

# Operational Framework for a Quantum Database

Carla Rieger<sup>1,2</sup>, Michele Grossi<sup>2</sup>, Gian Giacomo Guerreschi<sup>3</sup>, Sofia Vallecorsa<sup>2</sup>, and Martin Werner<sup>1</sup>

<sup>1</sup>School of Engineering and Design, Technical University of Munich, Germany

<sup>2</sup>European Organization for Nuclear Research (CERN), Geneva 1211, Switzerland

<sup>3</sup>Intel Labs, Intel Corporation, 2200 Mission College Blvd, Santa Clara, CA 95054, United States of America

May 27, 2024

## Abstract

Databases are an essential component of modern computing infrastructures and allow efficient access to data stored persistently. Their structure depends on the type and relationships of the stored data elements and on the access pattern. Extending the concept of databases to the quantum domain is expected to increase both the storage efficiency and the access parallelism through quantum superposition. In addition, quantum databases may be seen as the result of a prior state preparation ready to be used by quantum algorithms when needed. On the other hand, limiting factors exist and include entanglement creation, the impossibility of perfect copying due to the no-cloning theorem, and the impossibility of coherently erasing a quantum state.

In this work, we introduce quantum databases within the broader context of data structures using classical and quantum data and indexing. In particular, we are interested in quantum databases' practical implementation and usability, focusing on the definition of the basic operations needed to create and manipulate data stored in a superposition state. Specifically, we address the case of quantum indexing in combination with classical data. For this scenario, we define the operations for database preparation, extension, removal of indices, writing, and read-out of data, as well as index permutation. We present their algorithmic implementation and highlight their advantages and limitations. Finally, we introduce steps toward defining the same operations in the more general context of quantum indexing and quantum data.

## 1 Introduction

Managing and analyzing large amounts of data is a fundamental task in the big-data age when most decisions are expected to be data-driven. At the same time, new technologies like quantum computing have been developed to solve computational bottlenecks. It is natural to ask whether the advantages also extend to the database space. One of the key reasons for quantum computational advantage, e.g., in Grover's algorithm [1], is the presence of quantum superposition states that allow for inherent parallelism and interference. Furthermore, quantum systems have access to a state space that is exponentially large with respect to the number of qubits, which permits a compressed representation of data. Intuitively, to take full advantage of this representation, one needs to process the quantum state without converting it to classical information, e.g., by avoiding measurement. Prominent algorithms which may operate on quantum databases include pattern recognition [2, 3], collision finding [4], quantum search of an unstructured database [1], solving linear systems of equations [5] and many more.

Generally, we can characterize a database formalized by the data type stored and how elements

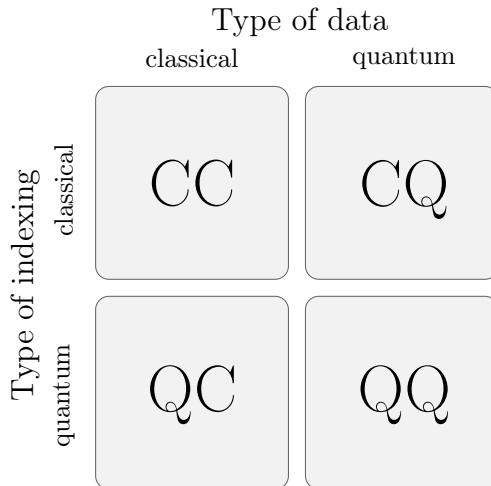


Figure 1: The different types of database structured with respect to classical and quantum states as index and data ( $C$ : classical,  $Q$ : quantum). Operations defined for each possible combination of type of data and indexing for a database differ vastly.

are indexed. Thus, we can distinguish four distinct settings as shown in Figure 1. This work focuses on designing database manipulation operations for the case of quantum indexing, mainly with respect to classical data encoded in a set of qubits. The community has discussed this data structure within the framework of a quantum random access memory (QRAM) [6, 7]. The QRAM query performance is reduced with respect to a classical random access memory call as fewer logic gates need to be activated. In this model, a register state indexed by the state  $|j\rangle$  points to the data element  $|d_j\rangle$ . A QRAM memory call returns a superposition of *data* states upon providing a superposition of indices [6]:

$$\sum_j \psi_j |j\rangle \xrightarrow{\text{QRAM call}} \sum_j \psi_j |j\rangle |d_j\rangle. \quad (1)$$

In (1), the amplitudes are normalized, i.e.,  $\sum_j |\psi_j|^2 = 1$ . The framework on which we base our work, as introduced in the following, is aligned with the QRAM model.

Recently, the definition of quantum data centers (QDC) [8, 9] as a QRAM combined with a quantum network has been proposed. In this proposal [8], the applications range from usage as a  $T$ -gate resource for multiparty private quantum communication to distributed sensing through data compression. Thus, states structured as in the QRAM model in (1) are relevant in many ways. In the context of quantum databases, recent works aim to define database manipulation operations acting on a quantum superposition state [10, 8, 11] hinting at the differences with respect to classical and quantum data.

Several fundamental quantum mechanical phenomena limit the operations that can be done within the framework of a quantum database, including the *no-cloning* [12], and the *no-deletion theorem* [13, 14]. Other works [10, 8, 11] consider database operations such as *select*, *extend* and *delete* in a more simplified manner or do not provide an implementation. An example concerning the QRAM model is given by the Flip-Flop QRAM [15] which demonstrates a specific implementation for *writing* classical data in a superposition state.

In this work, we introduce the specific set of operations that we believe are essential to operating a database in practice and that, thus, constitute the basis of our extension from classical to quantum databases. Their definition is presented in Table 1, and their implementation will be discussed for specific cases. In the general case of quantum indexing, we define a quantum

database as being the state:

$$|\text{QDB}^{(k)}\rangle = \sum_{j=0}^{k-1} \psi_j |j\rangle_I |d_j\rangle_D \in \mathcal{H}_I \otimes \mathcal{H}_D, \quad (2)$$

with  $\sum_j |\psi_j|^2 = 1$ . In the following, we often consider the restricted case of balanced database entries, namely in (2) we have  $\psi_j = 1/\sqrt{k}$  for  $\forall j$ . For certain algorithms, we exclude the amplitude  $\psi_0$  corresponding to the index  $|0\rangle_I |0\rangle_D$  from being balanced and use this state as a *probability reservoir* that can be used to add new data elements or be replenished when data are removed. The superposition state in (2) consists of a separate index ( $I$ ) and a data register ( $D$ ). The index states are orthonormal and belong to the Hilbert space  $\mathcal{H}_I$  of dimension at least  $k$ . The data states are part of the Hilbert space  $\mathcal{H}_D$  of dimension  $m$  composed of  $\tilde{m}$  qubits with  $m = 2^{\tilde{m}}$ . For simplicity, we assume that both spaces are formed by qubits. The superposition state in (2) is a linear combination of  $k$  elements. For the data elements, we used the shorthand notation  $|d_j\rangle = |d_{\alpha_j}\rangle$  indicating that the  $j$ -th data is chosen from a set of states  $\{|d_{\alpha}\rangle\}_{\alpha=0,1,\dots}$  in  $\mathcal{H}_D$ . The properties of such a set of states determine if we consider the data as quantum or classical. Specifically, classical data correspond to orthogonal states  $\langle d_{\alpha} | d_{\beta} \rangle = \delta_{\alpha\beta}$  that can be mapped to the computational basis by a known transformation  $U_D$ . Notice that the data elements in  $|\text{QDB}^{(k)}\rangle$  do not have to be chosen uniquely, meaning that it is not necessary that  $\alpha_j \neq \alpha_i$  for  $j \neq i$  in (2). The number of qubits is chosen to be the smallest integer that fulfills  $k = \lceil \log_2(k) \rceil$ . By  $\lceil x \rceil$  we denote the ceiling function that returns the smallest integer greater than or equal to  $x$ . We consider the different scenarios for orthogonal  $|d_j\rangle_D$  and non-orthogonal data entry states  $|\tilde{d}_j\rangle_D$ , discuss the former situation in detail and then present an outlook for the operations on the latter.

The rest of this article is structured as follows. We start by introducing the case of classical indexing in Section 2 and contrast it with the case of quantum indexing in Section 3.1. For the case of quantum indexing and classical data states, we formally define the operations:

*prepare, extend, remove,  
write, read-out and permute*

as summarized in Table 1. We then focus on the quantum database extension algorithm presented in Section 3.1.2. Following that, we summarize implications on the defined operations in the case of quantum data in Section 3.2 as some of the operations are equivalent to the classical data case and conclude this work in Section 4.

## 2 Data structured by classical indices (CC and CQ)

In the classical computer science domain, database models encompass structured frameworks for efficiently organizing, storing, and managing data, including hierarchical, network, relational, and object-oriented models. Briefly, the hierarchical model arranges data in a tree-like structure, while the network model uses a graph structure to handle complex many-to-many relationships. The relational model stores data in normalized tables, which are accessed via structured query languages (see, e.g., [16]), offering robust querying and indexing capabilities. Database models emphasize the importance of individual data relationships in maintaining the integrity of data and facilitating complex queries. Due to the maturity of this field, we focus on a form of relational databases that organize and access the data by an integer index. The classical setup is well known and commonly used and thus, we do not go further into detail but highlight the relevant quantum extension from now on.

If we consider quantum data accessed by classical indexing, we think of the following setup:

$$\bigotimes_{i=0}^p |d_i\rangle_D \in (\mathcal{H}_D)^{\otimes p}, \quad (3)$$

where the data element  $|d_i\rangle_D$  associated with the index  $i$  resides in a dedicated register. In (3) there are  $p$  individual registers consisting each of  $\tilde{m}$  qubits such that  $\dim(\mathcal{H}_D) = 2^{\tilde{m}}$ . The data may be drawn from any orthogonal or non-orthogonal set of states. For this approach, operations that act on a system as given in (3) are well defined and can be formulated as the individual registers are in product form. An extension corresponds to adding a new quantum state to the system. One can swap individual registers without creating entanglement. Due to the product form of (3), individual registers can be traced out (and thus neglected and removed) without affecting the remaining states. This setup is based on a linear number of quantum resources. Due to this *uncompressed* way of storing the data, the framework of (3) is inefficient in the number of resources corresponding here to the number of qubits. Thus, in the following section, we consider the case of *quantum indexing*.

Table 1: Set of defined quantum database (QDB) operations that we consider to be essential QDB operations. We summarize their respective actions on the QDB state and draw an analogy to operations in the classical database (DB) scenario to illustrate possible correspondences.

QDB operation	QDB action	Name	DB Analog
Prepare	Prepare empty QDB with $k$ index elements.	$P_{(k)}$	CREATE
Extend	Extend $ \text{QDB}^{(k)}\rangle$ by $l$ new index elements.	$E_{(l)}$	RESERVE
Remove	Remove index $f$ with data element from $ \text{QDB}^{(k)}\rangle$ .	$R_{(f)}$	DELETE
Write	Write data element with specific index $f$ .	$W_{(f)}$	INSERT
Read-out	Read-out the data element indexed by $f$ .	$G_{(f)}$	SELECT
Permute	Permute the index elements based on $\pi$ .	$P_\pi$	-

### 3 Data structured by quantum indices

When the index is itself a quantum state, we allow for superpositions of indices, each correlated to a data state. In general, we consider balanced superposition states as in (2). For some of the algorithms introduced below, we exclude the  $|0\rangle_I$ -index state from being balanced as it forms a reference and is used as a *probability reservoir*. This term comes from the fact that measuring the index register collapses the database to a single index-data pair with a probability given by the squared absolute value of the amplitude  $\psi_j$ .

