Simulation Paths for Quantum Circuit Simulation with Decision Diagrams

What to Learn from Tensor Networks, and What Not

Lukas Burgholzer Student Member, IEEE, Alexander Ploier and Robert Wille Senior Member, IEEE

Abstract—Simulating quantum circuits on classical computers is a notoriously hard, yet increasingly important task for the development and testing of quantum algorithms. In order to alleviate this inherent complexity, efficient data structures and methods such as tensor networks and decision diagrams have been proposed. However, their efficiency heavily depends on the order in which the individual computations are performed. For tensor networks the order is defined by so-called contraction plans and a plethora of methods has been developed to determine suitable plans. On the other hand, simulation based on decision diagrams is mostly conducted in a straight-forward, i.e., sequential, fashion thus far.

In this work, we study the importance of the path that is chosen when simulating quantum circuits using decision diagrams and show, conceptually and experimentally, that choosing the right simulation path can make a vast difference in the efficiency of classical simulations using decision diagrams. We propose an open-source framework (available at github.com/cda-tum/ddsim) that not only allows to investigate dedicated simulation paths, but also to re-use existing findings, e.g., obtained from determining contraction plans for tensor networks. Experimental evaluations show that translating strategies from the domain of tensor networks may yield speedups of several factors compared to the state of the art. Furthermore, we design a dedicated simulation path heuristic that allows to improve the performance even furtherfrequently yielding speedups of several orders of magnitude. Finally, we provide an extensive discussion on what can be learned from tensor networks and what cannot.

I. Introduction

Over the last couple of years, quantum computers have evolved from a theoretical computational model to practical devices aimed for pushing beyond the horizon of classically tractable problems. Even though actual quantum computers have already been built, their availability is still rather limited and devices are heavily affected by noise. Moreover, in order to develop and test potential applications, complete representations of the respective quantum states are needed—information which is fundamentally unavailable on actual devices. Hence, the simulation of quantum computations on classical machines plays a vital role in the ongoing race to realize useful applications for quantum computing.

Such simulations entail computing a representation of the state resulting from the application of a sequence of operations (typically described as a quantum circuit) to the initial state of a quantum system. Conceptually, this corresponds to a sequence of matrix-vector multiplications. While simple in principle, the underlying vectors and matrices grow exponentially with respect to the number of simulated qubits (the quantum analogue to the bit)—quickly requiring powerful supercomputing clusters to feasibly conduct the classical simulations [1]–[3]. Clever data structures such as tensor networks [4]–[6] or decision diagrams [7]–[10] have been

demonstrated to alleviate this complexity in many practically relevant cases.

To translate the problem of classically simulating a quantum circuit into the tensor network domain, each gate of the circuit as well as the initial state is represented as a tensor and individual tensors are connected via shared indices. Then, simulating the quantum circuit entails contracting all connected tensors along the shared indices until only a single tensor remains. It is commonly known that the complexity of such a simulation is extremely sensitive to the order in which the individual tensors are contracted. Accordingly, a plethora of methods have been developed to efficiently determine suitable contraction paths [11]–[14]—a task proven to be NP-hard [15]. The general idea of such contraction plans is to exploit the topological structure of the quantum circuit.

Decision diagrams, on the other hand, try to compactly represent the individual operations and quantum states by exploiting redundancies in the underlying representation. To this end, they represent these quantities as directed, acyclic graphs with complex-valued edge weights. Similar to tensor networks, the initial quantum state and each gate of a quantum circuit are first translated to their (typically linearly-sized) decision diagram representation. Simulation of such a system is then conducted by multiplying the respective decision diagrams until a decision diagram representation of the final state vector remains. Therein, the complexity of multiplying decision diagrams scales with the product of their sizes, i.e., their number of nodes. Whenever the respective intermediate decision diagrams remain rather compact, an efficient scheme for classical simulation is obtained.

Therefore, data structures such as tensor networks and decision diagrams can help to alleviate the complexity of simulating quantum circuits. In both cases, the efficiency heavily depends on the order in which computations are performed, namely the contraction plan for tensor networks and the order of matrix-matrix or matrix-vector multiplications for decision diagrams—called *simulation path* in the following. While thoroughly investigated for tensor networks, this effect has hardly been studied for decision diagrams yet.

In this work¹, we investigate this issue and propose an open-source framework that allows to exploit arbitrary simulation paths for decision diagram-based quantum circuit simulation. Instead of reinventing the wheel, we establish a flow that not only allows to investigate dedicated paths but also to re-use existing techniques, e.g., from the tensor network domain, also for decision diagrams.

¹A preliminary version of this work has been published in [16].

Considering the verification of quantum circuit compilation flow results as a particularly important use case for quantum circuit simulation, we show, both conceptually and experimentally, that choosing the right simulation path can make a vast difference in the efficiency of classical simulations using decision diagrams. To this end, we demonstrate that translating strategies from the domain of tensor networks allows for speedups of several factors compared to the state of the art in many cases. Additionally, we design a dedicated simulation path heuristic that allows to improve the performance even further—frequently yielding speedups of several orders of magnitude. Based on these evaluations, we discuss the resulting consequences on what can be learned from tensor networks and what cannot be learned from them. This eventually provides the basis for future research on quantum circuit simulation using decision diagrams.

The rest of this paper is structured as follows: Section II introduces the necessary background for the rest of this work. Then, Section III illustrates the degrees of freedom and the potential impact of arbitrary simulation paths. Motivated by that, Section IV presents the framework that allows to evaluate these arbitrary simulation paths and describes how existing techniques from the tensor network domain can be used to obtain "good" simulation paths without starting from scratch. Afterwards, Section V summarizes our experimental evaluations, followed by a discussion of their implications in Section VI. Section VII concludes the paper.

II. BACKGROUND

To keep this paper self-contained, this section briefly covers the basics on quantum circuit simulation followed by a brief review of decision diagrams—which provide the basis of the simulation approach considered in the rest of this work.

