

Linear optics quantum computing – construction of small networks and asymptotic scaling

Konrad Kieling

Thesis submitted in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy of the University of London
and the Diploma of Membership of Imperial College.

September 2008

Imperial College
London



Institute for

Mathematical Sciences

Abstract

In the paradigm of linear optics, quantum states of optical field modes are manipulated by means of photon sources, beam splitters, and photo-detectors. This restriction of the allowed interactions introduces non-unit success probabilities into every non-trivial computation irrevocably. The resulting probabilistic nature of this architecture opens up questions about scalability of this approach as well as about basic limitations like optimal resource consumption and the construction of optical networks for given tasks.

The scalability question is addressed within the framework of cluster state computing which allows to shift the problem towards the problem of scalable state production. Several schemes with different constraints, resulting in different experimental feasibility, are shown to exhibit optimal scaling behaviour. These schemes include restrictions to certain easy-to-implement gate sets, and the banning of re-routing, which is a major obstacle in optical quantum computing.

In the small-scale regime various methods for constructing and analysing linear optics networks are presented. Rather than constructing the gates in the general quantum gate model, a direct construction from linear optics' basic elements – beam splitters and phase shifters – will be considered. This can result in higher success probabilities and smaller resource consumption. Due to the complexity of the problems, these techniques are applicable only to quantum gates which involve a small number of optical modes and photons. A couple of examples ranging from state preparation over quantum gates to state discrimination are used to illustrate these tools.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 9 |
| 2 | Foundations | 13 |
| 2.1 | Quantum computing | 13 |
| 2.2 | Linear optics | 14 |
| 3 | Finite size linear optics – techniques | 20 |
| 3.1 | General network construction | 21 |
| 3.1.1 | Action of linear optics | 21 |
| 3.1.2 | Photon detection | 22 |
| 3.1.3 | Objective state | 23 |
| 3.1.4 | Solving polynomial equations | 24 |
| 3.1.5 | Manifold of solutions | 26 |
| 3.1.6 | Unitary extensions | 27 |
| 3.1.7 | Success probability | 29 |
| 3.2 | State transformations | 33 |
| 3.2.1 | A convenient way of specifying bosonic state vectors | 33 |
| 3.2.2 | $k = 2$ | 34 |
| 3.2.3 | State transformation | 34 |
| 3.2.4 | Classes of two-photon states in dual-rail encoding | 36 |
| 3.2.5 | Increasing the rank | 36 |
| 3.2.6 | Higher dimensional spaces | 38 |
| 3.3 | Polynomial factorisation | 39 |
| 4 | Examples of small-scale networks | 41 |
| 4.1 | NOON state generation | 42 |
| 4.2 | Dual-rail qutrits and ququads | 44 |
| 4.2.1 | Qutrits | 44 |

| | | |
|----------|---|------------|
| 4.2.2 | Ququads | 45 |
| 4.3 | Controlled phase gates | 50 |
| 4.3.1 | Single beam splitter | 51 |
| 4.3.2 | Arbitrary phases | 52 |
| 4.3.3 | Experimental implementation | 55 |
| 4.3.4 | Event-ready gates | 57 |
| 4.3.5 | Toffoli gates | 58 |
| 4.4 | State discrimination | 59 |
| 4.4.1 | Tetrahedral basis | 60 |
| 4.4.2 | Bell states | 62 |
| 4.4.3 | Photon number resolution | 64 |
| 4.5 | Discussion | 73 |
| 5 | Cluster states | 75 |
| 5.1 | The one-way computer | 77 |
| 5.2 | Resource state generation by probabilistic gates | 78 |
| 5.2.1 | Probabilistic gates used in cluster state production | 79 |
| 5.2.2 | Concepts: configurations and strategies | 82 |
| 5.2.3 | Simple strategies | 86 |
| 5.2.4 | Computer-assisted results | 94 |
| 5.2.5 | Lower bounds | 97 |
| 5.2.6 | Upper bounds – the razor model | 98 |
| 5.2.7 | An inverse question | 110 |
| 5.2.8 | Two-dimensional cluster states | 112 |
| 5.3 | Renormalisation and percolation: Banning re-routing | 118 |
| 5.3.1 | Percolation in a nutshell | 120 |
| 5.3.2 | Renormalisation | 121 |
| 5.3.3 | Path identification | 127 |
| 5.3.4 | Reduction to a renormalised lattice | 129 |
| 5.3.5 | Practical considerations – decreasing the vertex degree | 130 |
| 5.3.6 | Loss tolerance | 135 |
| 5.4 | Discussion | 136 |
| 6 | Discussion and outlook | 138 |
| A | Notation | 141 |
| B | Bounds to singular values | 142 |

C Irreducibility of a dual-rail EPR pair

143

LOQC – construction of small networks and asymptotic scaling

5

List of Figures

| | | |
|------|--|----|
| 2.1 | Decomposition of an $SU(n)$ unitary in terms of a general network of beam splitters. | 17 |
| 3.1 | Construction of a non-unitary linear mode transformation by using linear optics and vacuum extensions. | 29 |
| 4.1 | Decomposition of a general linear transformation on three modes. | 45 |
| 4.2 | General network on three non-trivial modes in polarisation encoding. | 46 |
| 4.3 | Experimentally feasible state transformation between dual-rail qutrits. | 47 |
| 4.4 | Optimal success probabilities of controlled phase gates. | 53 |
| 4.5 | Basic spatial modes based setup for a controlled phase gate . . . | 55 |
| 4.6 | Setup for controlled phase gate on two polarisation encoded dual-rail qubits. | 56 |
| 4.7 | Equivalent implementations of a partially polarising beam splitter. | 57 |
| 4.8 | Compact implementation of a controlled phase gate. | 58 |
| 4.9 | Optimal success probabilities of generalised Toffoli gates. | 59 |
| 4.10 | Linear optics network for EPR pair production. | 69 |
| 4.11 | Ratio of false count probability with respect to two photon detection probability. | 72 |
| 5.1 | Linear optics “fusion” gates. | 80 |
| 5.2 | Action of a fusion gate on the end qubits of two linear cluster states. | 81 |
| 5.3 | Bell state measurement by using a parity check gate. | 82 |
| 5.4 | Example of a tree of successive configurations under application of a strategy. | 84 |

| | | |
|------|---|-----|
| 5.5 | Expected length for MODESTY, GREED, STATIC, the optimal strategy, and lower and upper bounds to the latter. | 87 |
| 5.6 | Evolution of a configuration under GREED seen as a tree describing a balanced random walk. | 88 |
| 5.7 | Influence of the parameters of the different entangling gates on GREED. | 89 |
| 5.8 | Expected length for MODESTY and the optimal strategy and different upper bounds for different razor parameters. | 90 |
| 5.9 | Optimal expected length for different entangling gates with $p_s = 1/2$ | 95 |
| 5.10 | Intervals of reachable expected lengths and bounds to the respective maximum for fusion gates operating at different success probabilities. | 99 |
| 5.11 | Performance of the optimal strategy in the razor model and convergence to the full model for increasing razor parameter. | 100 |
| 5.12 | Configuration space and allowed actions in the $R = 2$ razor model. | 106 |
| 5.13 | Action of σ_z and σ_x measurements, and of the type-II fusion gate. | 113 |
| 5.14 | Arrangement of $n + 1$ chains to build a two-dimensional cluster state of width n | 114 |
| 5.15 | Overhead of two-dimensional cluster state production for different elementary success probabilities p_s | 117 |
| 5.16 | Overhead of two-dimensional cluster state production for different objective overall success probabilities P_s | 118 |
| 5.17 | Spanning probability for site percolation on a square lattice. | 119 |
| 5.18 | Examples of bond, site, and mixed percolation on the square lattice. | 121 |
| 5.19 | Blocks A_y and their overlap in the first dimension, B_y | 124 |
| 5.20 | Blocks A_y and their overlap in the second dimension, C_y | 124 |
| 5.21 | Effect of a σ_y measurement on a triangular junction of a cluster. | 130 |
| 5.22 | Numerical results on the scaling of the renormalisation procedure on the diamond lattice. | 131 |
| 5.23 | Dependence of the block size on the site occupation probability. | 132 |
| 5.24 | Measurements on the initial resources for the diamond and the pyrochlore lattices. | 133 |
| 5.25 | Parallel fusion on five-qubit stars for producing seven-qubit stars with an enhanced success probability. | 134 |

List of Tables

| | | |
|-----|--|-----|
| 4.1 | Output signals of photon counters given zero, one, or two photons. | 68 |
| 5.1 | Four types of probabilistic entangling gates. | 83 |
| 5.2 | Critical percolation probabilities for some lattices. | 122 |

With the advent of computers for information processing and communication, the demand for faster hardware and optimised algorithms led to a wide range of possible computer architectures. However, from the theorist’s point of view today’s computers are equivalent in that they can solve the same classes of problems efficiently. *Classical* (here used as non-quantum) information theory is based on the assumption that the concept of information [7] is independent of its physical carrier. Being implemented with classical devices, the smallest unit of information that is used in classical computer science is a binary digit (*bit*), so a dichotomic variable that can be in *either* of two states. These binary signals are carried by systems obeying the laws of classical physics. Together with a universal set of gates (basic logical operations such as the NAND gate), the so-called *circuit model* allows for the construction of computers which can simulate a Turing machine [8] up to polynomial overhead.

The famous quote by Rolf Landauer, “Information is physical” [9], summarises one of the key insights that led to quantum information theory. By abandoning classical information carriers, some restrictions of classical information theory could be avoided. Two-state systems ruled by quantum mechanics are able to adopt not only their two basis states, but also any complex (and normalised) superposition of these two. In analogy to the classical case, these two-level quantum systems are called *qubits* and they are manipulated by *quantum gates* (now in the *quantum circuit model*) which implement unitary evolutions on the quantum systems. Together with the structure of composite quantum systems another completely new feature enters information theory: *entanglement*, which is a type of non-classical correlation between sub-systems.

With these new features of information carriers at hand, Peter Shor showed in 1994 that factorisation of prime number products – which is believed to be a hard problem classically – can be solved *efficiently* [10] (*i.e.*, running time and size of the circuit depend at most polynomially on the problem size). Due to its importance to cryptography, this solution led to increased interest in possible quantum information processing architectures and the following in-depth investigation of possible algorithms identified some which are provably more efficient on a quantum computer

than any solution to the problem on a classical one (*e.g.*, searching an unstructured database [11]).

To pave the way to a quantum computer, systems that show intrinsic quantum behaviour – such that they can be brought into coherent superpositions – have to be identified. On the one hand, a theoretical understanding of the internal structure and possible decoherence channels, as well as tailored algorithms for these specific quantum systems are needed. On the other hand, experimental experience of how to initialise qubits carried by these systems in a defined state, how to implement unitary rotations on single and composite systems and how to read out the state of these qubits is necessary. Architectures that have been studied intensively in recent years include the ones based on atoms in optical cavities, trapped ions, superconductors, coherent light beams, nuclear magnetic resonance in molecules, nitrogen vacancy centres in quantum dots, trapped electrons, and single photons.

Qubits carried by single excitations of the electromagnetic field (*photons*) in the optical regime have the advantage of exhibiting only a small amount of decoherence. Furthermore, light is the natural choice when it comes to transmission of information over large distances as is necessary in quantum key distribution [12] or when coupling computers. On the downside, the lack of interaction implies that photons will not easily be persuaded to allow for non-trivial unitary evolutions on composite systems. An experimental advantage is given by the fact that today’s communication structure is based on fibre- and integrated optics, so some experience and infrastructure is readily available. Furthermore, energy-preserving operations (*linear optics* [13]) can be implemented experimentally at virtually no cost compared to other architectures where environments with close-to-zero temperatures as well as good vacuum are essential. The biggest experimental challenge is to keep all elements aligned with respect to each other within a wavelength to allow for interference. The most promising route here seems to be integration on optical chips [14], which also paves the way for miniaturisation and mass production.

At the theory side, the foremost open problems include the tailoring of circuits for linear optics and the optimisation of quantum gates and whole algorithms – with respect to resource requirements or the success probability. Being of central interest when it comes to implementations are ways of encoding information in a robust fashion with respect to phase mismatch and photon loss. Further research fields necessary for linear optics quantum computing lie in the construction of reliable photon sources and high efficiency detectors. In this thesis we are going to address the first two aspects, but will assume lossless optics with perfect phase matching. This allows for identification of general limitations and bounds to the linear optics

architecture, carrying over especially to the non-perfect case as well. Owing to the complexity of the problems, we will discuss our solutions in two different parts.

Firstly, a brief review of some basic physical concepts and methods is given in Chapter 2.

Secondly, in Chapter 3, methods will be introduced to explicitly construct linear optics networks for given unitaries and investigate their success probabilities. It turns out, that solutions for small problems (say, up to four modes) can be found explicitly, while solving large problems appears to be intractable.

These methods will be applied in Chapter 4 to a range of problems that are both – within the reach of experimental realisation, and interesting for the understanding of linear optics – as they show some unexpected behaviour. Starting from networks concerned with state preparation and -transformation (Sections 4.1 and 4.2), we will continue with the construction of two-qubit quantum gates (Section 4.3) and finally discuss some examples of measurement devices (Section 4.4).

Chapter 5 is dedicated to arguments concerning the scalability of linear optics. Instead of the explicit construction of large scale networks, bounds based on a fixed set of gates with known success probabilities will be derived. The generic production of entanglement, which is an essential resource for quantum computing, can be separated from the consumption of entanglement during the computation. This is achieved using the concept of a one-way computer [15] which is based on the preparation of highly entangled resource states (the *cluster states*) before the actual computation takes place and consumes the state successively. In fact, the latter part can easily be implemented with linear optical means, leaving only the question of scalable cluster state production, two different variants of which will be discussed.

A first scheme for cluster state generation will be introduced in Section 5.2, yielding good bounds confirming that scalable linear optics quantum computing could be possible in the framework of one-way computation. Afterwards, a more constrained approach, utilising less freedom, will be shown (Section 5.3). It aims at abandoning some aspects which are especially difficult to implement and is able to adopt almost the same scaling behaviour as is known from the unconstrained scheme.

Finally, in Chapter 6, we will deliver a brief summary of the results within the context of current research and will mention open problems in the field that are related to this work.

The algorithms presented can be implemented using computer algebra systems such as `Mathematica`¹ or `Singular`². Where exact results could not be obtained, numerics was carried out utilising the programming language `C`.

¹ <http://www.wolfram.com/products/mathematica/index.html>

² <http://www.singular.uni-kl.de>

2.1 Quantum computing

Quantum information processing takes the idea seriously that when storing or processing information, it matters whether the underlying physical system follows classical or quantum laws. In classical information theory, one is used to the fact that it hardly makes sense to think of the physical carrier of information, as one can transform the information stored in one form to another carrier in a lossless fashion anyway. This abstraction from the physical carrier is challenged when one thinks of *single quantum systems* forming the elementary processing units. Indeed, the very task of transforming the “information stored in a quantum system” into classical information and back is impossible. Quantum information processing is however not so much concerned with limitations due to quantum effects, but rather thinks of new applications in computing and communication when the carriers of information are single quantum systems.

A *quantum computer* [16, 17] is such an envisioned device: One thinks of having an array of n quantum systems – for example spins, referred to as *qubits*. This system, associated with a Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, is initially prepared in a known, preferably pure quantum state described by a *state vector* $|\psi\rangle$ ¹. Then one manipulates the state by means of unitary dynamics or by means of measurements. Acknowledging that not every unitary evolution is accessible on a quantum many-body system, in the *circuit model*, this step of computation is broken down to *quantum gates* [16, 17] – basic operations $\mathcal{U} : \mathcal{H} \mapsto \mathcal{H}$ with $\mathcal{U}^{-1} = \mathcal{U}^\dagger$, usually acting on a small set of sub-systems. One hence implements a sequence of unitary gates that have trivial support on all sites except single sites – giving rise to single qubit gates – and pairs of sites – two-qubit gates. The state vector after the unitary time evolution is then

$$|\psi\rangle \mapsto \prod_j \mathcal{U}_j |\psi\rangle. \quad (2.1)$$

¹We will usually not consider mixed states in this thesis. Therefore, where no confusion can arise, we will use the term “state” to refer to state vectors.

This is followed by local individual measurements on the spins. The measurement outcomes at the end then deliver (typically statistical) data, from which the outcome of the computation can be estimated. Importantly, the quantum systems follow the laws of quantum mechanics and can be prepared in a superposition exploring an Hilbert space of exponentially large dimension. This indicated that some problems could be solved with significantly less effort on this envisioned device than on any classical computer. In fact, some of them – such as factoring – which are believed to be computationally hard classically could be solved with only polynomial effort.

2.2 Linear optics

The underlying physical systems affect the way information can be processed and the techniques used to manipulate these systems can be quite different depending on the actual physical systems at hand. Just imagine internal states of trapped atoms compared to states of the light field as the information carrier. Consequently, procedures to implement the required unitary evolutions (*i.e.*, quantum gates) have to be identified separately for each of the different architectures.

The architecture we will focus on uses single photons to carry the information and energy-preserving optical elements (*linear optics*) and detectors to transform quantum states. It therefore allows for relatively cheap experimental setups: linear optical elements (beam splitters) are state-of-the-art devices that can be bought off-the-shelf. The real difficulty in experiments is the alignment of these elements with respect to each other to ensure interference (both, first and second order) of the photons. To overcome these drawbacks, future linear optics circuits will be banned on chips, rather than free-space in the laboratory, thereby ensuring stability [14]. Production of optical chips is already an every-day business in the telecommunication sector.

Capturing the action of beam splitter networks in general is as easy a task as it is to calculate a product of matrices describing the beam splitters. However, it is well known, that auxiliary vacuum modes, additional photons and measurements [18] are inevitable for linear optics to access all unitaries in the usual encodings. Together with the state space structure induced by the chosen encoding of logical qubits into suitable sub-spaces of the physical multi-photon system, the situation now looks quite complicated and offers a plethora of open questions.

In the following paragraphs we will revisit some basic concepts which will be used in the subsequent chapters where the actual problems are tackled.

Modes. One term that will be used quite often within this work is the notion of a *mode*. In the most general case it is defined as an elementary solution of the wave equation of the electromagnetic field with respect to the boundary conditions dictated by the chosen geometry. After quantisation, the excitations of a mode are called *photons*. To characterise an optical mode, all physical degrees of freedom (*i.e.*, all parameters that determine this particular solution) have to be specified. Because linear optics does not depend on any specific degree of freedom (such as polarisation or direction), the physical nature of the modes will be neglected in the theoretical considerations, and only abstract modes labelled by numbers $1, \dots, n$ will be used.

These are usually imagined as spatial modes with the same finite transversal mode profile shifted with respect to each other, subject to mixing on beam splitters.

It follows from this abstract starting point, that the solution obtained with the methods shown will also live in this abstract mode space. This lack of physics within the solution implies that features such as inherent stability due to the use of different polarisations in the same spatial mode have to be incorporated by hand.

Encoding. By $\mathcal{H} = (\mathbb{C}^d)^{\otimes q}$ we will denote the computational Hilbert space of q *qudits* with basis elements $\{|i_1\rangle \otimes \dots \otimes |i_q\rangle : i \in [0, d-1]^{\times q} \subset \mathbb{N}^q\}^2$. Each of these logical states will be implemented by a certain state of the light field. In this work we will focus on product states with respect to the Fock basis, more precisely usually only those with a constant overall photon number for a fixed encoding.

Such a state of N photons in $n \in \mathbb{N}$ modes, the Hilbert space of which will be denoted by \mathcal{H}_N^n , is specified by the photon number in each mode. Collectively they can be written as a pattern $\Omega \in \mathbb{N}^n$, such that the overall photon number satisfies $N = \sum_{i=1}^n \Omega_i$. By \mathbb{N}_N^n we will denote the set of all possible patterns on n modes with N photons, its cardinality being

$$\text{card}\mathbb{N}_N^n = \binom{n}{N} := \binom{N+n-1}{N}. \quad (2.2)$$

The state vector of the final modes corresponding to the pattern Ω will be denoted by $|\Omega\rangle = \otimes_{i=1}^n |\Omega_i\rangle$. We will also use the notation $e_i \in \mathbb{N}$ for the pattern describing exactly one photon in the i -th mode.

Now, an *encoding* is a map $\mathcal{H} \rightarrow \mathcal{H}_N^n$. For each logical basis state $|i\rangle$ we choose a pattern $\Omega^{(i)} \in \mathbb{N}_N^n$. Then, every logical state $|x_1, \dots, x_k\rangle_{\text{logical}}$ can be translated into a number state $\otimes_{i=1}^k \otimes_{j=1}^n |\Omega_j^{(x_i)}\rangle$.

²Where confusion with the physical states of modes might arise we will distinguish the logical qudit states by appropriate indices or Fraktur letters such as $|\circ\rangle$ and $|\mathbf{1}\rangle$, *etc.*

The most common types of encodings are *dual-rail* qudits, so qudits with patterns from $\mathbb{N}_{d-1}^2 = \{(0, d-1), \dots, (d-1, 0)\}$. In contrast to *single-rail* (patterns from $\{(0), \dots, (d-1)\}$), the fixed photon number in dual-rail encoding is tailored to linear optics which is inherently photon number preserving.

Transformation of creation operators. With n being the number of modes under consideration, $|\text{vac}\rangle := |0\rangle^{\otimes n}$ will denote the vacuum of all modes. Let us fix some notation: $n \in \mathbb{N}$ will be used for the number of modes in the linear optics problem we are considering, and $\mathbf{a}^\dagger = (a_1^\dagger, \dots, a_n^\dagger)^T$ the vector of *creation operators* on these modes. The creation operators and their conjugate, the *annihilation operators* a_i , satisfy the usual commutation relations for bosons,

$$[a_i, a_j^\dagger] = \delta_{i,j}, \quad (2.3)$$

and their action on number states (Fock states) is

$$a |m\rangle = \sqrt{m} |m-1\rangle \quad (2.4)$$

$$a^\dagger |m\rangle = \sqrt{m+1} |m+1\rangle. \quad (2.5)$$

Therefore, Fock states in the Hilbert space of N photons on n modes, \mathcal{H}_N^n , can be written as

$$|n_1, \dots, n_n\rangle = \prod_{i=1}^n \frac{(a_i^\dagger)^{n_i}}{\sqrt{n_i!}} |\text{vac}\rangle. \quad (2.6)$$

The term *linear optics* refers to systems described by Hamiltonians which preserve the overall photon number (*i.e.*, passive optics). It turns out [19], that the effect of the evolution is a linear transformation of the creation operators

$$U : \mathbf{a}^\dagger_{\text{in}} \mapsto \mathbf{a}^\dagger_{\text{out}} = U \mathbf{a}^\dagger_{\text{in}} \quad (2.7)$$

described by a unitary matrix $U \in SU(n)$. Corresponding to such a transformation with $U = e^{\imath B}$, its action on the Hilbert space of states is given by the unitary operator $\mathcal{U} = e^{\imath \mathbf{a} B \mathbf{a}^\dagger}$.

Transformations in the smallest non-trivial case, $n = 2$, are physically realised by beam splitters. Depending on the context, different notions of beam splitters are used in the literature. Owing to experimental devices with fixed phase relations, transformations with only one free parameter ϑ are often referred to as beam splitters as well. Here we will, however, use the most general form which can be described

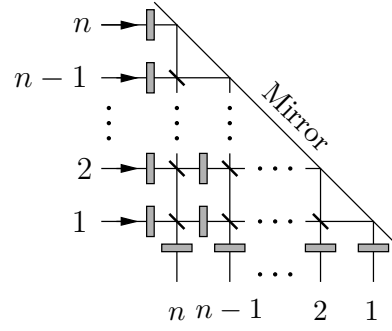


Figure 2.1: Decomposition of an $SU(n)$ unitary in terms of a general network of beam splitters according to Ref. [20]. Here, beam splitters (thick lines) are viewed as one-parameter rotations without phases to avoid double phase shifts in the in- and output ports of subsequent beam splitters. Phase shifts are accomplished by phase plates (grey boxes). Together these elements amount to the $2(n(n-1)/2) + n = n^2$ real parameters of $SU(n)$.

by matrices of the form

$$U_{\text{BS}} = \begin{pmatrix} e^{i(\varphi_1+\varphi_2)} \cos \vartheta & e^{i(-\varphi_1+\varphi_2)} \sin \vartheta \\ e^{i(\varphi_1-\varphi_2)} \sin \vartheta & -e^{i(-\varphi_1-\varphi_2)} \cos \vartheta \end{pmatrix}. \quad (2.8)$$

Not only is $U_{\text{BS}} \in SU(2)$, but any $U \in SU(n)$ can be decomposed into a product of at most $n(n-1)/2$ two-mode beam splitters by means of a Householder-type decomposition [20]. See Fig. 2.1 for more details on this decomposition. This means, the notions of linear optics and beam splitter networks are actually equivalent.

Plugging the linear transformation of the creation operators into Eqn. (2.6), the action of the induced unitary on Hilbert space, \mathcal{U} , can be written as

$$\begin{aligned} \mathcal{U} : \quad |\psi_{\text{in}}\rangle &= |n_1, \dots, n_n\rangle = \prod_{i=1}^n \frac{(a_i^\dagger)^{n_i}}{\sqrt{n_i!}} |\text{vac}\rangle \\ \mapsto |\psi_{\text{out}}\rangle &= \mathcal{U} |\psi_{\text{in}}\rangle = \prod_{i=1}^n \frac{1}{\sqrt{n_i!}} \left(\sum_{j=1}^n U_{i,j} a_j^\dagger \right)^{n_i} |\text{vac}\rangle. \end{aligned} \quad (2.9)$$

We will see, that unitarity of U is not necessary in some situations. Then, the symbols A (for the mode transformation) and \mathcal{A} (for its action on the Hilbert space) will be used to highlight the difference.

State vectors are not the only way of capturing the state of bosonic field modes. A different way of looking at them is the following: Obviously, any state $|\psi\rangle \in \mathcal{H}_N^n$ can be described by a homogeneous polynomial of degree N in the creation operators,

so

$$|\psi\rangle = p_\psi(a_1^\dagger, \dots, a_n^\dagger) |\text{vac}\rangle \quad (2.10)$$

with $p_\psi \in \mathbb{C}[a_1^\dagger, \dots, a_n^\dagger]$ (so a polynomial in the creation operators with complex coefficients). Yet another representation will be introduced in Section 3.2.

Permanents. An elegant tool that we will use is the permanent of a square matrix $A \in \mathbb{C}^{n \times n}$ [21, 22], defined as

$$\text{per} A := \sum_{\pi \in \text{Sym}(n)} \prod_{i=1}^n A_{i, \pi(i)} \quad (2.11)$$

where $\text{Sym}(n)$ is the set of all permutations of n elements. Further, for $\omega, \Omega \in \mathbb{N}_N^n$ we will need the sub-matrix of A obtained by taking Ω_i times the i -th row and ω_j times the j -th column, denoted by $\Lambda = A[\omega|\Omega]$, so³

$$\Lambda_{\alpha, \beta} := A_{k, l} \quad \text{such that} \quad \sum_{i=1}^{k-1} \Omega_i < \alpha \leq \sum_{i=1}^k \Omega_i \quad (2.12)$$

$$\text{and} \quad \sum_{j=1}^{l-1} \omega_j < \beta \leq \sum_{j=1}^l \omega_j.$$

For the sake of readability we will drop brackets where there is no danger of confusion. Similar to determinants, permanents can be expanded in terms of sub-matrices. Keeping the j -th row or column fixed, the respective expansions are

$$\text{per} A[\omega|\Omega] = \sum_i A_{j, i} \text{per} A[\omega - e_i | \Omega - e_j] \quad (2.13)$$

$$= \sum_i A_{i, j} \text{per} A[\omega - e_j | \Omega - e_i]. \quad (2.14)$$

With this notation, matrix elements of \mathcal{U} can be written in terms of matrix elements of U as

$$\langle \omega | \mathcal{U} | \Omega \rangle = \frac{\text{per} U[\omega|\Omega]}{\sqrt{\omega! \Omega!}}, \quad (2.15)$$

using the convention $\omega! := \prod_i \omega_i!$ [22].

An intuitive interpretation of the permanent follows from a stochastic point of view. Considering distinguishable systems instead of photons, each term in the sum (2.11) represents the amplitude of observing a specific output pattern, given a

³Note, that this corresponds to the sub-matrix $A^T[\omega|\Omega] = A[\Omega|\omega]$ with the definition from [22].

certain input pattern. Due to indistinguishability of the photons, there are a number of patterns which are equivalent in the output as well as the input. This is taken care of by effectively summing over all of those pattern. Therefore, one obtains the sum of amplitudes for all combinations of paths the photons can possibly take for given input- and output patterns.

This chapter will give an overview over methods that can be used to construct and analyse beam splitter networks comprised of just a few modes and photons. Although the basic transformations are easy to describe by unitary matrices, analysing their effective action on the quantum states is impeded by the choice of encoding. The size of the state space ($\dim \mathcal{H}_N^n = \text{card} \mathbb{N}_N^n$) grows rapidly in the number of modes as well as the number of photons. Also, the polynomials describing the states exhibit a complicated structure when it comes to several photons in a non-trivial encoding.

Techniques for explicit construction of beam splitter networks for given problems on the computational Hilbert space, based on the polynomial representation of photonic states (related to the ones used in Ref. [31]), will be explained in the first part. More precisely, unitary matrices will be constructed, the decomposition of which into beam splitter networks is already known [20]. This is different to earlier approaches where each problem was solved with individual techniques, rather than systematically. However, in general the translation back into an experimentally feasible network cannot be provided in this abstract setting. What is “feasible” in terms of experiments and what kind of simplifications (*e.g.*, for stabilising interferometers) are available is not included in the algorithm. These questions will have to be addressed on a per-problem basis, as will be demonstrated in the successive examples chapter.

The second part addresses another possible representation of photonic states by means of super-symmetric tensors which might allow to exploit results on higher dimensional tensors. In the case of two photons on four modes this representation was already used in Refs. [35, 36]. Small examples will be considered to find statements on state transformations with linear optical means.

After that, results from algebraic geometry will be applied to the polynomial describing the quantum state, $p_\psi \in \mathbb{C}[a_1^\dagger, \dots, a_n^\dagger]$. Statements on decompositions of a state by means of linear optics from the perspective of polynomial factorisation may help building intuition on the problems relevant in linear optics.

3.1 General network construction

Before we go into detail, we shall sketch the general outline of the scheme [4]. In order to find a network (described by the mode transformation $A \in \mathbb{C}^{n \times n}$) for a given task of state preparation, unitary evolution, or measurement, we proceed in the following steps:

1. Fix some variables $A_{i,j}$ due to known dependences of the probability of success (p_s) on certain gauge transformations (depending on the encoding and the actual problem). This can simplify steps 2–3 considerably.
2. Write down the polynomial equations describing the problem in terms of the $A_{i,j}$.
3. Check for existence of solutions. If none exist, different auxiliary states are necessary. Change the input- and detection patterns to account for them and proceed with step 1. If solutions exist, find the Gröbner basis (explained in more detail in Section 3.1.4).
4. Solve for $A_{i,j}$ by back-substitution.
5. Try to maximise p_s over all (including the previously fixed) variables. This might only succeed if the number of modes and the number of photons are sufficiently small and the set of solutions is well-behaved.
6. Construct the unitary extension and decompose it into $SU(2)$ rotations, the experimental counterpart of which are beam splitters.

3.1.1 Action of linear optics

For finding optical networks for a task that is given on the level of logical qubits in \mathcal{H} , let us recall how to write photonic Fock states after passing through a linear optical device $A \in SU(n+m)$:

$$|\psi^{\text{in}}\rangle = |n_1, \dots, n_{n+m}\rangle = \prod_{i=1}^{n+m} \frac{(a_i^\dagger)^{n_i}}{\sqrt{n_i!}} |\text{vac}\rangle \quad (3.1)$$

$$\mapsto |\psi^{\text{out}}\rangle = \prod_{i=1}^{n+m} \frac{1}{\sqrt{n_i!}} \left(\sum_{j=1}^{n+m} A_{i,j} a_j^\dagger \right)^{n_i} |\text{vac}\rangle. \quad (3.2)$$

Formally this works for any $A \in \mathbb{C}^{(n+m) \times (n+m)}$ and we forget about unitarity in the next sections, showing later how to recover it. The first n modes will be the signal

modes (as used in a gate or measurement), the last m ones the auxiliary modes which are required to implement effective non-linear transformations on the signal modes (the initial state and the post-selection condition being independent of the signal). In the abbreviation $|\psi^{\text{out}}\rangle = F(\mathbf{a}^\dagger)|\text{vac}\rangle$ we use $F \in \mathbb{C}[a_1^\dagger, \dots, a_{n+m}^\dagger]$ for a homogeneous polynomial of degree $N + M = \sum_{i=1}^{n+m} n_i$ in the creation operators $\mathbf{a}^\dagger = (a_1^\dagger, \dots, a_{n+m}^\dagger)^T$. In turn, the coefficients in front of the monomials in F are homogeneous polynomials of degree $N + M$ in the matrix elements of A . In other words, they are elements of the polynomial ring $\mathbb{C}[A_{i,j}]$.

3.1.2 Photon detection

For measurements where the photons come from the input state, or in the cases of state preparation or gates which require auxiliary photons¹, the state of some of the output modes will be conditioned on containing a certain number of photons.