#### 3.1 Classical Data (QC)

In this section, we consider classical data states indexed by quantum indices. As a reminder, by  $|d_j\rangle$  we denote a state non-uniquely drawn from an orthogonal set of states  $\{d_\alpha\}_\alpha \in \mathcal{H}_D$ . We assume that *classical* data is defined as being drawn from the aforementioned set of orthogonal states, and it is given that we know the unitary transformation  $U_D$  that maps each element in this set to the computational basis.

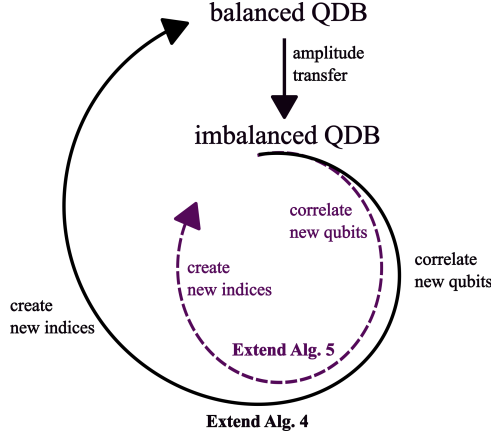


Figure 2: Schematic of the quantum database extension procedure consisting of adding new qubits and creating new indices that are correlated to the empty data string. In contrast to Algorithm 4, Algorithm 5 does not include the *amplitude transfer* step and uses a specifically imbalanced database with the zero-index state acting as a probability reservoir. Both protocols are described in detail in Section A.3.

### 3.1.1 Preparation of an empty QDB

To prepare a database with balanced amplitudes, quantum indexing, and an empty data register, we apply the following operation:

$$\begin{aligned}
 \mathcal{H}_I \otimes \mathcal{H}_D &\rightarrow \mathcal{H}_I \otimes \mathcal{H}_D \\
 |0\rangle_{I \otimes D} &\xrightarrow{P_{(k)}} \\
 \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |0\rangle_D &=: |\text{QDB}_{\text{empty}}^{(k)}\rangle.
 \end{aligned} \tag{4}$$

In (4),  $P_{(k)}$  is particularly simple when  $\log_2(k) \in \mathbb{N}_{>0}$ . This case corresponds to the Walsh-Hadamard transform applied only to the index register  $I$  with a circuit depth of  $\mathcal{O}(1)$ . The indices remain in a product state with respect to the empty data register  $D$ . Thus, the operation is given as follows:

$$P_{(k)} = H^{\otimes \log_2(k)} \otimes \mathbb{I}_D. \tag{5}$$

By creating the state  $|\text{QDB}_{\text{empty}}^{(k)}\rangle$ , we have prepared an empty database with  $k$  indices in superposition. All indices are correlated to the empty data string given by  $|0\rangle_D$ . The general case of the preparation operation for a balanced superposition state with  $k$  indices and  $\log_2(k) \notin \mathbb{N}_{>0}$  is presented in Appendix A.1 together with the case including the probability reservoir.

### 3.1.2 Extend the QDB by new indices correlated to an empty data string

The operation of extending the quantum database corresponds to expanding the number of indices in the superposition and, crucially, correlating them to the empty data string. This operation seems straightforward, but we will see that the possibility of having entanglement between index and data registers is very consequential. In general, we may correlate a number of new ancilla qubits to the index basis for the index extension procedure and aim to add as many new indices as possible. If the database initially holds a superposition of  $k$  indices, adding a qubit to the index basis doubles the maximal number of possible indices to  $2k$ . Next to the demand that the new indices be correlated to the empty data string, the database should remain in a balanced superposition with equally distributed amplitudes. The extension operation for

adding  $l > 0$  new elements is described by:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D &\longrightarrow \mathcal{H}_{I'} \otimes \mathcal{H}_D := (\mathcal{H}_0 \otimes \mathcal{H}_I) \otimes \mathcal{H}_D \\ |\text{QDB}^{(k)}\rangle &= \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_j\rangle_D \\ \xrightarrow{E^{(l)}} |\text{QDB}^{(k+l)}\rangle &= \frac{1}{\sqrt{k+l}} \sum_{j=0}^{k+l-1} |j\rangle_{I'} |d_j\rangle_D, \end{aligned} \quad (6)$$

with

$$|d_j\rangle_D = |0\rangle_D \text{ for } \forall j > (k-1) \text{ and } j = 0.$$

The transformation in (6) is non-trivial to define as there is no general unitary map  $E^{(l)}$  fulfilling it. A proof of this fact is presented in Appendix A.2. Thus, the unitary operation  $E^{(l)}$  depends on the specific database and is not generally applicable.

In the simplest case of (6) that still requires adding an extra index register  $\mathcal{H}_0$ , we add a single qubit to the index basis and consider  $\log_2(k) \in \mathbb{N}_{>0}$ , that forms an initial database with all available indices already in use. Thus, we would have  $\dim(\mathcal{H}_0) = 2$  and  $\mathcal{H}_{I'} = \mathcal{H}_0 \otimes \mathcal{H}_I$ . In this case, we can add  $l$  new index elements with  $0 < l \leq k$ . The data string corresponding to the newly added indices is empty. We present two different protocols for the extension of the quantum database:

1. Insert a QDB containing data, transfer amplitude to zero-index, and create new indices.
2. Initialize an empty database with an adjusted amplitude distribution, write data in the data registers, and create new indices afterward.

Regarding the first version, the quantum index extension protocol consists of two steps: an amplitude transfer operation to ensure the amplitudes remain balanced and an index creation step. The amplitude transfer operation is built on the generalized amplitude amplification algorithm with modified step sizes as presented by Brassard *et al.* [17]. The modified amplitude amplification algorithm generally allows us to find a good solution with certainty [17] and can be adapted to set a known amplitude to any other desired value. As this modification allows the transfer of small amplitudes onto a specific subspace, we may use it to re-balance the amplitudes of the superposition. Specifically, we transfer amplitude to the zero index state that we use as a *probability reservoir*. As this state is correlated to the empty data state, we create new indices from it based on the generalized prepare method described in Section A.1 after correlating an ancilla register. The whole procedure is costly as it involves knowing and applying the unitary  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}$  (and its inverse  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}^\dagger$ ) that creates the whole database (including the data elements) during the amplitude transfer step. There is a need for a less demanding approach that is not based on amplitude amplification.

The second protocol starts from an unbalanced database and can add new elements only until the probability reservoir is fully depleted. Thus, it is important to know an estimate or at least an upper bound on the number of data elements that one desires to store in the database (denoted here by  $k+l$ ) This version initializes a database with a *higher* amplitude on the zero index state. Next, the data registers are filled, an ancilla register is added, and new index states are created that are correlated to the empty data state. In the Appendix A.3, we go into the details for both variants of the extension protocol. The second approach appears more practical, and we observe that it uses a controlled *Prepare* operation. Thus, any improvements in the implementation of that operation directly translate to an improvement of the *Extend* operation, too.

### 3.1.3 Removing an index and corresponding data element from the QDB

Similarly to adding new indices to the database, we aim to define an operation that allows removing indices and the respective indexed data element from the database. When the index to remove is not the highest one, we obtain a database with non-continuous indexing. If this is undesirable, one can subsequently apply the *Permute* operation to do re-indexing appropriately. In general, the removal operation, which we name  $R_{(f)}$ , reduces the number of indices with a nonzero amplitude from the superposition, and the state would be re-normalized. In its simplest form, this operation would, e.g., remove the index state  $f$  and the corresponding data register from the superposition. Thus, by index removal, we understand the following map:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D &\longrightarrow \mathcal{H}_I \otimes \mathcal{H}_D \\ |\text{QDB}^{(k)}\rangle &= \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_j\rangle_D \xrightarrow{R_{(f)}} \frac{1}{\sqrt{k-1}} \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D, \end{aligned} \quad (7)$$

with  $\dim(\mathcal{H}_I) = k$ . Similarly to the extension operation presented in (6), the transformation  $R_{(f)}$  presented in (7) cannot be implemented in general by a unitary transformation. The proof is similar to the one of the extension presented in A.2 and can be intuitively understood considering the case of removing all indices and their data. This would correspond to an erasure operation that is clearly not unitary since it is a many-to-one function. Thus, the operation depends on the specific database and its indexed orthogonal data elements. We consider the case where the element  $|f\rangle_I |d_f\rangle_D$  is removed. To keep the superposition balanced, we need to take into account the re-normalization of the amplitudes, i.e.,  $\frac{1}{\sqrt{k}} \mapsto \frac{1}{\sqrt{k-1}}$  as in (7). In a quantum information setting, the no-cloning theorem applies [12] for a general quantum state. Thereof follows the no-deletion theorem [13]. If one aims to remove an index from the superposition, one intends to achieve the following reset operation:

$$|j\rangle_I |d_j\rangle_D \mapsto |j\rangle_I |0\rangle_D. \quad (8)$$

This can be achieved by knowing the state preparation unitary  $|d_j\rangle_D = U|0\dots 0\rangle_D$ . Since in this section we consider classical data in the form of orthogonal data states, another possible option would be to correlate a sensor state equivalent to  $|d_j\rangle_D$  and delete the data in register  $D$  utilizing CNOT gates with additional controls on the index register being in  $j$ . This procedure only works for orthogonal states with a known unitary that maps the orthogonal basis states to the computational basis, but this is assumed to be our definition of classical data. Next, the amplitude of the “resetted” index state can be transferred back to the  $|0\rangle_{I \otimes D}$  state. This requires knowing the specific index element  $j$  that is removed. This operation has the following effect:

$$(|0\rangle_I + |j\rangle_I) |0\rangle_D \mapsto |0\rangle_I |0\rangle_D. \quad (9)$$

The transformation described in (9) can be achieved by a specific rotation in the two-dimensional space spanned by  $\text{span}\{|0\rangle_I |0\rangle_D, |j\rangle_I |0\rangle_D\}$ . Thus, it is a non-trivial operation to remove elements in the quantum superposition state and requires specific data-dependent operations for each state that fulfills the requirements listed above. Finally, we comment on a non-deterministic implementation that works in the general case but has a finite probability of fully erasing the database. One can measure the index with a binary observable with an eigen-subspace corresponding to the span of the indices to remove and their respective orthogonal subspace. This solution may be acceptable when the database is easy to prepare or multiple copies are available but is not scalable.

### 3.1.4 Writing data in the QDB

In the following, we consider the process of writing individual *data* elements  $d_j$  (with  $1 \leq j \leq (k-1)$ ) in the data register that are each associated with a specific index in the index register.

We assume the data register states associated with the new elements are initially empty, as is the case after the *Extend* operation, e.g., the all-zero state  $|0\rangle_D$ . By *writing* the data element  $|d_f\rangle_D$  indexed by  $f$  we understand the following operation:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D &\longrightarrow \mathcal{H}_I \otimes \mathcal{H}_D \\ |\text{QDB}^{(k)}\rangle &= |\text{QDB}^{(k)}\rangle_{\neq f} + \frac{1}{\sqrt{k}} |f\rangle_I |0\rangle_D \xrightarrow{W_{(f)}} |\text{QDB}^{(k)}\rangle_{\neq f} + \frac{1}{\sqrt{k}} |f\rangle_I |d_f\rangle_D, \end{aligned} \quad (10)$$

with

$$|\text{QDB}^{(k)}_{\neq f}\rangle := \frac{1}{\sqrt{k}} \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D.$$

For the case of data states drawn from an orthogonal set with a known unitary transformation to the computational basis, we can use a writing procedure based on CNOT gates. Conceptually, we can view the writing process based on the sensor state  $|d_f\rangle_S$  as follows. We transform the orthonormal basis states to the computational basis and apply CNOT gates to copy the data into the superposition of the quantum database. Thus, the operation is schematically given as:

$$W_{(f)} \left( \left( |\text{QDB}^{(k)}\rangle_{\neq f} + \frac{1}{\sqrt{k}} |f\rangle_I |0\rangle_D \right) |d_f\rangle_S \right) \mapsto \left( |\text{QDB}^{(k)}\rangle_{\neq f} + \frac{1}{\sqrt{k}} |f\rangle_I |d_f\rangle_D \right) |d_f\rangle_S, \quad (11)$$

with

$$W_{(f)} = \left( (\mathbb{I}_I \otimes U_D \otimes U_D) \Lambda_\tau(\sigma_X) \left( \mathbb{I}_I \otimes U_D^\dagger \otimes U_D^\dagger \right) \right).$$

The central operation  $\Lambda_\tau(\sigma_X)$  consists of  $\tilde{m}$  multi-controlled X gates, each controlled on  $\tau := \tilde{k}+1$  qubits. The gate  $\Lambda_\tau(\sigma_X)$  behaves like a CNOT gate with  $\tau$ -control qubits. An adapted implementation of  $\Lambda_\tau(\sigma_X)$  is presented in [15]. Decompositions of  $\Lambda_\tau(\sigma_X)$  are considered in [18]. The operation defined in (11) keeps the additional system  $S$  in the product with respect to the database. This allows for tracing out and removing the subsystem  $S$  without any information loss in the database, as no entanglement is created between the QDB and system  $S$ .