A. Quantum Circuit Simulation

A quantum state $|\varphi\rangle$ of an *n*-qubit quantum system can be described as a linear combination of 2^n basis states, i.e.,

$$|\varphi\rangle = \sum_{i \in \{0,1\}^n} \alpha_i \, |i\rangle \ \text{ with } \, \alpha_i \in \mathbb{C} \, \, \text{and } \, \, \sum_{i \in \{0,1\}^n} |\alpha_i|^2 = 1.$$

This state is commonly represented as a vector $[\alpha_{0...0},\ldots,\alpha_{1...1}]^{\top}$, referred to as *state vector*. Measuring this state leads to a collapse of the system's state to one of the basis states $|i\rangle$ —each with probability $|\alpha_i|^2$ for $i\in\{0,1\}^n$. In the following, we will always identify $|\varphi\rangle$ with its corresponding state vector, i.e.,

$$|\varphi\rangle \equiv [\alpha_{0...0}, \dots, \alpha_{1...1}]^{\top}.$$

Example 1. An important example of a quantum state is the 3-qubit Greenberger–Horne–Zeilinger or GHZ state [17]:

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle) \equiv \frac{1}{\sqrt{2}}[1, 0, 0, 0, 0, 0, 0, 1]^{\top}$$

The state of any quantum system can be manipulated by quantum operations, also called *quantum gates*. Any such gate g acting on k qubits can be identified by a unitary matrix U of size $2^k \times 2^k$, i.e., $g \equiv U$. Its action on a quantum

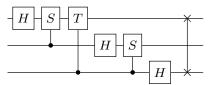


Fig. 1. Circuit for the 3-qubit quantum Fourier transform

state corresponds to the matrix-vector product of the matrix with the vector representing the state², i.e., $|\varphi'\rangle = U |\varphi\rangle$.

A quantum circuit is now described as a composition of quantum gates. Consequently, the evolution of an initial quantum state $|\varphi\rangle$ through a quantum circuit $G=g_1\dots g_{|G|}$ is described by the subsequent application of the individual gates to this initial state, i.e.,

$$|\varphi\rangle G = |\varphi\rangle g_1 \dots g_{|G|} \equiv U_{|G|} * \dots * U_1 * |\varphi\rangle.$$

If this task is conducted on a classical computer, it is commonly referred to as *quantum circuit simulation*.

Example 2. Fig. 1 shows a 3-qubit quantum circuit realizing an important quantum algorithm, namely the quantum analog to the Fourier transform. It consists of three single-qubit Hadamard gates (indicated by boxes labeled H), three two-qubit controlled-phase rotations (indicated by boxes labeled S and T connected to ●), and a two-qubit SWAP gate (indicated by ×). Given an initial state in the computational basis, this circuit outputs the state's representation in the Fourier basis.

B. Decision Diagrams

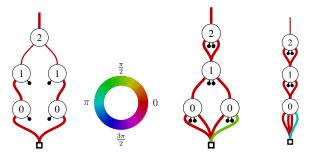
The representations of a quantum system state and the operations manipulating it are exponentially large with respect to the number of qubits involved. This quickly limits straight-forward approaches for representing (and manipulating) even moderately sized state vectors, such as arrays, without resorting to supercomputing clusters. For example, representing the dense state vector of a 32-qubit system already requires 64 GiB of memory (assuming 128 bit complex numbers).

Decision diagrams [7]–[10] have been proposed as a complementary approach for efficiently representing and manipulating quantum states by exploiting redundancies in the underlying representation. A decision diagram representing a quantum state (or operation) is a directed, acyclic graph with complex edge weights. To this end, a given state vector with its complex amplitudes α_i for $i \in \{0,1\}^n$ is recursively decomposed into sub-vectors according to

$$\begin{bmatrix} \alpha_{0\dots 0},\dots,\alpha_{1\dots 1} \end{bmatrix}^\top \\ \begin{bmatrix} \alpha_{0x} \end{bmatrix}^\top & \begin{bmatrix} \alpha_{1x} \end{bmatrix}^\top \\ \begin{bmatrix} \alpha_{00y} \end{bmatrix}^\top & \begin{bmatrix} \alpha_{01y} \end{bmatrix}^\top & \begin{bmatrix} \alpha_{11y} \end{bmatrix}^\top,$$

with $x \in \{0,1\}^{n-1}$ and $y \in \{0,1\}^{n-2}$, until only individual amplitudes remain. The resulting graph has n levels of nodes,

²Technically, the matrix first needs to be extended to the full system size (by forming appropriate tensor products with identity matrices) for the multiplication to make sense.



(a) $|GHZ\rangle$ (b) Color wheel (c) Controlled-S gate (d) H gate

Fig. 2. Decision diagrams for 3-qubit states and gates

labelled n-1 down to 0. Here, each node i has exactly two successors indicating whether the path leads to an amplitude where qubit i is in state $|0\rangle$ or $|1\rangle$.

By extracting common factors into edge weights (and employing suitable normalization schemes, see [7], [9]), any two sub-vectors that only differ by a constant factor can be unified and need not be represented by separate nodes in the decision diagram. Exploiting such redundancies frequently allows to obtain rather compact representations (in the best case linear with respect to the number of qubits) for the, in general, exponentially large state vectors.

Example 3. Fig. 2a shows a graphical representation (as proposed in [18]) of the decision diagram for the GHZ state considered previously in Example 1. To this end, the thickness of an edge indicates the magnitude of the corresponding weight, while the color wheel shown in Fig. 2b is used to encode its phase. Furthermore, edges with a weight of 0 are denoted as \bullet -stubs. In general, the decision diagram for an n-qubit GHZ state requires 2n-1 nodes for representing the 2^n -dimensional state vector.

Decision diagram representations for quantum gates are obtained by extending the decomposition scheme for state vectors by a second dimension. This corresponds to recursively splitting the respective matrix into four equally sized sub-matrices according to the basis

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

Example 4. Consider again the circuit shown in Example 2. Then, Fig. 2c and Fig. 2d show the decision diagram representations for the $2^3 \times 2^3$ matrices of the controlled-S and the Hadamard gate at the end of the circuit, respectively.

As described above, applying a gate to a quantum system entails the matrix-vector multiplication of the corresponding matrix with the current state vector. This operation can be recursively broken down according to

$$\begin{bmatrix} U_{00} & U_{01} \\ U_{10} & U_{11} \end{bmatrix} * \begin{bmatrix} \alpha_{0...} \\ \alpha_{1...} \end{bmatrix} = \begin{bmatrix} (U_{00} * \alpha_{0...} + U_{01} * \alpha_{1...}) \\ (U_{10} * \alpha_{0...} + U_{11} * \alpha_{1...}) \end{bmatrix},$$

with $U_{ij} \in \mathbb{C}^{2^{n-1} \times 2^{n-1}}$ and $\alpha_{i...} \in \mathbb{C}^{2^{n-1}}$ for $i, j \in \{0, 1\}$. Since the U_{ij} and $\alpha_{i...}$ directly correspond to the successors in the respective decision diagrams, matrix-vector (as well as matrix-matrix) multiplication is a native operation on decision

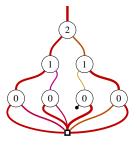


Fig. 3. Maximally large decision diagram resulting from $QFT \mid GHZ \rangle$

diagrams and its complexity scales with the product of the number of nodes of both decision diagrams. Thus, whenever the decision diagrams remain compact throughout the computation, the simulation of quantum circuits can be efficiently conducted using decision diagrams [19]–[22]. While many practical examples lead to compact decision diagram representations [23], their worst case complexity remains exponential.