Assuming m photon detectors placed in the output modes $n+1, \dots, n+m$ report the detection of m_1, \dots, m_m photons, respectively, the resulting state is given by

$$|\phi\rangle = \langle m_1, \dots, m_m | \psi^{\text{out}} \rangle = \langle \text{vac} | \left(H(a_{n+1}^\dagger, \dots, a_{n+m}^\dagger) \right)^\dagger |\psi^{\text{out}}\rangle = G(\mathbf{a}^\dagger) |\text{vac}\rangle. \quad (3.3)$$

Here $H \in \mathbb{C}[a_{n+1}^\dagger, \dots, a_{n+m}^\dagger]$ is the analogue of F for the target state – a homogeneous polynomial of degree $M := \sum_{i=n+1}^{n+m} n_i$ used to construct the desired state by $H(a_{n+1}^\dagger, \dots, a_{n+m}^\dagger)|\text{vac}\rangle$. Further, G is a homogeneous polynomial of degree $N := \sum_{i=1}^n n_i$ in the creation operators remaining after the detection, $a_1^\dagger, \dots, a_n^\dagger$, again with coefficients being homogeneous polynomials of degree $N + M$ belonging to $\mathbb{C}[A_{i,j}]$. By only considering $a_{n+1}^\dagger, \dots, a_{n+m}^\dagger$ as variables, G can be obtained as the coefficient of F in front of H ,

$$G = \text{Coeff}[F, H] = \left(\prod_{i=1}^m m_i! \right)^{1/2} \text{Coeff} \left[F, \left(a_{n+1}^\dagger \right)^{m_1} \cdots \left(a_{n+m}^\dagger \right)^{m_m} \right]. \quad (3.4)$$

Note, that $G(\mathbf{a}^\dagger)|\text{vac}\rangle$ is not necessarily normalised. Because $\langle \phi | \phi \rangle$ in fact gives the probability of detecting m_i photons in the i -th detector, the resulting state will only be normalised in the trivial case of this detector event being the only possible one.

¹These cases can be handled in the same way as measurements by incorporating the auxiliary photons into the input state $|\psi^{\text{in}}\rangle$.

3.1.3 Objective state

By demanding the final state to be equal – up to global scaling by $\gamma \in \mathbb{R}^+$ corresponding to a lowering of the success probability of the process – to a given target state $|\psi^{\text{tar}}\rangle = Q(\mathbf{a}^\dagger)|\text{vac}\rangle$, we obtain the constraint

$$G - \gamma Q = 0 \quad (3.5)$$

for the optical network A to produce the target state, given the input state $|\psi^{\text{in}}\rangle$ with success probability $p_s = |\gamma|^2$. γ will be used as an additional variable and maximised in the last step. The expression in (3.5) has to vanish for every monomial of creation operators independently. Therefore, by comparison of coefficients, Eqn. (3.5) can be read as a system of polynomial equations in the $A_{i,j}$,

$$P_k(\{A_{i,j}\}, \gamma) = 0, \quad (3.6)$$

where $k = 1, \dots, \binom{N+n-1}{N}$ labels patterns $\omega^{(k)} \in \mathbb{N}_N^n$ of N photons in the n signal modes (the physical basis states).

In case of a state preparator, Eqn. (3.5) is the defining equation for A . If the device should realise a quantum gate, the equation has to be specified for each state of the input basis separately, so

$$G_l - \gamma Q_l = 0 \quad (3.7)$$

with $l = 1, \dots, d^a$ enumerating the logical basis states of the input, $|\psi^{\text{in},(l)}\rangle = |\Omega^{(l)}\rangle|\delta\rangle$, $\Omega^{(l)} \in \mathbb{N}_N^n$. Independent of the signal state, δ is the pattern of photons in the auxiliary modes, the state of which will be measured later and used for the purpose of post-selection. For an event-ready quantum gate, k has to run over all possible patterns and l over all allowed input patterns (the logical basis states), but different notions are possible².

² Recent experiments considered so-called *post-selected gates* [23, 24, 25], which refers to post-selection of the signal modes in the output onto the logical sub-space. Although this post-selection is not easily implemented with linear optics (after all, photo-detection destroys the quantum system under observation), it allows for simpler gates to start with (we will consider this type of gates as examples in Section 4.3).

By post-selecting, l only has to enumerate the logical basis states, which results in fewer constraints. The consequence is a set of solutions to problems, where the full gate does not have linear optics solutions without using auxiliary photons (*e.g.*, the controlled phase gate), thus allowing for easier theoretical description and optimisation, and simpler experiments with fewer modes and photons.

This leads to $P_{k,l}(\{A_{i,j}\}, \gamma) = 0$, $P_{k,l} \in \mathbb{C}[\{A_{i,j}\}, \gamma]$, which can be translated to a variety of equivalent forms,

$$P_{k,l}(\{A_{i,j}\}, \gamma) = \text{Coeff} [G_l(\mathbf{a}^\dagger) - \gamma Q_l, p_{\omega^{(k)}}(\mathbf{a}^\dagger)] \quad (3.8)$$

$$= \langle \omega^{(k)} | \langle m_1, \dots, m_m | \mathcal{A} | \Omega^{(l)} \rangle | \delta \rangle - \gamma \langle \omega^{(k)} | \psi^{\text{tar},(l)} \rangle \quad (3.9)$$

$$= \frac{\text{per} A[\omega^{(k)}, m_1, \dots, m_m | \Omega^{(l)}, \delta]}{\sqrt{\omega^{(k)}! \Omega^{(l)}! \delta! \prod_i m_i!}} - \gamma \langle \omega^{(k)} | \psi^{\text{tar},(l)} \rangle \quad (3.10)$$

$$= 0 \quad (3.11)$$

where \mathcal{A} is the unitary on Hilbert space corresponding to the (not necessarily unitary) mode transformation A .

Furthermore, an optical projective measurement device can be seen as a gate followed by projective photon number measurements. Thus, a network implementing a measurement can be described in the same way as a set of polynomial constraints $P_{k,l}(\{A_{i,j}\}, \gamma) = 0$.

If the algorithm to be mentioned in the following section does not return a solution, then different patterns of, or more auxiliary photons have to be included in order for the device to work. Additional photons are taken care of by using modified auxiliary states $|\delta'\rangle$ and modified detection patterns described by m'_i . Note, that by increasing the numbers of modes, the size of A will also increase.

3.1.4 Solving polynomial equations

Interlude: Introduction to Gröbner bases

In order to solve a set of polynomial equations $\{f_1 = 0, \dots, f_s = 0\}$ (with the elements of $f = \{f_1, \dots, f_s\}$ being elements of the polynomial ring over \mathbb{C} with variables x_1, \dots, x_n , $\mathbb{C}[x_1, \dots, x_n]$) we consider the *ideal* I generated by these polynomials,

$$I = \langle f_1, \dots, f_s \rangle := \left\{ \sum_{i=1}^s h_i f_i : h_i \in \mathbb{C}[x_1, \dots, x_n] \right\}. \quad (3.12)$$

Between the mutual set of zeroes of the f_i (also called the *algebraic variety*) and I there exists a kind duality such that we can investigate generating sets of I instead of the solutions of the original polynomials. Among generators of I there exist some with rather special properties that prove suitable for finding the solutions.

A set $g = \{g_1, \dots, g_h\} \subset \mathbb{C}[x_1, \dots, x_n]$ will be called a *Gröbner basis* [26] of I , iff multivariate polynomial division of any $f \in I$ does not leave a remainder. By this definition the first important property is given: when a Gröbner basis for I is

given, the question whether a given polynomial $f \in \mathbb{C}[x_1, \dots, x_n]$ belongs to I can be answered by looking at the result of division of f by g . Algorithms of finding a Gröbner basis of $\langle f_1, \dots, f_s \rangle$ are known, for example *Buchberger's algorithm*. Starting with f as the basis, it proceeds by constructing suitable combinations of pairs of polynomials f_k and f_l , $S(f_k, f_l) \in I$ and adding the remainder that is obtained upon division by f to the basis. This step is iterated until the remainder is zero; the defining condition is satisfied. If the system does not have a solution (there is no common divider), the Gröbner basis consists of only the constant polynomial, $g = \{1\}$.

Loosely speaking, this algorithm is a generalisation of the *Gaussian elimination* known for linear equations. Just like in the linear case, the Gröbner basis has another important property that we will use in the following. There exists a hierarchy in the dependence on the x_i which allows for consecutive elimination of variables. The variables a polynomial g_k depends on are a subset of the variables g_l depends on for all $k < l$. This feature allows for solving the system by back-substitution and solving only one polynomial equation at a time, which is the analogue of the back-substitution performed in Gaussian elimination.

There exist implementations in many computer algebra systems including **Mathematica**, **Singular**, and **Magma**³.

Application to the linear optics problem

Solving (3.7) for A can be quite a demanding task. Moreover, optimising the success probability, *i.e.*, finding A such that γ is the maximum one fulfilling Eqn. (3.7), can quickly become a hopeless enterprise. However, for certain examples of small networks this technique can be applied to solve the problem on a desktop computer in a reasonable amount of time.

For this purpose one calculates the Gröbner basis of the ideal $\langle P_{k,l} \rangle$ generated by the polynomials $\{P_{k,l}\} \subset \mathbb{C}[A_{i,j}]$. This will be another set of polynomials $g = \{g_1, \dots, g_h\} \subset \mathbb{C}[A_{i,j}]$ with the same set of roots and the special properties mentioned above.

Note, that a symbolic back-substitution will only be feasible for Gröbner bases of small degree, and numerical methods must provide a very high accuracy due to the subsequent dependence of later polynomials on earlier solutions. As pointed out in Ref. [27], there are difficult cases where numerical back-substitution is very instable. Still, for certain instances of small (say, two-qubit) gates, this is a doable task.

³ <http://magma.maths.usyd.edu.au/magma>

That this procedure can run into problems already with a few photons on only a couple of modes is not very surprising. Although many questions regarding the complexity of constructing Gröbner bases are still not answered, it is known [28] that this problem is $\mathcal{EXPSPACE}$ -complete. The best known algorithm [29] uses exponential space in the size of the input problem. In fact, the degrees of the polynomials in the Gröbner basis of a problem of N photons in n modes can be bounded from above by

$$2 \left(\frac{N^2}{2} + N \right)^{2^{n-1}}, \quad (3.13)$$

which depends doubly exponentially (can be thought of as doubly exponential running time with exponential work space) on the number of modes, but only polynomially on the number of photons. For general ideals the scaling cannot be improved because there exist sets of polynomials the Gröbner bases of which consist of a doubly exponential number of elements of doubly exponential degree. One can hope that the special structure (*e.g.*, the encoding) in this problem can help to reduce the complexity. However, it is known that the homogeneity of the polynomials cannot lift the complexity from $\mathcal{EXPSPACE}$.

For a given instance, the *term order* also plays a crucial role. This is an artificial and arbitrarily chosen ordering of monomials which has a huge impact on the running time and the resulting polynomials.

3.1.5 Manifold of solutions

Apart from no solution at all, the character of the set of solutions \mathfrak{A} can vary a lot: discrete as well as continuous sets of solutions might occur. Thus, in the worst case, the solution (and also the success probability) depends on some free parameters, in general in a non-polynomial way. The set of solutions is discrete, iff the variety has dimension zero. This property can be determined by a \mathcal{PSPACE} algorithm [30]. Therefore, whether the set of polynomials (3.7) has no solutions at all, a discrete, or a continuous set of roots, can be decided more efficiently than calculating the whole Gröbner basis itself.

Some of the parameters encountered in continuous sets of solutions depend on the encoding: As an example, consider a two-qubit gate in dual-rail encoding. The relevant states are of the form $|n_1, n_2, n_3, n_4\rangle$. Due to the photon-number constraints $n_1 + n_2 = n_3 + n_4 = 1$, rescaling the amplitudes of the input modes by $X = \text{diag}\{x_1, x_1, x_2, x_2\}$ changes the success probability according to $p_s \mapsto |x_1 x_2|^2 p_s$. For the same reasoning, a rescaling of the output by $Y = \text{diag}\{y_1, y_1, y_2, y_2\}$ results

in a similar rescaling of p_s . Because unitary diagonal transformations can be implemented by means of phase shifters deterministically, the x_i and y_i can be assumed to be real and positive.

For any gate acting on a fixed number of photons, these transformations especially include a global rescaling by a number $\beta \in \mathbb{R}^+$, changing the success probability according to $p_s \mapsto |\beta|^{2N} p_s$.

The parameters x_i and y_i will be fixed in the very beginning, decreasing the number of variables and the degree of the polynomials, thus simplifying the steps before considerably. Because the dependence of p_s on these parameters is known, their effect can be accounted for in the following optimisation step anyway.

3.1.6 Unitary extensions

In this section we will focus on unitary implementations of non-unitary mode transformations by addition of vacuum modes (compared to implementing non-linear transformations by adding auxiliary photons). To avoid too many symbols, n will be used for the total number of modes from the previous steps, rather than solely signal modes. ω and Ω are defined accordingly.

Assuming that all $A_{i,j}$ have been determined, the whole matrix is not necessarily unitary. The constraints imposed by the action of the gate do not include unitarity of the mode transformation. Moreover, unitarity actually cannot be encoded in terms of polynomial constraints over \mathbb{C} which is the condition for the Gröbner basis technique to be applicable to this problem.

Still, there is a way to use the matrix found in the previous step to construct a linear optics network with the desired effect. Whenever a matrix element of the form $\langle \omega | \mathcal{A} | \Omega \rangle$ enters the calculation, it can be written in terms of permanents [22] as

$$\langle \omega | \mathcal{A} | \Omega \rangle = \text{per} A[\omega | \Omega]. \quad (3.14)$$

Note, that embedding A as a block inside another matrix U does not influence the matrix element for the same input and detection patterns:

$$\text{per} A[\omega'' | \Omega''] = \text{per} U[\omega | \Omega] \quad \text{for} \quad \begin{aligned} U &= \left(\begin{array}{c|c} A & \cdot \\ \cdot & \cdot \end{array} \right), \\ \omega'' &= (\omega, 0, \dots, 0), \\ \Omega'' &= (\Omega, 0, \dots, 0). \end{aligned} \quad (3.15)$$

Because these new patterns can be realised by adding vacuum modes to the input and conditioning on vacuum in the output, this implies that a bigger network U – which can be chosen to be unitary – can be used to achieve the effect of the non-unitary mode transformation A^4 . Vacuum projections lead to non-unitary mode transformations ($A \in GL(n)$ or even $A \in \mathbb{C}^{n \times n}$ instead of $A \in SU(n)$) similar to the effect of effective non-linear mode transformations when using photon detection.

Lemma 1 (Unitary extension). *For any matrix $A \in \mathbb{C}^{n \times n}$, with its largest singular value $\sigma_1 \leq 1$, there exists an extension*

$$U = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \in SU(N) \quad (3.16)$$

such that $N \leq 2n$.

Proof. Let $A = V\sigma W^\dagger$, $\sigma = \text{diag}(\sigma_1, \dots, \sigma_n)$ with $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ be a singular value decomposition (SVD) [32] of A .

Due to the assumption, all singular values of A lie between 0 and 1. Let ν be the multiplicity of the largest singular value, or $\nu = 0$ for $\sigma_1 = 0$. Then, there exist 2×2 matrices $U^{(\nu+1)}, \dots, U^{(n)} \in SU(2)$ such that $U_{1,1}^{(i)} = \sigma_i$. Let us set

$$U = (V \oplus \mathbb{1}_{n-\nu}) \left(\mathbb{1}_\nu \oplus \bigoplus_{i=\nu+1}^n U^{(i)} \right) (W^\dagger \oplus \mathbb{1}_{n-\nu}) \quad (3.17)$$

with the direct sum arranged such that

$$\left(\mathbb{1}_\nu \oplus \bigoplus_{i=\nu+1}^n U^{(i)} \right)_{kk} = \sigma_k \quad \text{for } k = 1, \dots, n. \quad (3.18)$$

U is unitary by construction and fulfils the embedding criterion stated in the lemma. Also, $N = 2n - \nu$. \square

In the case that the singular values of A exceed 1, we have to make use of the gauge freedom mentioned earlier. By rescaling the whole matrix in a suitable way, the singular values are ensured to be at most 1. For a gate with an encoding with a constant number of photons, a possible rescaling reads $A \mapsto \alpha A$ with

$$\alpha = \begin{cases} 1 & \text{for } \sigma_1 = 0 \\ \sigma_1^{-1} & \text{otherwise.} \end{cases} \quad (3.19)$$

⁴This idea was also sketched in Ref. [31]

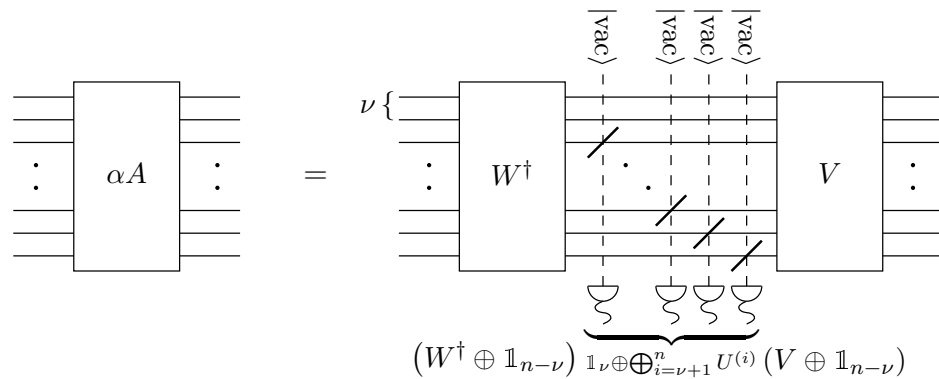


Figure 3.1: Construction of a linear mode transformation αA , with singular values bounded from above by 1, which is not unitary by using linear optics and vacuum extensions. V and W^\dagger are the unitaries obtained from the SVD of αA , ν is the multiplicity of the largest singular value, and the transmittivity of the k -th beam splitter is given by the singular value $\sigma_{\nu+k}$. A was applied successfully if all detectors signalled “no photon”.

More generally one could use $A \mapsto YAX$ with diagonal matrices X and Y describing the rescaling of input and output modes, respectively. Further, all parameters left after solving the polynomial equations will have an impact on the singular values.

The procedure can be easily generalised to rectangular matrices $A \in \mathbb{C}^{m \times n}$ by making the diagonal matrix at the core of the SVD square. This is achieved by filling it up with zeroes, which corresponds to coupling in $m - n$ additional vacuum modes ($m > n$), or post-selecting $n - m$ modes in the vacuum state ($m < n$).

In terms of linear optics there can be given a constructive interpretation of Lemma 1. The unitaries V and W^\dagger can already be implemented as linear optics networks. Further, each of the $U^{(i)}$ can be interpreted as a single beam splitter put into mode i , coupling it with mode $n + i - \nu$ which is initialised in the vacuum and post-selected in the vacuum state afterwards. All modes $i \leq \nu$ are left unchanged between application of V and W^\dagger . See Fig. 3.1 for more details.

While this result gives rise to a linear optics implementation, it also shows that more vacuum modes will not improve the probability of success. A more precise justification of this statement is postponed to the next section.

3.1.7 Success probability

To see how the success probabilities are affected by the rescaling mentioned above, assume they were fixed in the beginning such that $|\langle \psi^{\text{out}} | \psi^{\text{tar}} \rangle|^2 = p_s$.

Then, the solution of the polynomial equations (3.7) led to $n \times n$ matrices $A \in \mathfrak{A}$, which are still subject to rescaling as $A \mapsto YAX$. Due to Lemma 1 there always

exists a physical solution $U \in SU(2n)$ as long as the rescaled singular values are smaller than or equal to 1. Let $\Omega = (n_1, \dots, n_n)$ be a pattern of photons in the input modes, corresponding to the state $|n_1, \dots, n_n\rangle$. The probability of detecting the pattern $\omega = (m_1, \dots, m_n)$ in the output modes is then given by

$$p(\omega|\Omega) = |\langle m_1, \dots, m_n | \mathcal{U} | n_1, \dots, n_n \rangle|^2 \quad (3.20)$$

$$= \left(\prod_{i=1}^n n_i! \right)^{-1} \left(\prod_{i=1}^n m_i! \right)^{-1} |\text{per} U[\omega|\Omega]|^2. \quad (3.21)$$

By construction, the $n \times n$ block of U describing the non-trivial modes is YAX with $X = \text{diag}\{x_1, \dots, x_n\}$ and $Y = \text{diag}\{y_1, \dots, y_n\}$. Then the permanents can be written as

$$\text{per} U[\omega|\Omega] = \text{per} A[\omega|\Omega] \prod_{i=1}^n x_i^{n_i} \prod_{i=1}^n y_i^{m_i}. \quad (3.22)$$

Because all coefficients in the output state can be written as permanents in this form, the overall success probability will satisfy

$$p_s \mapsto p_s \left| \prod_{i=1}^n x_i^{n_i} \prod_{i=1}^n y_i^{m_i} \right|^2. \quad (3.23)$$

After suitable row- and column-wise rescaling, A is unitarily extendable. To maximise the probability of success one has to find the set of parameters maximising p_s subject to the constraint $\sigma_1 \leq 1$. For this being a necessary condition for a linear optical embedding to exist at all, and a sufficient condition for the construction above, more than n vacuum modes do not help to increase the success probability: Given a matrix $A \in \mathbb{C}^{n \times n}$ with sufficiently small singular values $\sigma_1(A) \leq 1$, the probability of success is already fixed and does not depend on the specific unitary extension, one of which with n vacuum modes is always possible. Thus, the last step would be to

$$\text{maximise} \quad \left| \prod_{i=1}^n x_i^{n_i} \prod_{i=1}^n y_i^{m_i} \right|^2 \quad (3.24)$$

$$\text{subject to} \quad YAX \leq \mathbb{1}. \quad (3.25)$$

When the dimension of the set of solutions after fixing Y and X in the beginning is non-zero, the constraint will not only depend on the parameters x_i and y_i . Via A other parameters will also have an influence. Only when the variety's dimension is zero, a discrete set of A -matrices has to be considered.

To give an explicit analytic solution to the optimisation problem is not easy in general. There are, however, instances when tools from semidefinite programming can be used to find the maximum. Let us highlight some cases. Firstly, there should only be a discrete set of solutions to the polynomial equations (up to application of X and Y). Secondly, due to the structure of the problem, only either row- or column-wise rescaling (either Y or X) should be allowed (in the following we assume $Y = \mathbb{1}$). Then, the optimisation can be written as

$$\begin{aligned} & \text{maximise} && p_s && (3.26) \\ & \text{subject to} && AX \leq \mathbb{1}. \end{aligned}$$

By using Schur complements the constraint can be written in the equivalent form

$$\begin{pmatrix} \mathbb{1} & X \\ X & (A^\dagger A)^{-1} \end{pmatrix} \geq 0 \quad (3.27)$$

as a semidefinite constraint. Still, the objective function

$$p_s = \prod_{i=1}^n x_i^{2n_i} \quad (3.28)$$

is non-linear. We introduce a hierarchy of additional variables $x_k^{(j)}$ and semidefinite constraints in order to write this as a convex optimisation problem. Let s be the smallest integer such that $2^s \geq N = \sum_i n_i$, and $\nu_k := \sum_{i < k} n_i$. We set $x_{\nu_k+1}^{(1)} = \dots = x_{\nu_{(k+1)}}^{(1)} = x_k$ for $k = 1, \dots, n$, $x_{N+1}^{(1)} = \dots = x_{2^s}^{(1)} = 1$, and

$$x_{k/2}^{(j+1)} = \sqrt{x_{k-1}^{(j)} x_k^{(j)}} \quad (3.29)$$

for $k = 2, 4, 6, \dots, 2^{s-j+1}$ and $j = 1, \dots, s$. The relation between the $x^{(j)}$ and $x^{(j+1)}$ will be relaxed to

$$x_{k-1}^{(j)} x_k^{(j)} \geq \left(x_{k/2}^{(j+1)}\right)^2, \quad (3.30)$$

but at the maximum, equality will hold. At the end of the hierarchy, we have $x_1^{(s)} = \sqrt{p_s}$. The full optimisation problem now reads

$$\begin{aligned}
 & \text{maximise} && x_1^{(s)} && (3.31) \\
 & \text{subject to} && \begin{pmatrix} \mathbb{1} & X \\ X & (A^\dagger A)^{-1} \end{pmatrix} \geq 0 \\
 & && \begin{pmatrix} 1 & x_k^{(1)} \\ x_k^{(1)} & 1 \end{pmatrix} \geq 0 && \text{and} \\
 & && \begin{pmatrix} x_k^{(1)} & 1 \\ 1 & x_k^{(1)} \end{pmatrix} \geq 0 && \text{for } k = N + 1, \dots, 2^s \\
 & && \begin{pmatrix} x_{k-1}^{(j)} & x_{k/2}^{(j+1)} \\ x_{k/2}^{(j+1)} & x_k^{(j)} \end{pmatrix} \geq 0 && \text{for } k = 2, 4, 6, \dots, 2^{s-j+1} \\
 & && && \text{and } j = 1, \dots, s.
 \end{aligned}$$

It can be solved in an efficient way using methods from convex optimisation [33, 34], which are available, for example, in form of the **SeDuMi**⁵ package for **MatLab**⁶.

Another instance which can be handled analytically is the case $n \leq 4$. Then, the singular values can be given in a closed form as solutions to a fourth-order polynomial equation. Therefore, the constraints $\sigma_i \leq 1$ can be written using only polynomials in the optimisation variables. Solutions to polynomial constraints can be found by using a variety of algorithms (not necessarily efficient). Solvers are available, *e.g.*, **Mathematica** can compute the exact extremum when polynomial constraints are used. Also, polynomial relaxation techniques can lead to efficiently solvable convex optimisation problems which bound the original one.

To illustrate this method, let us consider the smallest case, so $n = 2$. The singular values of A (where A may include rescaling parameters already) are

$$2\sigma_{\pm}^2 = \|A\|_2^2 \pm \sqrt{\|A\|_2^4 - 4|\det A|^2}. \tag{3.32}$$

Because $\sigma_+ \geq \sigma_-$, the list of constraints shortens and the full optimisation problem can be cast into the form

$$\begin{aligned}
 & \text{maximise} && p_s && (3.33) \\
 & \text{subject to} && \|A\|_2^2 - 1 - |\det A|^2 \leq 0 \\
 & && \|A\|_2^2 - 2 \leq 0,
 \end{aligned}$$

⁵ <http://sedumi.mcmaster.ca/>

⁶ <http://www.mathworks.com/products/matlab>

where p_s is a monomial in the rescaling parameters (see Eqn. (3.23)).

Another method for arbitrary n which will result in a set of polynomial constraints will be shown in Section 4.3.5. This type of problems can either be solved with a lot of effort exactly, or relaxations can be applied to cast the polynomials in a form similar to the one sketched above, yielding bounds to the solutions.

For larger problems in general only approximations can be given with these techniques. The challenge in solving the optimisation problem is to bound the singular values of the matrix product YAX . Some crude bounds are known, summarised in Appendix B. In contrast to the singular values, however, some of these bounds are invariant under permutation of the matrices. Therefore, with these approximations, tight bounds can only be expected in trivial cases (*e.g.*, global rescaling).

3.2 State transformations

In experiments the sources for multi-photon states can usually produce only a restricted set of states. When quite different ones shall be used in the experiment, simple state transformation circuits rather than full quantum gates are interesting.

The corresponding theoretical problem is to find a circuit for transformation of a given input state to a given output state. Compared to a full quantum gate, the constraints are much weaker, giving hope to find optical networks within experimental means.

3.2.1 A convenient way of specifying bosonic state vectors

Any $|\psi\rangle$ state of k photons on n modes can be written as a homogeneous polynomial $p_\psi \in \mathbb{C}[a_i^\dagger]$ of degree k in the creation operators

$$|\psi\rangle = P(\mathbf{a}^\dagger) |\text{vac}\rangle. \quad (3.34)$$

Now, any homogeneous polynomial P can be written in terms of a symmetric tensor (*state tensor*, or *state matrix* for $k = 2$) of order k ,

$$P(\mathbf{a}^\dagger) = S^{i_1, \dots, i_k} a_{i_1}^\dagger \dots a_{i_k}^\dagger, \quad (3.35)$$

where the summation over i_1, \dots, i_k is taken implicitly from 1 to n . Symmetry here means the invariance of S under all permutations of the indices, $S_{i_1, \dots, i_k} = S_{\pi(i_1), \dots, \pi(i_k)}$ for all $i_1, \dots, i_k = 1, \dots, n$ and $\pi \in \text{Sym}(k)$. It represents the indistinguishability of bosons in this language.

3.2.2 $k = 2$

In the special case of two photons, S is a matrix and the polynomial can be expressed as

$$P(\mathbf{a}^\dagger) = (\mathbf{a}^\dagger)^T S \mathbf{a}^\dagger, \quad (3.36)$$

and as

$$P(U\mathbf{a}^\dagger) = (U\mathbf{a}^\dagger)^T S (U\mathbf{a}^\dagger), \quad (3.37)$$

after passing through a linear optics network the action on the modes of which is represented by the $n \times n$ matrix $U = (U^\dagger)^{-1}$. As a special case of the singular value decomposition, the Takagi factorisation [32] ensures that any symmetric matrix S can be written as $S = U^T \Sigma U$ with Σ diagonal, and positive semi-definite. The diagonal is composed of the singular values $2^{-1/2} \geq \sigma_1 \geq \dots \geq \sigma_n \geq 0$ (the non-negative square roots of the eigenvalues of SS^\dagger) of S . Therefore, any state of two photons can be transformed to the diagonal form

$$|\psi\rangle \mapsto \sum_{i=1}^n \frac{\sigma_i}{\sqrt{2}} (a_i^\dagger)^2 |\text{vac}\rangle \quad (3.38)$$

by means of linear optics⁷. This diagonal form can be seen as a Schmidt-decomposition of the state of modes. This becomes more clear when thinking of the photons as distinguishable quzits the state of which is the mode they occupy. Indistinguishability is then ensured by the symmetry of S . The rank of S , $\nu := \text{rank } S$ is the number of non-zero singular values, the Schmidt-rank. In the following we will use the term *state* to refer either to S , or the state vector $|\psi(S)\rangle$ induced by (3.35) and (3.34). Normalisation of the state vector translates into $\text{tr } \Sigma^2 = 1/2$.

3.2.3 State transformation

In this language the optimal state transformations for two-qubit states can be found:

Lemma 2 (Optimal state transformation). *Let $S^{(1)}$ and $S^{(2)}$ be two bi-photon states on n modes. Then, $S^{(1)}$ can be transformed into $S^{(2)}$ with linear optics, vacuum detectors and empty ancilla modes if and only if $\nu^{(1)} \geq \nu^{(2)}$. The optimal success probability of such a transformation is given by*

$$p_s(S^{(2)}|S^{(1)}) = \mu^2 \quad (3.39)$$

⁷The $k = 2$, $n = 4$ case was used already in Refs. [35, 36]. Thanks to Dmitri Uskov for pointing out the nice structure of this.

with

$$\mu := \max_{\pi \in \text{Sym}(n)} \min \left\{ 1, \sigma_{\pi(1)}^{(1)}/\sigma_1^{(2)}, \dots, \sigma_{\pi(\nu^{(1)})}^{(1)}/\sigma_{\nu^{(1)}}^{(2)} \right\}. \quad (3.40)$$

Proof. Let $\Sigma^{(1)} = (U^{(1)})^T S^{(1)} U$ and $\Sigma^{(2)} = (U^{(2)})^T S^{(2)} U$ be the diagonal forms of the two bi-photon states. Note, that the diagonalisations – the unitaries $U^{(1)}$ and $U^{(2)}$ – can be carried out deterministically with linear optics alone. Now, two diagonal matrices are T -congruent with the congruence matrix being a product of a diagonal matrix (phases can be neglected due to $\Sigma^{(1)}$ and $\Sigma^{(2)}$ being real and non-negative) and a permutation.

Linear optics and vacuum extensions allow exactly for matrices with singular values smaller than or equal to 1. Therefore, a necessary condition for such a transformation is $\nu^{(1)} \geq \nu^{(2)}$ (to increase the rank, non-linearities induced by ancilla photons are required). If this is fulfilled one still has to guarantee that the singular values do not increase during the transformation. Let $\pi \in \text{Sym}(n)$, so a permutation of the n input modes (which can be implemented deterministically), and denote by π^* one that achieves the maximum in Eqn. (3.40). Now we can rescale $\sigma^{(2)}$ by μ such that

$$\sigma_{\pi^*(i)}^{(1)} \geq \mu \sigma_i^{(2)}, \quad (3.41)$$

which ensures the correct relation between the singular values of the in- and the output state. Now, $\Sigma^{(1)}$ can be transformed into $\Sigma^{(2)}$ by using the mode transformation described by the $n \times n$ matrix

$$D = \mu^{1/2} \text{diag} \left\{ \left(\sigma_1^{(2)}/\sigma_{\pi^*(1)}^{(1)} \right)^{1/2}, \dots, \left(\sigma_{\nu^{(1)}}^{(2)}/\sigma_{\pi^*(\nu^{(1)})}^{(1)} \right)^{1/2}, 1, \dots, 1 \right\}, \quad (3.42)$$

so an attenuation achieved by coupling each mode separately to a vacuum ancilla. Note, that an attenuation is also possible if a $\sigma_k^{(2)}$ vanishes: the k -th mode will simply be replaced by the vacuum ancilla using a beam splitter with unit reflectivity.