One may wonder why not using a conditional SWAP operation to write data data in the database. If we condition the SWAP operation on a specific index  $f$ , we obtain the following:

$$\begin{aligned} W_{(f)} \left( \frac{1}{\sqrt{k}} \left( \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D + |f\rangle_I |0\rangle_D \right) |d_f\rangle_S \right) &= \frac{1}{\sqrt{k}} \left( \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D |d_f\rangle_S + |f\rangle_I |d_f\rangle_D |0\rangle_S \right) \\ &\neq \left( \frac{1}{\sqrt{k}} \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D + \frac{1}{\sqrt{k}} |f\rangle_I |d_f\rangle_D \right) |d_f\rangle_S. \end{aligned} \quad (12)$$

In general, this state does not remain in product form due to entanglement between the database and system  $S$ . Hence, one either accepts a system growth by keeping the ancilla state or takes a process of disentangling  $S$  from the database into account. If only classical correlations are present, one may rely on the disentangling operation introduced in [19]. Thereby, an ancilla qubit is *decorrelated* from the system by adding an ancilla qubit in the state  $|0\rangle$  and applying a combination of (multi-)controlled CNOT gates. The general process of disentanglement for systems, including non-classical correlations, is further described in [20, 21, 22, 23] and shortly summarized in Appendix A.4.



### 3.1.5 Read-out data in the QDB

The quantum database contains data elements indexed by orthogonal states that one aims to process in some form. On the one hand, one can directly apply a quantum algorithm on the database state itself, such as Grover's search algorithm for unstructured search [1], and operate on, and thus, modify the entire QDB state. On the other hand, one may think of a read-out operation that does or does not affect the remaining data states in the QDB. A read-out operation could have the following effect with  $|d_0\rangle_D \neq |d_j\rangle_D$  for some  $0 < j \leq (k-1)$  and an ancilla system with the same dimension as the data register ( $\dim(\mathcal{H}_D) = \dim(\mathcal{H}_A)$ ):

$$|\text{QDB}^{(k)}\rangle |0\rangle_A \not\rightarrow |\text{QDB}^{(k)}\rangle |d_j\rangle_A. \quad (13)$$

One correlates the ancilla system  $A$  to the database in order to transfer the data contained in the  $j$ th data state to the ancilla system. The transformation in (13) is not possible in general due to linearity. Instead, an exemplary copying procedure based on CNOT gates for orthogonal states would result in the following transformation:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D \otimes \mathcal{H}_A &\rightarrow \mathcal{H}_I \otimes \mathcal{H}_D \otimes \mathcal{H}_A \\ |\text{QDB}^{(k)}\rangle |0\rangle_A &\xrightarrow{G(f)} |\text{QDB}_{\neq f}^{(k)}\rangle |0\rangle_A + \frac{1}{\sqrt{k}} |f\rangle_I |d_f\rangle_D |d_f\rangle_A, \end{aligned} \quad (14)$$

again with

$$|\text{QDB}_{\neq f}^{(k)}\rangle := \frac{1}{\sqrt{k}} \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D.$$

In (14), the ancilla system  $A$  gets entangled with the quantum database. For orthogonal data states, one could proceed with the copying procedure described in (14) for every data element in the QDB and essentially would create a *copied* data register in a superposition. This state looks as follows:

$$\frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_j\rangle_D |d_j\rangle_A \in \mathcal{H}_I \otimes \mathcal{H}_D \otimes \mathcal{H}_A. \quad (15)$$

To obtain the state in (15), one would apply  $\tilde{m}$  CNOT gates that are controlled by the  $i$ th qubit in the data register  $D$  and apply the controlled not operation on the  $i$ th qubit in the ancilla register  $A$ . One notices in (15) that the registers  $D$  and  $A$  are also entangled if at least two states in  $A$  are different, i.e.,  $|d_j\rangle_A \neq |d_i\rangle_A$  for some  $i \neq j$ .

Another possibility to define the read-out of a data state would be via a *projective measurement*. Thereby, we aim to do the projective measurement [24] on the index register and return the corresponding data state indexed by  $f$ :

$$|d_f\rangle_D = \frac{\hat{P}_{|f\rangle} |\text{QDB}^{(k)}\rangle}{\sqrt{\langle \text{QDB}^{(k)} | \hat{P}_{|f\rangle} | \text{QDB}^{(k)} \rangle}}. \quad (16)$$

The idempotent projection operator  $\hat{P}_{|f\rangle}$  is given by

$$\hat{P}_{|f\rangle} = |f\rangle \langle f|_I \otimes \mathbb{I}_D. \quad (17)$$

The resulting state of this projective measurement is the data state  $|d_f\rangle_D$  corresponding to index  $f$ , which is retrieved with a probability of  $\frac{1}{k}$ . Hence, with the given probability of  $\frac{1}{k}$  we obtain the data state  $|d_f\rangle_D$  corresponding to index  $j$ . This projective measurement essentially

returns a specific data state corresponding to an index and thus forms a query to the quantum database. This query procedure returning the state in (16) “consumes” the whole database. This is the case as the applied projective measurement collapses the superposition state that forms the quantum database onto a single state. An interesting application could be that the qubits belonging to the index and the qubits belonging to the data registers could be located at different spatial positions. Hence, a projective measurement on the index register would simultaneously collapse the data register at another physical location [25].

### 3.1.6 Permute data in the QDB

For this operation, we would like to permute database elements according to their indices. We notate this operation by  $P_\pi$  and the permutation  $\pi(\cdot) \in S_k$ . Thereby,  $S_k$  denotes the symmetric group over a finite set of  $k$  elements. We write  $P_\pi$  for a quantum database with  $k$  elements as follows:

$$\frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_j\rangle_D \xrightarrow{P_\pi} \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |\pi(j)\rangle_I |d_j\rangle_D = \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_{\pi^{-1}(j)}\rangle_D. \quad (18)$$

A general permutation  $\pi(\cdot) \in S_k$ , or its inverse  $\pi^{-1}(\cdot)$ , can always be decomposed into transpositions  $(i, j)$  as two-element transpositions generate the group  $S_k$ . We denote a transposition of indices  $(i, j)$  by:

$$i\sigma_x^{i,j} \in \text{SU}(k) \quad (19)$$

with the individual entries given by

$$(\sigma_x^{i,j})_{n,m} := \delta_{n,m} - \delta_{n,i} \delta_{m,i} - \delta_{n,j} \delta_{m,j} + \delta_{n,j} \delta_{m,i} + \delta_{n,i} \delta_{m,j}. \quad (20)$$

In the special case with  $\dim(I) = 4$ , one obtains  $P_{(2,3)} = \text{CNOT}$ . For every transposition  $(i, j)$  with  $0 \leq i, j \leq (k-1)$ , one applies the following operation to the database

$$P_{(i,j)} := (i\sigma_x^{i,j})_I \otimes \mathbb{I}_D. \quad (21)$$

By concatenating adjacent transpositions as defined in (21), one may generate any permutation  $\pi(\cdot) \in S_k$ . A general permutation on the index register  $\pi^{-1}(\cdot)$  as presented in (18) can be represented by its permutation matrix  $M_{\pi^{-1}} = M_\pi^t \in U(k)$ . By  $M_\pi^t$  we denote the transpose of  $M_\pi$ . Hence, we have:

$$P_\pi := M_\pi \otimes \mathbb{I}_D \in U(k \cdot m). \quad (22)$$

As mentioned before, any permutation  $\pi(\cdot)$  can be decomposed into (adjacent) transpositions, which can be implemented by  $P_{(i,i+1)}$  defined in (21). It is important to observe that this construction produces a possible decomposition and not the most efficient one. For example, consider the action of a single Pauli-x gate on an index qubit. This changes all indices (by flipping one bit in their binary representation) and may be used as an additional permutation in the decomposition of  $\pi(\cdot)$  in addition to transpositions. Special cases of the permute operation  $P_\pi$  appear in the classical scenario, such as a *shift* or *swap* operation.

## 3.2 Outlook towards Quantum Data (QQ)

If we consider the bottom-right quadrant in Figure 1, the database contains both quantum indices and quantum data. We defined quantum data as being drawn from a non-orthogonal or orthogonal set of states with an *unknown* unitary transformation to the computational basis. This definition of “quantum” data leads to the fact that this type of data generally cannot be copied exactly as perfect cloning is only possible for states belonging to a known set of

orthogonal states [26]. Furthermore, from the no-cloning theorem [12] follows the non-deletion theorem [14]. This theorem states that even though information can be deleted perfectly in a reversible manner in classical computation with respect to a copy, the analogous task on quantum information cannot be done for an arbitrary quantum state [14]. It holds due to the linearity of quantum theory [14].

The database preparation does not depend on the type of data written in the respective data register. For the extension operation, the first algorithm based on amplitude amplification only works if we know and are able to implement the unitary that prepares the database state. However, the second version of the database extension algorithm works independently of the data type. Regarding the permutation operation, one has to take into account that each data element may be a linear combination of (orthogonal) states. Because of this, data in a superposition corresponds to the same index state. One must consider this fact when conditioning an operation as an individual index transposition on a single data register (corresponding to one element in the *data superposition*). Also, it is different that data can no longer be removed or written in the QDB, as was the case for orthogonal data states based on CNOT gates. For this, alternative methods that are described below can be considered. The Read-out operation differs as the operation in (14) cannot be done for non-orthogonal states; instead, one may consider a swap operation with the following effect:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D \otimes \mathcal{H}_A &\rightarrow \mathcal{H}_I \otimes \mathcal{H}_D \otimes \mathcal{H}_A \\ |\text{QDB}^{(k)}\rangle |0\rangle_A &\xrightarrow{G(f)} |\text{QDB}_{\neq f}^{(k)}\rangle |0\rangle_A + \frac{1}{\sqrt{k}} |f\rangle_I |0\rangle_D |d_f\rangle_A. \end{aligned} \quad (23)$$

As before, the database and the ancilla system  $A$  are entangled in (23). A projective measurement as a read-out operation would return a general quantum state.

