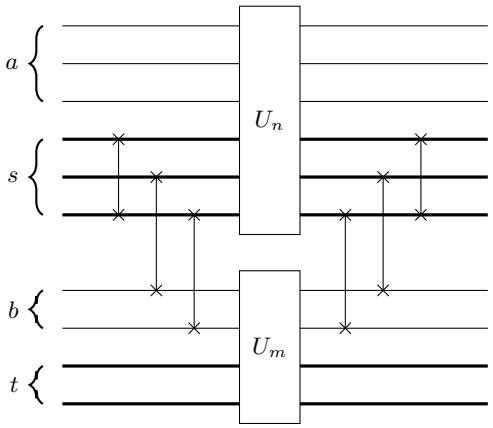
Lemma 1 *Let U_n and U_m be (α, a, ϵ_1)- and (β, b, ϵ_2)-block-encodings of A_s and A_t, respectively, and define $S_{n+m} = \prod_{i=1}^{s} \text{SWAP}_{a+b+i}^{a+i}$. Then,*

$$S_{n+m} \left(U_n \otimes U_m\right) S_{n+m}^\dagger \tag{1}$$

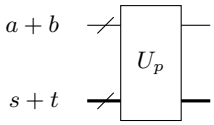
is an $(\alpha\beta, a+b, \alpha\epsilon_2 + \beta\epsilon_1 + \epsilon_1\epsilon_2)$-block-encoding of $A_s \otimes A_t$.

Lemma 1 shows how two individual block-encodings can be combined to encode the Kronecker product of two matrices. The method requires no additional ancilla qubits and the approximation error scales as a weighted sum of the individual errors up to first order. The operation requires only $2s$ additional SWAP operations.

Figure 2 shows the quantum circuit for a Kronecker product of block-encodings. This reveals the observation that in order to combine block-encodings into Kronecker products, the signal qubits of the leading block-encoding have to be swapped with the ancilla qubits of the second block-encoding in such a way that the $s + t$ signal qubits become the least-significant qubits in the combined circuit and that the mutual ordering of the signal qubits is preserved.



(a) Kronecker product of 2 block-encoded matrices



(b) Equivalent multi-qubit gate

FIG. 2. Block-encoding of the Kronecker product of 2 block-encoded matrices: (a) quantum circuit for $a = 3$, $s = 3$, $b = 2$, $t = 2$, and (b) equivalent multi-qubit gate U_p with $p = n + m$.

Lemma 1 trivially extends to Kronecker products of more than two block-encodings. Let U_{n_i} be $(\alpha_i, a_i, \epsilon_i)$-block-encodings of A_{s_i} for $i \in \{1, \ldots, d\}$. Define $n = \sum_i n_i$, and S_n as a SWAP register that swaps all signal qubits of each block-encoding U_{n_i} to the least significant

qubits of the n-qubit unitary while preserving the mutual ordering between the signal qubits. Then, ignoring the second order error terms,

$$S_n \left(U_{n_1} \otimes U_{n_2} \otimes \cdots \otimes U_{n_d}\right) S_n^\dagger \tag{2}$$

is an $(\prod_i \alpha_i, \sum_i a_i, \sum_i \epsilon_i \prod_{k\neq i} \alpha_k)$-block-encoding of $A_{s_1} \otimes A_{s_2} \otimes \cdots \otimes A_{s_d}$. In order for the subnormalization factor and approximation error on the Kronecker product not to grow too large, the subnormalization factors of the individual block-encodings should be small enough.

The second operation used in the proposed technique constructs a block-encoding of a linear combination of block-encodings. To this end, we review the notion of a *state preparation pair of unitaries* [12].