This gives rise to a well-defined version of D on the first $\nu^{(1)}$ modes. An attenuation on the other modes does not influence the output state because there is no contribution of the input state to these modes. The success probability of the whole transformation can be found by normalising the output state,

$$p_s(S^{(2)}|S^{(1)}) = 2 \text{tr} (\mu \Sigma^{(2)})^2 = \mu^2. \quad (3.43)$$

It is the optimal one in this setting due to the maximisation over all possible mode permutations in Eqn. (3.40). \square

3.2.4 Classes of two-photon states in dual-rail encoding

Let us fix $n = 4$, as this is the usual case used for dual-rail encoding. Then there are four equivalence classes, $\nu = 1, \dots, 4$ inside which states can be transformed into each other with non-zero success probability. These include

- a Fock-2-state ($\nu = 1$),
- two modes, each of which is occupied with a single photon ($\nu = 2$),
- the superposition of two photons being in either of three modes ($\nu = 3$), and
- an EPR pair in dual-rail encoding ($\nu = 4$).

The dual-rail states are of particular interest due to this type of encoding being especially suited for experiments. They can be written as

$$S_{\text{dr}} = \begin{pmatrix} & a & b \\ & c & d \\ a & c & \\ b & d & \end{pmatrix} \quad (3.44)$$

with $|a|^2 + |b|^2 + |c|^2 + |d|^2 = 1/4$. $\det S_{\text{dr}} = (ad - bc)^2$, where a, b, c , and d are the coefficients in front of the dual-rail basis states divided by 2. These states include two classes: $\nu = 2$ ($\det S_{\text{dr}} = 0$) and $\nu = 4$. Singular values of S_{dr} come in pairs of

$$\sigma_{\pm} = \frac{1}{\sqrt{8}} \sqrt{1 \pm \sqrt{1 - 64|\det S_{\text{dr}}|^2}}. \quad (3.45)$$

Hence, the optimal success probability in this framework of a transformation that does not increase the rank (so $\nu^{(1)} \geq \nu^{(2)}$) reads

$$p_{\text{s}} \left(S_{\text{dr}}^{(2)} | S_{\text{dr}}^{(1)} \right) = \begin{cases} \left(2\sigma_{+}^{(1)} \right)^2 & \det S_{\text{dr}}^{(2)} = 0 \quad (\nu^{(2)} = 2) \\ \left(\sigma_{+}^{(1)} / \sigma_{+}^{(2)} \right)^2 & 0 < |\det S_{\text{dr}}^{(2)}| < |\det S_{\text{dr}}^{(1)}| \\ \left(\sigma_{-}^{(1)} / \sigma_{-}^{(2)} \right)^2 & \text{else.} \end{cases} \quad (3.46)$$

3.2.5 Increasing the rank

The construction to transform a state into another one shown in Section 3.2.3 is possible iff the rank does not increase. In order to achieve this, ancilla photons are required. What are necessary and sufficient conditions on the photon numbers? Let

us consider a state S with $\nu < n$ in diagonal form:

$$S = \begin{pmatrix} \ddots & & & \\ & \sigma_\nu & & \\ & & 0 & \\ & & & \ddots \end{pmatrix}. \quad (3.47)$$

Applying a 50 : 50 beam splitter to the mode pair $(\nu, \nu + 1)$ results in

$$S' = \begin{pmatrix} \ddots & & & \\ & \sigma_\nu/2 & \sigma_\nu/2 & \\ & \sigma_\nu/2 & \sigma_\nu/2 & \\ & & & \ddots \end{pmatrix}. \quad (3.48)$$

In a further step, an NLS gate [37] on mode ν delivers

$$S' = \begin{pmatrix} \ddots & & & \\ & -\sigma_\nu/2 & \sigma_\nu/2 & \\ & \sigma_\nu/2 & \sigma_\nu/2 & \\ & & & \ddots \end{pmatrix}, \quad (3.49)$$

which has rank $\nu + 1$ and singular values $\sigma'_\nu = \sigma'_{\nu+1} = \sqrt{2}\sigma_\nu$. With this procedure the rank can be increased by 1 using one additional photon. We conjecture that this is also the maximum amount the rank can be increased by when using 1 photon: it can be shown that in order to generate an EPR pair out of single photons, a supply of two additional photons is required (see Section 3.3). Therefore, we conjecture the resources required to transform state $S^{(1)}$ into $S^{(2)}$ are $\max\{0, \nu^{(2)} - \nu^{(1)}\}$ single photons.

The KLM CZ has the ability to change the rank of a state by 2. This is achieved by two-fold application of an NLS gate. However, using NLS gates is by no means an optimal procedure concerning the success probability, as can be inferred from the existence of EPR producing circuits and CZ gates with $p_s > 1/16 = p_{\text{NLS}}^2$.

3.2.6 Higher dimensional spaces

$n > 4$

Keeping $k = 2$ fixed, the case $n > 4$ corresponds to using an encoding on more than two modes per qubit. In the physical world this can be realised by using more modes in the same degree of freedom, or by using additional degrees of freedom. The resulting states of higher rank $\nu > 4$ are often referred to as *hyper-entangled* states. Our framework which is centred around the modes and their excitations, rather than photons and their intrinsic properties (as in the first quantisation), allows to understand these states with a simple notation and without the danger of confusion about how the sub-systems are implemented.

For increasing the rank, special gates utilising additional photons can be applied. Let us mention a different approach, which is very similar to the generation of linear cluster states by means of fusion gates [38] (see Section 5.2). Simple projective measurements can be applied instead, building up higher rank states out of a supply of states with lower rank. Consider two diagonal states with respective ranks ν_1 and ν_2 on disjoint sets of modes. If two modes – one of each of the states – are interfered on a beam splitter and only outcomes of exactly two photons in total in these two output modes are accepted, the resulting state in the remaining modes will be one of rank $\nu_1 + \nu_2 - 2$. Given the interfering modes are labelled α and β , the probability of this process succeeding is

$$p_s = \sigma_\alpha^2 + (\sigma'_\beta)^2 - 4\sigma_\alpha^2(\sigma'_\beta)^2. \quad (3.50)$$

The actual (unnormalised) output state is described by

$$P(\mathbf{a}^\dagger) = \sigma'_\beta \sum_{i \neq \alpha} \sigma_i (a_i^\dagger)^2 + \sigma_\alpha \sum_{i \neq \beta} \sigma'_i (a_i^\dagger)^2. \quad (3.51)$$

In analogy to the generation of cluster states, these states can be imagined as chains with the vertices being the modes. Then, the beam splitter with the photon counters afterwards act on the two chains by either generating a longer one, or by totally destroying both.

Higher photon number

In the case of higher photon number ($k > 2$), matters are more complicated. A simple counting argument shows, that the state-tensor cannot be diagonalised in

general. While the dimension of the set of symmetric $k^{\times n}$ tensors,

$$\binom{\binom{n}{k}}{k} = \binom{n+k-1}{k}, \quad (3.52)$$

grows exponentially with n and k , the only handle one has are n^2 parameters (constant in k) from the unitary mode transformation. However, the diagonal only allows for n entries, making the two-photon case rather special. Similar problems were encountered while looking for a generalised (*i.e.*, for more than two parties in a system) Schmidt-decomposition [39, 40].

Nevertheless, there are possible generalisations to the singular value decomposition (SVD) used in the matrix case. These include the CP-decomposition [41, 42], the Tucker-decomposition [43], and the higher order SVD (HOSVD) [44]. While each of them takes some properties of the SVD, there does not exist a higher order counterpart with all the properties and such a well developed theory as for matrices. One possible way to generalise the SVD for our purposes will be shown in Section 4.2.

3.3 Polynomial factorisation

Abstracting from the physical interpretation behind the polynomials that describe a bosonic quantum state, we can interpret the problems from Section 3.2 from the viewpoint of polynomial factorisation: Let $f \in \mathbb{C}[x_1, \dots, x_n]$ be a multi-variate polynomial, can it be written as a product of polynomials $g, h \in \mathbb{C}[x_1, \dots, x_n]$, such that $f(\mathbf{x}) = g(\mathbf{x})h(\mathbf{x})$? In the case where f , g , and h are homogeneous (which is most interesting with respect to the encodings one usually considers in linear optics), the methods of Section 3.2 apply. Writing a state as a product of linear terms amounts to finding a linear optics network with vacuum modes and detectors such that, given product input, the output state is the objective one.

In the special case of degree 2, $f = p_\psi$ factorises iff $\text{rank} S \leq 2$: In this case g and h are linear in each of the variables and a suitably chosen transformation $A \in \mathbb{C}^{2 \times 2}$ (which can be generated by linear optics with vacuum extensions) affects them such that $g(A\mathbf{x}) = x_1$ and $h(A\mathbf{x}) = x_2$, giving rise to a state matrix of rank 2 or less. For a larger rank the state can be reduced to one with rank 2 by using a singular mode transformation $A \in \mathbb{C}^{n \times n}$. But it cannot be written as a product of linear combinations because A will not be invertible.

Results that are more general are known from algebraic geometry. In Ref. [45] a condition for bi-variate polynomials (so, two-modes states) $p_\psi \in \mathbb{C}[x, y]$ is given.

Lemma 3 (Ruppert [45]). *Let $p(x, y) = \sum_{i,j=0}^n p_{i,j} x^i y^j$ a bi-variate polynomial of degree n in both variables and M a block matrix $M \in \mathbb{C}^{2n(2n-1) \times (n+1)(2n-1)} = (B, C)$ where B and C are block matrices, the matrix elements of the blocks $B^{i,j} \in \mathbb{C}^{(2n-1) \times (n+1)}$ and $C^{i,j} \in \mathbb{C}^{(2n-1) \times (n-1)}$ of which are given by*

$$B_{k,l}^{i,j} = -(k - 2(l - 1))p_{i-j, k-l+1} \quad (k, l) \in [1, 2n - 1] \times [1, n + 1] \quad (3.53)$$

$$C_{k,l}^{i,j} = (i - 2(j - 1))p_{i-j+1, k-l} \quad (k, l) \in [1, 2n - 1] \times [1, n - 1]. \quad (3.54)$$

For convenience, we use the convention $p_{i,j} = 0$, if either of the indices does not lie in the range $0, \dots, n$.

Then, p is irreducible (i.e., it cannot be written as a product of non-trivial polynomials), iff M has full rank, so $\text{rank} M = 2n^2 + n - 1$.

This result allows us to test for reducibility by calculating the determinants of all $(2n^2 + n - 1) \times (2n^2 + n - 1)$ sub-matrices of M . By using the following lemma from Ref. [46], the test can be extended to general multi-variate polynomials.

Lemma 4 (Kaltofen [46]). *Let $p \in \mathbb{C}[x_1, \dots, x_n]$ and $L = \mathbb{C}(v_1, \dots, v_n, w_2, \dots, w_n, z_2, \dots, z_n)$. Then, p is irreducible over \mathbb{C} iff $\pi \in L[x, y]$ with $\pi(x, y) = p(x + v_1, w_2 x + z_2 y + v_2, \dots, w_n x + z_n y + v_n)$ is irreducible over L .*

In the multi-variate case, the $v_1, \dots, v_n, w_2, \dots, w_n, z_2, \dots, z_n$ are indeterminants, i.e., the rank of M has to be considered with these variables being unknown.

The construction of the *Ruppert matrix* M can easily be implemented in `Mathematica`, allowing for testing whether given photonic states can be generated from a product input and linear optics. For example, a Bell state in dual-rail encoding is described by a polynomial $p(x_1, x_2, x_3, x_4) = 2^{-1/2}(x_1 x_2 + x_3 x_4)$. After using Lemma 4, the corresponding Ruppert matrix (shown in Appendix C) is of size 12×9 and has full rank, so no factorisation of the polynomial p – and therefore of the state $|\psi\rangle = p(a_1^\dagger, a_2^\dagger, a_3^\dagger, a_4^\dagger) |\text{vac}\rangle$ by linear optical means – is possible.

Here we discuss small networks that may be of interest for experimental realisation. We use the techniques introduced before to explicitly construct specialised networks for a wide range of tasks (such as state preparation, quantum gates, and measurements) including only a small number of photons and modes. After solutions are obtained, they will be translated into the physical world. We try to optimise the networks for experimental feasibility by exploiting common tricks (such as using polarisation for dual-rail qubits).

While previous constructions of small networks in the literature were based on ad-hoc approaches specially tailored to the respective problems, our aim is to investigate this subject in a more systematic fashion. We will revisit networks for a variety of problems:

- State transformations for dual-rail qutrits (Section 4.2). A special class of transformations not requiring interferometers was already used in Ref. [51]. We will present general way of constructing such state transformations and will also generalise them to the three-photon (ququad) case.
- Controlled phase gates as introduced and experimentally implemented in Refs. [23, 53, 54] (Section 4.3). We will generalise them to arbitrary phases and prove their optimal probability of success. Further, such gates with two control qubits will be investigated. Here our construction will deliver the optimal success probability as well, in contrast to earlier proposals in Refs. [61, 62].
- Bell state measurements (Section 4.4.2). Although apparently settled (maximum success probability for with the help of auxiliary vacuum modes $p_s = 1/2$ [63] is achieved by the network introduced in Refs. [66, 67]), we will emphasise the importance of being able to apply easy-to-implement stochastic local unitaries (*i.e.*, beam splitters). Without it we will discover a much smaller optimal success probability.
- Photon number resolution with linear optics and bucket detectors (Section 4.4.3). As already shown in Ref. [71], this is not possible with non-perfect

detectors. A new proof will be shown and a network will be constructed for the purpose of photon number discrimination with perfect bucket detectors.

4.1 NOON state generation

In order to illustrate the methods introduced in the previous chapter, we start with the task that is important in the very beginning of every quantum computation: state preparation. As already mentioned for the case of the one-way computer, specially tailored multi-partite quantum states have a wide range of applications as resources in different protocols. One prominent example are the so-called NOON states, defined for $n = 2$ modes and $N \in \mathbb{N}$ photons by

$$|\psi^N\rangle = 2^{-1/2} (|N, 0\rangle + |0, N\rangle). \quad (4.1)$$

The interference pattern obtained by interfering the two modes shows a highly sensitive dependency on the relative phase between the two paths. Upon post-selection on N -photon events, the N -fold phase picked up by N in contrast to a single photon can be used to beat the classical diffraction limit in lithography [47]. This conditioning, however, requires N -photon absorbing material the lack of which still prevents applications of this protocol.

There are already various proposals for NOON state generation by linear optics [48, 49, 50], but here we aim at finding a network for a fixed size which shall be optimal with respect to its probability of success [4].

In our example we will fix $N = 5$, so

$$Q(\mathbf{a}^\dagger) = 240^{-1/2} \left((a_1^\dagger)^5 + (a_2^\dagger)^5 \right) \quad (4.2)$$

and start from a product of Fock states on three modes, $|\psi^{\text{in}}\rangle = |2, 2, 2\rangle$ or

$$F(\mathbf{a}^\dagger) = 8^{-1/2} \prod_{i=1}^3 \left(\sum_{j=1}^3 A_{i,j} a_j^\dagger \right)^2. \quad (4.3)$$

In this case the mode transformation matrix A is a complex 3×3 matrix. After passing through the network described by A , a photon number counter will be placed in the third mode to implement a post-selection on one photon in this mode, which is described by $H(\mathbf{a}^\dagger) = a_3^\dagger$.

The requirement $G = \alpha Q$ leads to a system of six polynomial equations for the coefficients of $(a_{1,\text{out}}^\dagger)^k (a_{2,\text{out}}^\dagger)^{5-k}$, $k = 0, \dots, 5$. After solving these equations, four free real parameters will correspond to scaling of the three input modes (x_1 , x_2 , and x_3), and the first output mode (y_1), respectively. To simplify the system of equations we eliminate these variables by setting $A_{1,1} = A_{2,1} = A_{3,1} = A_{1,3} = 1$.

It turns out, that the actual set of solutions now only consists of six different ones, each of which represents an equivalence class with respect to the rescaling parameters mentioned above. The one that will lead us to the optimal solution is

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{-\frac{4\pi i}{5}} & \frac{3+\sqrt{5}}{2} e^{-\frac{2\pi i}{5}} \\ 1 & e^{\frac{4\pi i}{5}} & \frac{3+\sqrt{5}}{2} e^{\frac{2\pi i}{5}} \end{pmatrix}. \quad (4.4)$$

It is easy to see that this matrix is not unitary and must be extended to a physically meaningful transformation by employing additional vacuum modes and post-selection onto vacuum.

Owing to the size of the problem, the optimal success probability $p_s = \alpha^2 (x_1 x_2 x_3)^4 (y_1)^2$, such that the YAX is unitarily extendable (see Sections 3.1.6 and 3.1.7), can be identified easily by numerical methods. To do this, the x_i are written in spherical coordinates – thereby separating the global rescaling from the ratio of the parameters – and the resulting parameters p_s depends on are varied subsequently [4].

The globally maximal success probability is attained by a matrix A which can be extended in a minimal way, *i.e.*, by exploiting only one additional vacuum mode. Containing a rescaled version of A as a sub-matrix, the unitary matrix

$$U = \begin{pmatrix} 0.5722 & 0.5722 & 0.1894 & 0.5561 \\ 0.5257 & 0.5257 e^{-\frac{4\pi i}{5}} & 0.4556 e^{-\frac{2\pi i}{5}} & 0.4895 e^{\frac{3\pi i}{5}} \\ 0.5257 & 0.5257 e^{\frac{4\pi i}{5}} & 0.4556 e^{\frac{2\pi i}{5}} & 0.4895 e^{-\frac{3\pi i}{5}} \\ 0.3461 e^{\pi i} & 0.3461 e^{\pi i} & 0.7409 & 0.4599 \end{pmatrix} \quad (4.5)$$

describes a linear optics network which generates a five-photon NOON state under the following conditions:

- a $|2, 2, 2, 0\rangle$ input state,
- detection of one photon in the third and
- detection of vacuum in the fourth output mode.

Its probability of success, $p_s \approx 0.05636$, is the maximum for this input and the objective state $|\psi^5\rangle$. By decomposing U into $SU(2)$ unitaries (which can easily be implemented as a `Mathematica` program), an explicit description of a beam splitter network on four modes can be obtained [20].

4.2 Dual-rail qutrits and ququads

Due to the hope for a compact implementation and the limited requirement of auxiliary photons, an encoding with $k = 2$ or $k = 3$ photons on two modes is a candidate for photonic implementation of higher dimensional Hilbert spaces. Some transformations along these lines have been implemented in experiments recently [51].

Using a dual-rail encoding for qudits – for $k = d - 1$ the space spanned by the Fock states $\{|k - 1, 0\rangle, |k - 2, 1\rangle, \dots, |0, k - 1\rangle\}$ – some local rotations can be readily implemented as d -dimensional representations inherited from $SU(2)$ beam splitters. These can be applied in a particularly robust way by resorting to polarisation encoding. However, to get all local rotations for $d > 2$, auxiliary modes and photons, as well as interferometers are required.

We will consider the problem of how to implement state transformations by applying the methods from Section 3.2. The problem will be separated into three parts – the diagonalisation of the state matrix, the changing of the “Schmidt coefficients”, and the rotation to the objective state. While the techniques presented in Section 3 could be used to find a network as well, the assessment of success probabilities and experimental simplifications seem to be easier to perform in this approach due to its separation into smaller problems.

4.2.1 Qutrits

Any qutrit state in dual-rail encoding ($n = k = 2$) can be written in terms of a 2×2 symmetric “state matrix” S as introduced in Section 3.2.1. Recalling the methods from Section 3.2, we can transform any state with rank ν to any state with rank $\nu' \leq \nu$ without using auxiliary photons. The necessary operations are a beam splitter between the two modes (or, for experimental robustness, an appropriate combination of wave-plates in polarisation encoding), separate coupling of the two modes to additional vacuum modes, which will be post-selected in the vacuum afterwards. The latter can be achieved by a partially polarising beam splitter. Another beam splitter on the signal modes is needed in order to obtain the objective state from its diagonal form (3.38).

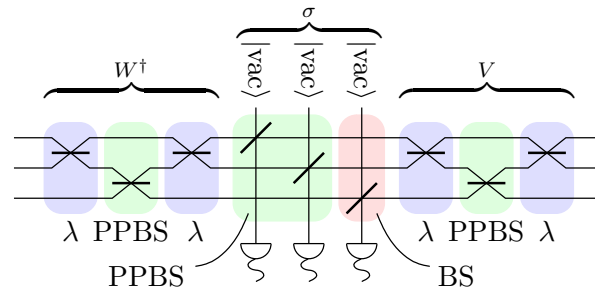


Figure 4.1: Decomposition of a general linear transformation of the creation operators of three modes, $A = V\sigma W^\dagger$. Because V and W^\dagger are unitary, they can be implemented by using solely linear optics without additional modes, so by networks consisting of three beam splitters each. σ is a real, positive diagonal matrix (w.l.o.g. all elements are not larger than 1) describing damping of the modes, so separate coupling to vacuum with post-selection. Using polarisation the structure can be simplified (λ shall denote a suitable combination of wave plates, PPBS a partially polarising beam splitter², and BS a beam splitter).

Increasing the rank, however, requires additional photons. In the $n = k = 2$ case, we know that one photon is sufficient. Hence, in general, we are dealing with a network on three modes (two signal modes and one containing the auxiliary photon) and additional vacuum modes for unitary extension. The general network adequate for any state transformation is shown in Fig. 4.1. Using polarisation encoding, requirements for interferometric stability can be relaxed a bit, giving rise to the network shown in Fig. 4.2.

If experimental simplicity is important, the interferometer based NLS gate or – more generally – the networks shown in Figs. 4.1 and 4.2 would not be the natural choice. A simple transformation without the need for interferometers can be used, making use of the inherent stability of polarisation encoding. The network for non-increasing rank described above is extended by inserting another beam splitter after the first set of wave-plates. Its purpose is to couple the two signal modes to two auxiliary modes which contain a photon and will be post-selected in the one-photon state (see Fig. 4.3). Depending on the actual source and target states, the network will usually not help in achieving the optimum success probability, but still allows for any state transformation.

4.2.2 Ququads

Increasing the dimension of the Hilbert space by one changes the state tensor to a $2 \times 2 \times 2$ tensor, so

$$|\psi\rangle = \sum_{i,j,k=1}^2 S_{i,j,k} a_i^\dagger a_j^\dagger a_k^\dagger |\text{vac}\rangle. \quad (4.6)$$

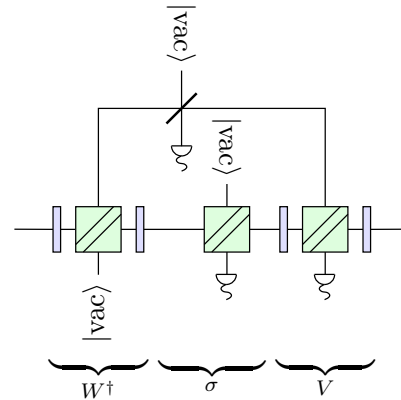


Figure 4.2: General network on three non-trivial modes in polarisation encoding. While the general network on two signal modes (so the general network on three modes) can be implemented by a PPBS (σ) in between a pair of wave plates (W^\dagger and V), the three-mode case is considerably more involved. See also the central parts (λ_1 , λ_V , and λ_2) of Figs. 4.5 and 4.6 for a two-mode example. Each of the boxes is a PPBS. In the qutrit case, one of the spatial input modes is initialised in the $|1\rangle$ -state and post-selected in the $|1\rangle$ -state. For ququads, it has to be a mode in the $|2\rangle$ -state. This is the brute-force version of a general 3-mode network with vacuum extension, there might exist much simpler networks in certain cases.

Note, that a stable, variable PPBS can be built with a Sagnac-type interferometer around a PBS with some wave plates (see Fig. 4.7).

Then, the concept of *rank* has to be discussed more carefully than in the matrix case where the number of independent rows or columns, the dimension of the image of the linear map induced by the matrix, and the minimal number of terms in a decomposition into rank-1 matrices all coincided. Also, the property that the set of matrices of non-maximal rank is of measure zero does not carry over to the notions of rank applied to tensors. Although it is known, that every $2 \times 2 \times 2$ tensor can be written as the sum of two rank-1 tensors [52],

$$S = U_1 \circ V_1 \circ W_1 + U_2 \circ V_2 \circ W_2, \quad (4.7)$$

where “ \circ ” denotes the outer product and the 2×2 matrices U , V , and W are not necessarily orthogonal. However, when imposing the symmetry of S (symmetric under any permutation of indices due to indistinguishability of the photons), we would require $U = V = W$ which might result in a decomposition using more terms. The notion of rank that we will adopt is the minimal number of symmetric rank-1 tensors in this sum.

²A beam splitter that has different reflectivities depending on the polarisation will be called partially polarising beam splitter (PPBS). A PBS is a special case of a PPBS with $(r_H, r_V) = (0, 1)$

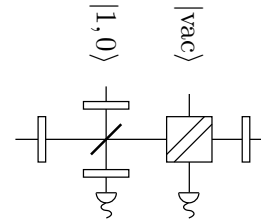


Figure 4.3: An experimentally feasible state transformation between dual-rail qutrit states. The first and last wave plate are used to diagonalise (see Eqn. (3.38)) the state and transform it from diagonal form into the objective state, respectively. Still not increasing the rank, the PPBS, coupling each signal mode to a vacuum mode, implements the mode-wise damping (3.42). To increase the rank, the signal interferes with an auxiliary photon (both in a pair of modes) on a beam splitter. The wave-plates distributing the auxiliary photon can be, for example, a 45° rotation and the beam splitter one with 75% transmittivity.

The problem of finding an appropriate decomposition – and the fact that it could be not reversible – forces us to find different methods for the state transformations we considered before.

From normal form to objective state

Finding a symmetric decomposition of a state tensor in terms of rank-1 tensors corresponds to finding a network that constructs this state from a given diagonal one, $|\psi(D)\rangle = \sum_i D_i (a_i^\dagger)^3 |\text{vac}\rangle$. This can be seen by generalising Eqn. (4.7) to

$$S = \sum_{i=1}^{\nu} D_i A_i \circ A_i \circ A_i \quad (4.8)$$

the components of which read

$$S_{i,j,k} = \sum_{r,s,t=1}^{\nu} D_{r,s,t} A_{i,r} A_{j,s} A_{k,t}, \quad (4.9)$$

where D is the diagonal core tensor with $D_{i,i,i} = D_i$. For large rank, the core tensor is larger than the original state which is achieved by making A rectangular. This corresponds to starting with $\nu > 2$ modes and post-selecting $\nu - 2$ of them in the vacuum after the network, allowing for embedding of A in a square, unitary matrix. The network A acts on a state of the form $|\psi(D)\rangle$, after post-selecting the spare modes in the vacuum state, the output in the remaining modes is $|\psi(S)\rangle$. Note, that

and a BS corresponds to $(r_H, r_V) = (r, r)$. See Fig. 4.7 for different ways of implementing a tunable PPBS.

this scheme is not efficiently scalable – in general the maximum rank could scale exponentially with the number of photons.

An algorithm which decomposes any $2 \times 2 \times 2$ tensor is given below. It turns out, that the maximum symmetric rank for such a tensor is 3, as the algorithm can always succeed in this case. Therefore, $A \in \mathbb{C}^{2 \times 4}$ acts on at most three non-trivial modes – the class of networks depicted in Figs. 4.1 and 4.2 is sufficient to implement them. Here they are acting on three signal modes (instead of two and an auxiliary one) in the input, and one output mode is post-selected in the vacuum.

Algorithm Because vacuum extensions are inherent in the scheme and we are only interested in the existence of a decomposition, rather than its optimality, we can absorb D_i into A_i , such that we always start with a state $|\psi(D)\rangle = 3^{-1/2} (|3, 0, 0\rangle + |0, 3, 0\rangle + |0, 0, 3\rangle)$ and

$$S_{i,j,k} = \sum_{s=1}^{\nu} A_{i,s} A_{j,s} A_{k,s} \quad , 1 \leq i \leq j \leq k \leq 2 \quad (4.10)$$

(there are only four independent entries of S). In our case of two signal modes, one can substitute $X = A_{1,1}^3$, $x = A_{2,1}/A_{1,1}$ and similarly for Y , y , Z , and z , yielding

$$S_{i,j,k} = Xx^\alpha + Yy^\alpha + Zz^\alpha \quad (4.11)$$

with $\alpha = i + j + k - 3$. Now it is always possible to find x , y , and z such that

$$\text{rank} \begin{pmatrix} 1 & 1 & 1 \\ x & y & z \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{pmatrix} = 3 \quad (4.12)$$

(which is easy due to the matrix structure) and

$$B = \begin{pmatrix} S_{1,1,1} & 1 & 1 & 1 \\ S_{1,1,2} & x & y & z \\ S_{1,2,2} & x^2 & y^2 & z^2 \\ S_{2,2,2} & x^3 & y^3 & z^3 \end{pmatrix} \quad \text{such that} \quad \det B = 0. \quad (4.13)$$

Then, a solution to the linear system $B(X, Y, Z)^T = (S_{1,1,1}, S_{1,1,2}, S_{1,2,2}, S_{2,2,2})^T$ in X , Y , and Z always exists.

How to find the inverse transformation (*i.e.*, from an arbitrary state into diagonal form) will be shown later.

Changing the rank of states in normal form

Again, only transformations that increase the rank require additional photons. However, one is not enough in this case – two photons are needed. The Gröbner basis for the network performing the transformation $|3, 0\rangle \mapsto 2^{-1/2}(|3, 0\rangle + |0, 3\rangle)$ is non-trivial only for at least two additional photons. Fortunately, these two can reside in one spatial mode which can be post-selected in the $|2\rangle$ -state afterwards. Again, the networks shown in Figs. 4.1 and 4.2 are sufficient for this purpose as well, now with the ancilla in state $|2\rangle$ instead of $|1\rangle$. Note that, although each of the operations corresponds only to a single interferometer, they are required to be stable with respect to each other.

With experience in the 2- and 3-photon cases, one could speculate about higher order tensors:

Conjecture 5 (Auxiliary photons for entanglement generation). *In order to increase the symmetric rank of a k -photon state by $\Delta\nu > 0$, $(k - 1)\Delta\nu$ photons are necessary and sufficient.*

Transformation into normal form

Intuitively, this transformation is the inverse of the one above. Let $A \in \mathbb{C}^{2 \times 3}$ be the transformation from diagonal form to objective state with symmetric rank $\nu = 3$. For the inverse transformation $\tilde{A} \in \mathbb{C}^{3 \times 2}$ we would require $\tilde{A}A = \mathbb{1}_3$, which cannot be satisfied in general.

Therefore, we have to use an alternative approach. The defining equations for the inverse transformation A' read

$$\delta_{i,j,k} = \sum_{r,s,t=1}^2 S_{r,s,t} a'_{i,r} a'_{j,s} a'_{k,t} \quad \text{with } i, j, k = 1, \dots, \nu, \quad (4.14)$$

and can be solved easily using Gröbner bases for the interesting cases $\nu \leq 3$.

Again, the transformation can be implemented by the class of networks shown in Figs. 4.1 and 4.2 with three signal modes in the output. The ancilla has to be prepared in the vacuum state.

Although the forward transformation might prove useful for dual-rail ququad state preparation, the equivalence classes shown in Section 3.2 do not carry over

to the three-photon case: there are examples of states where $\nu \neq \nu'$, which shows again, that these transformations are not easily inverted using linear optics.

4.3 Controlled phase gates

Two-partite quantum gates delivering a controlled phase-shift of $\varphi = \pi$ have already been experimentally demonstrated [23, 53, 54]. Here, we are concerned with the question whether arbitrary phases can be realised as well and what the optimum probabilities of success are.

The realisation may be interesting from the perspective of (i) gaining an understanding of the probabilistic character of quantum gates as well as (ii) serving as a proof of principle realisation of a kind of quantum gate that has several applications in linear optical quantum information processing.

Concerning the first aspect, one may well expect that there is a trade-off between the notorious problem of having a small probability of success and the phase that is being realised in the gate. In fact, the study in Ref. [55] suggests exactly such a behaviour: the presented upper bounds to the probability of success increase from the minimum at $p_s(\varphi = \pi) = 1/4$ to $p_s(0) = 1$. To investigate such a trade-off should be interesting in its own right and help in building intuition concerning the probabilistic behaviour of linear optics.

Then, concerning the second aspect, there are several applications for which such a trade-off is relevant. In linear optical architectures, it may be a good idea to have a smaller phase, if one only has higher success probabilities. The new measurement-based quantum computational models [56] for example offer this perspective: One does not have to have controlled π phase gates to prepare cluster states, but one would in principle also get away with smaller phases. This may well (but does not have to be) a significant advantage when preparing resources for measurement-based quantum computing different from cluster states [15, 56].