Due to the aforementioned limitations, we highlight possible modifications that allow us to operate on quantum data. Firstly, one may consider the application of approximate or probabilistic cloning that we summarize in the following [26, 27, 28]. In the case of approximate cloning, one creates an approximate copy by means of a unitary transformation. With regard to the non-trivial approximate cloning strategy, the optimal case would be the universal cloner by *Bužek and Hillery* [29]. Thereby, the transformation done by a general unitary is the following [29, 30]:

$$|\psi\rangle_A |R\rangle_B |\mathcal{M}\rangle_M \mapsto |\Psi\rangle_{ABM}. \quad (24)$$

This forms an optimal and symmetric universal quantum cloner [29, 30] with a single qubit ancilla in this simple example. In (24), the qubit to be cloned is the A-system, the system that contains the approximate clone in the B-system and the ancilla marked by  $M$ . In the optimal case, this transformation obtains the following [29, 30]:

$$\rho_A = \text{Tr}_{BM}(|\Psi\rangle\langle\Psi|_{ABM}) = \rho_B = \text{Tr}_{AM}(|\Psi\rangle\langle\Psi|_{ABM}) = F |\psi\rangle\langle\psi| + (1 - F) |\psi^\perp\rangle\langle\psi^\perp|, \quad (25)$$

with  $F = \frac{5}{6}$ . Thus, we have a mixed state if one traces out the ancilla in (24). For the case of the quantum database, this procedure means that the database would get entangled with the qubit containing the data to be written in the database. Thus, either one allows the quantum database system to essentially “grow,” or one has to rely on a disentanglement mechanism aiming to disentangle the state from the quantum database [20, 21, 22, 23].

Secondly, the case of probabilistic cloning describes the procedure of creating a perfect copy of an unknown quantum state with a success probability smaller than 1. In [31], it was shown that states that were secretly chosen from a set

$$\mathcal{S} = \{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_t\rangle\} \quad (26)$$

can be cloned probabilistically by a general unitary reduction operation if and only if the individual states in  $\mathcal{S}$  are linearly independent. Thereby, the unitary transformation  $U$  that is later followed by a post-selective measurement operation looks as follows [31, 26]:

$$U(|\psi_i\rangle |0\rangle |A_0\rangle) = \sqrt{p_i} |\psi_i\rangle |\psi_i\rangle |A_0\rangle + \sum_{j=1}^n c_{ij} |\Phi_j\rangle |A_j\rangle, \text{ for } i = 1, 2, \dots, t. \quad (27)$$

In (27), the ancilla states are all orthogonal to each other, i.e.,  $\langle A_k | A_l \rangle = \delta_{kl}$  for  $0 \leq k, l \leq t$ . Therefore, the desired perfect clone is achieved with a probability of  $p_i$  when measuring the ancilla in the basis  $\{A_k\}$ . In [31], a simple example of the cloning efficiency for a linearly independent set of two states  $\mathcal{S} = \{|\psi_1\rangle, |\psi_2\rangle\}$  is given. In this case, the success probabilities are bound as follows [31, 26]:

$$\frac{1}{2}(p_1 + p_2) \leq \frac{1}{1 + |\langle \psi_1 | \psi_2 \rangle|} \quad (28)$$

with  $|\langle \psi_1 | \psi_2 \rangle| \neq 1$ . Let us consider this procedure for writing data items in the data states (with all the data elements being part of a linearly independent set). As a simple example of a one-qubit data register, we have the following database state:

$$\frac{1}{\sqrt{k}} \left( \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D + |f\rangle_I |0\rangle_D \right) |d_f\rangle |A_0\rangle. \quad (29)$$

Aiming to write in the data register indexed by  $f$ , we apply the unitary presented in (27) conditioned on index  $f$  and obtain:

$$\frac{1}{\sqrt{k}} \left( \sum_{\substack{j=0 \\ j \neq f}}^{k-1} |j\rangle_I |d_j\rangle_D |d_f\rangle |A_0\rangle + |f\rangle_I (\sqrt{p_i} |d_f\rangle_D |d_f\rangle |A_0\rangle + \sum_{j=1}^n c_{ij} |\Phi'_j\rangle |A_j\rangle) \right). \quad (30)$$

If one now post-selectively applies a measurement on the ancilla in the basis  $\{A_k\}_k$ , measuring  $|A_0\rangle$  we end up successfully in the subspace containing the QDB with the copy of  $|d_f\rangle$  or failed the cloning when measuring the ancilla in any other state than  $|A_0\rangle$ . This means that this ancilla measurement either returns the preferred QDB state or a failure state when measuring  $\{|A_j\rangle\}_j$  for  $\forall j \neq 0$ .

## 4 Conclusion

We presented a framework of quantum databases that distinguishes between classical and quantum indexing and data values. Our definition of a quantum database consists of a balanced superposition state storing data correlated to specific orthogonal index states in a quantum superposition. We have highlighted essential database operations and presented possible implementations focusing on the case of quantum indexing. We note that multiple possible implementations exist for each operation, and we leave it as an open problem to refine and improve each of them individually.

The operations we showcase here already allow for general recommendations for best practices when working on a quantum database. For example, for the index extension operation, differently from the classical indexing case, extending the database is very costly, and one should rather set the total (or maximum) number of items to be added in advance and prepare the database accordingly. Furthermore, the available resources should be exhausted by adding all possible index states during the extension operation. We observe that new index states are released in bunches

most efficiently, and one should use that to have a low circuit depth. Furthermore, to unlock quantum advantage from using databases, we recommend avoiding projective measurements in order to keep the superposition state and employ its quantum nature algorithmically.

In this work, we considered a noise-free scenario. However, amplitude noise would affect the balanced amplitudes in the superposition state, especially when many states are present in the superposition, and the amplitudes get small. Bit-flip errors in the computational basis would disturb the correspondence of data and index elements, but one can rely on efficient error correcting codes, see, e.g., [32], that are expected to improve in the future. A detailed study on the impact of noise in the context of a QDB is left for future work.

We believe it is important to examine the differences between databases in the classical and quantum scenarios due to the naturally different conditions dictated by quantum mechanics. In this context, one needs to rethink even basic operations, which is essential to formalize a database version based on quantum data. The presented framework gives a more precise outlook on the practicality and usefulness of each of the operations. Furthermore, for applying a set of algorithms operating on a quantum database before a state collapse (e.g., through measurement), the presented operations are applicable as a state preparation and allow for dynamical manipulation after initialization of a certain quantum state.

## Acknowledgements

CR and MG are supported by CERN through the CERN Quantum Technology Initiative. CR has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA). CR would like to thank Tobias Duswald for his invaluable support throughout this work.

## References

- [1] Lov K. Grover. “A fast quantum mechanical algorithm for database search”. In Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing. Page 212–219. STOC ’96 New York, NY, USA (1996). Association for Computing Machinery.
- [2] Carlo A Trugenberger. “Quantum pattern recognition”. *Quantum Information Processing* **1**, 471–493 (2002). url: <https://doi.org/10.1023/A:1024022632303>.
- [3] Pradeep Niroula and Yunseong Nam. “A quantum algorithm for string matching”. *npj Quantum Information* **7**, 37 (2021). url: <https://doi.org/10.1038/s41534-021-00369-3>.
- [4] Gilles Brassard, Peter Høyer, and Alain Tapp. “Quantum cryptanalysis of hash and claw-free functions: Invited paper”. Page 163–169. Springer Berlin Heidelberg. (1998).
- [5] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. “Quantum algorithm for linear systems of equations”. *Phys. Rev. Lett.* **103**, 150502 (2009).
- [6] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. “Quantum random access memory”. *Phys. Rev. Lett.* **100**, 160501 (2008).
- [7] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. “Architectures for a quantum random access memory”. *Phys. Rev. A* **78**, 052310 (2008).
- [8] Junyu Liu, Connor T. Hann, and Liang Jiang. “Data centers with quantum random access memory and quantum networks”. *Phys. Rev. A* **108**, 032610 (2023).
- [9] Junyu Liu and Liang Jiang. “Quantum data center: Perspectives” (2023). [arXiv:2309.06641](https://arxiv.org/abs/2309.06641).

- [10] Ahmed Younes. “Database manipulation operations on quantum systems”. *Quantum Information Review* **1**, 9–17 (2013). url: <https://www.naturalspublishing.com/files/published/332fckl634r9zt.pdf>.
- [11] Rihan Hai, Shih-Han Hung, and Sebastian Feld. “Quantum data management: From theory to opportunities” (2024). url: <https://arxiv.org/pdf/2403.02856.pdf>.
- [12] William K Wootters and Wojciech H Zurek. “A single quantum cannot be cloned”. *Nature* **299**, 802–803 (1982). url: <https://doi.org/10.1038/299802a0>.
- [13] Arun K. Pati and Barry C. Sanders. “No-partial erasure of quantum information”. *Physics Letters A* **359**, 31–36 (2006).
- [14] Arun Kumar Pati and Samuel L Braunstein. “Impossibility of deleting an unknown quantum state”. *Nature* **404**, 164–165 (2000). url: <https://doi.org/10.1038/404130b0>.
- [15] Daniel K Park, Francesco Petruccione, and June-Koo Kevin Rhee. “Circuit-based quantum random access memory for classical data”. *Scientific reports* **9**, 3949 (2019). url: <https://doi.org/10.1038/s41598-019-40439-3>.
- [16] Chris J Date. “A guide to the sql standard”. Addison-Wesley Longman Publishing Co., Inc. USA (1986).
- [17] Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. “Quantum amplitude amplification and estimation”. *Contemporary Mathematics* **305**, 53–74 (2002). url: <https://doi.org/10.1090/conm/305>.
- [18] Adriano Barenco, Charles H. Bennett, Richard Cleve, David P. DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A. Smolin, and Harald Weinfurter. “Elementary gates for quantum computation”. *Physical Review A* **52**, 3457–3467 (1995).
- [19] John A. Cortese and Timothy M. Braje. “Loading classical data into a quantum computer” (2018). [arXiv:1803.01958](https://arxiv.org/abs/1803.01958).
- [20] Tal Mor. “Disentangling quantum states while preserving all local properties”. *Phys. Rev. Lett.* **83**, 1451–1454 (1999).
- [21] P. J. Dodd and J. J. Halliwell. “Disentanglement and decoherence by open system dynamics”. *Phys. Rev. A* **69**, 052105 (2004).
- [22] Mario Berta and Christian Majenz. “Disentanglement cost of quantum states”. *Phys. Rev. Lett.* **121**, 190503 (2018).
- [23] Somshubhro Bandyopadhyay, Guruprasad Kar, and Anirban Roy. “Disentanglement of pure bipartite quantum states by local cloning”. *Physics Letters A* **258**, 205–209 (1999).
- [24] Michael A Nielsen and Isaac L Chuang. “Quantum computation and quantum information”. *Phys. Today* **54**, 60 (2001).
- [25] A. Einstein, B. Podolsky, and N. Rosen. “Can quantum-mechanical description of physical reality be considered complete?”. *Phys. Rev.* **47**, 777–780 (1935).
- [26] Dagmar Bruß and Chiara Macchiavello. “Approximate quantum cloning”. *Chapter 4, pages 53–71*. John Wiley & Sons, Ltd. (2006).
- [27] Daowen Qiu. “Combinations of probabilistic and approximate quantum cloning and deleting”. *Phys. Rev. A* **65**, 052329 (2002).