Definition 2 *Let $y \in \mathbb{C}^m$, with $\|y\|_1 \leq \beta$, and define $\underline{y_b} = \left[y^T\, 0\right]^T \in \mathbb{C}^{2^b}$, where $2^b \geq m$. Then the pair of unitaries (P_b, Q_b) is called a (β, b, ϵ)-state-preparation-pair for y if $P_b \left|0\right\rangle^{\otimes b} = p_b$ and $Q_b \left|0\right\rangle^{\otimes b} = q_b$, such that*

$$\sum_{j=0}^{2^b-1} \left|\beta(p_j^* q_j) - \underline{y_j}\right| \leq \epsilon.$$

The following lemma is a known result [23], but we provide a sharper upper bound on the approximation error compared to [12].

Lemma 2 *Let $B_s = \sum_{j=0}^{m-1} y_j A_s^{(j)}$ be an s-qubit operator and assume that (P_b, Q_b) is a (β, b, ϵ_1)-state-preparation-pair for y. Further, let $U_n^{(j)}$ be (α, a, ϵ_2)-block-encodings for $A_s^{(j)}$ for $j \in [m]$ and define the following select oracle*

$$W_{b+n} = \sum_{j=0}^{m-1} \left|j\right\rangle\left\langle j\right| \otimes U_n^{(j)} + \sum_{j=m}^{2^b-1} \left|j\right\rangle\left\langle j\right| \otimes I_n.$$

Then,

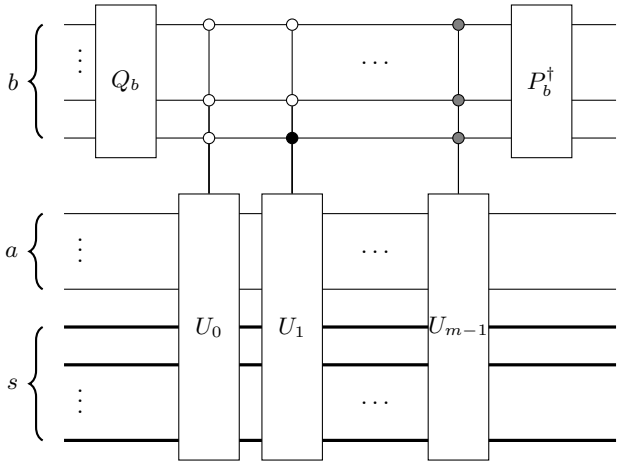
$$U_{b+n} = \left(P_b^\dagger \otimes I_a \otimes I_s\right) W_{b+n} \left(Q_b \otimes I_a \otimes I_s\right),$$

is an $(\alpha\beta, a+b, \alpha\epsilon_1 + \beta\epsilon_2)$-block-encoding of B_s.

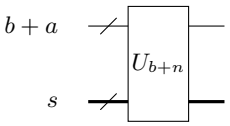
Lemma 2 shows that, if an efficient state preparation pair exists for the coefficient vector y, then we can efficiently implement a linear combination of block-encodings from the individual block-encodings. Figure 3 shows the corresponding quantum circuit. Note that this operation requires b additional ancilla qubits. The approximation error again scales as a weighted sum of the (maximum) error on the block-encodings and the error on the state-preparation pair.

The combination of Lemma 2 and Eq. (2) shows that we can directly construct a block-encoding of an s-qubit operator of the form

$$B_s = \sum_{j=0}^{m-1} y_j \; A_{s_1}^{(j)} \otimes A_{s_2}^{(j)} \otimes \cdots \otimes A_{s_{d_j}}^{(j)}, \tag{3}$$



(a) Linear combination of m block-encodings



(b) Equivalent multi-qubit gate

FIG. 3. Block-encoding of linear combinations of block-encodings: (a) quantum circuit where the gray control nodes for U_{m-1} encode the bitstring for $m-1$, and (b) equivalent multi-qubit gate.

if $\sum_{i=1}^{d_j} s_i = s$ for $j \in [m]$, i.e., all terms in the sum in Eq. (3) are of the same dimension, and if we have a block-encoding $U_{n_i}^{(j)}$ for each $A_{s_i}^{(j)}$ where $j \in [m]$, and $i \in \{1, \ldots, d_j\}$.