Then, of course, in standard gate-based quantum computing, one will typically encounter all kinds of controlled phase gates. For example, in the quantum Fourier transform [16], one has to implement several controlled phase gates. Of course, they can again be decomposed into other sets of universal gates (like CNOT or CPHASE and local unitaries). But, in terms of resource requirements, it is obviously an advantage to directly implement the relevant quantum gates with phases in the range $0 < \varphi < \pi$. There are also interesting trade-offs between resource requirements and success probabilities in a number of related contexts, like non-local gates in distributed quantum computation [57, 58, 59]. Refs. [58, 59], for example, study

distributed controlled phase-gates which would need less entanglement and succeed with a higher probability.

Here we will study post-selected gates, so not genuine “event-ready” quantum gates, but those where one measures the output modes and whether the gate actually succeeded is determined only *a posteriori* by accepting only those outcomes which lie in the computational subspace of \mathcal{H}_N^n . These gates concern only a smaller number of modes and are still within reach of current experiments. In principle non-demolition measurements of the output would be required for an event-ready gate. Already these gates themselves – together with special inputs – can be utilised to this end.

4.3.1 Single beam splitter

In a post-selected phase gate on four modes in the standard dual-rail encoding, two of the modes are merely involved as “by-standers”, in that their phase is compensated in exactly the same fashion as in Refs. [23, 53, 54]. In this section, we will hence concentrate on two modes forming the “core” of the scheme, giving rise to a two-qubit dual-rail phase gate on four physical modes. The core itself may be regarded as a single-rail phase gate in its own right.

We will investigate the consequences of simply having a single beam splitter forming the core of the quantum gate. The action on the photonic creation operators of the two involved modes it is mixing is described by the matrix

$$U = \text{diag} \{ e^{i\phi_1}, e^{-i\phi_1} \} \cdot B \cdot \text{diag} \{ e^{i\phi_2}, e^{-i\phi_2} \} \quad (4.15)$$

with

$$B = \begin{pmatrix} \sin \vartheta & \cos \vartheta \\ -\cos \vartheta & \sin \vartheta \end{pmatrix} \quad (4.16)$$

and appropriate phases ϕ_1 , ϕ_2 , and mixing angle ϑ . The phases can also be realised by local operations on the dual-rail qubits, which leaves the relevant part of the gate $U' = B$. The matrix elements of the unitary for vacuum, single photon operation and the two-photon component read

$$\langle 0, 0 | \mathcal{U}' | 0, 0 \rangle = 1 \quad (4.17)$$

$$\langle 1, 0 | \mathcal{U}' | 1, 0 \rangle = A_{1,1} = \sin \vartheta, \quad (4.18)$$

$$\langle 0, 1 | \mathcal{U}' | 0, 1 \rangle = A_{2,2} = \sin \vartheta, \quad \text{and} \quad (4.19)$$

$$\langle 1, 1 | \mathcal{U}' | 1, 1 \rangle = \text{per} A = \cos(2\vartheta), \quad (4.20)$$

respectively. Since we are restricted to $n \leq 2$, these four quantities determine the action of the core completely. With the constraint

$$2 \sin^2 \vartheta - 1 = \sin^2 \vartheta \quad (4.21)$$

that ensures equal single- and two-photon amplitudes, this is the beam splitter used in Refs. [53, 54], realising a sign-flip ($\varphi = \pi$).

Hence, one finds that in this way, one *can* implement quantum phase gates, but only two different ones: One is not doing anything, and the other ones effect is a controlled phase of π . This is exactly the gate of Refs. [53, 54]. In other words, without invoking at least a single additional mode, one can in this fashion not go beyond the known π -phase.

4.3.2 Arbitrary phases

However, we can go beyond such a description: The restriction to unitary two-mode beam splitters can be relaxed. Instead of starting with $U \in SU(2)$, we use an arbitrary $A \in \mathbb{C}^{2 \times 2}$. Then we will embed the two-mode matrix into a higher dimensional unitary. Only a single additional mode is required, so the full setup would consist of a transformation on three modes involving at most three beam splitters. In this class of gates, for each φ , the one with the optimal probability of success $p_s(\varphi)$ can be identified.

Theorem 6 (Optimal post-selected dual-rail controlled phase gate). *Consider linear optics, vacuum modes and detectors. When post-selecting on the state of the signal modes in the computational sub-space, the optimal network on four modes implementing the gate represented in the computational basis of two dual-rail qubits by $U = \text{diag}\{1, 1, 1, e^{i\varphi}\}$ has a success probability (shown in Fig. 4.4) of*

$$p_s(\varphi) = \left(1 + 2 \left| \sin \frac{\varphi}{2} \right| + 2 \sin \frac{\pi - \varphi}{4} \sqrt{\left| \sin \frac{\varphi}{2} \right|} \right)^{-2}. \quad (4.22)$$

Proof. In order to find p_s we will first construct the linear transformation of the relevant creation operators and identify the optimal unitary extension afterwards.

Two-mode transformation The two-mode transformation resulting from solving the equations (4.17)–(4.20) imposed by the gate we want to build is

$$A = p_s^{1/4} \begin{pmatrix} x & (e^{i\varphi} - 1)x/y \\ y/x & 1/x \end{pmatrix}. \quad (4.23)$$

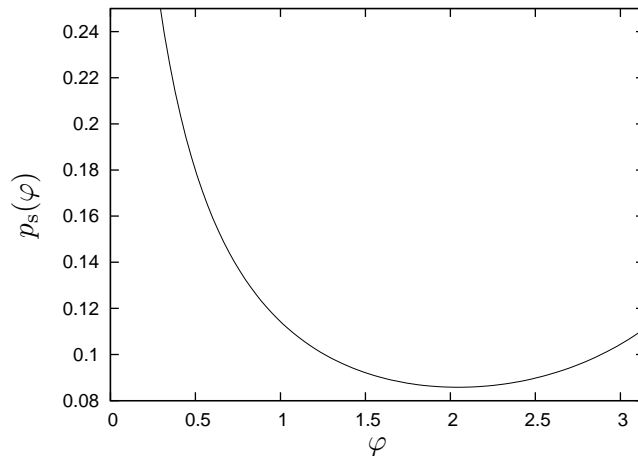


Figure 4.4: Optimal success probability $p_s(\varphi)$ of phase gates with vacuum ancillas (one vacuum ancilla is already optimal) *vs.* the phase φ (solid line). At $\varphi = \pi$ the result of Refs. [53, 54], $p_s(\pi) = 1/9$, is reproduced.

The intuitive assumption of a monotonous $p_s(\varphi)$ is not fulfilled: indeed, the success probability is worse than $1/9$ in the interval $\pi/3 < \varphi < 1$. Due to phases $\varphi < \pi$ not being implementable with a single beam splitter, the additional unitary extension requires further measurements and therefore decreases the probability of success near $\varphi = \pi$.

x and y are free non-zero complex parameters. By writing

$$A = p_s^{1/4} \text{diag} \{a, a^{-1}\} \cdot \begin{pmatrix} 1 & e^{i\varphi} - 1 \\ 1 & 1 \end{pmatrix} \cdot \text{diag} \{b, b^{-1}\}, \quad (4.24)$$

with $a = xy^{-1/2}$ and $b = y^{1/2}$ we see that, the singular values of A only depend on $|a|^2$ and $|b|^2$, so not on the phases of a, b , and x and y at all.

The general solution to the dual-rail problem is actually composed of the transformation A together with appropriate damping of the by-standers. Hence, single-rail imposes here no restriction at all. Further, this is the most general post-selected two-qubit gate with a 2×2 kernel.

Optimal extension Given the 2×2 matrix A that realises the transformation we are looking for, the optimal extensions can be identified. Let us extend the first and second row vectors (denoted by A_1 and A_2) to dimension 3 by appending $A_{1,3}$ and $A_{2,3}$, respectively, in such a way as to allow for unitarity of the extended matrix, $A' \in SU(3)$. As pointed out in Section 3.1.7, no larger extension has to be investigated. We choose $A_{1,3}$ and $A_{2,3}$ such that the new row vectors are orthogonal. By multiplying them by the root of the inverse of their respective norms, $|A'_1|$ and $|A'_2|$, they will be normalised. Finding a third orthogonal vector to fill the unitary matrix

can be done with the complex cross product $(A'_1 \times A'_2)^*$, or in general by choosing a vector at random and orthogonalising it with respect to the given ones.

The dependence of the success probability on the extension is $p_s = (|A'_1||A'_2|)^{-2}$. Therefore, the objective is to

$$\text{minimise} \quad f = |A'_1|^2 |A'_2|^2 \quad (4.25)$$

$$\text{subject to} \quad A'_1(A'_2)^\dagger = A_1 A_2^\dagger + A_{1,3} A_{2,3}^* = 0. \quad (4.26)$$

The first observation is, that the row-scaling by x is already included in the norm of the row vectors, leaving us with one parameter less. By using the phase of y , we can assure that $A_1 A_2^\dagger$ is real and positive and also $\arg(A_{1,3}) - \arg(A_{2,3}) = \pm\pi$. Using Lagrange multipliers to solve the constrained minimisation problem in $A_{1,2}$ and $A_{1,3}$ we find $|y| = (2(1 - \cos \varphi))^{1/4}$. Then an optimal solution (phases chosen conveniently) is

$$A_{1,3} = A_{2,3}^* = e^{\frac{i\pi}{2}} \left(\sqrt{2 \left| \sin \frac{\varphi}{2} \right|} \sin \left(\frac{\pi - \varphi}{4} \right) \right)^{1/2} \quad (4.27)$$

with the probability of success given by

$$p_s(\varphi) = \left(1 + |y|^2 + |y| \sqrt{2 - |y|^2} \right)^{-2} \quad (4.28)$$

$$= \left(1 + 2 \left| \sin \frac{\varphi}{2} \right| + 2 \sin \frac{\pi - \varphi}{4} \sqrt{\left| \sin \frac{\varphi}{2} \right|} \right)^{-2}. \quad (4.29)$$

The reflectivities of the compensating beam splitters in the by-passed modes have to be chosen such that the success probability is constant for all dual-rail states, *i.e.*, $r = p_s^{1/4}$. \square

The success probability $p_s(\varphi)$ of this gate is shown in Fig. 4.4. Interestingly – and quite surprisingly – the worst success probability is not achieved for the sign-flip ($\varphi = \pi$), but for $\varphi \approx 2.05$. That means, gates delivering a phase shift slightly smaller than π and thereby generation less entanglement will not give rise to a larger, but to a smaller success probability. As expected, the success probability for very small phases increases and reaches unity for $\varphi = 0$: one can always do nothing at all with unit probability.

In this restricted setting, one can also answer an interesting related question: Could the use of a full transformation on all four modes have enhanced the probability of success? In fact, the answer is no and the solution as such is already optimal.

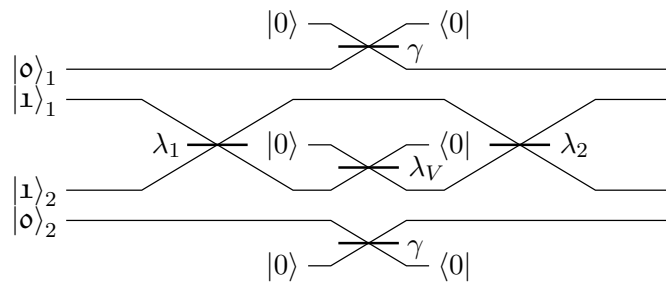


Figure 4.5: Basic spatial modes based setup obtained from translating an arbitrary 2×2 core into linear optics. The core extension is provided by mixing with a vacuum mode on the central beam splitter. This mode, in turn, has to be post-selected in the vacuum state afterwards. The upper and lower beam splitters implement the appropriate compensation by damping the by-passed modes (which is the same for both modes for the optimal solution which we consider). The labels at the beam splitters will be used to identify them with the respective optical elements in Figs. 4.6–4.8. In general, the parameters (*i.e.*, reflectivity and phases) of these elements depend on the non-linear phase φ .

Constrained singular values To illustrate the approach mentioned in the end of Section 3.1.7, we substitute the matrix A from (4.23) with $x = 1$ in Eqn. (3.33). The solution obtained with `Mathematica` can be written as a solution of

$$1 + 4\alpha + 6\alpha^2 + (20 - 16 \cos \varphi)\alpha^3 + (25 - 40 \cos \varphi + 16 \cos^2 \varphi)\alpha^4 = 0 \quad (4.30)$$

where we used $\alpha := -\sqrt{p_s}$. The validity of the solution can be easily confirmed by substituting p_s with (4.28). Other roots can be excluded because the success probability has to be real and $0 \leq p_s(\varphi) \leq 1$ for all values of φ .

4.3.3 Experimental implementation

In order to make the proposed gate experimentally feasible, some simplifications would have to be done. Even though this might result in lower success probabilities, the issue of interferometric stability is more significant in this context. We will hence discuss a number of strategies that could achieve this goal.

The straightforward setup on dual-rail encoding that realises a three-mode unitary and compensates the amplitudes in the remaining modes is shown in Fig. 4.5. It includes one interferometer, but the whole gate would sit inside a double interferometer, because local unitaries on the input and output qubits would require classical interference. Thus, the complexity of this gate is best described as a nested three-fold interferometer. In this first stage, the parameters (reflectivities and phases) of all five beam splitters depend on φ .

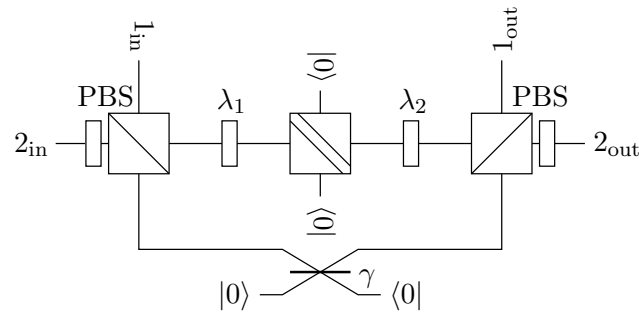


Figure 4.6: Setup for a controlled phase gate on two polarisation encoded dual-rail qubits. The logical modes of the two qubits are separated and united by means of polarising beam splitters (PBS). Replacing the left and right beam splitters in Fig. 4.5 by wave plates λ_1 and λ_2 , they become easier to tune to different φ , and provide better stability. The lower beam splitter, γ , implements compensation of both, $|0\rangle_1$ and $|1\rangle_1$, modes. λ_V is taken care of by the PPBS in the centre.

Additionally, one of the qubits has to be flipped prior to and after the circuit, which here is done by acting with a wave plate on the second qubit.

To get rid of some of the interferometers, polarisation encoding is convenient. Two modes can be united in one spatial mode, resulting in inherent stability (neglecting birefringence of the optical medium) of some interferometers. Rotations on these modes can be carried out easily using wave plates. Because the core acts on modes coming from different dual-rail qubits, they have to be combined into a single spatial mode before. This is achieved with a PBS, thus permuting H and V modes. Damping of the by-passed modes can be done simultaneously by a single beam splitter coupled to the vacuum. A straightforward translation of Fig. 4.5 into polarisation encoding, thereby collapsing the network into a single interferometer, is shown in Fig. 4.6.

Due to the asymmetry of the core, still one PPBS is used, the reflectivity of one polarisation component of which actually depends on φ , the other one being 1. Fig. 4.7 shows how a tunable PPBS can be constructed, introducing another interferometer.

Iterating the ideas that led to the compact PPBS implementation once more yields a collapsed form of the phase gate based on only two PBS and a couple of wave plates. Due to the paths of light being very similar, this setup should also be more robust. See Fig. 4.8 for more details.

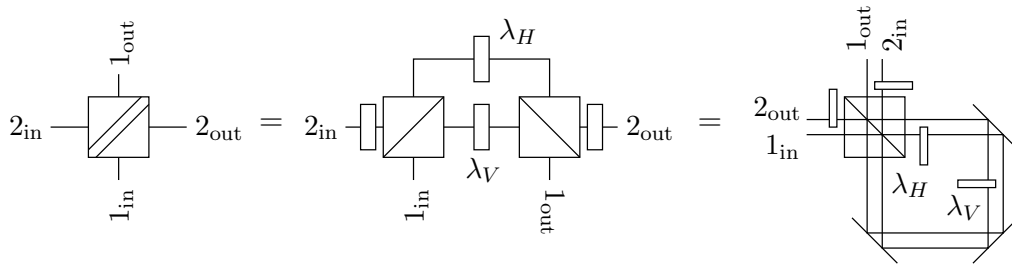


Figure 4.7: From left to right: (i) A PPBS implementing a beam splitter with polarisation dependent reflectivity. (ii) It is equivalent to an interferometer between two PBS where the reflectivities are implemented by means of wave plates, λ_1 and λ_2 . (iii) By identifying the two PBS, the interferometer collapses into a closed loop (which is more compact and more robust in experimental implementations), leaving only one PBS.

In the second and third circuit, a polarisation flip of the second qubit before and after the circuit is implemented by a wave plate.

4.3.4 Event-ready gates

To build an actual gate without measurements on the output modes, at least two additional photons are needed. This can be seen from the fact that a controlled π -phase gate is even stronger than a device that creates EPR pairs from single photons in the sense that it not only amounts to a state transformation from two single photons to an EPR pair, but a full unitary transformation on the whole state space (and creating an EPR pair when applied to two single photons). However, the latter task can only be achieved with at least two additional auxiliary photons (*e.g.*, the setup in Ref. [60]).

That a construction of an EPR pair out of single photons by means of linear optics, vacuum and detectors only is not possible is fairly obvious: It follows from the rank of the states and the bounds to the resources mentioned in Section 3.2.

An alternative approach is the following: the polynomial describing the objective state,

$$2^{-1/2} \left(a_1^\dagger a_3^\dagger + a_2^\dagger a_4^\dagger \right), \quad (4.31)$$

does not factorise over \mathbb{C} . That can easily be seen by constructing the polynomial's Ruppert matrix (see Section 3.3) which has maximum rank. Using one additional photon does still not help: the ideal generated by the polynomials describing the linear optics mode transformation is empty. Only when using two additional photons a solution is possible (as proven by the existence of such a setup).

An example of how to solve the polynomial equations for two additional photons is shown in [31], and solutions are given for $\varphi = \pi$ and $\varphi = \pi/2$.

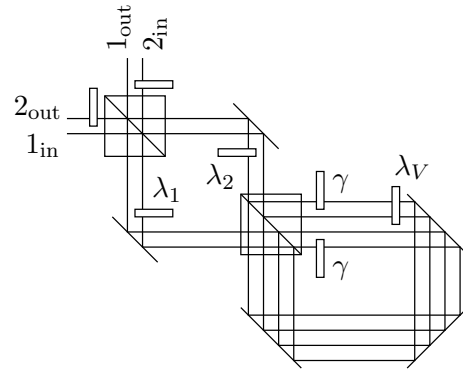


Figure 4.8: Compact implementation of a controlled phase gate by using a single loop to implement the central PPBS and the compensation beam splitters simultaneously. Additionally, the two PBS are identified, resulting in a second loop. All omitted modes are initialised in the vacuum and post-selected in the vacuum state (which will be achieved in practice by counting the photons in the other output).

4.3.5 Toffoli gates

In the same way as above, we can consider a generalised Toffoli gate, the effect of which on the computational basis can be described by the unitary $\text{diag}\{1, 1, 1, 1, 1, 1, 1, e^{i\varphi}\}$.

The solutions to the polynomial equations – up to mode permutations – can be parametrised by $x, y \in \mathbb{R}$ and are given by the matrix

$$A = p_s^{1/6} \begin{pmatrix} 1 & \frac{e^{i\varphi}-1}{xy} & 0 \\ 0 & 1 & y \\ x & 0 & 1 \end{pmatrix} \quad (4.32)$$

on the three modes which represent the $|1\rangle$ states (similar to the 2-mode core used by the controlled phase gate, and only global rescaling considered).

For a unitary extension all singular values of A have to be at most 1. In order to avoid to write down cubic singular values explicitly, we use the following constraints. Let $p_{AA^\dagger}(\lambda) = \det(AA^\dagger - \lambda\mathbb{1}_3)$ be the characteristic polynomial of AA^\dagger , the roots $\lambda_{1,2,3}$ of which are the squared singular values of A . By requiring $p_{AA^\dagger}(1) = 0$, one of the singular values has to be 1. That all the other singular values are not bigger than 1 is equivalent to the condition that all derivatives of p_{AA^\dagger} have the same sign at $\lambda = 1$. More formally, this results in further constraints of the form

$$(-1)^n p_{AA^\dagger}^{(k)}(1) = (-1)^n \left. \frac{dp_{AA^\dagger}(\lambda)}{d\lambda} \right|_{\lambda=1} \geq 0, \quad \text{for } 1 \leq k < n. \quad (4.33)$$

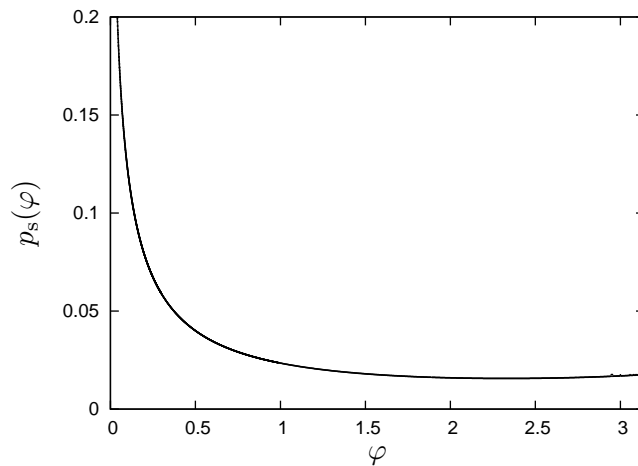


Figure 4.9: Optimal success probabilities of generalised Toffoli gates. The features exhibited by $p_s(\varphi)$ are similar to the ones observed at the controlled phase gate (Fig. 4.4): There is a shallow dip between $\pi/2$ and π below $p_s(\pi) = p_s(\pi/2)$ and a steep incline (more pronounced than for the controlled phase gate) for small phases towards $p_s(0) = 1$.

For $\varphi = \pi$ the optimal p_s compliant with these conditions is

$$p_s(\pi) = 1 + 3(2^{1/3} - 2^{2/3}) \approx 1/57. \quad (4.34)$$

See Fig. 4.9 for the maximum success probability in the range $0 \leq \varphi \leq \pi$. The corresponding networks could be constructed in the same way as above. However, they would consist of 3-mode cores (the class of networks shown in Figs. 4.1 and 4.2) inside separate interferometers for each of the 3 qubits. Only in terms of a single-rail gate this would be within current experimental reach. Experimentally more appealing approaches are shown in Refs. [61, 62], but leading only to success probabilities of at most $1/72$.

4.4 State discrimination

As an example of the methods discussed in Chapter 3, we will apply them to devices that shall be able to discriminate (i) between sets of two-partite states and (ii) between states of different photon number in a single mode. Keeping in mind the results from Section 3.3 and on Bell state discrimination [63, 64], we can expect interesting solutions if these states are entangled.

in the expressions

$$\alpha_1 = (156\gamma + 224\gamma^3 - 35\gamma^5 + 2\gamma^7)/144 \quad (4.38)$$

$$\alpha_3 = (1416\gamma + 380\gamma^3 - 74\gamma^5 + 5\gamma^7)/1152 \quad (4.39)$$

$$\alpha_2 = (-504\gamma - 724\gamma^3 + 118\gamma^5 - 7\gamma^7)/384 \quad (4.40)$$

$$\alpha_1\alpha_2 = (-456 + 20\gamma^2 + 10\gamma^4 - \gamma^6)/192 \quad (4.41)$$

$$\alpha_1\alpha_3 = (-104 - 108\gamma^2 + 18\gamma^4 - \gamma^6)/64 \quad (4.42)$$

$$\alpha_2\alpha_3 = (96 + 100\gamma^2 - 16\gamma^4 + \gamma^6)/48 \quad (4.43)$$

$$\alpha_1\alpha_2\alpha_3 = (-288\gamma - 76\gamma^3 + 16\gamma^5 - \gamma^7)/96 \quad (4.44)$$

which are substituted into U , thereby writing it over $\mathbb{Q}(\sqrt{2} + \sqrt{3} + i)$. Further substitution of (4.38)–(4.44) into $\gamma^8 = (\sum_i \alpha_i)^8$ delivers the primitive polynomial $144 + 192\gamma^2 + 88\gamma^4 - 16\gamma^6 + \gamma^8$ for this field extension.

To actually write down the equations that determine the optical networks implementing the desired measurement, detection patterns have to be agreed on. Given two photons in four modes, the state space has $\text{card}\mathbb{N}_N^n = 10$ dimensions. From the patterns representing the ten basis states in the Fock basis we choose four pairwise different ones, $(\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}) \subset \mathbb{N}_2^4$ as detection patterns. This leads to

$$\binom{10}{4} = 210 \quad (4.45)$$

different settings, each describing a set of 16 polynomial equations

$$\text{per}A [\omega^{(i)}|\Omega^{(j)}] = \sqrt{\omega^{(i)}!\Omega^{(j)}!}U_{i,j}, \quad i, j = 1, \dots, 4. \quad (4.46)$$

However, by exploiting the invariance under permutations of the output modes, there are only 105 equivalence classes. Further, if it is possible for linear optics to permute the input states, the number of equivalence classes can be reduced down to 17.

It turns out that the Gröbner bases of all 210 systems of polynomial equations are empty, *i.e.*, there is not a single detection pattern that could be achieved by linear optics with additional vacuum modes, when restricted to this gate-like setting.

Because the state of the modes is measured after the unitary matrix anyway, it should not matter whether a photon is detected in the auxiliary modes. It rather results in a larger set of 58905 detection patterns, which can be reduced to 260 equivalence classes by permuting the detector modes. The newly gathered freedom

helps to find a solution, indeed. We will not go into detail concerning these solutions. The networks are far too large to handle the optimisation analytically. Moreover, they include interferometers on more than four modes which has not been established as standard quantum information experiments, yet.

4.4.2 Bell states

Another case, which has been studied in much detail (*e.g.*, [63, 64]), is the discrimination of the four Bell states $B = \{|\psi_i\rangle, i = 1, \dots, 4\}$ (referred to as *Bell State Measurement* or *BSM*) in dual-rail encoding,

$$|\psi_{1,2}\rangle = \frac{1}{\sqrt{2}} (|\mathbf{01}\rangle \pm |\mathbf{10}\rangle) = \frac{1}{\sqrt{2}} (|1001\rangle \pm |0110\rangle) \quad (4.47)$$

$$|\psi_{3,4}\rangle = \frac{1}{\sqrt{2}} (|\mathbf{00}\rangle \pm |\mathbf{11}\rangle) = \frac{1}{\sqrt{2}} (|1010\rangle \pm |0101\rangle). \quad (4.48)$$

Let $\mathcal{I} \subset \mathbb{Z}$ be the set of labels of possible output signals of the measurement device. $p_i(|\psi\rangle)$ with $i \in \mathcal{I}$ shall be the probability that the output i is signalled, given the input state $|\psi\rangle$. In general, \mathcal{I} will contain more than 4 possible outputs, some of which shall be used to indicate an unknown event. In the following we will use $\mathcal{I} = [0, 4]$. The measurement will be called *unambiguous*, iff $p_i(|\psi_j\rangle) = p_i\delta_{i,j}$ for $1 \leq i \leq 4$, a restriction we will use throughout the whole chapter. Further, it will be called *complete*, iff $p_i = p_j \forall i, j = 1, \dots, 4$.

As a figure of merit for benchmarking the measurement device, the probability of success will be defined as

$$p_s := \frac{1}{4} \sum_{i=1}^4 p_i(|\psi_i\rangle), \quad (4.49)$$

which is a quantity constructed for the case of Bell states in the input mixed with equal probabilities.

Let us summarise what is known so far:

- Of course, when allowing for both, feed-forward and ancilla modes in arbitrary states, near-deterministic CNOT gates [37] can be constructed and used to transform the Bell states to orthogonal product states which can be distinguished perfectly. In this case an unambiguous BSM would trivially fulfil $p_s < 1$.

- Allowing for linear optics, photon counters, additional vacuum modes, feed-forward, and stochastic local unitaries, the maximum success probability of an unambiguous BSM satisfies $p_s \leq 1/2$ [63, 64].
- The latter bound is saturated by a network consisting of a single beam splitter [66, 67], the detection probabilities of which are $p_1 = p_2 = 1$ and $p_3 = p_4 = 0$. Usually it is made complete by using stochastic operations (*i.e.*, applying a σ_z or $\mathbb{1}$ on one of the qubits with $1/2$ probability).

Although the upper bound to p_s is saturated by an incomplete Bell measurement, it is not clear, how a complete BSM without stochastic elements would perform. This case can be analysed much in the same way as in the previous section. We can solve the equations over $\mathbb{Q}(\sqrt{2})^3$.

When restricted to four-mode detection patterns, there are 17 sets of equations to check. Only the one corresponding to the class of detection patterns represented by

$$\{(1, 1, 0, 0), (1, 0, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1)\} \quad (4.50)$$

has a non-trivial solution. We already know one solution, which also has to lie in this class – the controlled phase gate ($\varphi = \pi$) in Section 4.3, followed by a Hadamard gate and measurements in the computational basis. After eliminating the free parameters corresponding to rescaling of the output modes, there are still two free parameters left. Therefore, it is not evident, yet, where the controlled phase gate sits in this set of measurement devices.

The solution is the mode transformation given by the matrix

$$A = x \operatorname{diag} \{a, a, a^{-1}, a^{-1}\} \cdot \begin{pmatrix} 2^{-1/2} & 2^{-1/2} & 0 & 0 \\ -2^{-1/2} & 2^{-1/2} & -2^{-1/2} & 2^{-1/2} \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix} \cdot \operatorname{diag} \{b, b, b^{-1}, b^{-1}\} \quad (4.51)$$

³Although we could neglect the global pre-factor of $2^{-1/2}$ such that the resulting matrix only has entries of 1 and -1 , more square roots appear due to the factorial terms in Eqn. (4.46).

where $x, a, b \in \mathbb{R}^+$ are free parameters. The optimal measurement device is given by the solution to the problem

$$\text{maximise } x \tag{4.52}$$

$$\text{subject to } \sigma_i \geq 0 \tag{4.53}$$

$$\sigma_i \leq 1 \tag{4.54}$$

where $\sigma_i, i = 1, \dots, 4$ are the singular values of A . The constraints can be rewritten as polynomial constraints and hence the maximisation problem can be solved exactly (*e.g.*, by using `Mathematica`). The solution $a = -b = 2^{1/4}$, $x = 1/\sqrt{3}$ is obtained, which corresponds due to (3.23) to a success probability of $p_s = x^4 = 1/9$.

The $1/\sqrt{3}$ embodies the reflectivity of the compensation beam splitter, and the whole solution is equivalent to the post-selected controlled phase gate (see Section 4.3).

As a result, by refraining from stochastic local operations, the maximum success probability of a linear optics BSM with additional vacuum is $p_s = 1/9$.

Still, the restriction to four detector modes can be lifted. Surely, by allowing for an arbitrary number of detectors (the maximal useful number is 8), more solutions can be found, in particular ones with higher success probabilities. Apart from the four mode solution, there are five different classes. The one achieving the highest probability of success (identified by numerical optimisation), $p_s \approx 0.126$, uses 5 detector modes with patterns $\{(0, 0, 1, 1, 0), (0, 1, 0, 0, 1), (1, 1, 0, 0, 0), (1, 0, 1, 0, 0)\}$.

Rather than the details of how to actually build such a device, the crucial observation is that the highest probability for a complete BSM to succeed without using stochastic operations is only roughly a fourth of the probability attainable when allowing for stochastic operations. Furthermore, in contrast to a single 50 : 50 beam splitter, a nested network of interferometers is required in the case studied above. Even without stochastic unitaries an easy-to-implement network with a success probability not much below this p_s is induced by the post-selected controlled phase gate.

4.4.3 Photon number resolution

Apart from general schemes specially tailored to cope with photon loss (*e.g.*, Ref. [68]) and new methods for starting cluster generation with bucket detectors [69], most known linear optics quantum computation schemes condition on output modes having a certain photon number.

Unfortunately, reliable photon counter are still not available “off the shelf”. What is state-of-the-art today, are *dichotomic* (or *bucket*) detectors which only give a binary distinction between “vacuum” and “at least one photon”. By using only this type of detectors, linear optics and additional single photons, it is actually possible to count the number of photons in a mode in a way that is sufficient for LOQC. These photon number resolving detectors will not suffer from cross-counts, *i.e.* they will signal an unknown input rather than the wrong number of photons (it is an unambiguous state discrimination). However, this property vanishes when the used bucket detectors are lossy. We prove that no photon counter that is perfect in this sense can be built upon the use of lossy bucket detectors.

We are looking for a setup that should only consist of single photon sources, vacuum, beam splitters, and detectors. It should have one input mode (the signal, mode 0) and the pattern of firing detectors shall distinguish between two, one, and zero photons in the signal mode.

Single photon ancillas

In the first step we are interested in the case where only single photon ancilla modes are available. To this end, we will analyse the output pattern $\omega = (1, \dots, 1)$ on n modes for the ability to distinguish between one and two photons in the signal mode. We require the probability of detecting one click per detector given one signal photon to be strictly positive,

$$\langle 1, 1, \dots, 1 | \mathcal{A} | 1, 1, \dots, 1 \rangle = \text{per} A \neq 0. \quad (4.55)$$

This pattern shall be the indicator for one signal photon. As usual, $A \in \mathbb{C}^{n \times n}$ is the transformation of the creation operators and \mathcal{A} its action on the Hilbert space.