- [28] Marius Lemm and Mark M. Wilde. “Information-theoretic limitations on approximate quantum cloning and broadcasting”. *Phys. Rev. A* **96**, 012304 (2017).
- [29] V. Bužek and M. Hillery. “Quantum copying: Beyond the no-cloning theorem”. *Phys. Rev. A* **54**, 1844–1852 (1996).
- [30] Valerio Scarani, Sofyan Iblisdir, Nicolas Gisin, and Antonio Acín. “Quantum cloning”. *Rev. Mod. Phys.* **77**, 1225–1256 (2005).
- [31] Lu-Ming Duan and Guang-Can Guo. “Probabilistic cloning and identification of linearly independent quantum states”. *Phys. Rev. Lett.* **80**, 4999–5002 (1998).
- [32] Sergey Bravyi, Andrew W Cross, Jay M Gambetta, Dmitri Maslov, Patrick Rall, and Theodore J Yoder. “High-threshold and low-overhead fault-tolerant quantum memory”. *Nature* **627**, 778–782 (2024). url: <https://doi.org/10.1038/s41586-024-07107-7>.
- [33] G. Brassard and P. Høyer. “An exact quantum polynomial-time algorithm for simon’s problem”. In Proceedings of the Fifth Israeli Symposium on Theory of Computing and Systems. *ISTCS-97*. IEEE Comput. Soc (1997).
- [34] Maria Schuld and Francesco Petruccione. “Machine learning with quantum computers”. Springer. (2021). url: <https://doi.org/10.1007/978-3-030-83098-4>.
- [35] Dagmar Bruß, David P. DiVincenzo, Artur Ekert, Christopher A. Fuchs, Chiara Macchiavello, and John A. Smolin. “Optimal universal and state-dependent quantum cloning”. *Phys. Rev. A* **57**, 2368–2378 (1998).
- [36] Daniel R Terno. “Nonlinear operations in quantum-information theory”. *Physical Review A* **59**, 3320 (1999). url: <https://link.aps.org/doi/10.1103/PhysRevA.59.3320>.
- [37] Rafaella Vale, Thiago Melo D. Azevedo, Ismael C. S. Araújo, Israel F. Araujo, and Adenilton J. da Silva. “Decomposition of multi-controlled special unitary single-qubit gates” (2023). [arXiv:2302.06377](https://arxiv.org/abs/2302.06377).

## A Appendix

### A.1 General implementation of the Prepare operation

In this section, we generalize the *prepare* operation described in Section 3.1.1. This generalized algorithm allows for preparing a superposition state with balanced amplitudes that contains a number of elements  $k$  for the case of  $\log_2(k) \notin \mathbb{N}$ . Furthermore, we present a modified algorithm that may be used for preparing a balanced superposition state, with the exception that the zero index  $|0\rangle_I$  is used as a probability reservoir relevant with respect to database extension. Hence, we aim to prepare the general state with  $l \in \mathbb{N}$ :

$$\sqrt{\frac{l+1}{k+l}} |0\rangle_I + \sqrt{\frac{1}{k+l}} \sum_{j=1}^{k-1} |j\rangle_I. \quad (31)$$

This state reassembles the completely balanced case if  $l = 0$ . For creating the state in (31) one can use Algorithm 1. This procedure includes unitary gates  $Y(p)$  and  $\tilde{Y}(p)$  with  $0 \leq p \leq 1$ ,  $p \in \mathbb{R}$  defined below. The gate  $Y(p)$  plays a similar role to the Hadamard gate for  $p = 1/2$  when acting on the  $|0\rangle$  state (i.e.,  $H|0\rangle = Y(1/2)|0\rangle$ ), but, in general, it corresponds to a  $R_y$  rotation. It is defined as follows:

$$Y(p) = \begin{pmatrix} \sqrt{p} & -\sqrt{1-p} \\ \sqrt{1-p} & \sqrt{p} \end{pmatrix}. \quad (32)$$

such that  $Y(p) = R_y(\theta)$  for  $\theta = 2 \arccos(\sqrt{p})$  and  $Y(1) = \mathbb{I}$ . Furthermore, based on the previous

---

**Algorithm 1** Prepare unbalanced superposition of  $k$  computational basis states with extra weight for the  $|0\rangle$  state:  $\sqrt{l+1}/\sqrt{k+l}|0\rangle + 1/\sqrt{k+l} \sum_{j=1}^{k-1} |j\rangle$ .

---

**Require:**  $k > 1$ ,  $k \in \mathbb{N}$  and  $l \in \mathbb{N}_{\geq 0}$

**Ensure:**  $0 \leq p \leq 1$

```

1: procedure PREPARE( $k, l$ )                                ▷ prepare unbalanced superposition
2:   Compute  $t = \lceil \log_2 k \rceil$ 
3:   Express  $s = k - 1$  in binary form:  $s = \sum_{j=0}^{t-1} 2^j s_j$                                 ▷ one has  $s_{t-1} = 1$ 
                                                                      ▷ note that bit 0 is the least significant bit
4:   Compute  $p = \frac{2^{t-1}+l}{k+l}$ 
5:   Apply  $Y(p)$  to qubit  $t - 1$ 
6:   for  $j = t - 2, \dots, 0$  do
7:     Apply  $Y(1/2)$  to qubit  $j$ 
8:     if  $s_j = 0$  then
9:       Apply  $Y(1/2)^{-1}$  controlled on  $[s]_{t-1, \dots, j+1}$                                 ▷ controlled on qubits
    $t - 1, t - 2, \dots, j + 1$ 
                                                                      ▷ being equal to the corresponding bits of  $s$ 
10:    else
11:      Compute  $p = \frac{2^j}{[s]_{j, j-1, \dots, 0} + 1}$ 
12:      Apply  $\tilde{Y}(p)$  controlled on  $[s]_{t-1, \dots, j+1}$ 
13:    end if
14:    Compute  $p = \frac{2^j + l}{2^{j+1} + l}$ 
15:    Apply  $\tilde{Y}(p)$  controlled on  $[0]_{t-1, \dots, j+1}$                                 ▷ if  $l = 0$  then  $p = 1/2$  and  $\tilde{Y}(p) = I$ 
16:  end for
17: end procedure

```

---



definition of  $Y(p)$ , we define the following in order to simplify the notation

$$\tilde{Y}(p) := Y(p) \cdot Y(1/2)^{-1}. \quad (33)$$

The implementation of Algorithm 1 presents the most general case. In order to facilitate a more intuitive understanding, we showcase concrete examples.

**Example case of Algorithm 1 for  $k = 22$  and  $l = 0$ .**

For this, consider the case of initializing up to  $k = 22$  with  $l = 0$ . Hence, the circuit acts on a 5-qubit register  $I$ . The goal in this exemplary case is to create a balanced superposition state

$$\frac{1}{\sqrt{22}} \sum_{j=0}^{21} |j\rangle \quad (34)$$

in the register  $I$ . The first step is expressing  $k - 1$  as a binary number, here  $k - 1 = 21 =: b10101$ . Then, we apply unitary gates  $Y(p)$  starting from the qubit of the highest significance, noting that certain gates will be controlled on the higher-significance qubits according to the binary representation of  $(k - 1)$ . In general, for qubit  $a$ , consider the  $a$ -th bit of  $k - 1$  and denote it by  $(k - 1)_a$ . The initial layer of one-qubit gates creates a balanced superposition except with regard to the splitting of the most significant bit. In the later stages of the algorithm, balanced splitting is reverted in order to achieve an imbalanced distribution. Thereby, probabilities are distributed according to the final number of branches in the respective binary tree diagram. The corresponding quantum circuit for initializing  $k = 22$  is shown in Figure 3. The circuit depth scales linearly according to the number of qubits.

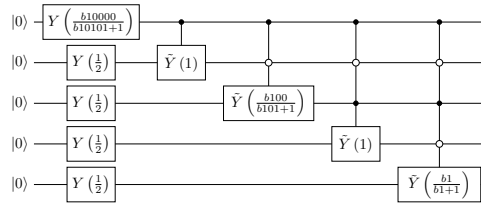


Figure 3: Circuit diagram for creating the balanced quantum superposition state with  $k = 22$  elements (where  $\log_2(k) \notin \mathbb{N}$ ). The presented quantum circuit acts on the index register of the database, and we apply the identity otherwise. The resulting state here is given by  $\frac{1}{\sqrt{22}} \sum_{j=0}^{21} |j\rangle$ .

**Example case of Algorithm 1 for  $k = 14$  and some  $l \in \mathbb{N}_{>0}$ .**

Furthermore, we also show an example case for initializing a superposition state with  $l > 0$  and  $k = 14$  given by:

$$\sqrt{\frac{l+1}{14+l}} |0\rangle + \frac{1}{\sqrt{14+l}} \sum_{j=1}^{13} |j\rangle. \quad (35)$$

We present the binary tree visualization for initializing up to element 1101 in Figure 5. Each branching step in the binary tree visualization corresponds to applying a  $Y(p)$  gate. If the branching is non-symmetric, symmetric gates must be reverted using controlled  $Y(p)$  gates. The corresponding quantum circuit is shown in Figure 4. The circuit depth scales linearly according to the number of qubits.

For the aforementioned exemplary cases we had  $\log_2(k) \notin \mathbb{N}_{>0}$ . If  $\log_2(k) \in \mathbb{N}_{>0}$ , the generalized prepare operation acts as the Walsh-Hadamard transformation, see Section 3.1.1. This holds true since  $H|0\rangle = Y(1/2)|0\rangle$ . Thus, the algorithm reduces to the balanced prepare operation introduced in Section 3.1.1 corresponding to the Walsh-Hadamard transformation, which has a circuit depth of  $\mathcal{O}(1)$ . This can be visualized by the circuit identity presented in Figure 6. By applying this circuit, a superposition state with 4 elements is prepared.

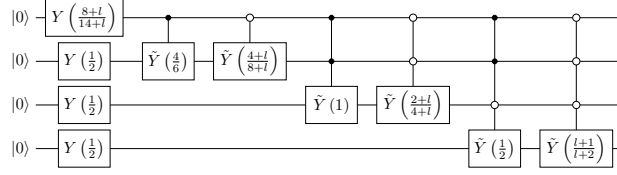


Figure 4: Circuit diagram for creating the quantum superposition state with  $k = 14$  elements (where  $\log_2(k) \notin \mathbb{N}$ ). The presented quantum circuit acts on the index register of the database, and we apply the identity otherwise. The resulting state here is given by  $\sqrt{\frac{1+l}{14+l}} |0\rangle + \sqrt{\frac{1}{14+l}} \sum_{j=1}^{14} |j\rangle$ .

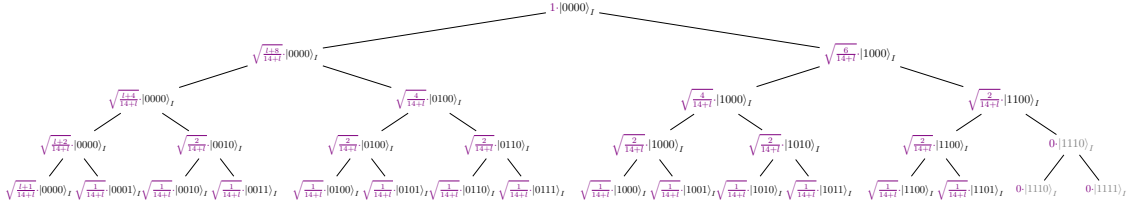


Figure 5: Binary tree that visualizes the step-by-step creation of new indices for a database with a probability reservoir. In this case, 14 elements are initialized (up to binary element 1101) and visualized in a binary tree structure. Each branching point visualizes the application of a (conditional)  $Y(p)$  gate that leads to the creation of a new state per branching, starting from the root towards the tree's leaves.

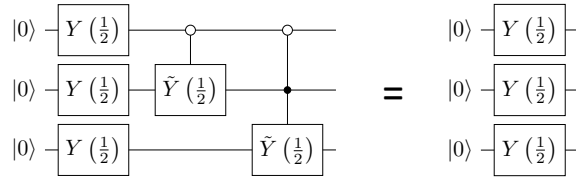


Figure 6: Circuit diagram for creating the quantum superposition state with  $k = 4$  elements (example for  $\log_2(k) \in \mathbb{N}$  and  $l = 0$ ). The operation reduces to the Walsh-Hadamard transformation, introduced in the *prepare* operation in Section 3.1.1 as the initial state is given by  $|0\rangle^{\otimes 3}$ . The presented quantum circuit acts on the index register of the database, and we apply the identity otherwise. The resulting state here is given by  $\frac{1}{2} \sum_{j=0}^3 |j\rangle$ .