To quantify the subnormalization factor, the number of ancilla qubits, and the approximation error in the block-encoding for Eq. (3), we assume that each $U_{n_i}^{(j)}$ is an $(\alpha_i^{(j)}, a_i^{(j)}, \epsilon_i^{(j)})$-block-encoding for $A_{s_i}^{(j)}$. Let

$$\alpha^{(j)} = \prod_i \alpha_i^{(j)}, \quad a^{(j)} = \sum_i a_i^{(j)}, \quad \epsilon^{(j)} = \sum_i \epsilon_i^{(j)} \prod_{k\neq i} \alpha_k^{(j)},$$

for $j \in [m]$. Then, using Eq. (2), we can combine these into $(\alpha^{(j)}, a^{(j)}, \epsilon^{(j)})$-block-encodings for each term

$$A_s^{(j)} = A_{s_1}^{(j)} \otimes A_{s_2}^{(j)} \otimes \cdots \otimes A_{s_{d_j}}^{(j)}.$$

Notice that while the number of signal qubits has to be the same for each term in the linear combination, we do not assume the same number of ancilla qubits here. If we define $a = \max_j a^{(j)}$, then each block-encoding for $A_s^{(j)}$ can simply be extended to a ancilla qubits by adding additional ones at the top of the register. This does not change the leading block of the unitary.

Finally, denote $\alpha = \max_j \alpha^{(j)}$ and $\epsilon_1 = \max_j \epsilon^{(j)}$ to be the maximum subnormalization factor and the maximum approximation error over all terms. By invoking Lemma 2, we can construct a unitary U_{b+n} that is an $(\alpha\beta, a+b, \alpha\epsilon_2 + \beta\epsilon_1)$-block-encoding of B_s from Eq. (3).

The subnormalization factors $\alpha^{(j)} \leq \alpha$ can be incorporated in the vector y encoding the coefficients of the linear combination.

By incorporating the SWAP registers from Eq. (2) in the select oracle of Lemma 2, generalized Fredkin gates [24] are introduced. Fredkin gates are difficult to realize experimentally [25] and can be avoided if every Kronecker product of the block-encodings in the linear combination uses the same SWAP register. In this case, the select oracle becomes

$$W_{b+n} = (I_b \otimes S_n)\, \tilde{W}_{b+n} \left(I_b \otimes S_n^\dagger\right),$$

where

$$\tilde{W}_{b+n} = \sum_{j=0}^{m-1} |j\rangle\langle j| \otimes \tilde{U}_n^{(j)} + \sum_{j=m}^{2^b-1} |j\rangle\langle j| \otimes I_n,$$

with $\tilde{U}_n^{(j)} = U_{n_1}^{(j)} \otimes \ldots U_{n_d}^{(j)}$.

IV. DISCUSSION

Our technique combines block-encodings of small matrices to create block-encodings of larger operators that can be represented as in Eq. (3). This decomposition is closely related to the CP decomposition of a tensor [20] and allows for more generality. The sizes of the individual block-encoded matrices can differ in each term of the linear combination but they must all have the same size when combined into a Kronecker product.

Optimization algorithms, such as for example alternating least squares, have been successfully used to compute approximations to CP decompositions in many applications. Even though exact CP decompositions are NP-hard to compute in general. The optimization algorithms can be extended to accommodate for the different sizes of block-encodings in each of the terms and could incorporate the flexibility in size of the terms in their objective. They can be used as such for approximate quantum circuit synthesis. As NISQ devices suffer from noise [13], the approximate nature of algorithms for CP-like decompositions can be exploited to obtain shorter circuits for less precise decompositions with fewer terms. Under a given noise level, the error on the approximate CP-like decomposition can be balanced with the error on the individual block-encodings to find a tradeoff with short circuit depth.

One of the major challenges with using block-encodings is the introduction of an ancilla register. This removes the constraint of strictly unitary approximations and allows for linear combinations, but at the same time it introduces a probabilistic nature in the synthesis process and requires that the circuit is repeatedly executed until success.