A necessary condition for the network to distinguish between one and two photons unambiguously is that the click pattern caused by ω shall not be reproduced by two signal photons (represented by $\Omega^{(2)} = (2, 1, \dots, 1)$):

$$\begin{aligned} \langle 2, 1, \dots, 1 | \mathcal{A} | 2, 1, \dots, 1 \rangle &= \text{per} A [2, 1, \dots, 1 | \Omega^{(2)}] &= 0 \\ \langle 1, 2, \dots, 1 | \mathcal{A} | 2, 1, \dots, 1 \rangle &= \text{per} A [1, 2, \dots, 1 | \Omega^{(2)}] &= 0 \\ & & \vdots \\ \langle 1, 1, \dots, 2 | \mathcal{A} | 2, 1, \dots, 1 \rangle &= \text{per} A [1, 1, \dots, 2 | \Omega^{(2)}] &= 0. \end{aligned} \quad (4.56)$$

By applying the expansion formula (2.13) to the permanent of the j -th equation in (4.56) with the j -th row fixed, the system (4.56) can be rewritten as the matrix

equation

$$A\alpha = A \begin{pmatrix} 2\text{per}A \\ \text{per}A[\omega|2, 0, 1, \dots, 1] \\ \text{per}A[\omega|2, 1, 0, \dots, 1] \\ \vdots \\ \text{per}A[\omega|2, 1, 1, \dots, 0] \end{pmatrix} = 0. \quad (4.57)$$

This equation has nontrivial solutions iff $\det A = 0$. Because of unitarity of A , α has to vanish which implies $\text{per}A = 0$, which, in turn, poses a contradiction to Eqn. (4.55). Therefore, this restricted setup cannot be used for photon number discrimination, auxiliary vacuum modes have to be considered as well.

Including vacuum modes

The next step will be to allow for empty ancilla modes that are projected onto vacuum afterwards. One straightforward way is to adapt the previous *ansatz* by allowing for a non-unitary A in the following (especially $\det A = 0$). Then, the constraints we want A to fulfil are (4.56) combined with $\det A = 0$.

On one and two modes with photons being present in the auxiliary modes there are no matrices satisfying these conditions. There are six solutions to the problem with three modes. One of them reads

$$A^{(1)} = \begin{pmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ -\frac{29-45\sqrt{3}i}{76} \frac{A_{1,1}A_{2,3}}{A_{1,3}} & \frac{2-\sqrt{243}i}{38} \frac{A_{1,2}A_{2,3}}{A_{1,3}} & A_{2,3} \\ -\frac{3}{2} \frac{A_{1,1}A_{3,3}}{A_{1,3}} & \frac{1-\sqrt{27}i}{4} \frac{A_{1,2}A_{3,3}}{A_{1,3}} & A_{3,3} \end{pmatrix} \quad (4.58)$$

with the five free parameters being not zero.

Having only product input states and detectors in every mode, we still have the gauge freedom of multiplying A by real diagonal matrices Y and X from the left and from the right, respectively. This actually already includes the five parameters $A_{i,j}$, which becomes obvious when substituting $A_{i,j} = A_{1,3}A'_{i,j}$. So, a normal form of this solution is

$$A^{(1)} = \begin{pmatrix} 1 & 1 & 1 \\ -\frac{29-45\sqrt{3}i}{76} & \frac{2-\sqrt{243}i}{38} & 1 \\ -\frac{3}{2} & \frac{1-\sqrt{27}i}{4} & 1 \end{pmatrix}. \quad (4.59)$$

The other solutions are $A^{(2)} = (A^{(1)})^*$, and after appropriate rescaling

$$A^{(3)} = \begin{pmatrix} 1 & 1 & 1 \\ -\frac{1+\sqrt{3}i}{2} & 1 & 0 \\ \frac{1+\sqrt{27}i}{4} & \frac{1-\sqrt{27}i}{4} & 1 \end{pmatrix}, \quad A^{(5)} = \begin{pmatrix} 1 & \frac{2a^2-7a-3(1-a)\tilde{a}-13}{14-4a+8a^2} & \frac{\tilde{a}-a-1}{2} \\ 1 & \frac{a^2-7a-3(1-a)\tilde{a}+1}{14-4a+8a^2} & a \\ 1 & 1 & 1 \end{pmatrix}^T, \quad (4.60)$$

and their complex conjugates, where $\tilde{a} = i\sqrt{3+a(2+3a)}$ and $a \in \mathbb{R}$ is another free parameter.

Unitary extensions

As the $A^{(k)}$ are certainly not unitary, they have to be extended by using additional vacuum modes (see Section 3.1.6). For a three-mode non-unitary transformation, the smallest extension uses four modes. A sufficient condition for A to be unitarily extendable with only one vacuum mode is

$$\langle A_1|A_2\rangle\langle A_2|A_3\rangle\langle A_3|A_1\rangle < 0, \quad (4.61)$$

where A_i is the i -th row- or column vector of A . Using `Mathematica`, the network achieving the maximum probability for detecting one photon given one photon in the signal mode, $p(1|1) = |\text{per}(YAX)[1, 1, 1|1, 1, 1]|^2$ can be found. It is in the class $A^{(5)}$ with $p(1|1) = 1.6 \cdot 10^{-2}$, but violates the condition (4.61). For current experimental implementations the complexity of the network has a huge impact, rather than small deviations from the optimal probability. Therefore, we chose another solution – $A^{(1)}$ – which can be embedded with only one additional vacuum mode, resulting in the mode transformation

$$U = \frac{1}{\sqrt{159}} \begin{pmatrix} \sqrt{46} & \sqrt{46} & \sqrt{46} & \sqrt{\frac{3}{2}(-13+i\sqrt{27})} \\ \frac{-29-45i\sqrt{3}}{2\sqrt{38}} & \frac{2+i\sqrt{243}}{\sqrt{38}} & \sqrt{38} & -\sqrt{\frac{23}{76}}(\sqrt{147}+9i) \\ -\frac{9}{\sqrt{2}} & \frac{3}{4}(\sqrt{2}+i\sqrt{54}) & \sqrt{18} & \sqrt{69} \\ \sqrt{27} & -\frac{5}{2}(\sqrt{3}-3i) & -\frac{1}{2}(\sqrt{3}+15i) & 0 \end{pmatrix} \quad (4.62)$$

with $p(1|1) = \text{per}U[1, 1, 1, 0|1, 1, 1, 0]^2 = 2254/446631 \approx 5 \cdot 10^{-3}$. Patterns for two photon detection can be discriminated from false one-photon outcomes by subsequently splitting modes 2 and 3 by means of 50 : 50 beam splitters and conditioning on four clicks in the detectors. This leads to a two photon detection probability of $p(2|2) = |\text{per}U[0, 2, 2, 0|2, 1, 1, 0]|^2 / 32 = 251275/15780962 \approx 1.6 \cdot 10^{-2}$.

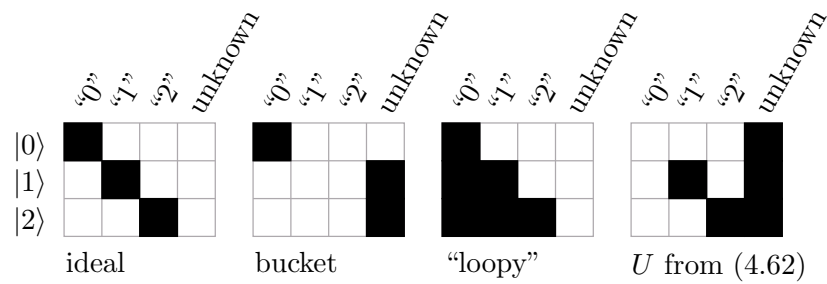


Table 4.1: Output signals of photon counters given zero, one, or two photons. Black squares shall symbolise non-zero probabilities, the detection events marked by white squares do not occur at all. Dark counts and photon loss would result in more cross-counts in all the cases.

With the construction of the extension in mind (*c.f.* Fig. 3.1) this network consists of two general three-mode interferometers with an additional vacuum mode coupled in between them. The additional beam splitters used to spread the photons to additional output modes also do not need any interferometric stabilisation as they only mix photons with the vacuum in front of the detectors.

Applications

Using such a photon counter allows us to discriminate between one and two photons. However, for all possible applications a distinction of zero photons in the signal mode is necessary. As the one- and two-photon inputs are detected by a threefold and fourfold coincidence count, respectively, a zero-photon input cannot result in a false pattern and will always be detected as unknown (see Table 4.1 for comparison of commonly used types of detectors). It turns out that these properties are sufficient for the counter to be used in the type-I fusion gate [38] and in a circuit for EPR preparation using only single photon resources [60]. Still, it is not enough to implement the nonlinear sign shift gate [37]. For this primitive element of a circuit based model of computation, a distinction between zero, one and “more than one” photons is required.

Fusion The two measured modes of a type-I fusion gate contain $s_1 = (1, 0)$ or $s_2 = (0, 1)$ photons on success, in contrast to $f_1 = (2, 0)$, $f_2 = (0, 2)$, or $f_3 = (0, 0)$ photons on failure. By placing a bucket detector in the second mode we can filter out the events s_1 , f_1 , and f_3 by post-selecting on a vacuum detection. Now, these three patterns can be discriminated by our counting circuit attached to the first mode in such a way that it only claims a successful outcome if there was exactly one photon, so the success indicating pattern s_1 . For not having a detector that

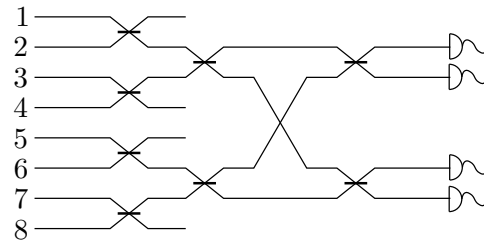


Figure 4.10: The circuit from [60] for EPR pair production. Which type of detectors are placed in modes 3, 4, 7, and 8 is explained in the text. The output state will be generated in modes 1, 2, 5, and 6. All the beam splitters are 50 : 50 and single photons are put in modes 1, 4, 5, and 8, all other modes are initialised in the vacuum state.

can fully discriminate between zero, one, and two photons, we only used one of the success outcomes and have therefore lowered the success probability even more.

With this fusion gate at hand, cluster states can be generated using only EPR pairs and single photons (another technique to start from EPR pairs without the need for photon counters is described in [69]).

EPR production The circuit for EPR production is shown in Fig. 4.10. Again, we concentrate on a single success outcome. An EPR pair is produced in the modes (1, 2, 5, 6) when there is a single photon detected in both of the modes 3 and 4, and vacuum in the latter pair (7 and 8). We place bucket detectors in the vacuum modes and counting circuits in the single photon ones. What false patterns could occur? On the one hand, there could be two or none instead of single photons, but these errors can be filtered out with the circuit introduced above. On the other hand, higher Fock layers (like the events (3, 1) and (1, 3)) could lead to false success. However, as can be easily seen from Fig. 4.10, the output of the EPR circuit does not contain the monomials $(a_3^\dagger)^3 a_4^\dagger$ or $a_3^\dagger (a_4^\dagger)^3$. As there are not more than four photons in the circuit, we already checked all possible patterns. Therefore, these ingredients are sufficient to produce EPR pairs with single photons, linear optics, and bucket detectors.

Non-demolition measurements Another application is the implementation of non-demolition photon counters. In Ref. [70] a network is proposed which is able to discriminate between zero, one, and two photons in a signal mode non-destructively. From the signals of photon counters, the events of zero and one photon in the signal mode can be inferred. The two-photon event results in a failure-outcome, which is sufficient for the KLM scheme [37], for example. What is required to decide between one and two photons are detectors able to tell the state $|0111\rangle$ (one-photon input)

apart from the set of states $\{|0022\rangle, |1111\rangle, |1012\rangle, |0121\rangle\}$ (two-photon input), with the signal carried by the second mode. A bucket detector in the first mode and two instances of the counting device introduced above in modes three and four already achieve this goal.

Realistic detectors efficiencies

In the previous section we assumed perfect detectors, *i.e.*, every photon that enters the detector will result in a detection event. However, this is still far from reality and lossy detectors have to be considered. Actually, the detection patterns introduced before are already tolerant against loss. It is only the vacuum extension that leads to problems: the post-selection onto a vacuum detection cannot tolerate photon loss because bright events can be mistaken for darkness. This might cause the two-photon pattern $(1, \dots, 1, 1)$ to be recognised as one photon (the pattern $(1, \dots, 1, 0)$). Unfortunately, the $(1, \dots, 1, 1)$ -pattern cannot be excluded from the two-photon events. In the example above, the false count probability is $p(1|2) = \text{per}U[1, 1, 1, 1|\Omega^{(2)}]^2/2 = 644/8427 \approx 7.6 \cdot 10^{-2}$.

Not being able to discriminate between photon numbers when relying on lossy bucket detectors is actually not restricted to this network. Rather, it is a general property of linear optics. To prove this quite intuitive statement, the following observation will be useful.

Let $\omega^{(1)}$ be the detection pattern such that a one photon input can be inferred from it unambiguously. By using linear optics (mode permutations and mixing with the vacuum), $\omega^{(1)}$ can be transformed into the pattern with each mode being populated by at most one photon. Although the detection probabilities decrease, the detector will still discriminate unambiguously between the same photon numbers.

Now, a no-go theorem for unambiguous photon counting can be stated⁴:

Theorem 7 (No photon counters with linear optics). *Consider linear optics, lossy bucket detectors, single photon sources and vacuum. With these tools alone it is not possible to discriminate between one- and two-photon states unambiguously.*

Proof. Let $n - 1$ be the number of auxiliary photons, $n + k$ the total number of modes available to the network (w.l.o.g. we assume $k \geq 0$) and $A \in \mathbb{C}^{(n+k) \times (n+k)}$ the linear transformation of the creation operators. The aim will be to show that $A \notin SU(n + k)$ for all product input states for all $n \in \mathbb{N}$ and $k \in \mathbb{N}$.

⁴This result is already known from [71]. However, using the methods introduced before, another approach might be insightful as well.

Further, let $\Omega^{(l)} \in \mathbb{N}^\nu$ with $\Omega_1^{(l)} = l$, $\Omega_2^{(l)} \geq \dots \geq \Omega_\nu^{(l)} > 0$ and $\sum_i \Omega_i^{(l)} = n - 1 + l$, $l = 0, 1, 2$ be the possible input patterns (the remaining $n + k - \nu$ modes are initialised in the vacuum). Concerning detection patterns for a single input photon, as mentioned above we only have to care about $\omega^{(\nu)} = (1, \dots, 1)$, and we demand

$$\text{per}[\omega^{(1)}|\Omega^{(1)}] \neq 0. \quad (4.63)$$

In order to exclude any two-photon events which are wrongly interpreted as one-photon ones, A has to satisfy the $n + k$ equations

$$\text{per}A[\omega + e_i|\Omega^{(2)}] = 0, \quad i = 1, \dots, n + k. \quad (4.64)$$

We expand the permanent in Eqn. (4.64) in the i -th row and obtain the matrix equation

$$(A_{i,j})_{\substack{i=1,\dots,n+k \\ j=1,\dots,\nu}} \left(\Omega_j^{(2)} \text{per}A[\omega^{(1)}|\Omega^{(2)} - e_j] \right)_{j=1,\dots,\nu}^T = 0. \quad (4.65)$$

By taking any subset π of ν of these equations we can compose $\binom{n+k}{\nu}$ sets of equations, each of which has no trivial solution due to (4.63). In turn, the existence of a non-trivial solution imposes the constraint

$$\det A \left[\sum_{i=1}^{\nu} e_{\pi(i)} \left| \sum_{j=1}^{\nu} e_j \right. \right] = 0 \quad (4.66)$$

on the coefficient matrices. By successively expanding $\det(A)$ in the last $n + k - \nu$ columns and substituting (4.66) for all π , $\det A = 0$ follows immediately. This contradicts the requirement of unitarity for a physical implementation in terms of linear optics for any choice of n and $\Omega^{(0)}$, and any combination of detection patterns. \square

An intuitive explanation of this result is that there is no big difference between the detectors themselves being lossy, some of the linear optics components in front of them, or rather the input to the whole network being constituted of less photons. Therefore, it seems plausible, that there is no detection event that allows lossy detectors to tell apart one photon from two photons with certainty.

Comparison with time-multiplexed loop detectors An alternative approach that can be found in the literature [72, 73] is that of a multiplexed bucket detector, or “*loopy*” detector. To describe this device we use the following model: The signal mode is distributed into k detector modes with equal amplitudes. All detectors are assumed

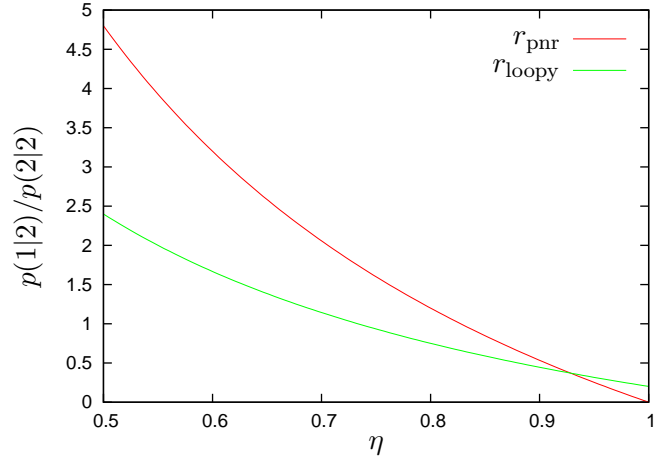


Figure 4.11: Ratio of the false count probability $p(1|2)$ with respect to the success probability of two photon detection $p(2|2)$. The two curves represent the network introduced in the text and a “loopy” detector, both based on lossy bucket detectors.

to work with the same efficiency $\eta \in (0, 1]$. Obviously, even in the case of perfect detectors, the probability of false counts is strictly larger than zero. Given two input photons, the probability of detecting only one is

$$p_{\text{loopy}}(1|2) = \frac{1}{2} [\eta^2 + \eta(1 - \eta)(2k - 1)]. \quad (4.67)$$

On the other hand, the false count probability of our setup is given by

$$p_{\text{pnr}}(1|2) = \frac{\text{per}A[1, 1, 1, 1|2, 1, 1, 0]^2}{2} (1 - \eta)\eta^3. \quad (4.68)$$

For a comparison of this network with “loopy” we allow for the same number of detectors – $k = 6$ – and operate the “loopy” detector under the same conditions (*i.e.*, output “0” will be interpreted as “unknown”), lowering the rate of successful detections, but lowering the cross-count rates to $p_{\text{loopy}}(0|2) = p_{\text{loopy}}(0|1) = 0$. A fair figure of merit should also take the success probability itself into account.

Therefore, we will consider the ratio $r := p(1|2)/p(2|2)$, which is shown in Fig. 4.11. For today’s detector efficiencies (which can be assumed to reach up to 70%) the relative failure rate of our setup turns out to be higher than the “loopy” one. But when going to large η , r_{loopy} converges to the constant $(k - 1)^{-1}$, while r_{pnr} tends to zero. So even with imperfect detectors $r_{\text{pnr}} < r_{\text{loopy}}$ for $\eta > 85199/91754 \approx 0.93$ the presented setup will provide advantages over multiplexed counting although it suffers from detector inefficiencies as well.

On the same number of modes it remains, however, significantly harder to experimentally construct a network that uses interferometers compared to a cascade of beam splitters. Please note, that leaving the set of devices we restricted ourselves to in the beginning can mean also allowing for qualitatively better detection devices. For example counting detectors which are essentially integrated cascades of a large number of small bucket detectors on one chip – splitting up the signal mode into a multitude of detector modes – allow for significant improvements such as discrimination between 1 and more photons with a cross count rate of less than 1% (see *e.g.*, [74]).

4.5 Discussion

This chapter was concerned with the application of the methods outlined in Chapter 3 to a range of small problems in linear optics.

Some of the examples were not discussed in the literature and are merely used to illustrate the techniques. We showed the construction of highly specialised networks for (i) the generation of 5-photon NOON states and (ii) the projective measurement of states in the tetrahedral basis for the purpose of state estimation.

Of more interest are the examples dealing with problems already introduced earlier. While it is known that linear optics with feed-forward is able to distinguish between all four Bell states with an arbitrarily high success probability $p_s < 1$ [37], it relies on a supply with highly entangled auxiliary states. With only auxiliary vacuum a bound of $p_s \leq 1/2$ was shown before [63]. We discussed this measurement problem in an even more static setting. Without prior knowledge about the input state it turns out that the application of random local unitaries plays a crucial role: without it only $p_s \approx 0.126$ (obtained by numerical optimisation) can be achieved.

Further, the importance of photon number discrimination is highlighted by analysing the possibility of simulation of a fair photon counter with linear optics and bucket detectors. No such circuit is possible without auxiliary vacuum modes or non-perfect detectors. This is a result shown in Ref. [71] and we give a new version of the proof with our methods. Also, a network is constructed which achieves this job in the case of perfect detectors.

A prominent example appearing quite often in recent experiments [23, 53, 54] is the post-selected controlled phase gate. Within the framework of linear optics, auxiliary vacuum modes and photon detectors, we constructively show the optimal success probability of networks implementing such gates for arbitrary phases $0 \leq \varphi \leq \pi$. In the special case $\varphi = \pi$ usually discussed in the literature it is shown that

the original network is already an optimal one. Furthermore, and quite surprisingly, we discover that $p_s(\varphi)$ is not monotonous, therefore limiting the procedure of using specially tailored gates instead of the usual gate set in the circuit model to improve the probabilities of success. We also apply the same methods to the phase gate with two control qubits (for $\varphi = \pi$ equivalent to the Toffoli gate) and observe a similar behaviour. For $\varphi = \pi$, the previously known maximum of $p_s = 1/72$ [62] is shown to be obsolete when considering the full class of these gates, where we show the maximum to be $p_s \approx 1/57$.

Finally, we considered a class of networks with a complexity allowing for experiments today. Being small and using only few photons we show how networks for the purpose of state transformation of dual-rail qutrits and ququads (so states of two or three photons on two modes) can be constructed.

We now turn to the preparation of cluster-, or more generally, graph states – resource states for measurement-based quantum computation. The measurement-based model is a desirable route to quantum computation in a number of architectures, for example, where nearest-neighbour interactions or the topology of clusters are inherent in the architecture. Also, when multi-qubit gates are difficult to apply, this route can be advantageous over schemes based on the gate model due to the very clear distinction between the preparation of entanglement, and the consumption of it by means of local measurements.

Although, in principle, an efficient scaling of linear optics quantum computation was proved already in the gate model [37], alternative routes might offer far less demanding schemes. Independent of the actual algorithm to be run, one can concentrate on the construction of cluster states with a fixed set of gates in the preparation step, rather than optimising over all possible networks in general – quite a hopeless enterprise for a large number of qubits. In the computation stage, only single qubit rotations, which correspond to suitable phase plates in the usual polarisation based dual-rail encoding and which work deterministically, are required in front of detectors.

A *graph state* [75, 76, 77] is a quantum state described by an undirected graph $G = (V, E)$, with V being the set of vertices, and E the set of edges. The vertices are embodied by physical systems, so single spins or qubits (in the optical case usually single photons). The edges represent interactions. More specifically, a *graph state* is the simultaneous eigenstate to the set of *stabilisers*

$$K_G^{(a)} := \sigma_x^{(a)} \bigotimes_{b:\text{dist}(a,b)=1} \sigma_z^{(b)} \tag{5.1}$$

for all $a \in V$ with eigenvalue $+1$, *i.e.*,

$$K_G^{(a)} |\psi_G\rangle = |\psi_G\rangle \quad \forall a \in V. \tag{5.2}$$

Here, $\text{dist}(x, y)$ is the graph-theoretical distance [78] between vertices x and y on G , *i.e.*, $\text{dist}(x, y) = 1$ exactly for neighbouring vertices. From here on, $\sigma_{x,y,z}^{(b)}$ denote Pauli operators with support on the Hilbert space of the physical system labelled b .

Equivalently, this state can be thought of as having each qubit prepared in the state $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$ and applying an interaction leading to a controlled-Z-gate

$$U_{CZ} := \sum_{i,j} (-1)^{ij} |ij\rangle\langle ij| \quad (5.3)$$

to neighbouring vertices $a, b \in V$, so to vertices that are connected by an edge $(a, b) \in E$. Because these gates are diagonal in the computational basis, they commute, and hence the order in which they are applied does not influence the final state.

Now, a *cluster state* is a special graph state with the square lattice ($V \in \mathbb{Z}^2$, $E = \{(a, b) : |a - b| = 1\}$) as the underlying graph. In the following chapters we will also consider graph states on other regular lattices (such as hexagonal-, diamond-, and pyrochlore lattices). Note, that this definition is in no way restricted to two or three dimensions. The dimensionality only refers to the abstract underlying graph that defines the correlations, rather than actual physical dimensions.

However, one can think of the above lattice structure emerging from a physical lattice as well. This may be an optical lattice generated by standing wave laser light, where atoms are located at individual lattice sites [79]. By means of interactions or controlled collisions, implementing (5.2) or (5.3), respectively, a cluster state can in principle be prepared in such systems. Different states that might provide similar computational resources can be obtained as ground states of a wide range of many-body Hamiltonians [56].

The simple description of a graph state by its stabilisers is supported by simple rules how Pauli measurements of single qubit transforms the underlying graph. Let us briefly summarise the effective action of these operations when applied to a qubit constituting a vertex $a \in V$ of a graph state $|\psi_G\rangle$ with $G = (V, E)$ [76]. The neighbourhood of a vertex a will be denoted by $N_a := \{b : (a, b) \in E\}$, the subgraph of G induced by $A \subset V$ is $G[A] := (A, \{(a, b) \in E : a, b \in A\})$, and the complement of a graph G with respect to the set of all possible edges by $G^c := (V, \{(a, b) : a, b \in V\} \setminus E)$. The measurement rules now read

- A σ_z “cuts” the vertex a . The graph representing the resulting state is $G' = G[V \setminus \{a\}]$.

- σ_y measurements perform local complementations: $G' = (G[V \setminus \{a\}] \setminus G[N_a]) \cup G[N_a]^c$. So, it swaps all possible edges in its immediate neighbourhood.
- Further, we will encounter σ_x measurements only in the special case where the state in the neighbourhood of the affected qubit looks like a cluster chain, so $V \subset \mathbb{Z}$ and $E = \{(a, b) : |a - b| = 1\}$. Upon measurement of the $\sigma_x^{(a)}$ operator, the stabilisers which result from those having support on a are $\sigma_z^{(a-2)} \sigma_x^{(a-1)} \sigma_x^{(a+1)} \sigma_z^{(a+2)}$ and $\sigma_z^{(a-1)} \sigma_z^{(a+1)}$. This will either be interpreted as a redundantly encoded qubit (qubits $a - 1$ and $a + 1$ locally look like a GHZ state), or – after application of a Hadamard $H^{(a-1)}$ or $H^{(a+1)}$ – as a T -junction with a single dangling edge, $G' = (G[V \setminus \{a\}] \setminus G[\{a - 1, a + 1, a + 2\}]) \cup G[a - 1, a + 1, a + 2]^c$. Measurement rules for the redundancy qubits follow from the rotated rules for cluster states, *e.g.*, a σ_x measurement removes one qubit from the encoding.

The resulting graphs G' are actually only obtained after applying some local unitaries (*by-product operators*) to the neighbours of a , depending on the measurement outcomes. Instead of applying the unitaries right after the measurement, the cheaper way is to keep track of them and instead rotate future measurement bases accordingly.

5.1 The one-way computer

It has been shown that local single-qubit measurements in the x - y -plane on constituents of such a cluster state are just as powerful as the gate model for quantum computation, and the one-way model can hence efficiently simulate any other quantum computer [15, 80, 81]. That is, for a given finite computation there exists a finite graph with $V \subset \mathbb{Z}^2$ such that the computation can be simulated by a sequence of suitable local single-qubit measurements on the cluster state. The usual mapping from a circuit model to a cluster state allows to label one dimension as time which comes in handy because only time slices of a cluster actually have to exist at a given point during the computation. We will also frequently encounter finite subsets of the hexagonal lattice which is a computational resource in the same sense.

Similar to the outcome-dependent local rotations after Pauli measurement, such updates are necessary after general measurements as well. Keeping the accumulated rotations (*by-product operators*) in mind, changing the bases of subsequent measurements appropriately can be accomplished efficiently with a classical computer.

5.2 Resource state generation by probabilistic gates

Especially in linear optics, success probabilities are inherent in entangling gates. Still, whether the gate succeeded is known, in contrast to the case of blind losses (as in absorption). In this heralded setting, efficient quantum computing in the gate model with an arbitrarily high success rate $P_s < 1$ was shown to be possible with linear optics [37].

Although efficient in the sense of polynomial scaling of the overhead required, this circuit-based scheme with thousands of auxiliary qubits for a single gate should be seen as a proof of principle, rather than manual how to build optimal linear optics circuits. Merging the best of two worlds, it was first introduced in Ref. [82] how to apply linear optics' probabilistic entangling gates to the cluster computation model to achieve much more promising resource requirements. Arguments towards the general scaling included subsequent probabilistic addition of single qubits – thereby endangering the whole structure created so far upon failure of the gates (as also discussed in Section 5.2.8) – as well as “micro-clusters” – star-like structures allowing for multiple attempts and thereby featuring a stable growth. The latter approach utilises $O(N^2 \log(N^2))$ qubits to generate a cluster state of size $N \times N$. So far it was the usual method – also applied in Refs. [83, 84] and also Ref. [85] for higher dimensional states. Also, highly redundantly encoded vertices (as proposed in Ref. [38]) give rise to essentially the same scaling behaviour.

In contrast, proceeding in two stages – first building chains and connecting them afterwards (as introduced in Ref. [38] and discussed for small chains in Ref. [86]) – shows to be less wasteful with respect to the invested resource. This new strategy, which we will follow here, allows to identify a scheme with the optimal $O(N^2)$ scaling behaviour. This is possible in the chain-based approach because the required overhead is allocated “per-chain” rather than “per-site”. How to distribute this buffer turns out to be crucial as it can lead to schemes which do not allow for general scaling (fixed overhead per site as in Ref. [86]), require a super-optimal amount of resources ($O(N^2 \log(N^2))$) as in Refs. [83, 84, 85]), or an optimal scaling of $O(N^2)$.

During the first stage it will show that “recycling” chains from failed applications of the entangling gates (as introduced and discussed in Ref. [38] and Ref. [86]) is a very complicated but crucial detail. We will formalise it in the notion of “strategies” that will be of central interest during the first stage. This will allow us to exceed the linear performance indicated in Ref. [83] and find the optimal procedure within the chosen restrictions.

The full potential of such a cluster state based scheme was not known so far in full generality, rather than in examples for small N . Although the possibility for a linear growth was already indicated in Refs. [38, 83], we will attempt to prove it in a rigorous way. Especially the various averages occurring in such a probabilistic process – implicitly merged and interchanged in earlier proposals – will be highlighted and handled with care.

When probabilistic gates are employed in the production of cluster states, a variety of effects can be observed that influence the actual scaling behaviour of the production process. Here we will start with a fixed reservoir of EPR pairs and study the size of the states that can be produced under the action of probabilistic two-qubit gates. In fact, the very choice of where and in which particular order these gates are applied turns out to be crucial for the scaling of the final state's size.

Given a certain gate probability, we will aim at finding the optimal procedure, *i.e.* the one that finishes with the largest state on average [1, 2, 5]. In the first step, converging hierarchies of linear upper and lower bounds to the maximal size of linear cluster states will be given. After that, a scheme that processes these chains to two-dimensional cluster states which are the actual resources for the one-way computer will be given, the performance of which achieves the optimal scaling behaviour.

5.2.1 Probabilistic gates used in cluster state production

For the task of joining intermediate cluster states together to form longer chains, we restrict ourselves to two qubit gates that act symmetrically and with the same action for chains of all lengths – most suggested quantum gates do have this property.

Fusion

Mainly we will employ entangling gates such as the type-I fusion gate [38], which is essentially a destructive parity check [87] with suitable error outcomes (see also Fig. 5.1).