Example 5. Let $|\varphi\rangle$ denote the n-qubit GHZ state and let G be the circuit for the n-qubit quantum Fourier transform. As demonstrated in Example 3 and Example 4 for n=3, both, the decision diagram representations for the initial state $|\varphi\rangle$ as well as the individual gates are linear. However, it can be shown that the decision diagram of the final state resulting from the simulation of G with initial state $|\varphi\rangle$ is maximally large, i.e., consists of 2^n-1 nodes. Fig. 3 shows the corresponding decision diagram for n=3.

III. MOTIVATION AND RELATED WORK

In this section, we consider the question of how the order in which the respective multiplications are conducted influences the complexity of decision diagram-based simulation a topic hardly considered thus far. Afterwards, we discuss correspondingly related work including how other types of quantum circuit simulators address this problem.

A. Considered Problem

As reviewed above, the simulation of a quantum circuit $G=g_1\dots g_{|G|}$ given an initial state $|\varphi\rangle$ entails the sequence of computations

$$|\varphi\rangle G = |\varphi\rangle g_1 \dots g_{|G|} \equiv U_{|G|} * \dots * U_1 * |\varphi\rangle.$$

Since matrix-matrix and matrix-vector multiplication is associative, the order in which the individual multiplications are conducted can, in principle, be chosen arbitrarily. We refer to such an order of computations as a *simulation path*. Due to matrix-vector multiplication, in general, being far less complex than matrix-matrix multiplication, the most natural simulation path is to *sequentially* compute the matrix-vector product of the individual (and compact) gate matrices with the current state vector. However, for a circuit G with |G| gates, there are

$$|G| * (|G| - 1) * \cdots * 1 = |G|!$$

i.e., exponentially many, unique simulation paths—raising the question whether the most natural path indeed is always the best path.

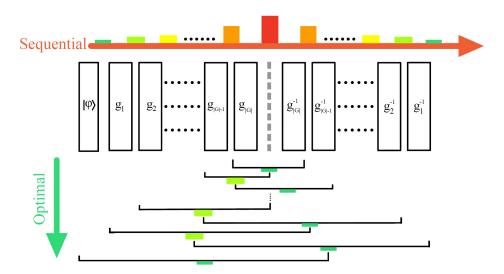


Fig. 4. Comparison of the sequential and the optimal simulation path for the problem of verifying the equivalence of two quantum circuits

In order to demonstrate the impact of the simulation path on the simulation complexity, we consider the following typical use case for quantum circuit simulation: Given two quantum circuits $G = g_1 \dots g_{|G|}$ and $G' = g_1' \dots g_{|G'|}'$, it shall be checked whether both circuits are equivalent—an essential question when, e.g., verifying the results of quantum circuit compilation flows [24]. Due to quantum circuits being inherently reversible, this can be checked by concatenating one circuit with the inverse of the other, i.e.,

$$\tilde{G} = GG'^{-1} = g_1 \dots g_{|G|} \ g'^{-1}_{|G'|} \dots g'^{-1}_1,$$

and simulating the resulting circuit with various initial states $|\varphi\rangle$. Whenever G and G' are equivalent, $\tilde{G}\equiv I$ holds (with I denoting the identity transformation) and, hence, \tilde{G} maps $|\varphi\rangle$ to itself. However, as the following example will show, choosing the right simulation path for the simulation of \tilde{G} can make the difference between linear and exponential complexity.

Example 6. Consider the scenario as in Example 5 and, for the sake of the argument, assume that G' = G, i.e., it naturally holds $\tilde{G} = GG^{-1} \equiv I$ for any $|\varphi\rangle$. Then, following the discussion in Example 5, simulating \tilde{G} in a sequential fashion leads to an intermediate decision diagram that is maximally large—implying an exponential memory complexity and, hence, exponential runtime. If however, the simulation path is chosen to start "in between" G and G^{-1} and alternate between applying gates from G and G^{-1} , any computation (except the last matrix-vector multiplication) has the form

$$U_i^{-1} * U_i$$
 or $I * U_i$,

for $i=1,\ldots,|G|$. Since, in general, the complexity of decision diagrams representing individual gates is linear, the overall runtime and memory complexity is linear as well. Fig. 4 illustrates this scenario, i.e., sketches the respectively applied matrix-matrix or matrix-vector multiplications, in addition to a color palette indicating the size of the correspondingly resulting decision diagrams (with green denoting a small and red a large size).

Obviously, the previous example is specifically constructed to show the extremes—there is a one-to-one correspondence between gates from G and G' and, hence, an easy way to define an "optimal" strategy. In practice, i.e., when $G' \neq G$, no such natural correspondence exists and, as a consequence, it is generally hard to determine an "optimal" strategy for conducting the simulation. While this is hardly surprising, given that equivalence checking of quantum circuits has been shown to be computationally hard [25], it underpins the importance of efficient and automated methods for determining suitable simulation paths.

B. Related Work

Decision diagrams are not the only data structure for simulating quantum circuits that suffers from the exponential difference in best and worst case complexity. The connection to tensor networks [4]–[6] has already been pointed out in Section I. In general, a tensor can be understood as a multi-dimensional array of complex numbers—the tensor's rank being the number of dimensions (or indices) while its shape specifies the number of elements in each dimension. Two tensors sharing common indices can be *contracted* into a single tensor by summing over repeated indices. A *tensor network* is a countable set of tensors connected by shared indices.