## A.2 Problem with the Extend operation

The operation  $E_{(l)}$  described in (6) cannot be implemented by a unitary operation. This holds as the overlap is not preserved by  $E_{(l)}$  (as it would be if  $E_{(l)}$  is unitary) for  $l > 0$ . We formally write this observation in the following Lemma.

**Lemma 1.** *There does not exist a general unitary operation  $E_{(l)}$  with  $l \in \mathbb{N}_{>0}$  that fulfills the transformation in (6) for a general QDB state with  $k \geq 2$  elements.*

*Proof.* As before, we denote the newly added ancilla by  $A$ , and the index and data register by  $I$  and  $D$ , respectively. We extend two individual databases containing different data elements,

assuming there exists a unitary  $V := E_{(l)}$  as in (6) with the following effect:

$$\begin{aligned} \frac{1}{\sqrt{k}}(|0\rangle_I |0\rangle_D + \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D) &\xrightarrow{V} \frac{1}{\sqrt{k+l}} \left( |0\rangle_A (|0\rangle_I |0\rangle_D + \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D) + |1\rangle_A \sum_{i=0}^{l-1} |i\rangle_I |0\rangle_D \right), \\ \frac{1}{\sqrt{k}}(|0\rangle_I |0\rangle_D + \sum_{j=1}^{k-1} |j\rangle_I |d'_j\rangle_D) &\xrightarrow{V} \frac{1}{\sqrt{k+l}} \left( |0\rangle_A (|0\rangle_I |0\rangle_D + \sum_{j=1}^{k-1} |j\rangle_I |d'_j\rangle_D) + |1\rangle_A \sum_{i=0}^{l-1} |i\rangle_I |0\rangle_D \right). \end{aligned} \quad (36)$$

If we assume the map in (6), which has the effect as given in (36), can be implemented by a unitary  $V$ , the overlap has to be preserved. Thus, we compare the overlap before and after applying  $V$  for the two databases for general data states  $\{|d_j\rangle_D, |d'_j\rangle_D\}_j$  and  $\sum_{j=1}^{k-1} \langle d'_j | d_j \rangle_D < (k-1)$  in (36):

$$\frac{1}{k} \left( 1 + \sum_{j=1}^{k-1} \langle d'_j | d_j \rangle_D \right) \neq \frac{1}{k+l} \left( 1 + l + \sum_{j=1}^{k-1} \langle d'_j | d_j \rangle_D \right), \quad (37)$$

where we took advantage of the orthogonality of the index states. As seen in (37), the overlap is not preserved except if and only if  $\sum_{j=1}^{k-1} \langle d'_j | d_j \rangle_D = (k-1)$ , meaning that the two databases must be equal. Hence, there does not exist a unitary  $V$  that does the transformation in (6) for an arbitrary database.  $\square$

### A.3 Extension of the quantum database size

In this section, we present in detail the two extension algorithms summarized in Section 3.1.2. We introduce the individual operations and include the specific algorithmic implementation in Algorithm 4 and Algorithm 5. We remind ourselves that there does not exist a general unitary for the extension operation as in (6); hence, each proposed extension mechanism is data-dependent as seen in Section A.2.

#### A.3.1 Transfer and unfold

First, we present a protocol for extending a database state using amplitude transfer. In general, the operation is defined as:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D &\longrightarrow \mathcal{H}_{I'} \otimes \mathcal{H}_D := (\mathcal{H}_A \otimes \mathcal{H}_I)_{I'} \otimes \mathcal{H}_D \\ |\text{QDB}^{(k)}\rangle &\mapsto |\text{QDB}^{(k+l)}\rangle. \end{aligned} \quad (38)$$

Thereby, the newly created indices  $k \leq i < (k+l)$  are all correlated to the empty data string  $|0\rangle_D$ . For the first approach, the amplitudes are adjusted so that the database will be a balanced superposition of equally distributed amplitudes after the index creation procedure is applied (i.e., after  $l$  indices are added, with  $0 < l \leq k$ ). If we tolerate an error in the amplitude distribution, we may apply the general amplitude transfer algorithm presented by Brassard *et al.* [17]. On the other hand, if the amplitudes have to be transferred *exactly*, one applies the *exact* amplitude amplification algorithm [17] that includes two additional parameters  $\phi, \rho$ . Thus, the former is a particular case of the latter with  $\phi, \rho = \pi$  for the following step operator with the angles  $0 \leq \phi, \rho < 2\pi$  [17]:

$$\mathbf{Q} = \mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \phi, \rho) = -\mathcal{U}_{|\text{QDB}^{(k)}\rangle} \mathbf{S}_0(\phi) \mathcal{U}_{|\text{QDB}^{(k)}\rangle}^\dagger \mathbf{S}_\chi(\rho). \quad (39)$$

This is equivalent to the operator for Grover search if the unitary  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}$  is substituted by the Walsh-Hadamard transform [1]. In our case, the unitary  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}$  corresponds to the operations needed to prepare the database state and satisfies:

$$\mathcal{U}_{|\text{QDB}^{(k)}\rangle} (|0\rangle_I |0\rangle_D) = \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_j\rangle_D, \quad (40)$$

while the operator  $\mathbf{S}_\chi(\rho)$  is defined as:

$$|x\rangle \mapsto \begin{cases} e^{i\rho} |x\rangle, & \text{if } \chi(x) = 1, \\ |x\rangle, & \text{if } \chi(x) = 0. \end{cases} \quad (41)$$

Thereby,  $\chi(x) = 1$  if and only if  $|x\rangle = |0\rangle_{I \otimes D}$  and  $\chi(x) = 0$  otherwise. Furthermore,  $\mathbf{S}_0(\phi)$  is acting as a phase multiplication by  $e^{i\phi}$  to the all-zero string and equals to the identity operation otherwise. Hence, in the approximate case, one would apply the step operator  $m$ -times (i.e., this is given by  $(\mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \pi, \pi))^m$ ) with  $m = \lfloor m^* \rfloor$  and  $\lfloor x \rfloor$  denoting the floor function that returns the greatest integer  $\leq x$  for  $x \in \mathbb{R}$ . Thereby,  $m^* \in \mathbb{R}$  is given by the relation:

$$m^* = \frac{\mp \arcsin\left(\frac{1}{\sqrt{k+l}}\right) - \arcsin\left(\frac{1}{\sqrt{k}}\right) + \pi n}{2 \arcsin\left(\frac{1}{\sqrt{k}}\right)} \quad (42)$$

with  $n \in \mathbb{Z}$ ,  $0 < l \leq k$  and  $k \geq 2$  such that  $m^* > 0$ . As the step operator  $(\mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \pi, \pi))^m$  is only applied in integer multiples, the last step is done with respect to the generalized step operator given in (39) to obtain an exact amplitude transfer. To transfer the correct amount of amplitude, one can make arbitrarily small rotations by varying the angles  $\rho, \psi$  in (39). This is because if parameters  $\rho, \psi$  are varied continuously between 0 and  $2\pi$ . Hence, to achieve a correct amplitude transfer in the last step of the operation, the parameters are chosen accordingly to fulfill the following equation:

$$\langle 0 | (\mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \phi, \rho) \cdot (\mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \pi, \pi))^m | 0 \rangle_{I \otimes D} = \sqrt{\frac{l+1}{k+l}}. \quad (43)$$

The normalization condition implies:

$$\langle \text{QDB}'_{\neq 0}^{(k)} | (\mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \phi, \rho) \cdot (\mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \pi, \pi))^m | \text{QDB}'_{\neq 0}^{(k)} \rangle = \sqrt{\frac{k-1}{k+l}}, \quad (44)$$

with

$$|\text{QDB}'_{\neq 0}^{(k)}\rangle := \frac{1}{\sqrt{k-1}} \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D \quad (45)$$

being the re-normalized superposition state of  $k$  indices and data elements without the state  $|0\rangle_{I \otimes D}$ . The best practice for the index extension scenario is to add as many indices as possible and thus be resource-efficient concerning each newly added ancilla qubit. In this case, we have  $l = k$ . Thus, the total number of indices doubles during the operation. After the amplitude transfer is applied, the imbalanced amplitudes are now given in the exact case:

$$|\psi_1\rangle_{I \otimes D} := \sqrt{\frac{l+1}{k+l}} |0\rangle_{I \otimes D} + \frac{1}{\sqrt{k+l}} \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D = \sqrt{\frac{l+1}{k+l}} |0\rangle_{I \otimes D} + \sqrt{\frac{k-1}{k+l}} |\text{QDB}'_{\neq 0}^{(k)}\rangle. \quad (46)$$

---

**Algorithm 2** Transfer amplitude to zero-index state of  $|\text{QDB}^{(k)}\rangle$  (an application of Quantum Amplitude Amplification [33, 17]).

---

**Require:**  $k > 1$ ,  $k \in \mathbb{N}$  and  $l, m \in \mathbb{N}_{>0}$

- 1: **procedure**  $\text{TRANSFER}(k, l)$  ▷ transfer amplitude to the  $|0\rangle_I$
  - 2:  $|\text{QDB}^{(k)}\rangle \leftarrow \frac{1}{\sqrt{k}} \sum_{j=0}^{k-1} |j\rangle_I |d_j\rangle_D$  ▷ insert  $|\text{QDB}^{(k)}\rangle$
  - 3:  $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow \left( \mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \phi = \pi, \rho = \pi) \right)^m |\text{QDB}^{(k)}\rangle$  ▷ apply  $m$  times
  - 4:  $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow \left( \mathbf{Q}(\mathcal{U}_{|\text{QDB}^{(k)}\rangle}, \chi, \phi, \rho) \right) |\text{QDB}_{imb}^{(k)}\rangle$  ▷ apply with  $\chi, \rho$  as in (43)
  - 5: **end procedure**
- 

As the amplitude transfer is achieved, an ancilla qubit in the state  $|0\rangle$  is added as a next step. A conditional  $Y$ -rotation gate is applied that acts non-trivially on the added ancilla qubit. The  $Y$ -rotation is conditioned on the index being the zero-string and the angle  $\theta$  is given by  $\theta = 2 \arccos(\frac{1}{\sqrt{l+1}})$ . The operation is explicitly given by:

$$CR_y(\theta) := R_y(\theta) \otimes |0\rangle\langle 0|_{I \otimes D} + \mathbb{I}_A \otimes (\mathbb{I}_{I \otimes D} - |0\rangle\langle 0|_{I \otimes D}). \quad (47)$$

Applying this operation from (47), one obtains the state:

$$CR_y(\theta) (|0\rangle_A |\psi_1\rangle_{I \otimes D}) = \left( \sqrt{\frac{1}{k+l}} |0\rangle_A + \sqrt{\frac{l}{k+l}} |1\rangle_A \right) |0\rangle_{I \otimes D} + \frac{1}{\sqrt{k+l}} \sum_{j=1}^{k-1} (|0\rangle_A |j\rangle_I) |d_j\rangle_D. \quad (48)$$

The state given in (48) can be rewritten as:

$$\sqrt{\frac{l}{k+l}} |1\rangle_A |0\rangle_{I \otimes D} + \frac{1}{\sqrt{k+l}} \sum_{j=0}^{k-1} (|0\rangle_A |j\rangle_I) |d_j\rangle_D. \quad (49)$$

Thereby, we used the following notation  $|0\rangle_{I \otimes D} := |0\rangle_I |0\rangle_D$  as before. Given the state in (49), the index of the zero-index  $I$  part in the state  $|1\rangle_A |0\rangle_I |0\rangle_D$  is *unfolded*. If we read the index basis from left to right, the new indices correlated to empty data strings will be all *odd* numbers. To create  $l$  new *odd* indices correlated to the empty data state  $|0\rangle_D$ , we apply the following operation:

$$|1\rangle\langle 1|_A \otimes (P_{(l)})_I \otimes \mathbb{I}_D + |0\rangle\langle 0|_A \otimes \mathbb{I}_I \otimes \mathbb{I}_D. \quad (50)$$

As before, the operation  $P_{(l)}$  reduces to the Walsh-Hadamard transformation for  $\log_2(l) \notin \mathbb{N}_{>0}$ . One observes that it makes sense to fully use the available resources by adding all possible new indices per added ancilla qubit with regard to the circuit depth for each index extension operation. The circuit depth of  $P_{(l)}$  is  $\mathcal{O}(1)$  if  $\log_2(l) \in \mathbb{N}_{>0}$  and scales linearly for  $\log_2(l) \notin \mathbb{N}_{>0}$  according to  $\mathcal{O}(n)$  with  $n = \lceil \log_2(l) \rceil$ .