The asymptotic gate complexity of the resulting quantum circuit synthesis technique depends on two factors: the number of terms m in the CP-like decomposition in Eq. (3) and the gate count of each individual block-encoding. If we assume that in Eq. (3) m is $\mathcal{O}(\text{poly}(s))$, then b is $\mathcal{O}(\text{polylog}(s))$ and quantum circuits with $\mathcal{O}(\text{poly}(s))$ gates for the state-preparation unitaries always exist [26]. Also the select oracle W_{b+n} of Lemma 2 can in this case be implemented with $\mathcal{O}(\text{poly}(s))$ gates in the ancilla register [27].

To get an estimate for the complexity of the synthesis of the individual block-encodings, we consider the edge case where every block $A_{s_i}^{(j)}$ is a 2×2 matrix. In this case, each block can be encoded as the leading block of a two qubit unitary, which adds one ancilla qubit for every signal qubit. Since every two qubit unitary can be synthesized with at most three CNOT-gates [28], the synthesis of all blocks individually requires at most $3ms$ CNOT gates. The Kronecker products and linear combination of the blocks only add $\mathcal{O}(\text{poly}(s))$ CNOT-gates, leading to an overall $\mathcal{O}(\text{poly}(s))$ CNOT complexity. Furthermore, the synthesis of ms two-qubit unitaries requires fewer classical resources than the synthesis of larger blocks and the total number of required ancilla qubits is $\mathcal{O}(s)$.

For the proposed technique to be efficient, it is crucial to have CP-like decompositions with $\mathcal{O}(\text{poly}(s))$ terms and with all individual block-encodings either small enough to be synthesized efficiently with small-scale synthesis algorithms or having a certain structure that admits an efficient synthesis. The strength of the technique lies in the ability to combine small-scale block-encodings to build larger operators.

We stress that unitariness of B_s is not required because of the embedding as a block-encoding and that even if B_s is unitary, the individual terms in Eq. (3) clearly are not unitary. One class of matrices that naturally exhibit the form of Eq. (3) are the Laplace-like operators [29]

$$\sum_{j=1}^{d} M^{(1)} \otimes \cdots \otimes M^{(j-1)} \otimes L^{(j)} \otimes M^{(j+1)} \otimes \cdots \otimes M^{(d)},$$

and they can directly be encoded from block-encodings of the individual terms. For example in the Laplace operator itself, all $M^{(j)}$ are identities and $L^{(j)} = L$ for $j \in \{1, \ldots, d\}$. In this case we only need one block-encoding of L, which is repeated d times, to encode the full operator. This is an improvement over the d^2 block-encodings that are required in general.

V. CONCLUSIONS

In this paper we showed how block-encodings of small matrices, which are easier to synthesize, can be combined together to create block-encodings of larger operators with CP-like structure. Under the assumption of $\mathcal{O}(\text{poly}(s))$ terms in the decomposition and small individual block-encodings, this scheme has a polynomial dependence on the number of signal qubits both for gate complexity and ancilla qubits.

Further research is required to study the class of operators with CP-like structure and operators that can be well-approximated in this form. The modification of optimization algorithms for CP decompositions [20] to admit decompositions like Eq. (3) is another interesting research direction.