Example 7. Let A, B, C be matrices in $\mathbb{C}^{N \times N}$. Further, let the matrix product C = AB be given by

$$C_{i,j} = \sum_{k=0}^{N-1} A_{i,k} B_{k,j},$$

with i, j = 0, ..., N-1. Then, this corresponds to the contraction of the rank-2 tensors $A = [A_{i,k}]$ and $B = [B_{k,j}]$ over the shared index k. This is conveniently represented graphically as:

$$i - C - j = i - A - B - j$$

The order in which a tensor network is contracted into a single tensor is called a *contraction plan*. In general, an efficient contraction plan tries to keep the size of intermediate tensors and the dimension of contracted indices in check. However, the problem of determining an optimal order of contractions has been proven to be NP-hard [15]. Accordingly, trying to efficiently solve this challenging task for tensor networks has been a heavily researched topic for years [11]–[14].

Both techniques—decision diagrams and tensor networks—efficiently represent the initial state as well as all the individual operations in the form of a dedicated data structure. Then, they choose a certain path to combine these individual descriptions in order to eventually form a representation of the final quantum state—either by multiplying decision diagrams or by contracting tensors. Hence, the problem of determining an optimal simulation path for decision diagrams poses a similar challenge as determining an optimal contraction order for a tensor network.

However, in case of decision diagrams, this question is hardly studied and almost no heuristics exist for determining an efficient simulation path. Initial works related to the problem considered in this work have been conducted in [26], [27]. In [26], it is shown that initially constructing the functionality of certain building blocks in prominent quantum algorithms such as Grover's [28] or Shor's [29] algorithm (by conducting, potentially expensive, matrix-matrix multiplications) can lead to significant runtime improvements compared to the sequential matrix-vector multiplication approach described above. In [27], the authors describe schemes for constructing the functionality of such building blocks in a more efficient fashion, which can be interpreted as very specific simulation paths. But still, only a significantly limited subset of the immense space of possibilities for very specific problems has been explored.

Besides the sizeable difference of research conducted on tensor networks compared to decision diagrams in general, one of the main reasons for this disparity in focusing on the order of execution during simulation can be identified from the fundamental properties of both techniques: The performance when contracting tensors only depends on their size and shape—not on the actual content (the data) in the tensors. On the one hand, this implies that an a-priori estimate of a particular contraction plan's performance can be efficiently inferred from the sizes and shapes of the tensors involved in all contractions. On the other hand, this also means that, without a proper contraction plan, there is nothing to be gained by employing tensor networks.

In contrast, decision diagrams explicitly try to exploit redundancies in the underlying representations rather relying on "external" characteristics. As discussed in Section II-B, this allowed them to efficiently represent and simulate even large quantum systems in many cases. Consequently, while tensor networks are in dire need of efficient contraction plans to achieve peak performance, simulating circuits in a sequential fashion using decision diagrams has been "good enough" in many cases.

First steps towards combining tensor networks and decision diagrams have been taken in [10]. There, the authors use a

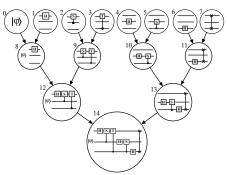


Fig. 5. Task dependency graph for one particular simulation path

variation of decision diagrams as illustrated in Section II-B to represent individual tensors in a more efficient fashion, i.e., they show what can be learned from decision diagrams in order to improve tensor networks. In contrast, this work investigates whether knowledge from the tensor network domain can lead to improvements for simulation based on decision diagrams.

IV. A SIMULATION PATH FRAMEWORK

In an effort to foster the understanding and development of simulation path heuristics for quantum circuit simulation based on decision diagrams, this section presents an open-source simulation path framework. This framework allows to execute arbitrary simulation paths via the powerful TaskFlow [30] library. Instead of reinventing the wheel and trying to compensate for years of research on tensor contraction methods, the framework includes a push-button flow to employ existing techniques from the tensor network domain while simultaneously providing the means to easily realize dedicated simulation path strategies for decision diagrams. In the following, we describe how the framework itself handles simulation paths and, afterwards, describe the flow for translating the problem from the domain of decision diagrams to the tensor network domain and back again.

A. Handling Simulation Paths

The simulation of a quantum circuit $G = g_1 \dots g_{|G|}$ with the initial state $|\varphi\rangle$ entails the computation of the expression

$$U_{|G|} * \cdots * U_1 * |\varphi\rangle$$
.

Initially, this requires the construction of decision diagrams for the initial state and the individual gates. Then, each multiplication in the above expression can be regarded as a task that takes two decision diagrams and returns the result of their multiplication. Thus, a path for the simulation of G corresponds to a sequence of (multiplication) tasks that eventually results in the final state vector. It is natural to represent such a sequence as a task dependency graph. An example illustrates the idea.

Example 8. Consider again the 3-qubit QFT circuit shown in Fig. 1. Then, the following sequence of tasks describes one particular simulation path of G:

$$[(0,1),(2,3),(4,5),(6,7),(8,9),(10,11),(12,13)].$$

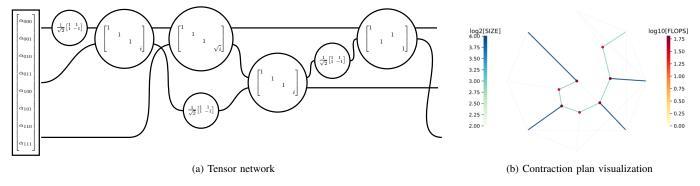


Fig. 6. Tensor network for the simulation of the 3-qubit QFT and visualization of a corresponding contraction plan

To this end, index 0 denotes the initial state, index 1 to |G| the individual operations, and the result of a task is indexed by the next largest integer not already in use. The corresponding graph is shown in Fig. 5.

This task-based formulation of quantum circuit simulation allows to employ powerful tools for asynchronous task-parallelism [30]–[32] to conduct the simulation. The resulting framework takes a circuit and a sequence of tasks as an input and uses the Taskflow [30] library to build the corresponding task dependency graph and execute it asynchronously³. The question remains how to determine suitable ones out of the |G|! options.

B. Utilizing Research on Tensor Network Contraction

As reviewed in Section III-B, a plethora of methods has been developed for determining efficient contraction plans for tensor networks and first steps have been taken to combine these two techniques. Due to the direct connection between the two domains, it would not make sense to try and reinvent the wheel when it comes to simulation using decision diagrams. Instead, we realized a flow in Python that connects both domains and, as a consequence, allows to make use of research conducted towards tensor network contraction. The following (rather technical) paragraphs give a detailed description of this process.