The extension operation presented here is expensive as the step operator in (39) includes the database creation unitary  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}$ , and its inverse  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}^\dagger$ , that one might not even have at hand. One has to apply this unitary repeatedly if  $m > 1$ . Hence, we considered an alternative approach that requires setting the maximal number of indices to be added beforehand. In the next section, we discuss a protocol in which one initializes a specifically imbalanced database during the preparation stage of the empty database and creates new indices later.

---

**Algorithm 3** Unfold zero index-string based on Walsh-Hadamard (WH) transformation with one ancilla qubit.

---

**Require:**  $k > 1$ ,  $k \in \mathbb{N}$  and  $l \in \mathbb{N}_{>0}$

- 1: **procedure** UNFOLD( $k, l$ ) ▷ unfold zero index string into  $l$  new index states
  - 2:    $|\text{QDB}_{imb}^{(k)}\rangle$  ▷ insert imbalanced QDB
  - 3:    $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow |0\rangle_A |\text{QDB}_{imb}^{(k)}\rangle$  ▷ add  $|0\rangle_A$
  - 4:    $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow CR_y(\theta) |\text{QDB}_{imb}^{(k)}\rangle$  ▷  $CR_y(\theta)$  as in (47)
  - 5:    $|\text{QDB}_{imb}^{(k+l)}\rangle \leftarrow (|1\rangle\langle 1|_A \otimes (P_{(l)})_I \otimes \mathbb{I}_D + |0\rangle\langle 0|_A \otimes \mathbb{I}_{I \otimes D}) |\text{QDB}_{imb}^{(k)}\rangle$  ▷ apply  $P_{(l)}$  on register  $I$
  - 6: **end procedure**
- 

---

**Algorithm 4** Extension of  $|\text{QDB}^{(k)}\rangle$  by arbitrary  $l > 0$ .

---

**Require:**  $k > 1$ ,  $k \in \mathbb{N}$  and  $l \in \mathbb{N}_{>0}$

- 1: **procedure** EXTEND( $k, l$ ) ▷ extend database by  $l$  indices correlated to  $|0\rangle_D$
  - 2:   **while**  $l > k$  **do**
  - 3:     Set  $l' \leftarrow k$  and  $l \leftarrow (l - k)$
  - 4:     Apply TRANSFER( $k, l'$ ) ▷ amplitude transfer using Alg. 2
  - 5:     Apply UNFOLD( $k, l'$ ) ▷ create new index states by using Alg. 3
  - 6:      $k \leftarrow k + l'$
  - 7:   **end while**
  - 8:   **if**  $l \leq k$  **then**
  - 9:     Set  $l' \leftarrow l$
  - 10:     Apply TRANSFER( $k, l'$ ) ▷ amplitude transfer using Alg. 2
  - 11:     Apply UNFOLD( $k, l'$ ) ▷ create new index states by using Alg. 3
  - 12:      $k \leftarrow k + l'$
  - 13:   **end if**
  - 14:   **return**  $|\text{QDB}^{(k)}\rangle$  ▷ balanced superposition as database with  $k + l$  index elements
  - 15: **end procedure**
- 

### A.3.2 Initialize and unfold

This extension technique does not depend on using the unitary  $\mathcal{U}_{|\text{QDB}^{(k)}\rangle}$  (or its inverse). In general, the extension mechanism is described by the following map:

$$\begin{aligned} \mathcal{H}_I \otimes \mathcal{H}_D &\longrightarrow \mathcal{H}_{I'} \otimes \mathcal{H}_D := (\mathcal{H}_A + \mathcal{H}_I)_{I'} \otimes \mathcal{H}_D \\ |\text{QDB}_{imb}^{(k)}\rangle &\mapsto |\text{QDB}_{imb'}^{(k+f)}\rangle. \end{aligned} \quad (51)$$

The dimensions are given as follows:  $\dim(\mathcal{H}_A) = z$ ,  $\dim(\mathcal{H}_I) = k$  and  $\dim(\mathcal{H}_D) = m$  as before. We make use of a modification of the generalized prepare operation  $P_{(l)}$  given in A.1 or alternatively general amplitude encoding (see e.g., [34]). Thereby, we know that we would like to add at most  $l$  new elements to the database. Thus, we initialize the following state:

$$|\text{QDB}_{imb}^{(k)}\rangle := \sqrt{\frac{l+1}{l+k}} |0\rangle_{I \otimes D} + \sqrt{\frac{1}{l+k}} \sum_{j=1}^{k-1} |j\rangle_I |0\rangle_D. \quad (52)$$

A corresponding quantum circuit for amplitude encoding operates on  $\tilde{k}$  qubits and has a depth of  $\mathcal{O}(\text{poly}(k))$  [34]. Following this operation, we write the first  $(k-1)$  data elements to the

empty database we just prepared, and we obtain the following:

$$|\text{QDB}_{imb'}^{(k)}\rangle = \sqrt{\frac{l+1}{l+k}} |0\rangle_{I\otimes D} + \sqrt{\frac{1}{l+k}} \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D. \quad (53)$$

Next, we correlate  $z$  ancilla qubits in the  $|0\rangle$ -state, denoted by  $|0\rangle_A$ , in order to create at most  $l \leq (2^z - 1) \cdot k$  new indices correlated to the empty data string:

$$|0\rangle_A \left( \sqrt{\frac{l+1}{l+k}} |0\rangle_{I\otimes D} + \sqrt{\frac{1}{l+k}} \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D \right). \quad (54)$$

In this case, we can read the binary indices from right to left, and thus, all newly added indices have an absolute value larger than  $k$ . Hence, if we add new indices in a parallelized manner, we proceed as follows for  $l' \leq (2^z - 1)$ :

$$CP_A(z) := \left( (P_{(l'+1)})_A \otimes |0\rangle\langle 0|_I + \mathbb{I}_A \otimes (\mathbb{I}_I - |0\rangle\langle 0|_I) \right) \otimes \mathbb{I}_D. \quad (55)$$

The best practice here is to create as many new indices as possible. This reduces the operation (55) to:

$$CH_A(z) := \left( (H^{\otimes z})_A \otimes |0\rangle\langle 0|_I + \mathbb{I}_A \otimes (\mathbb{I}_I - |0\rangle\langle 0|_I) \right) \otimes \mathbb{I}_D, \quad (56)$$

and we have  $l' = (2^z - 1)$  when using (56). Thereby,  $l'$  is chosen to be  $l' = (2^z - 1)$  if  $l \geq (2^z - 1)$  and  $l' = l$  otherwise. Applying the operator given in (55) on the database state in (54) results in:

$$\sqrt{\frac{l+1}{(l'+1) \cdot (l+k)}} \sum_{i=0}^{l'} |i\rangle_A |0\rangle_{I\otimes D} + \sqrt{\frac{1}{l+k}} |0\rangle_A \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D. \quad (57)$$

By applying (55) on the state (54) we bring the indices in the  $A$ -subspace in superposition, conditioned on the index in  $I$  being  $|0\rangle_I$ . As given in (57), the operation decreases the amplitude of the initial  $|0\rangle_{I\otimes D}$  state by a factor of  $\sqrt{1/(l'+1)}$ . We created a superposition of all possible combinations of orthogonal states in the  $A$ -subspace that are all in a product state concerning the empty old index string part  $I$  and the all-zero data part, i.e., the state  $|0\rangle_{I\otimes D}$ . The operation defined in (55) thus creates  $l'$  new index states correlated to the empty data register. Alternatively, if one aims to add more indices, one may bring the old index  $I$ -subspace (correlated to indices with a Hamming weight larger than one in the  $A$ -subspace) into superposition. In the case that  $l \leq (2^z - 1)$ , we do not need to apply the next steps as we have already added enough indices. If  $l > (2^z - 1)$ , we add  $l - l' =: l'' \leq (k-1) \cdot (2^z - 1)$  new indices. For this, we mark all states with a Hamming weight larger than one by an ancilla qubit based on the algorithm in Appendix A.5. Conditioned on this ancilla qubit being in  $|0\rangle$ , we create new indices in the *old* index subspace  $I$ . Alternatively to adding a marker qubit, one can apply the following conditional operation with the effect (with  $l' \leq (\log_2(k) - 1)$ ):

$$CP_I(k) := \left( |0\rangle\langle 0|_A \otimes \mathbb{I}_{I\otimes D} \right) + \left( (\mathbb{I}_A - |0\rangle\langle 0|_A) \otimes (P_{(l'')})_I \otimes \mathbb{I}_D \right). \quad (58)$$

This transformation decreases the amplitude of all states (except those correlated to  $|0\rangle_A$ ) by a factor of  $\frac{1}{\sqrt{l''}}$ . As a result, we added in total  $l' \cdot l''$  new index states correlated to the empty data string. A complete amplitude balance is only achieved in a special case for this operation. The final state after applying this extension procedure and adding all the indices is given as follows:

$$|\text{QDB}_{imb}^{(k+l' \cdot l'')}\rangle := \alpha \cdot |0\rangle_A |0\rangle_{I\otimes D} + \sum_{j=1}^{k-1} \beta \cdot (|0\rangle_A |j\rangle_I) |d_j\rangle_D + \sum_{h=0}^{l''-1} \sum_{j=1}^{l'+1} \gamma \cdot |j\rangle_A |h\rangle_I |0\rangle_D, \quad (59)$$

---

**Algorithm 5** Extend imbalanced database by  $l$  new index elements with multiple ancilla qubits.

---

**Require:**  $k > 1$ ,  $k \in \mathbb{N}$  and  $l, l', l'', z \in \mathbb{N}_{>0}$

- 1: **procedure** IMBALANCED EXTEND( $k, l$ ) ▷ extend imbalanced  $|\text{QDB}_{imb}^{(k)}\rangle$
  - 2:  $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow \sqrt{\frac{l+1}{k+l}} |0\rangle_{I \otimes D} + \frac{1}{\sqrt{k+l}} \sum_{j=1}^{k-1} |j\rangle_I \otimes |0\rangle_D$  ▷ prepare empty QDB
  - 3:  $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow \sqrt{\frac{l+1}{k+l}} |0\rangle_{I \otimes D} + \frac{1}{\sqrt{k+l}} \sum_{j=1}^{k-1} |j\rangle_I \otimes |d_j\rangle_D$  ▷ write data in register  $D$
  - 4:  $|\text{QDB}_{imb}^{(k)}\rangle \leftarrow |0\rangle_A |\text{QDB}_{imb}^{(k)}\rangle$  ▷ correlate  $z$  ancilla qubits to index
  - 5:  $|\text{QDB}_{imb}^{(k+l')}\rangle \leftarrow CP_A(z) |\text{QDB}_{imb}^{(k)}\rangle$  ▷ create ( $l'$ ) new indices
  - 6:  $|\text{QDB}_{imb}^{(k+l'.l'')}\rangle \leftarrow CP_I(k) |\text{QDB}_{imb}^{(k+l')}\rangle$  ▷ create ( $l'$ )  $\cdot$  ( $l'' - 1$ ) new indices
  - 7: **end procedure**
- 

with

$$\alpha := \sqrt{\frac{l+1}{(l'+1) \cdot (l+k)}}, \quad \beta := \sqrt{\frac{1}{l+k}}, \quad \gamma := \sqrt{\frac{l+1}{l'' \cdot (l'+1) \cdot (l+k)}}. \quad (60)$$

Therefore, we obtain a database that is balanced except with respect to the  $|0\rangle_A |0\rangle_{I \otimes D}$  state, if  $\frac{l+1}{(l'+1) \cdot l''} \equiv 1$ . As a note, we add here that, in general, one may add individual indices instead of applying (59) based on a combination of Hadamard and CNOT operations. When using this, one must keep track of the newly added indices to avoid adding the same index *twice*. This would “disturb” the correlated data state.