ACKNOWLEDGMENTS

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, New York, NY, USA, 2010).
[2] P. W. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994) pp. 124–134.
[3] L. K. Grover, in *Proceedings 28th Annual ACM Symposium on the Theory of Computing* (ACM, 1996) pp. 212–219.
[4] M. Szegedy, "Spectra of quantized walks and a $\sqrt{\delta\epsilon}$ rule," (2004), arXiv:0401053 [quant-ph].
[5] M. Szegedy, in *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '04 (IEEE Computer Society, USA, 2004) pp. 32–41.
[6] A. W. Harrow, A. Hassidim, and S. Lloyd, Phys. Rev. Lett. **103**, 150502 (2009).
[7] A. M. Childs, R. Kothari, and R. D. Somma, SIAM J. Comput. **46**, 1920 (2017).
[8] D. W. Berry, A. M. Childs, and R. Kothari, in *2015 IEEE 56th Annual Symposium on Foundations of Computer Science* (2015) pp. 792–809.
[9] G. H. Low and I. L. Chuang, Quantum **3**, 163 (2019).
[10] G. H. Low and I. L. Chuang, Phys. Rev. Lett. **118**, 010501 (2017).
[11] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, "Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics," (2018), arXiv:1806.01838 [quant-ph].
[12] A. Gilyén, Y. Su, G. H. Low, and N. Wiebe, in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019 (Association for Computing Machinery, New York, NY, USA, 2019) pp. 193–204.
[13] J. Preskill, Quantum **2**, 79 (2018).
[14] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Boston, MA, USA, 2002).
[15] V. V. Shende, S. S. Bullock, and I. L. Markov, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **25**, 1000 (2006).
[16] C. M. Dawson and M. A. Nielsen, Quantum Inf. Comput. **6**, 81 (2006).
[17] E. A. Martinez, T. Monz, D. Nigg, P. Schindler, and R. Blatt, New J. Phys. **18**, 063029 (2016).
[18] S. Khatri, R. LaRose, A. Poremba, L. Cincio, A. T. Sornborger, and P. J. Coles, Quantum **3**, 140 (2019).
[19] E. Younis, K. Sen, K. Yelick, and C. Iancu, "QFAST: Quantum Synthesis Using a Hierarchical Continuous Circuit Space," (2020), arXiv:2003.04462.
[20] T. G. Kolda and B. W. Bader, SIAM Rev. **51**, 455 (2009).
[21] F. L. Hitchcock, Stud. Appl. Math. **6**, 164 (1927).
[22] G. Alber, T. Beth, M. Horodecki, P. Horodecki, R. Horodecki, M. Rötteler, H. Weinfurter, R. Werner, and A. Zeilinger, *Quantum Information* (Springer-Verlag Berlin Heidelberg, 2001).
[23] A. M. Childs and N. Wiebe, Quantum Inf. Comput. **12**, 901 (2012).
[24] E. Fredkin and T. Toffoli, Internat. J. Theoret. Phys. **21**, 219 (1982).
[25] T. Ono, R. Okamoto, M. Tanida, H. F. Hofmann, and S. Takeuchi, Sci. Rep. **7**, 45353 (2017).
[26] M. Plesch and Č. Brukner, Phys. Rev. A **83**, 32302 (2011).
[27] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Phys. Rev. A **52**, 3457 (1995).
[28] G. Vidal and C. M. Dawson, Phys. Rev. A **69**, 10301 (2004).
[29] D. Kressner, M. Steinlechner, and A. Uschmajew, SIAM J. Sci. Comput. **36**, A2346 (2014).

Appendix A: Proof of Lemma 1

Proof. From Definition 1 and the mixed-product property of the Kronecker product $(A\otimes B)(C\otimes D) = (AC)\otimes(BD)$, we obtain

$$\tilde{A}_s \otimes \tilde{A}_t = \left(\langle 0|^{\otimes a} \otimes I_s \otimes \langle 0|^{\otimes b} \otimes I_t\right) (U_n \otimes U_m) \left(|0\rangle^{\otimes a} \otimes I_s \otimes |0\rangle^{\otimes b} \otimes I_t\right). \tag{A1}$$