Starting from an initial quantum circuit (provided in the form of an *OpenQASM* file [34] or Qiskit *QuantumCircuit* object [35]), the first step is to create a corresponding tensor network representation. To this end, each individual gate is transformed to a corresponding tensor representing the underlying matrix. We do not employ tensor slicing techniques (as demonstrated in [12], [36], [37]), which allow to split the tensors of multi-qubit gates into multiple smaller tensors, since these techniques are not yet widely adopted in the decision diagram domain (although first efforts towards this direction have been conducted in [38]).

Next, the initial state $|\varphi\rangle$ needs to be translated to the tensor network domain. In general, an n-qubit state is described by a rank-n tensor of size 2^n , i.e., the complete state vector. In case of product states, i.e., states that can be written as a product of single-qubit states such as the

all-zero state $|0...0\rangle = |0\rangle \otimes \cdots \otimes |0\rangle$, this can rather be represented as n rank-1 tensors of size 2. A similarly compact representation is achieved for decision diagrams of product states, which always consist of n nodes (as opposed to the general worst case of 2^{n-1} nodes). However, while tensor networks allow for arbitrary contractions between two tensors as long as they share a common index, decision diagrams for quantum computing, as considered in this work, do not support arbitrary kinds of (tensor) contractions as of now. Instead, they only support (proper) matrix-vector and matrix-matrix multiplication, i.e., it is, for example, not possible to contract the (vector) decision diagram representing a single-qubit state and the (matrix) decision diagram representing a two-qubit operation. As a consequence, the initial state in the translated tensor network needs to be represented as a full rank-n tensor (see Section VI for further discussions).

Contracting the resulting tensor network results in a tensor representing the complete output state vector of the simulation.

Example 9. Consider once more the 3-qubit QFT circuit shown in Fig. 1. Then, the corresponding tensor network representation for the simulation of the circuit is shown in Fig. 6a.

After the translation, the resulting tensor network can be fed into any available tensor network contraction tool in order to determine a suitable contraction plan. We used the hyper-optimized tensor network contraction tool CoTenGra [11] as a state-of-the-art representative. It allows to determine contraction plans for large tensor networks using various graph-based methods and is publicly available at github.com/jcmgray/cotengra. Furthermore, it provides means to visualize contraction plans and their complexity in meaningful ways.

Example 10. Feeding the tensor network shown in Fig. 6a into CoTenGra results in the following simulation path:

$$[(0,1),(2,8),(3,9),(4,10),(5,11),(6,12),(7,13)].$$

The resulting contraction tree can be visualized as shown in Fig. 6b. To this end, the color of each node represents the number of floating point operations required for a particular contraction, while the color of each edge represents the size of the respective tensors—the darker the color, the more complex the contraction or the larger the tensor.

³A similar effort has been conducted for the tensor network domain in [33].

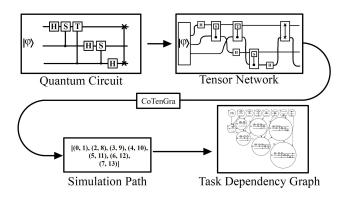


Fig. 7. Automated simulation path flow inspired by the tensor network domain

Overall, this results in a flow as shown in Fig. 7, where the initial quantum circuit is first translated to a tensor network and then fed into CoTenGra. Afterwards, a task dependency graph is constructed from the obtained contraction plan and used for the decision diagram simulation.

V. EXPERIMENTAL EVALUATIONS

The proposed simulation path framework has been implemented on top of the publicly available decision diagram-based simulator *DDSIM* (available at github.com/cda-tum/ddsim), which is part of the *Munich Quantum Toolkit* (MQT, formerly known as JKQ [39]). Afterwards, we used the resulting tool to conduct an extensive case study in order to (1) evaluate the effectiveness of strategies developed in the tensor network domain, and (2) determine the potential of dedicated heuristics for quantum circuit simulation using decision diagrams. All evaluations have been conducted on a machine equipped with an AMD Ryzen 9 3950X CPU and 128 GiB RAM running Ubuntu 20.04.

A. Experimental Setup

Inspired by the discussions in Section III-A, we primarily focused on the typical use case of verifying the equivalence of two quantum circuits, i.e., for two given quantum circuits G and G', we considered the simulation of the combined circuit $\tilde{G} = GG'^{-1}$. In order to create meaningful verification instances, a broad selection of circuits G and G' has been taken from the publicly available benchmark suite MQT Bench [40], which offers various quantum algorithms on different abstraction levels. More specifically, the circuits G have been taken from the algorithmic layer (the highest available abstraction), while the circuits G' have been taken from the native-gates layer (where circuits are compiled and optimized for a particular architecture). As discussed in Section III-A, this creates a non-trivial verification scenario where there no longer is a one-to-one correspondence between the gates of G and G' (as, e.g., shown in Example 6).

B. Dedicated Simulation Path Heuristic

In order to evaluate the potential of a non-trivial, dedicated simulation path scheme, we developed and implemented a heuristic that aims to efficiently solve this kind of verification tasks by exploiting some knowledge about the compilation flow itself (inspired by the ideas in [24]). For each gate g in the original circuit G, the resulting method estimates the corresponding number of operations in G' based on the decomposition of g into the native gate-set of G' (which can easily be computed a-priori using the same settings for the decomposition as were used for compiling G to G'). Then, starting in between both circuits, any application of a gate from G is followed by the application of the corresponding number of gates from G'^{-1} , until only the final multiplication with the initial state vector remains—yielding the final result of the simulation.

In all but the most trivial cases, any method based on this principle can only ever derive an approximation of the actual number of operations in the compiled circuit, since already the simplest optimizations employed during compilation frequently eliminate a significant amount of gates from the overall circuit. However, as witnessed by the evaluations, this type of simulation path still frequently allows to keep the intermediate decision diagrams throughout the simulation very compact (i.e., close to the identity structure) and, hence, allows for an efficient simulation.

C. Experimental Results

A representative subset of the obtained results is summarized in Table I—with the first three columns denoting the name of the benchmark, the number of qubits n, as well as the number of gates $|\tilde{G}|$ of the combined circuit. The remaining columns show

- the runtime t_{seq} of the state-of-the-art, i.e., sequential, simulation path strategy [19],
- the runtime t_{ten} of the proposed flow for translating strategies from the tensor network domain (split into the time t_{cot} spent on determining a simulation path using CoTenGra⁴ and the simulation time t_{sim}), as well as
- ullet the runtime t_{heur} of the proposed dedicated simulation path heuristic.

The framework as well as the benchmark script are publicly available at github.com/cda-tum/ddsim to conduct further evaluations.