**Example case of Algorithm 5 for  $z = 1$  and  $l \leq k$ .**

We present the extension Algorithm 5 for the simplified case in which we add one ancilla qubit ( $z = 1$ ) to the index register of the imbalanced database. Initially, we start off in an imbalanced database state (prepared for  $l$  new elements to be added) and have a large amplitude on the all-zero index state that we see as playing the role of a *reservoir*. The database contains already written data elements; thus, we start off in the imbalanced state that is prepared for  $l$  new index elements to be added:

$$|\text{QDB}_{imb}^{(k)}\rangle = \sqrt{\frac{l+1}{l+k}} |0\rangle_I |0\rangle_D + \frac{1}{\sqrt{l+k}} \sum_{j=1}^{k-1} |j\rangle_I |d_j\rangle_D. \quad (61)$$

Next, we add a single ancilla to the index register. We denote the ancilla register by  $A$ :

$$|\text{QDB}_{imb,A}^{(k)}\rangle = \sqrt{\frac{l+1}{l+k}} |0\rangle_A |0\rangle_I |0\rangle_D + \frac{1}{\sqrt{l+k}} \sum_{j=1}^{k-1} |0\rangle_A |j\rangle_I |d_j\rangle_D. \quad (62)$$

Following that, we apply a conditional  $Y$ -rotation to the state in (62) and obtain:

$$\begin{aligned} |\text{QDB}_{imb,A}^{(k)}\rangle &= \sqrt{\frac{l+1}{l+k}} \left( \frac{1}{\sqrt{l+1}} |0\rangle_A + \sqrt{\frac{l}{l+1}} |1\rangle_A \right) |0\rangle_I |0\rangle_D + \frac{1}{\sqrt{l+k}} \sum_{j=1}^{k-1} |0\rangle_A |j\rangle_I |d_j\rangle_D \\ &= \frac{1}{\sqrt{l+k}} |0\rangle_A |0\rangle_I |0\rangle_D + \sqrt{\frac{l}{l+k}} |1\rangle_A |0\rangle_I |0\rangle_D + \frac{1}{\sqrt{l+k}} \sum_{j=1}^{k-1} |0\rangle_A |j\rangle_I |d_j\rangle_D. \end{aligned} \quad (63)$$

Following that, we can create new indices from the state  $|1\rangle_A |0\rangle_I$  in order not to affect the already existing indices that correspond to written data elements. To implement this function,



we use the generalized prepare operation for  $l$  elements:

$$|1\rangle_A |0\rangle_I \xrightarrow{P(l)} |1\rangle_A \left( \frac{1}{\sqrt{l}} \sum_{j=0}^{l-1} |j\rangle_I \right). \quad (64)$$

Thus, we obtained the following superposition state with  $l$  new indices:

$$|\text{QDB}^{(k+l)}\rangle = \frac{1}{\sqrt{l+k}} |0\rangle_A |0\rangle_I |0\rangle_D + \frac{1}{\sqrt{l+k}} \sum_{i=0}^{l-1} (|1\rangle_A |i\rangle_I) |0\rangle_D + \frac{1}{\sqrt{l+k}} \sum_{j=1}^{k-1} |0\rangle_A |j\rangle_I |d_j\rangle_D. \quad (65)$$

In (65), we obtained a balanced quantum database state with  $l$  new index elements correlated to  $|0\rangle_D$ .

#### A.4 Note on disentangling quantum states

In order to describe the process of *disentanglement*, we first summarize *entanglement* conceptually. Hence, we consider the Schmidt decomposition for bipartite pure states of a composite system  $A + B$ . The Schmidt decomposition [24] of a pure quantum state is given by:

$$|\psi\rangle = \sum_{i=1}^{r_\psi} \lambda_i (|a_i\rangle_A \otimes |b_i\rangle_B) \quad (66)$$

with  $\lambda_i \in \mathbb{R}, \lambda_i \geq 0$  and  $\sum_{i=1}^{r_\psi} \lambda_i^2 = 1$ . Thereby,  $r_\psi$  is the Schmidt rank of  $|\psi\rangle$ , and a state is entangled if and only if  $r_\psi > 1$ . The reduced density matrices of the subsystems  $A$  and  $B$  are given by  $\text{Tr}_B(|\psi\rangle\langle\psi|) =: \rho_A = \sum_{i=1}^{r_\psi} \lambda_i^2 (|a_i\rangle\langle a_i|_A)$  and  $\text{Tr}_A(|\psi\rangle\langle\psi|) =: \rho_B = \sum_{i=1}^{r_\psi} \lambda_i^2 (|b_i\rangle\langle b_i|_B)$ . During the general read-in process to the quantum database, a respective sensor would get entangled with the database state. However, after the data is stored in the database, we no longer need the sensor and would like to trace out this additional system. This is non-trivial due to the entanglement between both subsystems (database and sensor). Thus, we consider the concept of disentanglement while preserving as many local properties as possible. In general, in the case of the database, we aim to maintain the database but not the sensor system.

Under the process of *disentanglement*, as introduced in [20, 35], we understand the following operation:

**Theorem 1** (Definition [20, 23]). *Disentanglement is the process that transforms a state of two or more subsystems into an unentangled state in general, i.e., a mixture of product states such that the reduced density matrices of each of the subsystems are unaffected.*

Mathematically, we aim to do the following transformation [36]:

$$\rho_{AB} = |\psi\rangle\langle\psi| \rightarrow \rho_A \otimes \rho_B. \quad (67)$$

It was shown that there exists no universal disentangling machine (describing the transformation for a general entangled state  $\rho_{AB}$  in (67)) that transforms an arbitrary entangled (or inseparable state) into a separable one [20, 36]. But what happens if we are only interested in a unitary transformation doing the following operation :

$$|\psi\rangle\langle\psi|_{AB} = \sum_{i,j=1}^{r_\psi} \lambda_i \lambda_j^* (|a_i\rangle\langle a_j|_A \otimes |b_i\rangle\langle b_j|_B) \rightarrow \left( \sum_{i,j=1}^{r_\psi} \lambda_i \lambda_j^* |a_i\rangle\langle a_j|_A \right) \otimes \sigma_B. \quad (68)$$

**Lemma 2.** *If the transformation presented in (68) is a unitary transformation, we have  $H(\sigma_B) \equiv 0$  with  $H(\cdot)$  being the von Neumann entropy.*

*Proof.* Let us consider the von Neumann entropies:  $H(|\psi\rangle\langle\psi|_{AB}) = \sum_{i=1}^{r_\psi} |\lambda_i|^2 \log_2(|\lambda_i|^2)$  and  $H\left(\sum_{i,j=1}^{r_\psi} \lambda_i \lambda_j^* |a_i\rangle\langle a_j|_A\right) = \sum_{i=1}^{r_\psi} |\lambda_i|^2 \log_2(|\lambda_i|^2) = H(|\psi\rangle\langle\psi|)$ . Suppose the transformation in (68) is done by a similarity transformation, i.e.,  $\rho \mapsto U\rho U^{-1}$  with  $U \in GL(\mathcal{H}_A)$  with  $\rho$  being an arbitrary state on  $S(\mathcal{H}_{AB})$ , the total von Neumann entropy is constant during the transformation. Thus,  $H(\sigma_B) = 0$  and then  $\sigma_B = |\Sigma\rangle\langle\Sigma|_B$  with  $|\Sigma\rangle_B$  being necessarily a pure state with von Neumann entropy zero.  $\square$

This relates to the preservation of local properties during the disentanglement process.

$$|\psi\rangle_{AB} = \sum_{i=1}^{r_\psi} \lambda_i (|a_i\rangle_A \otimes |b_i\rangle_B) \rightarrow \left( \sum_{i=1}^{r_\psi} \lambda_i |a_i\rangle_A \right) \otimes |\Sigma\rangle_B. \quad (69)$$

Here, system  $A$  would be the quantum database, and system  $B$  is the sensor from which we read the data. During the writing process, both systems get entangled. For specific cases, e.g., if only classical correlations are present, techniques such as those presented in [19] can be used.

## A.5 Algorithmic Check of Hamming weight

As a sub-algorithm for the QDB extension explained in Appendix A.3, we aim to mark all states with a Hamming weight larger than one based on an ancilla qubit. To implement this condition, we correlate an ancilla qubit (initially in the  $|0\rangle$  state) that is  $|1\rangle$  only if the Hamming weight of the state in the ancilla subspace  $A$  is larger than one (0th qubit here). Next, one would apply the Algorithm 6. This procedure realizes a function marking the computational basis state with a Hamming weight  $H(\mathbf{s})$  of 0 by taking the ancilla qubit to the  $|1\rangle$  state only in this particular case. Hence, we implement the following function on the ancilla qubit:

$$f(\mathbf{s}) = \begin{cases} 0, & \text{if } H(\mathbf{s}) = 0, \\ 1, & \text{if } H(\mathbf{s}) > 0. \end{cases} \quad (70)$$

$H(\mathbf{s})$  denotes the Hamming weight of the binary string  $\mathbf{s}$  in Eq. (70). The full algorithmic description is below (see Algorithm 6). The circuit depth scales as  $\mathcal{O}(|B|)$  if we aim to use only single-qubit and CNOT gates. This is due to the decomposition of the multi-controlled CNOT gate, which is part of the third step of the algorithm. For more details on the decomposition of this gate, see, e.g., Ref. [37].

---

**Algorithm 6** Mark if the Hamming weight of string  $\mathbf{s}$  is larger than zero.

---

**Require:**  $\mathbf{s} \in \{0, 1\}^{|B|}$

- 1: **procedure** HAMMING CHECK( $\mathbf{s}$ )
  - 2:   Insert  $|0\rangle_A |\mathbf{s}\rangle_B \in \mathcal{H}_A \otimes \mathcal{H}_B$
  - 3:    $|0\rangle_A |\bar{\mathbf{s}}\rangle_B \leftarrow \mathbb{I}_A \otimes \left( \bigotimes_{|B|} \sigma_x \right) |0\rangle_A |\mathbf{s}\rangle_B$     $\triangleright$  revert the state in the  $B$ -subspace
  - 4:    $|f(\bar{\mathbf{s}})\rangle_A |\bar{\mathbf{s}}\rangle_B \leftarrow \Lambda_{|B|}(\sigma_x) |0\rangle_A |\bar{\mathbf{s}}\rangle_B$   $\triangleright$  multi-controlled gate applying  $\sigma_x$  on  $A$  iff  $\bar{\mathbf{s}} = \{1\}^{|B|}$
  - 5:    $|f(\mathbf{s})\rangle |\mathbf{s}\rangle \leftarrow \sigma_x \otimes \left( \bigotimes_{|B|} \sigma_x \right) |f(\bar{\mathbf{s}})\rangle_A |\bar{\mathbf{s}}\rangle_B$     $\triangleright$  revert the state back in the  $B$ -subspace
  - 6: **end procedure**
-