The Kronecker product $\tilde{A}_s \otimes \tilde{A}_t$ is encoded in $U_n \otimes U_m$, but not as the leading principal block. We use the property,

$$\mathrm{SWAP}_2^1 \left(I_1 \otimes |0\rangle\right) = |0\rangle \otimes I_1,$$

to show that S_{n+m} recovers the correct order by swapping the s signal qubits:

$$\begin{aligned}
S_{n+m}\left(|0\rangle^{\otimes a}\otimes I_s\otimes|0\rangle^{\otimes b}\otimes I_t\right) &= \prod_{i=1}^{s}\mathrm{SWAP}_{a+b+i}^{a+i}\left(|0\rangle^{\otimes a}\otimes I_s\otimes|0\rangle^{\otimes b}\otimes I_t\right),\\
&= \prod_{i=1}^{s-1}\mathrm{SWAP}_{a+b+i}^{a+i}\,\mathrm{SWAP}_{a+b+s}^{a+s}\left(|0\rangle^{\otimes a}\otimes I_s\otimes|0\rangle^{\otimes b}\otimes I_t\right),\\
&= \prod_{i=1}^{s-1}\mathrm{SWAP}_{a+b+i}^{a+i}\left(|0\rangle^{\otimes a}\otimes I_{s-1}\otimes|0\rangle^{\otimes b}\otimes I_1\otimes I_t\right),\\
&= \dots\\
&= |0\rangle^{\otimes a}\otimes|0\rangle^{\otimes b}\otimes I_s\otimes I_t.
\end{aligned}$$

Taking the Hermitian conjugate yields

$$\left(\langle 0|^{\otimes a}\otimes I_s\otimes\langle 0|^{\otimes b}\otimes I_t\right)S_{n+m}^{\dagger} = \langle 0|^{\otimes a}\otimes\langle 0|^{\otimes b}\otimes I_s\otimes I_t.$$

Combining this with Eq. (A1) shows

$$\begin{aligned}
\tilde{A}_s\otimes\tilde{A}_t &= \left(\langle 0|^{\otimes a}\otimes I_s\otimes\langle 0|^{\otimes b}\otimes I_t\right)S_{n+m}^{\dagger}S_{n+m}\left(U_n\otimes U_m\right)S_{n+m}^{\dagger}S_{n+m}\left(|0\rangle^{\otimes a}\otimes I_s\otimes|0\rangle^{\otimes b}\otimes I_t\right),\\
&= \left(\langle 0|^{\otimes a}\otimes\langle 0|^{\otimes b}\otimes I_s\otimes I_t\right)S_{n+m}\left(U_n\otimes U_m\right)S_{n+m}^{\dagger}\left(|0\rangle^{\otimes a}\otimes|0\rangle^{\otimes b}\otimes I_s\otimes I_t\right),
\end{aligned}$$

such that Eq. (1) has $\tilde{A}_s\otimes\tilde{A}_t$ as principal leading block. The subnormalization and approximation error of $\tilde{A}_s\otimes\tilde{A}_t$ satisfy:

$$\begin{aligned}
\left\|A_s\otimes A_t-\alpha\beta\tilde{A}_s\otimes\tilde{A}_t\right\|_2 &\le \left\|\left(\alpha\tilde{A}_s+\epsilon_1 I_s\right)\otimes\left(\beta\tilde{A}_t+\epsilon_2 I_t\right)-\alpha\tilde{A}_s\otimes\beta\tilde{A}_t\right\|_2,\\
&= \left\|\alpha\tilde{A}_s\otimes\epsilon_2 I_t+\epsilon_1 I_s\otimes\beta\tilde{A}_t+\epsilon_1 I_s\otimes\epsilon_2 I_t\right\|_2,\\
&\le \alpha\epsilon_2\left\|\tilde{A}_s\right\|_2+\beta\epsilon_2\left\|\tilde{A}_t\right\|_2+\epsilon_1\epsilon_2,\\
&\le \alpha\epsilon_2+\beta\epsilon_1+\epsilon_1\epsilon_2,
\end{aligned}$$

where we used that $\|A_s\|_2\le\alpha\|\tilde{A}_s\|_2+\epsilon_1$, and $\|\tilde{A}_s\|_2\le 1$ and analogous results for $\tilde{A}_t$. This completes the proof. □