In a first series of evaluations, we used the design flow proposed in Section IV-B to make use of the plethora of available tensor network strategies to determine a suitable simulation path. As shown by these results, re-using or translating methods developed in the tensor network domain via the proposed flow can already speed up the simulation of quantum circuits using decision diagrams by a large margin compared to the state-of-the-art, i.e., sequential, approach [19] (runtimes of instances where $t_{ten} < t_{seq}$ are highlighted in **bold**). Interestingly, for some cases (such as the *Graph State* benchmark on 54 qubits), the substantially improved runtime (t_{sim}) is overshadowed by the time spent on searching for a suitable path (t_{cot}) —these instances are highlighted in *italic*. In other cases, nothing at all is to be gained by spending up to

⁴The tool allows to set a timeout and a maximum number of repetitions, which were set to 600 s and 65, respectively.

TABLE I EXPERIMENTAL EVALUATIONS

Benchmark	SOTA [19]	Proposed Flow (Section IV-B)			Section V-B		
Name	n	$ \tilde{G} $	t_{seq} [s]	$t_{cot} [s]$	t_{sim} [s]	$t_{ten} [s]$	t_{heur} [s]
GHZ State	64	130	0.00	6.92	0.00	6.92	0.00
GHZ State	96	194	0.01	14.35	0.01	14.36	0.01
GHZ State	128	258	0.01	29.34	0.01	29.35	0.02
Graph State	50	300	46.60	23.55	0.04	23.59	0.21
Graph State	52	312	1.47	11.59	0.06	11.64	0.14
Graph State	54	324	44.71	13.84	0.07	13.92	0.36
Graph State	56	336	8.09	30.22	36.37	66.59	0.15
Two Local Ansatz	14	826	2.79	46.60	2.51	49.11	0.02
Two Local Ansatz	15	930	12.72	38.31	12.37	50.69	0.02
Two Local Ansatz	16	1040	70.56	76.55	55.88	132.42	0.02
Two Local Ansatz	17	1156	210.21	122.91	231.75	354.66	0.03
Entangled QFT	16	714	105.25	48.81	0.24	49.05	0.04
Entangled QFT	17	799	529.37	55.33	113.47	168.81	0.09
Entangled QFT	18	893	6 821.64	75.58	2.36	77.94	0.18
Deutsch-Jozsa	64	614	0.03	36.61	0.03	36.63	0.05
Deutsch-Jozsa	96	952	0.03	94.21	0.03	94.24	0.11
Deutsch-Jozsa	128	1278	0.05	164.98	0.06	165.03	0.21
QPE inexact	15	668	0.68	61.02	0.59	61.60	1.93
QPE inexact	16	750	3.77	50.48	4.25	54.73	4.41
QPE inexact	17	843	20.22	47.47	13.85	61.33	14.37
QPE inexact	18	935	190.23	70.47	43.18	113.65	33.43
Efficient SU2 Ansatz	14	882	3.98	32.82	3.30	36.12	0.35
Efficient SU2 Ansatz	15	990	13.29	82.51	12.20	94.72	0.90
Efficient SU2 Ansatz	16	1104	43.47	68.41	40.19	108.60	4.11
Efficient SU2 Ansatz	17	1224	222.47	80.92	409.12	490.04	11.37
W-State	64	758	0.03	71.04	0.02	71.06	0.04
W-State	96	1142	0.05	133.44	0.05	133.49	0.09
W-State	128	1526	0.08	239.39	0.09	239.49	0.14
Real Amplitudes Ansatz	15	930	15.02	62.51	32.06	94.57	0.02
Real Amplitudes Ansatz	16	1040	104.56	96.86	47.08	143.94	0.02
Real Amplitudes Ansatz	17	1156	252.46	112.12	218.20	330.32	0.03

n: Number of qubits $|\tilde{G}|$: Gate count of GG'^{-1} t_{seq} : Total runtime of simulation with sequential path [19] t_{cot} : Runtime of CoTenGra [11] t_{sim} : Runtime of path determined by CoTenGra [11] t_{ten} : Total runtime of simulation using the proposed flow

 t_{heur} : Total runtime of simulation with heuristic simulation path

600 s on the search for a suitable path. This is further discussed in Section VI.

In a second series of evaluations, we studied the performance of the dedicated heuristic simulation path method proposed above. The results clearly underline the potential of dedicated simulation path schemes for decision diagrams. Although the proposed method is only a heuristic, it oftentimes yields several orders of magnitude faster runtimes compared to the state of the art (again, runtimes where $t_{heur} < t_{seq}$ are highlighted in **bold**). Most notably, for the *Entangled QFT* benchmark on 18 qubits, choosing the right simulation path makes the difference between waiting almost two hours for the result and having it available in the blink of an eye.

VI. DISCUSSION

While the problem of finding a suitable simulation path for decision diagrams has many apparent similarities to finding an efficient contraction plan for tensor networks, the results summarized above demonstrate that there are key differences between both data structures. Based on that, interesting conclusions can be drawn on what *can* be learned from tensor networks and what *cannot* for quantum circuit simulation based on decision diagrams.

For tensor networks, it is inevitable that the final result of the full state simulation is exponentially large. This inherently limits the usability of such simulations to the available system memory. In contrast, there are many known quantum algorithms whose state vectors during the simulation emit a very compact (in the best case, linear) decision diagram representation [19] (also witnessed in the evaluation above for the *GHZ State*, *W-State*, and *Deutsch-Jozsa* benchmarks). As such, there is no strict qubit limit for simulation based on decision diagrams. Furthermore, the margin between the simulation times of all three considered strategies is negligible and, hence, it does not make sense to spend any amount of time searching for an alternative simulation path in these cases.

One might be tempted to conclude that, given the same simulation path/contraction plan, simulation based on decision diagrams always has at most as large a memory footprint and takes at most as long as the corresponding tensor network contraction—which would imply that decision diagrams are strictly superior to tensor networks. However, in order to maintain the decision diagram structure, a data structure more advanced than simple multi-dimensional arrays (as commonly used for tensors) is necessary. As a consequence, each node in a decision diagram is significantly larger in memory than individual tensor entries. Since, in the worst case, the decision diagram consists of exponentially many unique nodes (2^{n-1}) for an n-qubit state), the simulation might incur a much higher memory footprint. In addition, decision diagram operations frequently make use of compute tables in order to avoid redundant computations and exploit any redundancy present in the representations. If the decision diagram contains no redundancies, all that constitutes a significant overhead compared to straight-forward tensor contraction (as, e.g., observed for the sequential simulation of the *Entangled QFT* benchmark). Overall, decision diagrams and tensor networks complement each other in many different ways and, therefore, should be employed under consideration of the specific task at hand.