Linear cluster states can be written in the form

$$|\psi_N\rangle = N^{-1/2} \left(\bigotimes_{i=1}^{N-1} (|0\rangle + |1\rangle \sigma_z^{(i+1)}) \right) \otimes (|0\rangle + |1\rangle). \quad (5.4)$$

A parity check which can be described by the Kraus operators $A_{\pm} = 2^{-1/2}(|0\rangle\langle 00| \pm |1\rangle\langle 11|)$ acts on the end qubits of two cluster chains [19] consisting of N and M

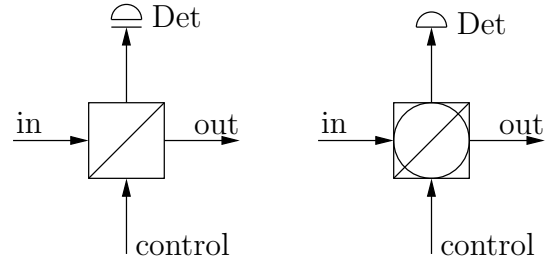


Figure 5.1: Linear optics “fusion” gates [87, 38]. Left: A quantum parity check gate (*type-I fusion gate*). The detector projects onto the diagonal basis’ elements. Right: The destructive CNOT gate. The PBS acts in the diagonal basis and is therefore an ordinary PBS embedded by four Hadamard beam splitters. With another detector in the output, this corresponds to the *type-II fusion gate*. Depending on the measurement outcomes, local rotations of the output (feed-forward) has to be applied in both cases.

qubits (so $N - 1$ and $M - 1$ edges, respectively) according to

$$A_{\pm}^{(N,N+1)} |\psi_N\rangle |\psi_M\rangle = \cdots (|0\rangle + |1\rangle \sigma_z) \quad (5.5)$$

$$A_{\pm}^{(N,N+1)} (|00\rangle + |01\rangle \sigma_z + |10\rangle + |11\rangle \sigma_z)$$

$$(|0\rangle + |1\rangle \sigma_z) \cdots$$

$$= \cdots (|0\rangle + |1\rangle \sigma_z) (|0\rangle \pm |1\rangle \sigma_z) / 2 (|0\rangle + |1\rangle \sigma_z) \cdots \quad (5.6)$$

$$= \sqrt{p_{\pm}} \sigma_{\pm}^{(N)} |\psi_{N+M-1}\rangle, \quad (5.7)$$

where $\sigma_+ = \mathbb{1}_2$ and $\sigma_- = \sigma_z$ are local corrections. The remaining $N + M - 1$ qubits constitute a cluster state of $N + M - 2$ edges. Further Kraus operators $B_1 = \langle 10|$ and $B_2 = \langle 01|$ describe failure of the gate (σ_z measurements on qubits N and $N + 1$), resulting in two cluster chains of $N - 1$ and $M - 1$ qubits (see Fig. 5.2 for an illustration). The total success probability is $p_s = p_+ + p_- = 1/2$.

In the case of linear optics, optimality of the fusion gate can be shown. The gate’s probability of success is $p_s = 1/2$ and the following theorem states that this cannot be increased in the setting of dual-rail encoded linear optical quantum computation without the use of auxiliary photons.

Theorem 8 (Maximum probability of success of fusion). *The optimal probability of success p_s of a type-I fusion quantum gate is $p_s = 1/2$. More specifically, the maximal $p = p_+ + p_-$ such that*

$$A_{\pm} = \sqrt{2p_{\pm}} (|H\rangle\langle H, H| \pm |V\rangle\langle V, V|) \quad (5.8)$$

are two Kraus operators of a channel that can be realised with making use of

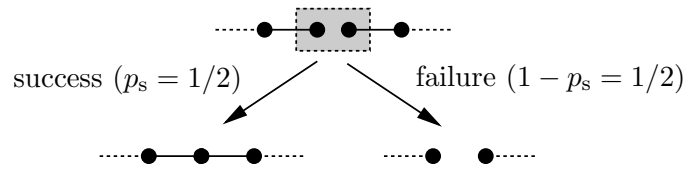


Figure 5.2: Action of a fusion gate on the end qubits of two linear cluster states. In case of a successful outcome, the two chains are connected to a larger one. Otherwise the ends are cut off.

(i) any number of auxiliary modes prepared in the vacuum,

(ii) linear optical networks acting on all modes, and

(iii) photon counting detectors

is given by $p_s = 1/2$.

Proof. Given the setup in Fig. 5.3, we notice a parity check described by these Kraus operators can be used to realise a measurement, distinguishing with certainty two from four two-partite Bell states: The following Hadamard gate and the measurement in the computational basis give rise to the Kraus operators

$$B_{\pm} = \langle \pm | = 2^{-1/2} (\langle H | \pm \langle V |). \quad (5.9)$$

On input of the symmetric Bell states with state vectors, $|\phi^{\pm}\rangle = 2^{-1/2}(|H, H\rangle \pm |V, V\rangle)$, the measurement results (A_-, B_-) and (A_+, B_+) indicate a $|\phi^+\rangle$, and (A_-, B_+) and (A_+, B_-) a $|\phi^-\rangle$, respectively. These two states can be identified with certainty. The anti-symmetric Bell states with state vectors $|\psi^{\pm}\rangle = 2^{-1/2}(|H, V\rangle \pm |V, H\rangle)$ in turn, will result in a failure outcome.

Applying a bit-flip on the second input qubit (therefore implementing the map $|\phi^{\pm}\rangle \mapsto |\psi^{\pm}\rangle$, $|\psi^{\pm}\rangle \mapsto |\phi^{\pm}\rangle$) at random, a discrimination between the four Bell states with uniform *a priori* probabilities is possible, succeeding in 50% of all cases (see Section 4.4.2 for more details on Bell state discrimination). Following Ref. [63] this is already the optimal success probability when only allowing for (i) auxiliary vacuum modes, (ii) networks of beam splitter and phase shifts and (iii) photon number resolving detectors. Thus, a more reliable parity check is not possible within the presented framework. \square

Note that one could in principle use additional single-photons from sources or EPR pairs to attempt to increase the success probability p_s of the individual gate. These additional resources would yet have to be included in the resource count. Such a generalised scenario will not be considered.

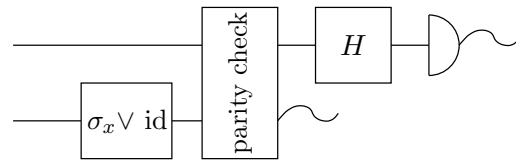


Figure 5.3: Diagram illustrating how a parity check gate can be employed to realise a Bell state discriminating device.

Other gates

How a controlled- Z gate can be used to connect cluster chains is obvious from their definition. At this point we will neglect the actual physics behind the gates. The effective actions on two chains of lengths l_1 and l_2 (now in the number of edges) of the family of gates shown in Table 5.1 can then be described by the outcomes on *success* and *failure*,

$$\{l_1, l_2\} \mapsto \{l_1 + l_2 + d_s\}, \quad \text{and} \quad (5.10)$$

$$\{l_1, l_2\} \mapsto \{l_1 - d_f, l_2 - d_f\}, \quad (5.11)$$

respectively, and are cast into the tuple $d = (d_s, d_f)$. This family embodies four gates: On the one hand it contains two with an undefined failure outcome. To obtain a proper cluster state, an additional Z -measurement on the neighbours has to cut off the end qubits, thus $d_f = 2$. Additionally, there are those gates with a “built-in” σ_z -measurement on failure, so $d_f = 1$. On the other hand, there are two “parity check” gates (no new edges are created, $d_s = 0$) and two controlled- Z gates (entanglement is created, $d_f = 1$). Unless stated otherwise we will assume the fusion gate ($d_{\text{fusion}} = (0, 1)$ and $p_s = 1/2$), which is most relevant for linear optics.

5.2.2 Concepts: configurations and strategies

The current section will set up a rigorous framework for the description and assessment of control strategies. All considerations concern the case of one-dimensional cluster states; the two-dimensional case will be deferred to Section 5.2.8. Note that having described the action of the elementary gate on the level of graphs, we may abstract from the quantum nature of the involved cluster states altogether.

Configurations

A *configuration* (in the *identity picture*) I is a list of numbers $I_k, k \in \mathbb{N}$. We think of I_k as specifying the length of the k -th chain that is available at some instance of

| Gate | d_f | d_s | Physical realisation |
|------------------|-------|-------|--|
| CZ | 2 | 1 | Distant atoms [88, 89, 86] |
| KLM CZ | 1 | 1 | Linear optics with $p_s = 1/16$ [37] or $p_s = 1/4$ [87]; weak non-linearities with $p_s = 3/4$ [90]; linear optics with $p_s = p_{\text{NDM}}/9$ [54, 53]; $p_s = 1/8$ [91] |
| DPC ¹ | 2 | 0 | Trapped atoms and frequency qubits [92], $p_s < 1/4$ |
| Fusion | 1 | 0 | Linear optics parity check [38], optimal $p_s = 1/2$ [2] |

Table 5.1: The four quantum gates described in the text. (d_s, d_f) denote the number of edges gained on success and the number of edges deleted per chain on failure, respectively. p_{NDM} is the probability of success of a photon number non-demolition measurement that has to follow the respective gate.

time. For most of the statements to come a more coarse-grained point of view is sufficient: in general we do not have to distinguish different chains of equal length. It is hence expedient to introduce the *anonymous representation* of a configuration C as a list of numbers $C_i, i \in \mathbb{N}$ with C_i specifying the number of chains of length i . We will always use the latter description unless stated otherwise. Trailing zeroes will be suppressed, *i.e.*, we abbreviate $C = \{1, 2, 0, \dots\}$ as $C = (1, 2)$. Define the *total number of edges* (total length) to be $L(C) := \sum_i i C_i$ (and $L(I) := \sum_i I_i$ in the identity picture) and the *number of non-trivial chains* as $|C| := \sum_i C_i$ ($|I|$ accordingly). The space of all configurations is denoted by \mathcal{C} . By $\mathcal{C}^{(N)}$ we mean the set of configurations C having a total length less than or equal to N . Lastly, let e_i be the configuration consisting of exactly one chain of length i . This definition allows us to expand configurations as $C = \sum_{i=1}^{\infty} C_i e_i$.

Elementary rule

Let us re-formulate the action of the gates introduced above in this language. An attempted fusion of two chains of length k and l gives rise to a map $C = \sum_{i=1}^{\infty} C_i e_i \mapsto C' = \sum_{i=1}^{\infty} C'_i e_i$ with

$$C' = C - e_k - e_l + e_{k+l+d_s} \quad (5.12)$$

in case of success with probability $p_s = 1/2$ (leading to a single chain of length $k + l - d_s$) and

$$C' = C - e_k + e_{k-d_f} - e_l + e_{l-d_f} \quad (5.13)$$

¹Destructive parity check.

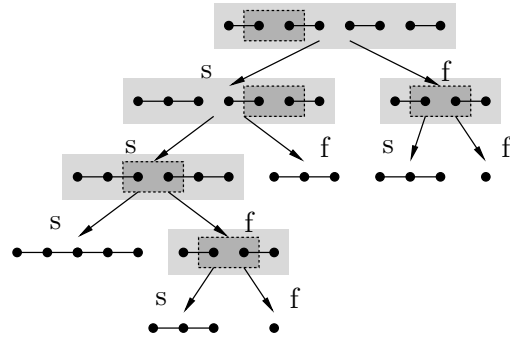


Figure 5.4: An example of a tree of successive configurations under application of a strategy. Light boxes group configurations. We start with $N = 4$. Dark boxes indicate where the strategy decided to apply a fusion gate. Possible outcomes are success (to the left) or failure (to the right), resulting in different possible future choices. The expected length of the final chain is $\tilde{Q}_M(4) = Q(4) = 13/8$.

in case of failure, meaning that one edge each is lost from the chains of length k and l . All other elements of C are left unchanged.

Strategies

A *strategy* defines what *action* to take when faced with a specific configuration (see also Fig. 5.4). Actions (in the anonymous picture) can be either “try to fuse a chain of length k with one of length l ” or “do nothing”. Formally, we will represent these choices by the tuple $\langle k, l \rangle$ and the symbol \emptyset , respectively. The expected gain per fusion attempt is $\bar{\Delta} = (d_s + 1)p_s - d_f(1 - p_s)$, which is zero for “balanced” gates such as the fusion gates, where it hence never pays off not to use all available resources. Due to the motivation by the fusion gate, we require a strategy to choose a non-trivial action as long as there is more than one chain in the configuration. This requirement will be relaxed for obtaining computer-assisted results in the $\bar{\Delta} < 0$ regime where it can be favourable to keep the longest chain and halt before reaching a single-chain configuration.

Formally, a strategy is said to be *valid* if it fulfils

1. (No null fusions): $S(C) = \langle k, l \rangle \Rightarrow C_l, C_k \neq \emptyset$
2. (No premature stops): $S(C) = \emptyset \Leftrightarrow C$ contains at most one chain.

We will implicitly assume that all strategies that appear are valid. Strategies in the identity picture are defined completely analogously.

An *event* E is a string of elements of $\{S, F\}$, denoting success and failure, respectively. The length of E is denoted by $|E|$, its i -th component by E_i , and the set of events with length k by \mathcal{E}_k . Now fix an initial configuration C_\emptyset and some strategy

S . We write C_E for the configuration which will be created by S out of C_\emptyset in the event E . Here, as in several definitions to come, the strategy S is not explicitly mentioned in the notation. It is easy to see that any strategy acting on some initial configuration will, in any event, terminate after a finite number of steps $n_T(C)$.

Recall that the outcome of each action is probabilistic and *a priori* we do not know which C_E with $|E| = n$ will have been obtained in the n -th step. It is therefore natural to introduce a probability distribution on \mathcal{C} , by setting

$$p_n(C) := 2^{-n} |\{E : |E| = n, C_E = C\}|. \quad (5.14)$$

In words: $p_n(C)$ equals 2^{-n} times the number of events that lead to C being created after n steps. The fact that S terminates after a finite number of steps translates to $p_{n_T+k} = p_{n_T}$ for all positive integers k . Expectation values of functions f on \mathcal{C} now can be written as

$$\langle f \rangle_{p_n} := \sum_C p_n(C) f(C). \quad (5.15)$$

The *expected total length* is

$$\langle L \rangle_{p_n} := \sum_{C,i} p_n(C) i C_i. \quad (5.16)$$

In particular, the *expected final length* is given by $\tilde{Q}(C_\emptyset) := \langle L \rangle_{p_{n_T}}$. Of central importance will be the best possible expected final length that can be achieved by means of any strategy:

$$Q(C_\emptyset) := \sup_S \tilde{Q}_S(C_\emptyset). \quad (5.17)$$

This number will be called the *quality* of C_\emptyset . For convenience we will use the abbreviations $\tilde{Q}(N) := \tilde{Q}(N e_1)$ and $Q(N) := Q(N e_1)$.

A last quantity that will play a role is the number of fusion attempts $T(C)$. Due to the nature of the gates involved, it holds that $L(C_\emptyset) + n_s(E) d_s - n_f(E) d_f = L(C_E)$ and $T(C_E) = n_s(E) + n_f(E)$ where n_s and n_f are the number of succeeded and failed fusion attempts, respectively. For the fusion gate it is particularly simple due to $d_s = 0$:

$$T(C_E) = L(C_\emptyset) - L(C_E). \quad (5.18)$$

5.2.3 Simple strategies

A priori, a strategy does not allow for a more economic description other than a ‘look-up table’, specifying what action to take when faced with a given configuration. If one restricts attention to the set of configurations $\mathcal{C}^{(N)}$ that can be reached starting from N EPR pairs, $|\mathcal{C}^{(N)}|$ actions have to be fixed.

The cardinality $|\mathcal{C}^{(N)}|$, in turn, can be derived from the accumulated number of integer partitions of $k \leq N$. Its asymptotic behaviour [93] can be identified to be

$$|\mathcal{C}^{(N)}| = \frac{1 + O(N^{-1/6})}{(8\pi^2 N)^{1/2}} e^{\pi(2N/3)^{1/2}}, \quad (5.19)$$

which is almost exponential in the number N of initially available EPR pairs [19].

However, there are of course strategies which do allow for a simpler description in terms of basic general rules that apply similarly to all possible configurations. It might be surmised that close-to-optimal strategies can be found among them. Also, these simple strategies are not only potentially accessible numerically, but also to analytical treatment. Subsequently, we will discuss three such reasonable strategies, referred to as GREED, MODESTY, and STATIC.

GREED

This is one of the most intuitive strategies. It can be described as follows: “Given any configuration, try to fuse the largest two available chains”. This actually is nothing but

$$S_G(C) := \begin{cases} \emptyset & \text{if } \sum_i C_i \leq 1 \\ \langle k, l \rangle & \begin{cases} k = \max\{i : C_i > 0\} \\ l = \max\{i : C_i - \delta_{i,k} > 0\} \end{cases} \end{cases}. \quad (5.20)$$

Alternatively, one may think of GREED as fusing the first two chains after sorting the configuration in descending order. The rationale behind choosing this strategy is the following: fusing is a probabilistic process which destroys entanglement on average. Hence it should be advantageous to quickly build up as long a chain as possible. Clearly, the strategy’s name stems from its pursuit of short-term success. From a theoretical point of view, GREED is interesting, as its asymptotic performance can easily be assessed (see Fig. 5.5):

Lemma 9 (Asymptotic performance of GREED). *The expected length of the final chain after applying GREED with fusion gates to N EPR pairs scales asymptotically*

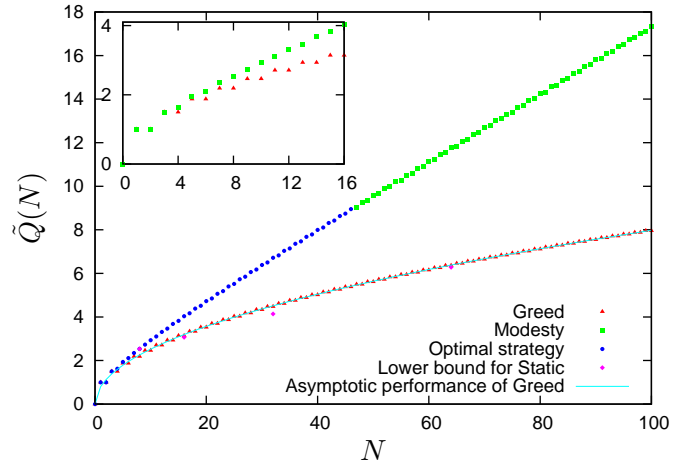


Figure 5.5: Expected length for the globally optimal strategy, for MODESTY (in this plot indistinguishable from the former), for GREED, its asymptotic performance, and the lower bound for STATIC, as functions of even number N of initial EPR pairs. The inset shows GREED and MODESTY for small N , revealing the parity-induced step-like behaviour.

as

$$\tilde{Q}_G(N) = (2N/\pi)^{1/2} + o(1). \quad (5.21)$$

Proof. It is easy to see that an application of GREED to $C_0 = Ne_1$ only generates configurations in $\{me_1 + e_l, m = 0, \dots, N; l = 0, \dots, N; l + m \leq N\}$. This set is parametrised by m (the number of EPR resources) and l (the size of the distinguished chain), giving rise to the notation $C = (l, m)$. By definition of S_G , whenever $l \geq 1$, the next fusion attempt is made on this longer chain and one of the other EPR pairs. As for the case $l = 0$ we identify $(0, m)$ with $(1, m - 1)$ (when encountering $(0, m)$ we distinguish one of the m pairs). Therefore, in this slightly modified notation we have with $C = (l, m), l > 0$, in case of success $C_S = (l + 1, m - 1)$ and in case of failure $C_F = (l - 1, m - 1)$, respectively.

The tree in Fig. 5.6 can be obtained by reflecting the negative half of a standard random walk tree at $l = 0$ and identifying the vertices with same m but opposite l . One can readily read off the expectation value of the final chain's length,

$$\tilde{Q}_G(N) = 2 \sum_{k=0}^{\lfloor (N-1)/2 \rfloor} p_s^{N-k} (1 - p_s)^k \binom{N}{k} (N - 2k). \quad (5.22)$$

The probabilities are twice the probabilities of the standard random walk tree, and the length-0 term has been omitted. $\tilde{Q}_G(N)$ adopts an especially simple form in the balanced case ($p_s = 1/2$).

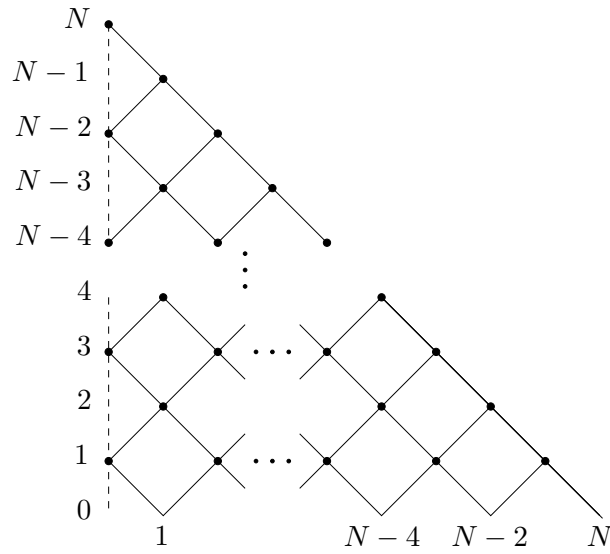


Figure 5.6: The process of fusion of the largest can be represented as a tree similar to a random walk. Reflection occurs at the dashed line (the largest string is lost and replaced with an EPR pair). Time evolves from top to bottom, thus decreasing the number of EPR resources. The horizontal dimension represents the length of the largest string.

With an approximation using a Gaussian distribution we easily find the asymptotic behaviour for large N (setting $\mu = pN$ and $\sigma^2 = p_s(1 - p_s)N$ with $p_s = 1/2$),

$$\begin{aligned}\tilde{Q}_G(N) &= \left(\frac{8}{N\pi}\right)^{1/2} \int_0^\infty 2x \exp\left(-\frac{2x^2}{N}\right) dx + r(N) \\ &= \left(\frac{2N}{\pi}\right)^{1/2} \Gamma(1) + r(N)\end{aligned}\quad (5.23)$$

with approximation error $r(N) = o(1)$. \square

The behaviour of GREED changes qualitatively upon variation of p_s : For $p_s > 1/2$, $\tilde{Q}_G(N)$ shows linear asymptotic behaviour in N , while in case of $p_s < 1/2$ the quality $\tilde{Q}_G(N)$ is not even unbounded as a function of N .

Also, variations of $d = (d_s, d_f)$ have a huge impact on the outcome. For $p_s = 1/2$, the other symmetric (*i.e.*, $C_{FS} = C_{SF}$, $\bar{\Delta} = (L(C_S) + L(C_F))/2 - L(C) = 0$) case, $d = (1, 2)$, Eqn. (5.22) can be rewritten by replacing the length $N - 2k$ with $2(N - 2k) - 1 + o(1)$, which results in $\tilde{Q}_G(N) = 2 \left(\frac{2N}{\pi}\right)^{1/2} - 1$. The other gates show qualitatively different behaviours: linear growth for $d = (1, 1)$, where the average overall length change per fusion attempt is positive ($\bar{\Delta} > 0$), and saturation for $d = (0, 2)$ where $\bar{\Delta} < 0$ (see Fig. 5.7).

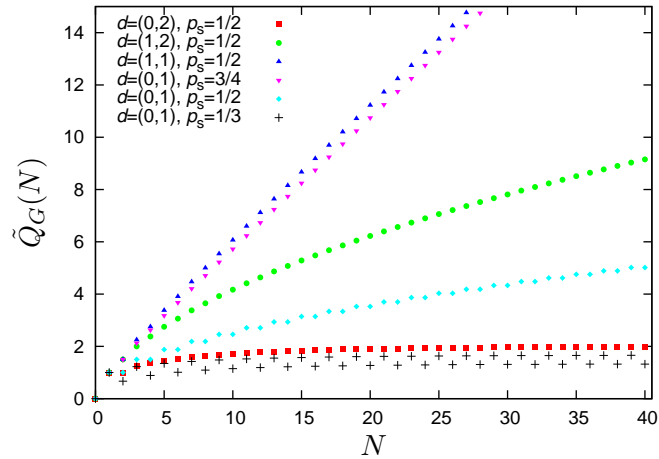


Figure 5.7: Influence of the parameters of the different entangling gates on GREED. The effects of the four gates from Tab. 5.1 at $p_s = 1/2$ as well as the fusion gate ($d = (0, 1)$) at $p_s = 3/4$ and $p_s = 1/3$ are shown.

There is a phenomenon present in the performance of many strategies, which can be understood particularly easy when considering GREED: \tilde{Q}_G displays a “smooth” behaviour when regarded as a function on either only *even* or only *odd* values of N . However, the respective graphs appear to be the same (only displaced with respect to each other), see the inset in Fig. 5.5. For simplicity, we will in general restrict our attention to even values and explore the reasons for this behaviour in the following lemma.

Lemma 10 (Parity and \tilde{Q}_G). *Let N be odd. Then $\tilde{Q}_G(N + 1) = \tilde{Q}_G(N)$.*

Proof. Let $C_\emptyset = Ne_1, C'_\emptyset = (N+1)e_1$, for N even. Now let E be such that $S_G(C_E) = \emptyset$ but $S_G(C_{E_1, \dots, |E|-1}) \neq \emptyset$. As GREED does not touch the i -th chain before the i -th step, it holds that $C'_E = C_E + e_1$ (with $C_E = e_k$). Further, since type-I fusion preserves the parity of the total number of edges, $C'_E \neq 0$. Hence C'_E is of the form $C'_E = e_1 + e_k$ and one computes:

$$\tilde{Q}_G(C'_E) = (k + 1)/2 + (k - 1)/2 = k = \tilde{Q}_G(C_E). \quad (5.24)$$

From here, the assertion is easily established by re-writing $\tilde{Q}_G(C'_\emptyset)$ as a suitable average over all events E fulfilling the assumptions made above. \square

Note that fusing an EPR pair to another chain does not, on average, increase its length ($\bar{\Delta} = 0$). Hence the fact that $\tilde{Q}_G(N)$ grows at all as a function of N is solely due to the asymmetric situation at length zero. This is actually a feature of any strategy which utilises fusion gates and more intuition on these effects can be found in Section 5.2.4.

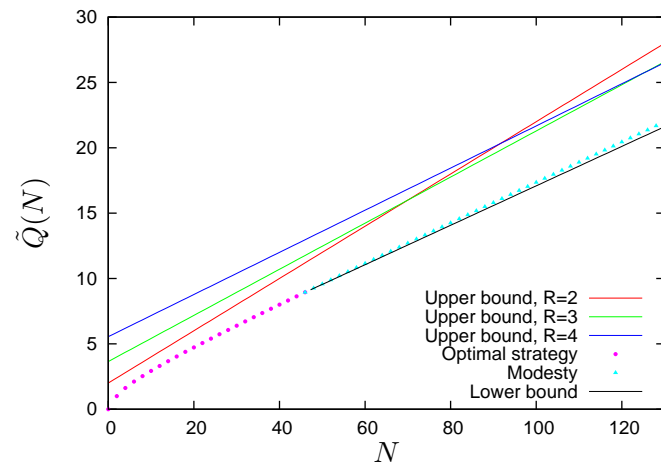


Figure 5.8: Expected length for MODESTY, the optimal strategy (where known), a lower bound to the quality as in Theorem 13, but with MODESTY and $N_0 = 46$ (for better visualisation), and the upper bound attained with the razor model with razor parameter R (see also Section 5.2.6) as functions of the number of initial EPR pairs N .

Lemma 10 explains the steps apparent in Fig. 5.5. Such steps are present also in the performance of MODESTY, to be discussed now, and several other strategies – albeit not in such a distinct manner.

MODESTY

There is a very natural alternative to the previously studied strategy. Instead of trying to fuse always the largest existing linear cluster states in a configuration, one could try the opposite: “Given any configuration, try to fuse the smallest two available chains”. In contrast to GREED this strategy intends to build up chains of intermediate length, making use of the whole EPR reservoir before trying to generate larger chains. Even though no long chains will be available at early stages, the strategy might nevertheless perform reasonably. Quite naturally, this strategy will be called MODESTY.

Formally, this amounts to replacing max by min, *i.e.* replacing descending order by ascending order:

$$S_M(C) = \begin{cases} \emptyset & \text{if } \sum_i C_i \leq 1 \\ \langle k, l \rangle & \begin{cases} k = \min\{i : C_i > 0\} \\ l = \min\{i : C_i - \delta_{i,k} > 0\} \end{cases} \end{cases} . \quad (5.25)$$

Maybe surprisingly, MODESTY will not only turn out to give better results than GREED, but it is actually close to being globally optimal, as can be seen in Figures 5.5 and 5.8. See Section 5.2.4 for further discussion.

STATIC

Another strategy of particular interest is called STATIC, S_S . To describe its action, we need to define the notion of an *insistent strategy*. The term is only meaningful in the identity picture, which we will employ for the course of this section. Now, a strategy is called insistent if, whenever it decides to fuse two specific chains, it will keep on trying to glue these two together until either successful or at least one of the chains was completely destroyed. Formally:

$$(S(I_E) = \langle k, l \rangle) \wedge ((I_{EF})_k(I_{EF})_l \neq 0) \Rightarrow S(I_{EF}) = S(I_E) \quad (5.26)$$

STATIC acts by insistently fusing the first chain to the second one; the third to the fourth and so on. After this first level, the resulting chains will be renumbered in the way that the outcome of the k -th pair is now the k -th chain. At this point, STATIC starts over again, using the configuration just obtained as the new input. This procedure is iterated until at most one chain of nonzero length has survived.

The proceeding of S_S is somehow related to MODESTY and GREED, just without sorting the chains between fusion attempts. This results in much less requirements on the routing of the physical systems actually carrying the cluster states. From an experimentalist's point of view, STATIC is a meaningful choice. It only requires a minimal amount of feed-forward for re-routing the qubits in the chains, but not on the level of routing the whole chains (for a completely re-routing free scheme see Chapter 5.3). STATIC performs, however, asymptotically already better than GREED (see Fig. 5.5).

It turns out that STATIC performs rather poorly when acting on a configuration consisting only of EPR pairs. To cure this deficit, we will proceed in two stages. Firstly, the input is partitioned into blocks of eight EPR pairs each. Then MODESTY is used to transform each block into a single chain. The results of this first stage are subsequently used as the input to STATIC itself, as described before. Slightly overloading the term, we will call this combined strategy STATIC as well. Note that, even when understood in this wider sense, STATIC still reduces the need for physically re-routing chains: no fusion processes between chains of different blocks are necessary during the first stage. The following theorem bounds STATIC's performance. For technical reasons, it is stated only for suitable N .

Theorem 11 (Linear performance of STATIC). *For any $m \in \mathbb{N}$, given $N = 2^{3+m}$ EPR pairs, STATIC will produce a single chain of expected length*

$$\tilde{Q}(N) \geq (137/2048)N + 2. \quad (5.27)$$

The proof of the above theorem utilises the following lemma which quantifies the average length one can expect when combining several configurations.

Lemma 12 (Combined configurations). *The following holds.*

1. *Let C be a configuration consisting of single chains of respective lengths l_1 and l_2 . Then one obtains*

$$Q(C) \geq l_1 + l_2 - 2(1 - p_s)/p_s. \quad (5.28)$$

which simplifies to

$$Q(C) \geq l_1 + l_2 - 2 \quad (5.29)$$

for $p_s = 1/2$.

2. *Let $C_{(1)}, \dots, C_{(k)}$ be configurations. Let S be a strategy that acts on $\sum_i C_{(i)}$ by first acting with S' on each of the $C_{(i)}$ and then acting insistently on the resulting chains. Then,*

$$\tilde{Q}\left(\sum_i C_{(i)}\right) \geq \sum_i \tilde{Q}_{S'}(C_{(i)}) - 2(1 - p_s)(k - 1)/p_s \quad (5.30)$$

$$\stackrel{p_s=1/2}{=} \sum_i \tilde{Q}_{S'}(C_{(i)}) - 2(k - 1). \quad (5.31)$$

3. *When substituting all occurrences of \tilde{Q} by Q , the above estimate remains valid.*

Proof. Firstly, any strategy will try to fuse the only two chains in the configuration together until it either succeeds or the shorter one of the two is destroyed (after $\min(l_1, l_2)$ unsuccessful attempts). In other words: in case of these special configurations any strategy is insistent. By Lemma 14 one could write $Q(l_1, l_2) = l_1 + l_2 - \langle T \rangle$

(with T being the number of fusion attempts), or

$$Q(l_1, l_2) = \sum_{k=1}^{\min(l_1, l_2)} \underbrace{p_s(1-p_s)^{k-1}}_{\text{prob. of success in } k\text{-th attempt}} \underbrace{(l_1 + l_2 - 2(k-1))}_{\text{length after } k\text{-th attempt}} \quad (5.32)$$

$$+ \underbrace{(1-p_s)^{\min(l_1, l_2)} |l_1 - l_2|}_{\text{residual after destroying the smaller chain}} \quad (5.33)$$

$$= l_1 + l_2 - 2(1-p_s) \left(1 - (1-p_s)^{\min(l_1, l_2)}\right) / p_s \quad (5.34)$$

$$\stackrel{p_s=1/2}{=} l_1 + l_2 + 2^{1-\min(l_1, l_2)} - 2. \quad (5.35)$$

Note, that this holds for any strategy S , so $\tilde{Q}_S(l_1, l_2) = Q(l_1, l_2)$ and it can formally be extended to $l_1, l_2 \in \mathbb{R}^+$ by replacing $\min(l_1, l_2)$ by $\lfloor \min(l_1, l_2) \rfloor$, preserving bounds (5.28) and (5.29).