Appendix B: Proof of Lemma 2

Proof. We have that the leading s-qubit block of U_{b+n} is given by

$$\begin{aligned}
\tilde{B}_s &= \left(\langle 0|^{\otimes b}\otimes\langle 0|^{\otimes a}\otimes I_s\right)U_{b+n}\left(|0\rangle^{\otimes b}\otimes|0\rangle^{\otimes a}\otimes I_s\right),\\
&= \left(\langle 0|^{\otimes b}\otimes\langle 0|^{\otimes a}\otimes I_s\right)\left(P_b^{\dagger}\otimes I_a\otimes I_s\right)W_{b+n}\left(Q_b\otimes I_a\otimes I_s\right)\left(|0\rangle^{\otimes b}\otimes|0\rangle^{\otimes a}\otimes I_s\right),\\
&= \left(\langle 0|^{\otimes b}P_b^{\dagger}\otimes\langle 0|^{\otimes a}\otimes I_s\right)W_{b+n}\left(Q_b|0\rangle^{\otimes b}\otimes|0\rangle^{\otimes a}\otimes I_s\right),\\
&= \left(p_b^{\dagger}\otimes\langle 0|^{\otimes a}\otimes I_s\right)W_{b+n}\left(q_b\otimes|0\rangle^{\otimes a}\otimes I_s\right).
\end{aligned}$$

Under the assumptions of the lemma, this yields

$$\begin{aligned}
\tilde{B}_s &= \sum_{j=0}^{m-1}p_b^{\dagger}|j\rangle\langle j|q_b\otimes(\langle 0|^{\otimes a}\otimes I_s)U_n^{(j)}(|0\rangle^{\otimes a}\otimes I_s) + \sum_{j=m}^{2^b-1}p_b^{\dagger}|j\rangle\langle j|q_b\otimes\langle 0|^{\otimes a}|0\rangle^{\otimes a}\otimes I_s,\\
&= \sum_{j=0}^{m-1}p_b^{\dagger}|j\rangle\langle j|q_b\otimes\tilde{A}_s^{(j)} + \sum_{j=m}^{2^b-1}p_b^{\dagger}|j\rangle\langle j|q_b\otimes\langle 0|^{\otimes a}|0\rangle^{\otimes a}\otimes I_s,\\
&= \sum_{j=0}^{m-1}p_j^{*}q_j\,\tilde{A}_s^{(j)} + \sum_{j=m}^{2^b-1}p_j^{*}q_j\,I_s.
\end{aligned}$$

By Definition 1 and Definition 2, we get that

$$\begin{aligned}
\left\| B_s - \alpha\beta\tilde{B}_s \right\|_2 &= \left\| \sum_{j=0}^{m-1} y_j A_s^{(j)} - \alpha\beta \sum_{j=0}^{m-1} p_j^* q_j \, \tilde{A}_s^{(j)} \;-\; \alpha\beta \sum_{j=m}^{2^b-1} p_j^* q_j \, I_s \right\|_2, \\
&= \left\| \sum_{j=0}^{m-1} y_j A_s^{(j)} - \alpha\beta p_j^* q_j \, \tilde{A}_s^{(j)} \;-\; \alpha \sum_{j=m}^{2^b-1} \beta p_j^* q_j \, I_s \right\|_2, \\
&\leq \alpha\epsilon_1 + \left\| \sum_{j=0}^{m-1} \underline{y_j} (A_s^{(j)} - \alpha\tilde{A}_s^{(j)}) \right\|_2 + \alpha \left\| \sum_{j=m}^{2^b-1} \underline{y_j} \, I_s \right\|_2, \\
&\leq \alpha\epsilon_1 + \beta\epsilon_2.
\end{aligned}$$

The penultimate inequality approximates all $\beta p_j^* q_j$ terms by $\underline{y_j}$ in the two sums. The error of each individual approximation is bounded by ϵ_1, such that the total error is bounded from above by $\alpha\epsilon_1$ as $\|\tilde{A}_s^{(j)}\|_2 \leq 1$ and $\|I_s\|_2 = 1$. The last term in the penultimate line is equal to zero by Definition 2. The final equality directly follows from the block-encoding property and $\|y\|_1 \leq \beta$. □