Although finding a suitable simulation path inspired by the tensor network domain may consume a significant amount of time (in some cases overshadowing the subsequent runtime improvements), the runtime (t_{sim}) of the simulation paths determined by CoTenGra demonstrates that there is something to be learned from the domain of tensor networks. However, the current state of the art in decision diagrams for quantum computing, as they are considered in this work, imposes several limitations on what can actually be learned from the domain of tensor networks.

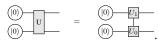
One of the main (technical) restrictions of decision diagrams at the moment is that they only allow for proper matrix-vector and matrix-matrix multiplication while tensor networks allow for arbitrary contractions between two tensors (as long as they share a common index). So, while the following contraction between a single-qubit state and a two-qubit operation makes perfect sense for tensor networks

decision diagrams currently only permit the following type of contraction

$$\begin{array}{c|c} \hline |0\rangle \\ \otimes \\ \hline |0\rangle \\ \end{array} = \begin{array}{c|c} U & = \end{array} = \begin{array}{c|c} U & |00\rangle \\ \end{array} = \begin{array}{c|c} \end{array}$$

This limits the degrees of freedom to explore during the search for an efficient contraction plan and eliminates some of the benefits of tensor networks as it essentially fixes the tensor of the initial state to be maximally large.

In a similar fashion, so-called tensor slicing techniques have been shown to accelerate and trivially parallelize tensor network contractions [12], [36], [37]. There, the main idea is to split multi-qubit tensors into multiple smaller tensors of higher order, as exemplary illustrated in the following



While first efforts towards employing such techniques for decision diagrams have been conducted in [38], they are not yet mature and flexible enough to be employed in the same way as for tensor networks.

As discussed above, employing tensor networks for full quantum state simulation inherently carries an exponential memory complexity. For that reason, tensor networks are commonly rather used to determine individual amplitudes of the resulting state. This is accomplished by attaching additional tensors to the end of the tensor network that describe the desired amplitude, as exemplary illustrated by



Since the result of the simulation no longer is an exponentially large vector, but rather a single scalar, the order of contractions plays a far greater role for the efficiency of such simulation. At the moment, the only established way to determine individual amplitudes using decision diagrams is to compute the decision diagram for the full state vector and, then, extract the desired amplitude by traversing the decision diagram from top to bottom. While there still are many instances where this procedure works just fine, this imposes a significant restriction on the available degrees of freedom for determining a suitable simulation path.

VII. CONCLUSIONS

In this work, we studied the importance of the path that is chosen when simulating quantum circuits using decision diagrams. The resulting framework allows to employ arbitrary simulation paths and to connect the domain of tensor networks with the domain of decision diagrams. Our experimental evaluations have shown that much can be learned from the domain of tensor networks-potentially allowing for runtime improvements of up to several factors compared to the state of the art. In addition, we demonstrated that the development of application-specific heuristics which are tailored for decision diagrams can achieve speedups of several orders of magnitude compared to the state of the art. Finally, we have shown, conceptually as well as experimentally, that decision diagrams and tensor networks differ in some key aspects and that not everything can be learned from tensor networks—at least not given the current state of the art in decision diagrams for quantum computing. As decision diagrams, which still are a very young data structure compared to tensor networks, are

developed further, the potential for taking advantage of all the research conducted towards tensor networks is expected to increase drastically. By making the developed framework publicly available, we hope to accelerate such endeavors.

ACKNOWLEDGMENTS

This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101001318), was part of the Munich Quantum Valley, which is supported by the Bavarian state government with funds from the Hightech Agenda Bayern Plus, and has been supported by the BMWK on the basis of a decision by the German Bundestag through project QuaST, as well as by the BMK, BMDW, and the State of Upper Austria in the frame of the COMET program (managed by the FFG).

REFERENCES

- G. G. Guerreschi, J. Hogaboam, F. Baruffa, and N. P. D. Sawaya, "Intel Quantum Simulator: A cloud-ready high-performance simulator of quantum circuits," *Quantum Sci. Technol.*, 2020.
- [2] T. Häner and D. S. Steiger, "0.5 petabyte simulation of a 45-Qubit quantum circuit," in *Int'l Conf. for High Performance Computing*, Networking, Storage and Analysis, 2017.
- [3] T. Jones, A. Brown, I. Bush, and S. C. Benjamin, "QuEST and high performance simulation of quantum computers," in *Scientific Reports*, 2018.
- [4] R. Jozsa, On the simulation of quantum circuits, 2006. arXiv: quantph/0603163.
- [5] I. L. Markov and Y. Shi, "Simulating quantum computation by contracting tensor networks," SIAM J. Comput., vol. 38, no. 3, pp. 963–981, 2008.
- [6] J. D. Biamonte and V. Bergholm, Tensor networks in a nutshell, 2017. arXiv: 1708.00006.
- [7] P. Niemann et al., "QMDDs: Efficient quantum function representation and manipulation," *IEEE Trans. on CAD of Integrated Circuits and Systems*, 2016.
- [8] L. Chin-Yung, W. Shiou-An, and K. Sy-Yen, "An extended XQDD representation for multiple-valued quantum logic," *IEEE Trans. Comput.*, pp. 1377–1389, 2011.
- [9] A. Zulehner, S. Hillmich, and R. Wille, "How to efficiently handle complex values? Implementing decision diagrams for quantum computing," in *Int'l Conf. on CAD*, 2019.
- [10] X. Hong et al., A tensor network based decision diagram for representation of quantum circuits, 2020. arXiv: 2009.02618.
- [11] J. Gray and S. Kourtis, "Hyper-optimized tensor network contraction," Quantum, vol. 5, p. 410, 2021.
- [12] C. Huang et al., Classical simulation of quantum supremacy circuits, 2020. arXiv: 2005.06787.
- [13] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, and H. Neven, Simulation of low-depth quantum circuits as complex undirected graphical models, 2018. arXiv: 1712.05384.
- [14] D. Lykov et al., Tensor Network Quantum Simulator With Step-Dependent Parallelization, 2020. arXiv: 2012.02430.
- [15] L. Chi-Chung, P. Sadayappan, and R. Wenger, "On optimizing a class of multi-dimensional loops with reduction for parallel execution," *Parallel Process. Lett.*, 1997.
- [16] L. Burgholzer, A. Ploier, and R. Wille, "Exploiting arbitrary paths for the simulation of quantum circuits with decision diagrams," in *Design*, *Automation and Test in Europe*, 2022.
- [17] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information. Cambridge University Press, 2010.
- [18] R. Wille, L. Burgholzer, and M. Artner, "Visualizing decision diagrams for quantum computing," in *Design, Automation and Test in Europe*, 2021.
- [19] A. Zulehner and R. Wille, "Advanced simulation of quantum computations," *IEEE Trans. on CAD of Integrated Circuits and Systems*, 2019.
- [20] V. Samoladas, "Improved BDD algorithms for the simulation of quantum circuits," in *Algorithms - ESA*, D. Halperin and K. Mehlhorn, Eds., 2008.