For the second part, we run S' on each $C_{(i)}$, resulting in k single chain configurations $C'_{(i)} = e_{l_i}$ with probability distributions p_i on $\mathcal{C}^{L(C_i)}$ obeying $\tilde{Q}_{S'}(C_{(i)}) = \langle l_i \rangle_{p_i}$. Let $K = \sum_i L(C_i)$, then the joint distribution on \mathcal{C}^K is given by $p = \prod_i p_i$. Now we fuse a pair of these chains together: After uniting them into one configuration $C' := C'_{(i)} + C'_{(j)}$, C' contains at most two chains which we fuse together as described in the first part of the Lemma. As Eqn. (5.28) is linear in the respective lengths of the chains in C' , the distribution $p' = p_i p_j$ fulfils on the one hand

$$\langle \tilde{Q} \rangle_{p'} \geq \langle l_i + l_j - 2(1-p_s)/p_s \rangle_{p'} = \langle L \rangle_{p_i} + \langle L \rangle_{p_j} - 2(1-p_s)/p_s \quad (5.36)$$

and on the other hand

$$\tilde{Q}(\langle l_i \rangle_{p_i}, \langle l_j \rangle_{p_j}) \geq \langle L \rangle_{p_i} + \langle L \rangle_{p_j} - 2(1-p_s)/p_s. \quad (5.37)$$

for any insistent strategy. Because these two quantities are bounded by the same value, we will use this bound and replace averages over \tilde{Q} with \tilde{Q} of configurations of average – not necessarily integer – lengths.

We now iterate this scheme to obtain a single chain. A moment of thought reveals that – as a result of our neglecting the $2^{1-\lfloor \min(l_1, l_2) \rfloor}$ -term – the order in which chains are fused together does not enter the estimate for \tilde{Q}_S . The claim follows.

As for the third point: It follows by setting S' to the optimal strategy. \square

Proof. (of Theorem 11) Consider a configuration consisting of $n = 2^m$ chains of length x each. Using Lemma 12 one sees that the second stage of STATIC will convert it into a single chain of expected length $\tilde{Q}(2^m e_x) \geq (x-2)n + 2$.

According to Section 5.2.4, MODESTY fulfils $\tilde{Q}_M(8) = Q(8) = 649/256$. With $x = \tilde{Q}_M(2^3)$ and $N = 2^{3+m} = 8n$ we find

$$\tilde{Q}_S(N) \geq \frac{649/256 - 2}{8}N + 2 = \frac{137}{2048}N + 2 \quad (5.38)$$

$$\approx 6.69 \cdot 10^{-2}N + 2. \quad (5.39)$$

□

In case of $p_s \neq 1/2$,

$$\tilde{Q}'_S(ne_x) \geq n(x - l_{\text{initial}}) + l_{\text{initial}} \quad (5.40)$$

can be obtained in the same way, where $l_{\text{initial}} = 2(1 - p_s)/p_s$ (similar to n_c in [83]) in order to ensure the bound (5.30) to be positive. Initial chains of length $\geq l_{\text{initial}}$ can be produced by employing various strategies (even GREED), aborting the process when $2(1 - p_s)/p_s$ is reached. Although large chains are produced with only a small overall success probability, this does not affect the linear asymptotic behaviour as this process only acts on blocks the size of which only depends on p_s , rather than N .

5.2.4 Computer-assisted results

Algorithm for finding the optimal strategy

Before passing from the concrete examples considered so far to the more abstract results of the next sections, it would be instructive to explicitly construct an optimal strategy for small N . Is that a feasible task for a desktop computer? Naïvely, one might expect it not to be. Since the number of strategies grows super-exponentially as a function of the total number of edges N of the initial configuration, a direct comparison of the strategies' performances is quickly out of reach. Fortunately, a somewhat smarter, recursive algorithm can be derived which will be described in the following paragraph.

The *number of vertices* in a configuration is given by $V(C) := \sum_i C_i(n_i + 1)$. An attempted fusion will *decrease* (at least not increase in case of the other gates) $V(C)$ regardless of whether it succeeds or not. Now fix a V_0 and assume that we know the value of Q for all configurations comprised of up to V_0 vertices. Let C be such that $V(C) = V_0 + 1$. It is immediate that

$$Q(C) = \max_{i,j} (Q(S_{i,j} C) + Q(F_{i,j} C))/2, \quad (5.41)$$

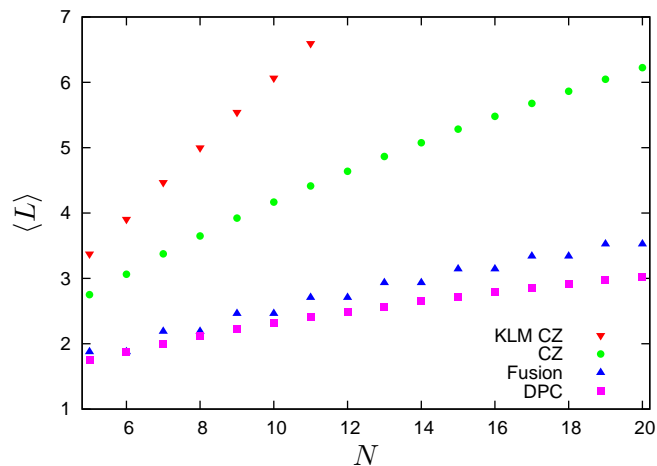


Figure 5.9: The optimal expected length $\langle L \rangle$ of the final cluster for the gates listed in Table 5.1. For comparison all gates are assessed at $p_s = 1/2$.

where $S_{i,j}C$ denotes the configuration resulting from successfully fusing chains of lengths i and j . $F_{i,j}C$ is defined likewise. As the r.h.s. involves only the quality of configurations possessing less than or equal to V_0 vertices, we know its value by assumption and we can hence perform the maximisation in $O(|C|^2)$ steps. One thus obtains the quality of C and the pair of chains that need to be fused by an optimal strategy.

The algorithm now works by building a *lookup table* containing the value of Q for *all* configurations up to a specific V_{\max} . It starts assessing the set of configurations with $V(C) = 2$ and works its way up, making at each step use of the previously found values. One needs to supply an anchor for the recursion by setting $Q(e_i) = i$.

By assessing a configuration in terms of its longest chain, one can allow for premature halting. Then p_s can be adjusted to investigate gates with $p_s \neq 1/2$ as well. The generalisation to gates with parameters other than $d = (0, 1)$ is straightforward. Because the gates under consideration decrease the number of vertices or the number of chains, adapting the recursion leads to an algorithm suitable for all classes of states that we introduced (*c.f.* Fig. 5.9).

Clearly, the memory consumption is proportional to $|\mathcal{C}^{(N)}|$, which is exponential in N and will limit the practical applicability of the algorithm before time issues do.

We have implemented this algorithm using the computer algebra system **Mathematica** and employed it to derive in closed form an optimal strategy for all configurations in $\mathcal{C}^{(46)}$, the quality of which is shown in Figures 5.5 and 5.8. A desktop computer is capable of performing the derivation in a few hours.

From the discussion above, it is clear that the leading term in the *computational complexity* of the algorithm is given by $|\mathcal{C}^{(N)}|$: every configuration needs to be

looked at at least once. A straightforward analysis reveals a poly-log correction; the described program terminates after $O(|C^{(N)}| (\log |C^{(N)}|)^5)$ steps.

Data, intuitive interpretation, and competing tendencies

Starting with $C_\emptyset = Ne_1$, MODESTY turns out to be the optimal strategy for $N \leq 10$. For configurations containing more edges, slight deviations from MODESTY can be advantageous. The relative difference with respect to $Q(N)$ is smaller than 1.1×10^{-3} for $N \leq 46$. More generally, two heuristic rules seem to hold:

1. It is favourable to fuse small chains (this is the dominant rule).
2. It is favourable to create chains of equal length.

Is there an intuitive model which can explain these findings? Several steps are required to find one. Firstly, note that every fusion attempt entails a $1/2$ -probability of failure, in which case two edges are destroyed. So “on average” the total length $L(C)$ decreases by one in each step and it is natural to assume that *the quality $Q(C)$ equals $L(C)$ minus the expected number of fusion attempts* a specific strategy will employ acting on C (this will be made precise in Lemma 14). Hence, a good strategy aims at *reaching a single-chain configuration as quickly as possible*, so as to reduce the expected number of fusions (this reasoning will be made precise in Section 5.2.6). Now, if there are k chains present in C , then *a priori* $k - 1$ successful fusions will be needed before a strategy can terminate. If, however, in the course of the process one chain is completely destroyed, then $k - 2$ successes would already be sufficient. Therefore – paradoxically – within the given framework *it pays off to destroy chains*. Since shorter chains are more likely to become completely consumed due to failures, they should be subject to fusion attempts whenever possible. This explains the first rule.

Secondly, there is one single scenario in which *two* chains can be destroyed in a single step; that is when one selects two EPR pairs to be fused together. Now consider the case where there are two chains of equal length in a configuration. If we keep on trying to fuse these two chains, then – in the event of repeated failures – we will eventually be left with two EPR pairs, which are favourable to obtain as argued before. Hence the second rule.

We have thereby identified two *competing tendencies* of the optimal strategy. Obtaining a quantitative understanding of their interplay seems extremely difficult: deviating from MODESTY at some point of time might open up the possibility of creating two chains of equal lengths many steps down the line. We hence feel it is

sensible to conjecture that *the globally optimal strategy allows not even for a tractable closed description*. A *proof* of its optimality seems therefore beyond any reasonable effort. One is left with the hope of obtaining appropriately tight analytical bounds – and indeed, the sections to come pursue this programme with perhaps surprising success.

5.2.5 Lower bounds

We will now turn to establishing rigorous upper and lower bounds to Q , the outcome of the optimal strategy. These bounds, in turn, give rise to bounds to the resource consumption any probabilistic scheme will have to face. Lower bounds are in turn less technically involved than upper bounds. In fact, rigorous lower bounds can be based on known bounds for given strategies: For not too large configurations, the performance of various strategies can be calculated explicitly on a computer (see Section 5.2.4). Any such computation in turn gives a lower bound to Q . The following theorem is based on a construction which utilises the computer results to build a strategy valid for inputs of arbitrary size. This strategy is simple enough to allow for an analytic analysis of its performance while at the same time being sufficiently sophisticated to yield a very tight lower bound for the quality, shown in Fig. 5.8. Notably, the resulting statement is *not* a numerical estimate restricted to small N , but a proven bound valid for all N .

Theorem 13 (Lower bound to the globally optimal strategy). *For any strategy S , fix $N_0 > 0$. If S satisfies for all $N_0 \leq N \leq 2N_0$*

$$\frac{\tilde{Q}_S(N) + 2(1 - 1/p_s)}{N} \geq \alpha := \frac{\tilde{Q}_S(N_0) + 2(1 - 1/p_s)}{N_0}, \quad (5.42)$$

and for all $N \leq 2N_0$

$$Q(N) \geq \tilde{Q}_S(N_0) + \alpha(N - N_0) \quad (5.43)$$

holds, then a lower bound to the quality is given by (5.43) for all $N \geq 2N_0$.

Proof. Assume we are given $N > 2N_0$ EPR pairs. Clearly, there are positive integers $k \geq 2$ and $M \leq N_0$ such that $N = kN_0 + M$. Set $n_i = N_0$ for $i = 1, \dots, k - 1$ and $n_k = N_0 + M$. The n_i fulfil $\sum_i n_i = N$ and $N_0 \leq n_i \leq 2N_0$. We partition the input

into k blocks of lengths n_i and compute

$$\begin{aligned}
Q\left(\sum_i n_i\right) &\geq \sum_{i=1}^k Q(n_i) + 2(1 - 1/p_s)(k - 1) \\
&\geq \sum_{i=1}^k \tilde{Q}(n_i) + 2(1 - 1/p_s)(k - 1) \\
&= \tilde{Q}(N_0) + \sum_{i=2}^k n_i \frac{\tilde{Q}(n_i) + 2(1 - 1/p_s)}{n_i} \\
&\geq \tilde{Q}(N_0) + \sum_{i=2}^k n_i \frac{\tilde{Q}(N_0) + 2(1 - 1/p_s)}{N_0} \\
&= \tilde{Q}(N_0) + \alpha \sum_{i=2}^k n_i \\
&= \tilde{Q}(N_0) + \alpha(N - N_0),
\end{aligned} \tag{5.44}$$

where we made use of Lemma 12 and the assumptions mentioned above. \square

In the case of MODESTY for $p_s = 1/2$, the function $\tilde{Q}_M(N)$ can be explicitly computed for not too large values of N . Indeed, the results for all $N \leq 2N_0 = 184$ were computed. They obey the assumptions in Theorem 13 with

$$\tilde{Q}_M(92) = \frac{175388528938590098714204473464373259356085}{10889035741470030830827987437816582766592} \approx 16.1069. \tag{5.45}$$

Therefore, a linear lower bound for $N \geq 92$ reads

$$Q(N) \geq 16.1069 + 0.153336(N - 92). \tag{5.46}$$

Another example of a lower bound uses $p_s = 3/4$, $\tilde{Q} = Q$ and $N_0 = 14$. All assumptions are fulfilled and the bound reads

$$Q(N) \geq 7.9066 + 0.517139(N - 14). \tag{5.47}$$

See Fig. 5.10 for comparison to upper bounds and the actual quality.

5.2.6 Upper bounds – the razor model

While the performance of any strategy delivers a lower bound for the optimal one, giving an upper bound is considerably harder. We will tackle the problem for fusion gates with $d = (0, 1)$ by passing to a family of simplified models. For every integer

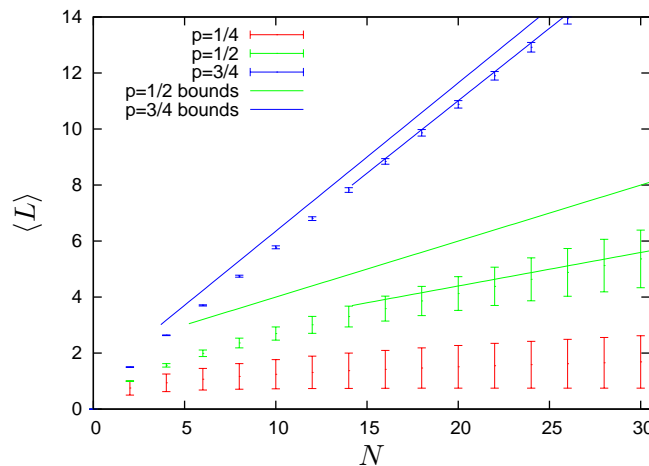


Figure 5.10: The graphs show the influence of the classical strategy on the expected length of the final cluster for type-I fusion gates operating at $p_s = 1/4, 1/2$ and $3/4$, respectively. For each probability, the range enclosed by the error bars indicates the spread between the best possible and worst possible strategy. Bounds are with respect to the optimal strategy, upper bounds use the razor parameter $r = 2$.

Note that for $p_s < 1/2$ no meaningful bounds can be derived from Theorem 13 and Corollary 22 because no premature halting was considered there, but it was used in the algorithm determining the optimal strategy.

$R \geq 2$, the *razor model with parameter R* is defined by introducing the following new rule: after every fusion step all chains will be cut down to a maximum length of R . Obviously, the full problem may be recovered with $R \geq N$. Given the complexity of the problem, it came as a surprise that even for parameters as small as $R = 2$ the essential features of the full setup seem to be retained by the simplification, in the sense that understanding the razor model yields extraordinarily good bounds for Q .

The razor model – outline

In the spirit of Section 5.2.2, a configuration in the razor model is specified by a vector in \mathbb{N}^R . Thus, the number of configurations with a maximum total number of N edges is certainly smaller than N^R , which is a polynomial in N . Adapting the techniques presented in Section 5.2.4, we can obtain the optimal strategy with polynomially scaling effort. *We have thus identified a family of simplified problems which, in the limit of large R , tend to become exact, and where each instance is solvable in polynomial time.*

How do the results of the razor model relate to the original problem? Clearly, for small values of R , $Q_{\text{razor}}(C)$ will be a very crude lower bound to $Q(C)$. However, as indicated in Section 5.2.4, the quality of a configuration C can be assessed in terms of the optimal strategy's expected number of fusion attempts $\langle T(C) \rangle$ when

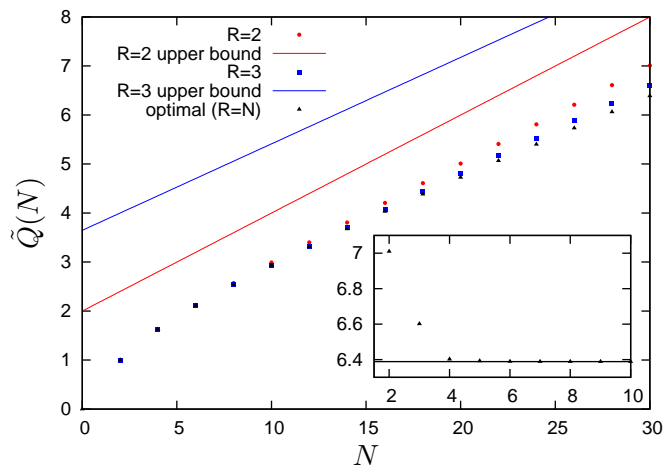


Figure 5.11: Performance of the optimal strategy in the razor model ($R = 2$ and $R = 3$), the full model ($R = N$) and the upper bound attained with the $R = 2$ razor model. The inset shows the convergence of the quality in the razor model *vs.* the razor parameter $R = 2, \dots, 10$ for $N = 30$ together with the optimal value in the full model, $Q(30)$. See also Fig. 5.8 for comparison.

acting on C . It is intuitive to assume that $\langle T \rangle_{\text{razor}} \leq \langle T \rangle$, as the “cutting process” increases the probability of early termination. We will thus employ the following argument: for a given configuration C , derive a lower bound for $\langle T(C) \rangle_{\text{razor}}$, which is in particular a lower bound for $\langle T(C) \rangle$, which in turn gives rise to the upper bound

$$Q(C) \leq L(C) - 2(1 - p_s)\langle T(C) \rangle \quad (5.48)$$

for Q .

The results of this ansatz are extremely satisfactory. Fig. 5.11 shows the performance of the optimal quality for various R , and the convergence when increasing R .

The intuitive explanation for the success of the model is the observation that the chance that a chain of length R is built up, and eventually disappears again, is strongly suppressed as a function of R . That is, the crucial observation is that the error made by this radical modification is surprisingly small. A rigorous justification for this reasoning is supplied by the following two propositions which will be proved in the next section.

Lemma 14 (Quality and attempted fusions). *The expected final length $\langle L \rangle$ equals the initial number of edges $L(C_0)$ minus the weighted expected number of attempted fusions $2(1 - p_s)\langle T \rangle$.*

Theorem 15 (Bound to the full model from the razor model). *Let $C_0 \in \mathcal{C}$ be a configuration. The optimal strategy in the setting of the razor model will use fewer*

fusion attempts on average to reach a final configuration starting from C_\emptyset than will the optimal strategy of the full setup.

The razor model – proofs

For the present section, it will prove advantageous to introduce some alternative points of view on the concepts used so far. Recall that a strategy is a function from *configurations* to *actions*. However, once we have fixed some initial configuration C_\emptyset , we can alternatively specify a strategy as a map from *events* to actions. Indeed, the configuration present after n steps is completely fixed by the knowledge of the initial configuration, the past decisions of the strategy and the sequence of failures and successes. We will call the resulting mapping the *decision function* D_{S,C_\emptyset} and will suppress the indices whenever no danger of confusion can arise. In the same spirit, we are free to conceive *random variables* on \mathcal{C} as real functions $f : \{S, F\}^n \rightarrow \mathbb{R}$. Expectation values are then computed as

$$\langle f \rangle := \langle f \rangle(p_{n_T}) = \sum_{E, |E|=n_T} p_s^{n_s(E)} (1 - p_s)^{n - n_s(E)} f(E). \quad (5.49)$$

Quantities of the form $\langle f \rangle(C)$ for some configuration C refer to expectation values $\langle f \rangle$ given the initial configuration $C_\emptyset = C$.

An interesting class of random variables can be written in the form

$$f(E) = \sum_{i=1}^{|E|} \phi_f(E_{1,\dots,i}) \quad (5.50)$$

where ϕ_f is some function of events and $E_{1,\dots,i}$ denotes the restriction of E to its first i elements. A simple example is the *amount of lost edges* $M(E)$ that was suffered as a result of E . Here,

$$\phi_M(E_{1,\dots,i}) = \begin{cases} 2d_f, & E_i = F \wedge D(E_{1,\dots,i-1}) \neq \emptyset, \\ -d_s, & E_i = S \wedge D(E_{1,\dots,i-1}) \neq \emptyset, \\ 0, & \text{else.} \end{cases} \quad (5.51)$$

Let us refer to observables as in Eqn. (5.50) as *additive random variables*. The following lemma states that when evaluating expectation values of additive variables, only their *step-wise mean*

$$\bar{\phi}(E_{1,\dots,i}) := p_s \phi(E_{1,\dots,i-1}, S) + (1 - p_s) \phi(E_{1,\dots,i-1}, F) \quad (5.52)$$

enters the calculation.

Lemma 16 (Expectation values of additive random variables). *Let f be an additive random variable. Set*

$$\bar{f}(E) := \sum_{i=1}^{|E|} \bar{\phi}(E_{1\dots i}). \quad (5.53)$$

Then $\langle f \rangle = \langle \bar{f} \rangle$.

Proof. Without loss of generality, we assume $p_s = s/t$, $s, t \in \mathbb{N}$ for readability. Set $n = n_T$. We then have, by definition,

$$\begin{aligned} \langle f \rangle &= \sum_{E, |E|=n} p_s^{n_s(E)} (1 - p_s)^{n - n_s(E)} \sum_{i=1}^n \phi(E_{1,\dots,i}) \\ &= \frac{1}{t^n} \sum_{\substack{|E|=n \\ i=1}}^n s^{n_s(E)} (t - s)^{n - n_s(E)} \phi(E_{1,\dots,i}) \\ &= \frac{1}{t^n} \sum_{i=1}^n \sum_{|E|=i} s^{n_s(E)} (t - s)^{n - n_s(E)} t^{n-i} \phi(E) \\ &= \frac{1}{t^n} \sum_{i=1}^n \sum_{|E|=i-1} s^{n_s(E)} (t - s)^{n - n_s(E)} t^{n-i} (s\phi(E, S) + (t - s)\phi(E, F)) \\ &= \frac{1}{t^n} \sum_{i=1}^n \sum_{|E|=i} s^{n_s(E)} (t - s)^{n - n_s(E)} t^{n-i} \bar{\phi}(E) \\ &= \langle \bar{f} \rangle. \end{aligned} \quad (5.54)$$

□

Proof. (of Lemma 14) Note that

$$\bar{\phi}_M(E_{1,\dots,i}) = \begin{cases} 2(1 - p_s)d_f - p_s d_s, & D(E_{1,\dots,i-1}) \neq \emptyset, \\ 0, & \text{else,} \end{cases} \quad (5.55)$$

for $d = (0, 1)$ counts the *weighted number of attempted fusions* $2(1 - p_s)T$. Using Lemma 16, we see that the expected number of lost edges equals the expected number of fusion attempts: $\langle M \rangle = 2(1 - p_s)\langle T \rangle$. This proves Lemma 14. □

In the following proof of Theorem 15, we will employ the identity picture introduced in Section 5.2.2. The argument is broken down into a series of lemmata.

Lemma 17 (More is better than less). *Let I be a configuration. Then, for all i , $Q(I + e_i) \geq Q(I)$.*

Proof. The proof is conducted by induction on two parameters: on the number of non-trivial chains $|I|$ and on the total length $L(I)$. To base the induction in both variables, we note that the claim is trivial if either $|I| \leq 1$ or $L \leq 2$.

Now consider any configuration I . Let S be the optimal strategy and denote by I_S and I_F the configurations created by $S(I)$ in case of success and failure, respectively. It is simple to check that $S(I)$ acting on $I + e_i$ yields $I_S + e_i$ or $I_F + e_i$. Hence

$$Q(I + e_i) \geq p_s Q(I_S + e_i) + (1 - p_s) Q(I_F + e_i). \quad (5.56)$$

But unless $|I| \leq 1$ it holds that in any event $E \in \{S, F\}$ either $|I_E| < |I|$ or $L(I_E) < L(I)$ and thus the claim follows by induction. \square

Lemma 18 (Winning is better than losing). *Let I be a configuration, and I_S the one resulting from the action of the optimal strategy on C in the case of success, let I_F be the obvious analogue. Then $Q(I_S) \geq Q(I_F)$.*

Proof. Let $\langle k, l \rangle$ be the action defined above. Clearly, $I_F = I - d_f e_k - d_f e_l$. By the last lemma, $Q(I_F) \leq Q(I)$. But $Q(I)$ is the average of $Q(I_F)$ and $Q(I_S)$; hence

$$Q(I_S) \geq Q(I) \geq Q(I_F). \quad (5.57)$$

\square

Lemma 19 (No catalysis). *Let I be a configuration. Then, for all i , $Q(I + e_i) \leq Q(I) + 1$.*

Proof. We show the equivalent statement: for I and i such that $I_i \neq 0$ it holds that $Q(I - e_i) \geq Q(I) - 1$. Once more, the proof is by induction on $|I|$, L , and the validity of the claim for $|I| \leq 1$ or $L \leq 2$ is readily verified.

Let I, S, I_S, I_F be as in the proof of Lemma 17. If the application of $S(I)$ and the subtraction of e_i commute, we can proceed as we did in Lemma 17. A moment of thought reveals that this is always the case if not $I_i = 1$ and $S(C) = \langle i, k \rangle$ (or, equivalently, $\langle k, i \rangle$) for some k . In fact, in this case we have

$$I_S = (\dots, l_i + l_k, \dots) \quad (5.58)$$

so that $I_F - e_i$ would take on a negative value at the i -th position. Note, however, that $I - e_i = I_S - e_i$. By induction it holds that $Q(I_S - e_i) \geq Q(I_S) - 1$ and further, by Lemma 18 $Q(I_S) - 1 \geq Q(I) - 1$, which concludes the proof. \square

Lemma 20 (Fewer edges – fewer fusions). *Let I be a configuration, i be such that $I_i \neq 0$. Then*

$$\langle T \rangle(I - e_i) \leq \langle T \rangle(I), \quad (5.59)$$

where the expectation values are taken with respect to the respective optimal strategies.

Proof. We will show that, for every I , the optimal strategy acting on $I' := I - e_i$ will content itself with a lower number of average fusion attempts $\langle T \rangle(I')$, than will the optimal strategy acting on I . Recall that Lemma 14 states

$$Q(I) = L(I) - 2(1 - p_s)\langle T \rangle(I). \quad (5.60)$$

Combining this and Lemma 19 we find

$$\begin{aligned} Q(I') &\geq Q(I) - 1 \\ \Leftrightarrow L(I) - 1 - 2(1 - p_s)\langle T \rangle(I') &\geq L(I) - 2(1 - p_s)\langle T \rangle(I) - 1 \\ \Leftrightarrow \langle T \rangle(I') &\leq \langle T \rangle(I). \end{aligned} \quad (5.61)$$

□

We are finally in a position to tackle the original problem.

Proof. (of Theorem 15) Let C_\emptyset be some configuration. We will build a strategy which is valid on C_\emptyset in the razor model and uses a fewer number of expected fusions than the optimal strategy in the full setup. Define the shaving operator $\hat{R} : \mathcal{C} \rightarrow \mathcal{C}$ which sets the length of each chain of length i in the configuration it acts on to $\min(i, R)$. By a repeated application of the relation stated in Lemma 20, we see that $\langle T \rangle(\hat{R}C) \leq \langle T \rangle(C)$.

We build the razor model strategy's decision function D' inductively for all events in \mathcal{E}_i , for increasing i . Consider an event $E \in \mathcal{E}_i$. Denote by C'_E the configuration resulting from C_\emptyset under the action of D' in the event of E . C'_E is well-defined as only the values of D' for events with length smaller than i enter its definition. Set $D'(E)$ to the action taken by the optimal strategy for $\hat{R}C'_E$.

It is simple to verify that D' defines a valid strategy for the razor model. By the results of the first paragraph, the expected number of fusions decreased in every step of the construction of D' . The claim follows. □

An analytical bound – random walk

Finally, we are in a position to prove an analytic upper bound on the yield of any strategy building one-dimensional cluster chains. Quite surprisingly, the description given by the razor model with a rather radical parameter of $R = 2$ is still faithful enough to deliver a good bound as will be explained now. Bounds based on higher razor parameters will be shown as well, up to a regime where they are indistinguishable from the quality within a reasonable accuracy.

In the $R = 2$ model, configurations are fully specified by giving the number of EPR pairs n_1 , and the number of chains of length two n_2 they contain. Hence the configuration space is $\mathbb{N}_0 \times \mathbb{N}_0$ and we can picture it as the positive quadrant of a two-dimensional lattice. In each step a strategy can choose only among three non-trivial actions:

- (a) Try to fuse two EPR pairs. We call this action $a := \langle 1, 1 \rangle$ for brevity. Let C_S be the configuration resulting from a successful application of a on C . Define the vector $a_S \in \mathbb{Z} \times \mathbb{Z}$ as $a_S := C_S - C$. An analogous definition for a_F and some seconds of thought yield

$$a_S := (-2, 1), \quad (5.62)$$

$$a_F := (-2, 0). \quad (5.63)$$

- (b) Try to fuse two chains of length one and two, respectively. In the same manner as above we have $b := \langle 1, 2 \rangle$ and

$$b_S := (-1, 0), \quad (5.64)$$

$$b_F := (0, -1). \quad (5.65)$$

- (c) Try to fuse two chains of length two, so $c := \langle 2, 2 \rangle$ and

$$c_S := (0, -1), \quad (5.66)$$

$$c_F := (2, -2). \quad (5.67)$$

The objective is to bound the minimum number of non-trivial actions taken on average from below. Initially, we start with N EPR pairs, so $C_0 = (N, 0)$. As the configuration space is a subspace of $\mathbb{N}_0 \times \mathbb{N}_0$, we can describe the situation by a random walk in a plane.

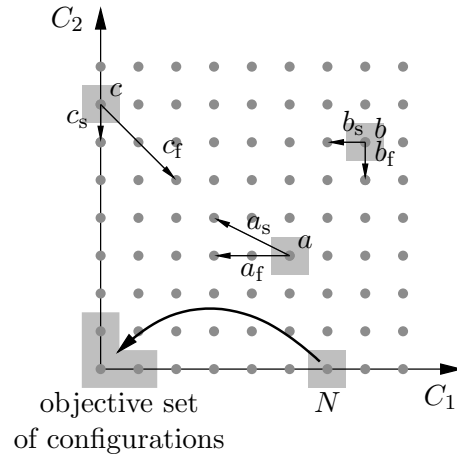


Figure 5.12: The configuration space of the $R = 2$ razor model is $\mathbb{N}_0 \times \mathbb{N}_0$. Only the three actions a , b and c are available to reach the final configurations (exactly one EPR pair, one GHZ state, or no chain at all), starting from the initial configuration that consists of N EPR pairs.

Any strategy will apply the rules a, b, c until one of the points $(0, 0)$, $(1, 0)$, $(0, 1)$ is reached (illustrated in Fig. 5.12). Our proof will be led by the following idea: by applying one of the three non-trivial actions to a configuration C , we will move “on average” by

$$\bar{a} := p_s a_S + (1 - p_s) a_F = (-2, p_s), \quad (5.68)$$

$$\bar{b} := (-p_s, p_s - 1), \quad \text{or} \quad (5.69)$$

$$\bar{c} := (2(1 - p_s), p_s - 2), \quad (5.70)$$

respectively. The minimum number of expected fusion steps should then be given by the minimum number of vectors from $\{\bar{a}, \bar{b}, \bar{c}\}$ one has to combine to reach the origin starting from $(N, 0)$. This procedure amounts to an interchange of two averages. The aim is to reach the origin or a point with distance one to it on average as quickly as possible.

To make this intuition precise, set

$$\phi_\delta(E_{1,\dots,i}) := D(E_{1,\dots,i-1})_{E_i}. \quad (5.71)$$

Recall that $D(E_{1,\dots,i-1})$ is one of $\{a, b, c, \emptyset\}$. Given the event E , $\phi_\delta(E)$ is the last action applied to the configuration. For any event $E = \{S, F\}^n$ we require that

$$\delta(E) := \sum_i \phi_\delta(E_{1,\dots,i}) \leq (-N + 1, 1), \quad (5.72)$$

which implies in particular that the same bound holds for $\langle \delta \rangle$. Define $a(E)$ to be the number of times the strategy will have decided to apply rule “ a ” in the chain of events $\{E_{1,\dots,i} \mid i = 1, \dots, |E|\}$ leading up to E . Formally

$$\phi_a(E_{1,\dots,i}) = \begin{cases} 1, & D(E_{1,\dots,i-1}) = a, \\ 0, & \text{else,} \end{cases} \quad (5.73)$$

and $a(E) = \sum_{i=1}^{|E|} \phi_a(E_{1,\dots,i})$. Further,

$$\bar{\phi}_\delta(E_{1,\dots,i}) = \phi_a(E_{1,\dots,i})\bar{a} + \dots + \phi_c(E_{1,\dots,i})\bar{c}, \quad (5.74)$$

where ϕ_b, ϕ_c are defined in the obvious way. It follows that

$$\langle \delta \rangle = \langle \bar{\delta} \rangle = \langle a \rangle \bar{a} + \langle b \rangle \bar{b} + \langle c \rangle \bar{c} \leq (-N + 1, 1), \quad (5.75)$$

and

$$\langle T \rangle = \langle a \rangle + \langle b \rangle + \langle c \rangle. \quad (5.76)$$

An analytical bound – convex optimisation program

If $\langle T \rangle$ originates from a valid strategy, it is necessarily subject to the constraints put forward in Eqn. (5.75) and (5.76). For each $N \in \mathbb{N}$, a lower bound for the minimum expected number of losses is thus given by a linear program, so a convex optimisation problem: We define

$$B := \begin{pmatrix} -2 & p_s \\ -p_s & p_s - 1 \\ 2(1 - p_s) & p_s - 2 \end{pmatrix}. \quad (5.77)$$

Then, this lower bound can be derived from the optimal solution of the linear program given by

$$\begin{aligned} & \text{minimise} && (1, 1, 1)x^T \\ & \text{subject to} && xB \leq (-N + 1, 1), \\ & && x \geq 0, \end{aligned} \quad (5.78)$$

where the latter inequality is meant as a component-wise positivity. This is a minimisation over a vector $x \in \mathbb{R}^3$. In this way, the performance of the razor model is reduced to solving a family of convex optimisation problems. According to Lemma 21,

the solution of this linear program delivers the optimal objective value satisfying

$$\langle T \rangle = \frac{1}{1 + (1 - p_s)^2} N - \frac{3 - p_s}{p_s(p_s - 2) + 2} \quad (5.79)$$

for $N \geq N_0(p_s) < \infty$.