- [21] G. F. Viamontes, I. L. Markov, and J. P. Hayes, "High-performance QuIDD-Based simulation of quantum circuits," in *Design*, *Automation and Test in Europe*, 2004.
- [22] S. Hillmich, R. Kueng, I. L. Markov, and R. Wille, "As accurate as needed, as efficient as possible: Approximations in DD-based quantum circuit simulation," in *Design, Automation and Test in Europe*, 2020.
 [23] T. Grurl *et al.*, "Arrays vs. Decision Diagrams: A case study on
- [23] T. Grurl et al., "Arrays vs. Decision Diagrams: A case study on quantum circuit simulators," in *Int'l Symp. on Multi-Valued Logic*, 2020.
- [24] L. Burgholzer, R. Raymond, and R. Wille, "Verifying results of the IBM Qiskit quantum circuit compilation flow," in *Int'l Conf. on Quantum Computing and Engineering*, 2020.
- [25] D. Janzing, P. Wocjan, and T. Beth, ""Non-identity check" is QMA-complete," Int. J. Quantum Inform., vol. 03, no. 03, pp. 463–473, 2005.
- [26] A. Zulehner and R. Wille, "Matrix-Vector vs. Matrix-Matrix multiplication: Potential in DD-based simulation of quantum computations," in *Design, Automation and Test in Europe*, 2019.
- [27] L. Burgholzer, R. Raymond, I. Sengupta, and R. Wille, "Efficient construction of functional representations for quantum algorithms," in *Int'l Conf. of Reversible Computation*, 2021.
- [28] L. K. Grover, "A fast quantum mechanical algorithm for database search," *Proc. of the ACM*, pp. 212–219, 1996.
- [29] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," SIAM J. Comput., 1997.
- [30] T.-W. Huang, D.-L. Lin, C.-X. Lin, and Y. Lin, "Taskflow: A lightweight parallel and heterogeneous task graph computing system," *IEEE Trans. Parallel Distrib. Syst.*, 2021.
- [31] H. Carter Edwards, C. R. Trott, and D. Sunderland, "Kokkos: Enabling manycore performance portability through polymorphic memory access patterns," *Journal of Parallel and Distributed Computing*, vol. 74, no. 12, pp. 3202–3216, 2014.
- [32] H. Kaiser et al., "HPX The C++ standard library for parallelism and concurrency," JOSS, vol. 5, no. 53, p. 2352, 2020.
- [33] T. Vincent et al., Jet: Fast quantum circuit simulations with parallel task-based tensor-network contraction, 2021. arXiv: 2107.09793.
- [34] A. W. Cross et al., OpenQASM 3: A broader and deeper quantum assembly language, 2021. arXiv: 2104.14722.
- [35] G. Aleksandrowicz et al., "Qiskit: An open-source framework for quantum computing," Zenodo, 2019.



Lukas Burgholzer Lukas Burgholzer (S'19) received his Master's degree in industrial mathematics (2018) and Bachelor's degree in computer science (2019) from the Johannes Kepler University Linz, Austria. He is currently a Ph.D. student at the Institute for Integrated Circuits at the Johannes Kepler University Linz, Austria. His research focuses on design automation and software for quantum computing. In these areas, he has published several papers on international conferences such as ASP-DAC, DAC, ICCAD, DATE, and QCE.

- [36] Z.-Y. Chen et al., "64-qubit quantum circuit simulation," Science Bulletin, vol. 63, no. 15, pp. 964–971, 2018.
- [37] E. Pednault et al., Pareto-efficient quantum circuit simulation using tensor contraction deferral, 2020. arXiv: 1710.05867.
- [38] L. Burgholzer, H. Bauer, and R. Wille, "Hybrid Schrödinger-Feynman simulation of quantum circuits with decision diagrams," in *Int'l Conf.* on Quantum Computing and Engineering, 2021.
- [39] R. Wille, S. Hillmich, and L. Burgholzer, "JKQ: JKU tools for quantum computing," in *Int'l Conf. on CAD*, 2020.
- [40] N. Quetschlich, L. Burgholzer, and R. Wille, "MQT Bench: Benchmarking software and design automation tools for quantum computing," 2022. arXiv: 2204.13719 [quant-ph].



Alexander Ploier Alexander Ploier received his Master's degree in industrial mathematics (2019) from the Johannes Kepler University Linz, Austria. He is currently a Ph.D. student at the Institute for Integrated Circuits at the Johannes Kepler University Linz, Austria. His research interests include design automation for quantum computing—currently focusing on different data structures.



Robert Wille Robert Wille is a Full and Distinguished Professor at the Technical University of Munich, Germany, and Chief Scientific Officer at the Software Competence Center Hagenberg, Austria. He received the Diploma and Dr.-Ing. degrees in Computer Science from the University of Bremen, Germany, in 2006 and 2009, respectively. Since then, he worked at the University of Bremen, the German Research Center for Artificial Intelligence (DFKI), the University of Applied Science of Bremen, the University of Potsdam, and the Technical University

Dresden. From 2015 until 2022, he was Full Professor at the Johannes Kepler University Linz, Austria, until he moved to Munich. His research interests are in the design of circuits and systems for both conventional and emerging technologies. In these areas, he published more than 400 papers and served in editorial boards as well as program committees of numerous journals/conferences such as TCAD, ASP-DAC, DAC, DATE, and ICCAD. For his research, he was awarded, e.g., with Best Paper Awards, e.g., at TCAD and ICCAD, an ERC Consolidator Grant, a Distinguished and a Lighthouse Professor appointment, a Google Research Award, and more.