Lemma 21 (Duality for linear program). *The optimal objective values of the family of linear programs for $N \geq N_0 = \lceil (2 + p_s)/p_s \rceil$*

$$\begin{aligned} & \text{minimise} && (1, 1, 1)x^T \\ & \text{subject to} && xB \leq (-N + 1, 1), \\ & && x \geq 0, \end{aligned} \quad (5.80)$$

are given by

$$(1, 1, 1)x_{opt}^T = \frac{1}{1 + (1 - p_s)^2} N - \frac{3 - p_s}{p_s(p_s - 2) + 2} =: mN + n \quad (5.81)$$

Proof. This can be shown making use of Lagrange duality for linear programs [33, 34]. The dual to the above problem, referred to as primal problem, is found to be

$$\begin{aligned} & \text{maximise} && (N - 1, -1)y^T \\ & \text{subject to} && -yB^T \leq (1, 1, 1), \\ & && y \geq 0. \end{aligned} \quad (5.82)$$

This is a maximisation problem in $y \in \mathbb{R}^2$ – again a linear program. By finding – for each N – a solution of the dual problem which is assumed by the primal problem, we have hence proven optimality of the respective solution. For all N , this family of solutions can be determined to be

$$y = (m, -m - n). \quad (5.83)$$

It is straightforward to show that these are solutions of the dual problem, and that the respective objective values are attained by appropriate solutions of the primal problem for $N \geq N_0$, *e.g.*,

$$x = \left(\frac{1 - p_s/2}{1 + (1 - p_s)^2} N - \frac{4 - 3p_s}{4 - 4p_s + 2p_s^2}, 0, \frac{p_s/2}{1 + (1 - p_s)^2} N - \frac{2 + p_s}{4 - 4p_s + 2p_s^2} \right). \quad (5.84)$$

The solutions yield the objective values stated in the lemma. \square

We subsequently highlight the consequence of this proof: we find the bound to the quality of the globally optimal strategy: this shows that asymptotically for $p_s = 1/2$ at least five EPR pairs have to be invested on average (see also the following section) to gain a single edge in the linear cluster state.

Corollary 22 (Upper bound to globally optimal strategy). *The quality of the optimal strategy for $N \geq \lceil (2 + p_s)/p_s \rceil$ is bounded from above by*

$$Q(N) \leq \frac{6 + p_s((2 + N)p_s - 8)}{2 + (p_s - 2)p_s}. \quad (5.85)$$

In case of $p_s = 1/2$ for $N \geq 5$ this reads

$$Q(N) \leq \frac{N}{5} + 2. \quad (5.86)$$

Increasing the razor parameter

Of course, the same techniques may be applied with arbitrary $R > 2$. Generating the allowed $R(R + 1)/2$ actions in the razor model and solving the linear program can be done in an automatic manner using **Mathematica**. Within a few minutes, a desktop computer can solve the razor model up to $R = 32$, which amounts to solving a linear programs in 528 variables. As expected, the slope of the upper bound converges rather quickly: for $R \geq 16$ and N sufficiently large, the quality is bounded by

$$Q(N) \leq 0.154438N + \text{const}. \quad (5.87)$$

Together with the lower bound (5.46) the slope of the quality could be determined up to a thin region:

$$0.153336 \leq \lim_{N \rightarrow \infty} \frac{Q(N)}{N} \leq 0.154438. \quad (5.88)$$

Due to the razor bounds becoming tighter when increasing R , statements on the slope and offset of these linear bounds with respect to R are of interest as well. Within the inspected range of R , the linear coefficient of $Q(N)$ can be written as $X(R, 4, 14, 26)/X(R, 5, 17, 31)$, where X fulfils the recursion relation

$$X(R, X_2, X_3, X_4) = \begin{cases} X_R & \text{for } R = 2, \dots, 4 \\ 2\alpha(R)X(R - 1, X_2, X_3, X_4) & \text{for } R > 4 \\ -X(\lfloor \frac{R-1}{2} \rfloor - 1, X_2, X_3, X_4) & \text{for } R > 4 \end{cases} \quad (5.89)$$

where $\alpha(R) := 2^{\lfloor \log_2(R+1) \rfloor - \lfloor \log_2 R \rfloor}$, so 2 iff $R + 1 = 2^k$.

Conjecture 23 (Converging upper bounds). *For $p_s = 1/2$ a converging family of upper bounds to the quality is given for increasing razor parameter R by*

$$Q(N) = N \left(1 - \frac{X(R, 4, 14, 26)}{X(R, 5, 17, 31)} \right) + \text{const}(R). \quad (5.90)$$

5.2.7 An inverse question

Recall that, so far, we treated the problem “given some fixed number of input pairs, how long a single chain can be obtained on average?”. It is also legitimate to ask “how many input pairs are needed to produce a chain of some fixed length with (almost) unit probability of success?”. After all, we might need just a specific length for a given task. In the present section we establish that both questions are *asymptotically equivalent*, in the sense that bounds for either problem imply bounds for the other one.

Note that a straightforward inversion of MODESTY – and therefore most likely also of the optimal strategy – is not possible because taking more and more “smallest chains” does not amount to a finite strategy. While this “duality of bounds” was assumed implicitly in recent works it now allows to explicitly compare schemes aiming for a fixed length (Refs. [82, 38, 86, 83, 84, 85]) and starting with a fixed number of resources (Refs. [1, 2, 5]) where the actual limit of large systems is considered.

Theorem 24 (Resources for given resulting length, upper bounds). *Let S be some strategy, let*

$$\tilde{Q}_S(N) \geq \alpha N + \beta \quad (5.91)$$

be a lower bound to its outcome for some $\alpha, \beta \in \mathbb{R}$ and all $N \geq N_0$. Choose an $\varepsilon > 0$. Then there exists a strategy S' such that, if S' acts on $(1/\alpha + \varepsilon)L$ EPR pairs, it will output a single chain not shorter than L with probability approaching unity as $L \rightarrow \infty$.

Proof. Choose a number $b \in \mathbb{N}$. Set $N = (1/\alpha + \varepsilon)L$. There are arbitrary large L such that b divides N and we will presently assume that L has this property. We comment on the general case in the end.

The strategy S' proceeds in two stages, labelled I and II, to be analysed in turn. Firstly, we divide the N input pairs into $B = N/b$ blocks of size b and let S run on each of these blocks.

Denote by N_i the random variable describing the final output length of the i -th block, $i = 1, \dots, B$. The N_i are independent, identically distributed variables

satisfying $\langle N_i \rangle \geq \alpha b + \beta$. Set $N_I = \sum_{i=1}^B N_i$ (the roman I signifies that we are dealing with the expected total length after the *first* stage of S'). As the N_i are independent, the variance of N_I equals $B\sigma^2$, where $\sigma^2 < \infty$ is the variance of any of the N_i . By Chebychev's inequality we have

$$P[|N_I - \langle N_I \rangle| \geq B^{3/4}] \leq \text{Var}(N_I)B^{-3/2} \quad (5.92)$$

$$= \sigma^2 B^{-1/2}. \quad (5.93)$$

In other words, the relation $|N_I - \langle N_I \rangle| < B^{3/4}$ holds almost certainly if we let L (and hence B) go to infinity for any fixed b . The same is true in particular for the weaker statement

$$N_I \geq \langle N_I \rangle - B^{3/4} \geq B(\alpha b + \beta) - B^{3/4}. \quad (5.94)$$

In the second stage, II, S' builds up a single chain out of the B ones obtained before. Irrespective of how S' goes about in detail, the process will stop after at most $B-1$ successful fusions². Now choose any $\delta > 0$. We claim that asymptotically no more than $(1/p_s - 1 + \delta)(B-1)$ failures will have occurred before the strategy terminates. Indeed, consider an event E of length $(1/p_s + \delta)(B-1)$. By the law of large numbers, E contains no fewer than $B-1$ successes and not more than $(1/p_s - 1 + \delta)(B-1)$ failures, almost certainly as $B \rightarrow \infty$. Hence, the final output length N_{II} fulfils

$$P[N_{II} > B(\alpha b + \beta) - B^{3/4} - 2(1/p_s - 1 + \delta)B] \rightarrow 1 \quad (5.95)$$

as $B \rightarrow \infty$. Plugging in the definitions of B, N , the r.h.s. of the estimate takes on the form

$$L + L \left(\varepsilon \alpha + \frac{1}{b} f_1(p_s, \alpha, \beta, \delta, \varepsilon) \right) - \left(\left(\frac{L}{b} f_2(p_s, \alpha, \varepsilon) \right) \right)^{3/4}, \quad (5.96)$$

where f_1, f_2 are some (not necessarily positive) functions of the constants. By choosing the block length b large enough, we can always make the second summand positive. For large enough L , the positive second term dominates the negative third one and hence $N_{II} > L$ almost certainly as $L \rightarrow \infty$.

Lastly consider the case where L is such that b does not divide N . Choose $L \geq b/\varepsilon$. We can decompose $N = kb + r$ where $r < b$ and hence $r/L < b/L \leq \varepsilon$. Set $\varepsilon' = \varepsilon - r/L$, which ensured to be positive for L being large enough. By construction

²If chains are destroyed during the process, less than $B-1$ would be sufficient. However, we can get a lower bound by requiring $B-1$ successes.

b divides $N' = (1/\alpha - \varepsilon')L$ and therefore already $N' < N$ input pairs are enough to build a chain of length L asymptotically with certainty. \square

Theorem 25 (Resources for given resulting length, lower bounds). *Let*

$$Q(N) \leq \alpha N + \beta \quad (5.97)$$

be some upper bound to the optimal strategy's performance. Choose an $\varepsilon > 0$. Then there exists no strategy S' such that, if S' acts on $(1/\alpha - \varepsilon)L$ EPR pairs, it will output a single chain not shorter than L with probability approaching unity as $L \rightarrow \infty$.

Proof. Assume there is such a strategy S' . Then

$$\lim_{N \rightarrow \infty} \frac{\tilde{Q}_{S'}(N)}{N} \geq (1/\alpha - \varepsilon)^{-1} > \alpha. \quad (5.98)$$

Hence $\tilde{Q}_{S'}(N)$ is eventually larger than $Q(N)$, which is a contradiction. \square

Suppose one aims to build a linear cluster state of length N with $p_s = 1/2$ fusion gates. Combining the results of the present section with the findings of Sections 5.2.5 and 5.2.6 yields that the goal is achievable with unit probability as $N \rightarrow \infty$ if more than $6.63N$ EPR pairs are available. Similarly, one will face a finite probability of failure in case there are less than $6.47N$ input chains. Both statements are valid asymptotically for large N .

5.2.8 Two-dimensional cluster states

Preparation prescription

We finally turn to the preparation of two-dimensional cluster states, which are the actual universal resources for cluster computation [15]. To build up a two-dimensional $n \times n$ cluster state clearly requires the consumption of $O(N^2)$ EPR pairs. That this bound can actually be met constitutes the main result of this section; previous schemes exhibited worse scalings such as $O(N^2 \log(N^2))$ [82, 83, 84]. The approach shown here is motivated by the one introduced in Ref. [38] where the action of the type-II fusion gate was introduced in the context of connecting independent cluster chains, but no general treatment of the resource scaling was given. In this class of schemes (building two-dimensional structures by connecting chains) we will identify one that can achieve an $O(N^2)$ scaling.

From our previous derivations, we already know that length- N linear cluster chains can be built consuming $O(N)$ entangled pairs. Hence it suffices to prove that

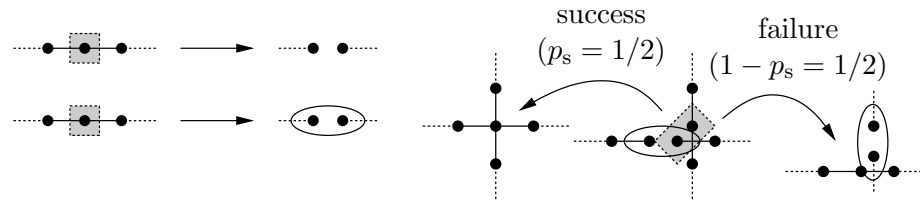


Figure 5.13: The elementary linear optics tools for building two-dimensional structures from linear cluster chains. From top to bottom: a σ_z measurement to remove unneeded nodes, a σ_x measurement to create a redundantly encoded qubit in preparation of type-II fusion. The last figure shows the action of a type-II fusion attempt.

linear chains with an accumulated length of $O(N^2)$ can be combined to an $N \times N$ -cluster. Consequently, for the constructions to come, we will employ linear chains – as opposed to EPR pairs – as the basic building blocks.

Again, to actually connect two chains to form a two-dimensional structure, probabilistic gates from arbitrary architectures may be utilised. The following claim will hold for gates that delete a constant amount of edges from the participating chains on failure (maybe unequal for the two chains), but not splitting them (no σ_z error outcome). In case of success it shall create cross-like structures, again deleting a certain amount of edges (see Fig. 5.13). In particular, the quadratic scaling as such is not altered by a possibly small probability of success $p_s < 1/2$.

The main problem faced is to find a preparation scheme that does not ‘tear apart’ successfully prepared intermediate states in case of a failed fusion. The challenge will be met by (a) switching from type-I to type-II fusion and (b) employing the pattern shown in Fig. 5.14.

As for linear optics fusion gates, an error outcome in the type-I gate would tear each chain apart where we tried to fuse. Hence the related type-II fusion gate [38] with a more suitable error outcome will be used. How this one actually acts is shown in Fig. 5.13.

Linear optical type-II fusion gate

In preparation of a fusion attempt, a “redundantly encoded” qubit with two photons (see Ref. [38]) is produced in one chain by a σ_x measurement, which consumes two edges (giving rise to another $2N^2$ edges). Now, the type-II fusion gate creates a two-dimensional cross-like structure on success when being applied to one of the photons in the redundantly encoded qubit and one of the other chain’s qubits. In case of failure it acts like a σ_x measurement on both qubits, therefore decreasing

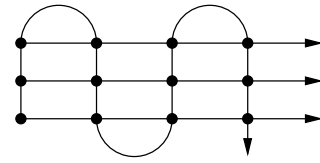


Figure 5.14: A possible pattern of how to arrange $n + 1$ linear clusters to build a two-dimensional cluster of width n . Fusion operations have to be applied at the black circles along the long linear cluster state. Free ends carrying spare overhead are shown as arrows.

the encoding level of the redundantly encoded qubit by one and deleting two edges from the other chain, leaving us with a redundantly encoded qubit there. Hence, we may apply the type-II fusion again, without any further preparation, deleting two edges on successive failures from the two chains in alternation. For convenience we assume that we lose two edges per involved chain per failure instead. This increases the overhead requirement roughly by a factor of two but allows us to forget about the asymmetry in the fusion process and also allows for different types of failure outcomes (as long as they can be corrected by local operations). Hence, in the following any resource requirements will be given in terms of double edges instead of single ones.

Similar to the type-I case, the optimal success probability can be found. Actually, this type of fusion gate should perform a Bell state measurement, hence $p_s \leq 1/2$ [63]. In fact, the gate proposed in Ref. [38] consists of the parity check, the Hadamard rotation and measurement of the second qubit (see Fig. 5.3) with two additional Hadamard gates applied before (which only map Bell states onto Bell states).

Asymptotic resource consumption for near-deterministic cluster state preparation

Theorem 26 (Quadratic scaling of resource overhead). *For any success probability $p_s \in (0, 1]$ of type-II fusion, an $N \times N$ cluster state can be prepared using $O(N^2)$ edges in a way such that the overall probability of success approaches unity*

$$P_s(N) \rightarrow 1 \quad (5.99)$$

as $N \rightarrow \infty$.

Proof. The aim is to prepare an $N \times N$ cluster state, starting from $N + 1$ chains. For any positive integer l , the starting point is a collection of N one-dimensional states of length $m = N + l$, and a single longer chain of length $L = ml + N - 1$, referred to subsequently as thread. In order to achieve the goal, a suitable choice for a pattern of fusion attempts is required. One such suitable “weaving pattern” is

depicted in Fig. 5.14. Here, solid lines depict cluster chains, whereas dots represent the pairs of vertices along them where fusion gates are being applied.

The objective will be to identify a function $N \mapsto g(N)$ such that the choice $m = g(N)$ leads to the appropriate scaling of the resources. In fact, it will turn out that a linear function is already suitable, so for an $a > 1/p_s$ we will consider $g(N) = aN$. The number

$$m - N = g(N) - N = (a - 1)N \quad (5.100)$$

quantifies the resource overhead: in case of failure, one can make use of this overhead to continue with the prescription without destroying the cluster state. If this overhead is too large, we will fail to meet the strict requirements on the scaling of the overall resource consumption. On the other hand, if it is too small, the probability of an overall failure becomes too large. Note that there is an additional overhead reflected by the choice of L . This, however, is suitably chosen not to have an implication on the asymptotic scaling of the resources (it solely results in another factor of 2).

Given the above prescription, depending on N , the overall probability $P_s(N)$ of successfully preparing an $N \times N$ cluster state can be written as

$$P_s(N) = \pi_s(N)^N. \quad (5.101)$$

Here,

$$\pi_s(N) = p_s^N \sum_{k=0}^{(a-1)N} (1 - p_s)^k \binom{N + k - 1}{k} \quad (5.102)$$

is the success probability to weave a single chain of length aN into the carpet of width N , with the binomial quantifying the number of ways to distribute k failures on N nodes [94]. $p_s > 0$ and $1 - p_s$ are the success and failure probabilities for a fusion attempt, respectively. $\pi_s(N)$ can be rephrased as the probability to find at least N successful outcomes in aN trials,

$$\pi_s(N) = \sum_{k=N}^{aN} (1 - p_s)^{aN-k} p_s^k \binom{aN}{k} \quad (5.103)$$

$$= 1 - F(N - 1, aN, p_s). \quad (5.104)$$

F denotes the cumulative distribution function of the binomial distribution,

$$F(k, n, p) = \sum_{l=0}^k \binom{n}{l} p^l (1-p)^{n-l}. \quad (5.105)$$

Since $N-1 < aNp_s$ for all N – as $a > 1/p_s$ is assumed – we can hence bound $\pi_s(N)$ from below by means of *Hoeffding's inequality* [95] which states

$$F(k, n, p) \leq \exp\left(\frac{-2(np-k)^2}{n}\right) \quad (5.106)$$

for $k < np$, so an exponentially decaying upper bound of the tails of the cumulative distribution function. This gives rise to the lower bound

$$P_s(N) \geq \left(1 - \exp\left(-\frac{2(aNp_s - N + 1)^2}{aN}\right)\right)^N. \quad (5.107)$$

Now, again since $a > 1/p_s$, we have that

$$\pi_s^N \geq (1 - e^{-cN})^N \quad (5.108)$$

with $c := 2(ap_s - 1)/a > 0$. Further, for any $k \in \mathbb{N}$ there exists an $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$

$$(1 - e^{-cN})^N > \left(1 - \frac{1}{kN}\right)^N. \quad (5.109)$$

Noticing

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{kN}\right)^N = e^{-1/k} \quad (5.110)$$

we can find for any $\varepsilon > 0$ a suitable k , satisfying $1 - e^{-1/k} < \varepsilon$. Therefore, for any $\varepsilon > 0$ it holds that $\lim_{N \rightarrow \infty} P_s > 1 - \varepsilon$. This ends the argument leading to the appropriate scaling. \square

Even within the class of quadratic resources, the appropriate choice for a does have an impact: If the probability of success p_s is too small for a given a ,

$$1/p_s > a > 1, \quad (5.111)$$

then this will lead to $\lim_{N \rightarrow \infty} P_s(N) = 0$. The preparation of the cluster will eventually fail, asymptotically with certainty. This sudden change of the asymptotic behaviour of the resource requirements, leading essentially to either almost unit or almost vanishing success probability is a simple threshold phenomenon as in *perco-*

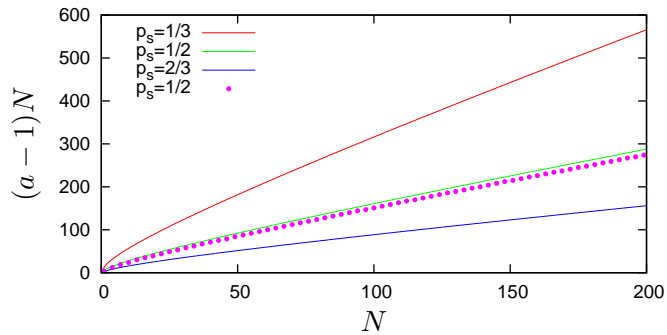


Figure 5.15: Overhead $(a - 1)N$ in the weaving process with elementary probability $p_s = 1/3, 1/2, 2/3$ to succeed for cluster size N with a probability of at least $P_s = 19/20$ (Eqn. (5.107) solved for a). In addition to the upper bounds, exact values (Eqn. (5.103)) for $p_s = 1/2$ are given.

lation theory (see also Section 5.3). In turn, for a given a , $p_c = 1/a$ can be taken as a threshold probability: above this threshold almost all preparations will succeed, below it they will fail.

This number a essentially dictates the coefficient in front of the quadratic behaviour in the scaling of the resource requirements. Summing up all the terms mentioned above, an upper bound for the pre-factor in front of the quadratic term reads

$$4 + 2(1/p_s - 1). \quad (5.112)$$

See Figs. 5.15 and 5.16 for the required overhead in the finite case. For $p_s = 1/2$ this amounts to 9 edges per site in the two-dimensional cluster to ensure success certainty in the limit of $N \rightarrow \infty$.

This pre-factor can possibly still be improved. For example, one may aim at not preparing a two-dimensional cluster state, but a state that is equivalent to such a state, up to local unitary rotations. Specifically, one could aim at preparing a graph state that is equivalent to a cluster up to local Clifford operations. The orbit under such local Clifford operations is reflected on the level of graphs by the orbit under *local complementations* [76, 77]. Ref. [96] already exploits such local complementations to prepare two-dimensional structures. It would be interesting to see whether a systematic exploration of these tools gives rise to a significant improvement of the above pre-factor in the optimal quadratic scaling.

To give some intuition on the scaling behaviour, the crucial difference to other schemes is the recycling of overhead: When adapting the number of arms when using stars [84] – or the length of the arms when using crosses [83] – as the ingredients, these quantities determine the available overhead (and therefore the maximum number of failures) per site. This buffer is available only locally and unused overhead

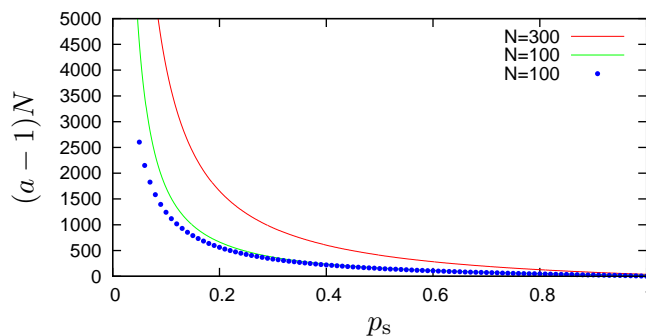


Figure 5.16: Overhead $(a-1)N$ that is needed for the weaving process to succeed for a fixed cluster size $N = 100, 300$ with a probability of at least $P_s(N) = 19/20$. For comparison with the upper bounds, exact values are given for $N = 100$ as well.

cannot be carried over to other sites. In contrast, our scheme minimises the number of buffers ($N + 1$ chains with overhead), and unused buffer can be reused in later steps by drawing in the thread.

5.3 Renormalisation and percolation: Banning re-routing

So far, we were allowing for arbitrary re-shuffling of qubits, potentially depending on results of all probabilistic gates employed before. Tolerating such extensive dynamics results in the need for

- good photon storage,
- fast read-out, classical post-processing and switching, and
- interferometric stability between all possible paths the photons could take.

Further, how to split the state preparation into smaller sub-tasks without sacrificing the scalability is not obvious.

Although feed-forward on a single qubit has already been demonstrated in the lab [97, 98], these requirements pose a major threat to such a scheme not only in the light of today’s experimental situation, but also with respect to its feasibility in general.

However, all these restrictions can be overcome or at least lifted significantly by a static procedure which only allows for nearest-neighbour³ interactions, as will be shown in this section [3, 6]. More precisely, the improvements offered by this scheme are:

³Here, the term “neighbour” refers to neighbourhood on the graph defining the cluster state, rather than spatial dimensions.

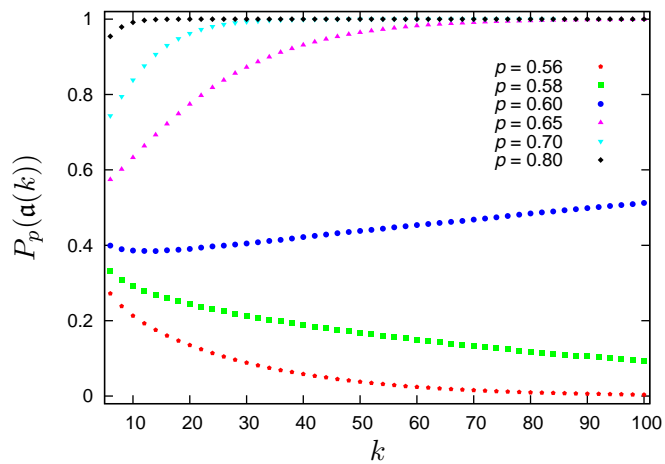


Figure 5.17: Probability of the event $\mathbf{a}(k)$: there exists an open cluster connecting all four faces of a block of size $k \times k$ on a square lattice with site percolation as obtained from Monte Carlo simulations (for every combination of k and p , $5 \cdot 10^5$ realisations were generated). The critical probability for this percolation problem is $p^{(c)} \approx 0.592$. Above the threshold, $\mathbf{a}(k)$ occurs asymptotically almost with certainty, below $p^{(c)}$ almost no spanning cluster is found in the limit of large k .

- Photon storage is only needed to the same extent as it is required in cluster computing anyway. Depending on the actual algorithm to be run after the preparation stage, the time-complexity of the measurement pattern is fixed. By postponing all single-qubit measurements to the computation step, only a constant amount of time is needed for the initial application of entangling operations.
- Fast read-out is still required during the computation stage to adapt future measurement bases. However, no path switching has to be done anymore. In the preparation stage only a constant amount of time has to be reserved for the read-out of detectors in the fusion gates. Efficient classical algorithms for determining the subsequent measurement patterns (which will be combined with the computation) will be shown later.
- Optical paths only have to be stabilised to the extent that the initial entangling operations can be performed between neighbours. Further multi-qubit operations are not needed, and therefore no further interferometric stability when sticking to suitable encoding (*i.e.*, dual-rail encoded qubits in polarisation).

Due to the intrinsically probabilistic behaviour of the entangling gates, such a procedure has to be robust against holes in the cluster (*i.e.*, missing edges in the graph where the gates failed). We can achieve this by partitioning the lattice into

blocks, each of which is entangled in such a way as to allow for these blocks to be used as renormalised qubits. Such a high tolerance against loss of bonds is ensured by *percolation theory* when using gate probabilities above the *percolation threshold* of the underlying lattice. The central insight from percolation theory is that for every lattice there exists a finite *critical probability* (percolation threshold) $p^{(c)}$ such that if the edges exist with a probability $p > p^{(c)}$, there exist paths connecting the faces of the lattice almost with certainty in the limit of a large lattice.

A similar effect occurs in many-body systems such as optical lattices where nearest-neighbour interactions are inherent and the entangling operations work deterministically [79]. Yet, Mott hole defects – where sites are left unoccupied in a random fashion – lead to similar defects. Although this situation is described by *site percolation*, rather than *bond percolation* as in the case of probabilistic gates, the bounds presented below can be derived in a similar way (see also Ref. [99] for an application of bond percolation for cluster state preparation). Similar to bond percolation, a threshold exists here as well. For illustration purposes this is shown in Fig. 5.17.

5.3.1 Percolation in a nutshell

To make the central results of percolation theory we are going to use more accessible, let us write down the basic quantities in a more formal way. The first structure one has to introduce is the lattice. The most common one is the square lattice in two dimensions, or more generally, the hyper-cubic lattice in d dimensions. It can be represented as the undirected graph [78] $G = (V, E)$ with vertex set $V = \mathbb{Z}^d$ and edges $e = \langle x, y \rangle$ between pairs of points $x, y \in \mathbb{Z}^d$ where $\delta(x, y) = 1$. Here,

$$\delta(x, y) = \sum_{i=1}^d |x_i - y_i| \quad (5.113)$$

is the distance between the points x and y on the lattice.

Now, we call each site (bond) *open* with probability p_{site} (p_{bond}) and *closed* otherwise. The special case $p_{\text{site}} = 1$ ($p_{\text{bond}} = 1$) will be called *bond (site) percolation*, the general case will be called *mixed percolation* with $p = (p_{\text{site}}, p_{\text{bond}})$. Examples of bond, site and mixed percolation on \mathbb{Z}^2 are shown in Fig. 5.18.

We use the usual graph-theoretic notion of a *path* [100], which is an alternating sequence $x_0, e_0, \dots, e_{n-1}, x_n$ of distinct vertices x_i and edges $e_i = \langle x_i, x_{i+1} \rangle$. A path is *open (closed)* iff all of its sites and edges are open (closed). We write $x \leftrightarrow y$ iff there exists an open path connecting x and y . Further, $C(x)$ is the *open cluster* at

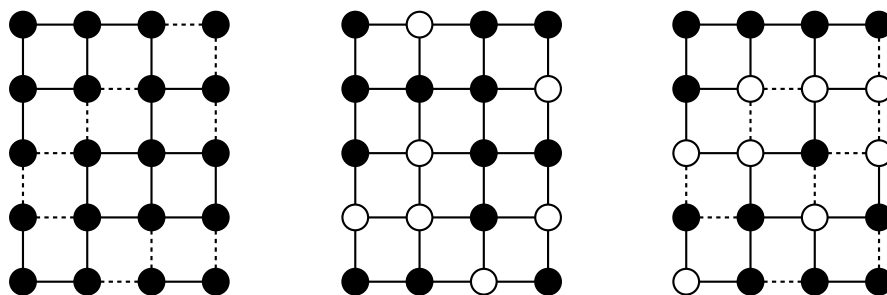


Figure 5.18: Bond, site, and mixed percolation on the square lattice. Open (closed) sites are symbolised by black (white) circles, open (closed) bonds by solid (dashed) lines.

x which is a sub-graph with the set of vertices given by $V_C = \{y \in V : x \leftrightarrow y\}$ and the set of edges given by the open edges between these vertices. The size $|C|$ of an open cluster C is defined by the number of sites it contains, $|C| := |V_C|$.

With the so called *percolation probability*

$$\theta(p) := P_p(|C(0)| = \infty) \quad (5.114)$$

an important phenomenon can be investigated: for every regular lattice there exists a threshold below which there exist only open clusters of finite size. Above threshold, however, the probability of existence of infinite open clusters is strictly larger than zero. This threshold is characterised by the *critical probabilities* $p^{(c)} = \left(p_{\text{site}}^{(c)}, p_{\text{bond}}^{(c)}\right)$:

$$\theta(p) \begin{cases} = 0 & , p < p^{(c)} \\ > 0 & , p > p^{(c)} \end{cases} , \quad (5.115)$$

where the inequalities are meant component-wise⁴. Table 5.2 shows some examples of critical probabilities for a choice of common lattices.

This property implies probabilities of various events when scaling the size of the lattice for given probabilities above threshold. In the following we will only consider bond percolation and unless stated otherwise will use $p := p_{\text{bond}}$. Please note, however, that the results can be translated into the site-percolation setting as well in a straightforward manner as was already shown in Ref. [99].

5.3.2 Renormalisation

The goal of this section will be to show how to generate a cluster state of a given size with probabilistic entangling gates (succeeding with probability p) almost certainly

⁴Note, that $p_{\text{site}}^{(c)}$ and $p_{\text{bond}}^{(c)}$ are not independent.

| lattice name | $\bar{\Delta}$ | $p_{\text{site}}^{(c)}$ ($p_{\text{bond}}^{(c)} = 1$) | $p_{\text{bond}}^{(c)}$ ($p_{\text{site}}^{(c)} = 1$) |
|------------------------|----------------|---|---|
| hexagonal (\odot) | 3 | 0.697 | $1 - 2 \sin(\pi/18) = 0.653$ |
| square | 4 | 0.592 | $1/2 = 0.500$ |
| triangular | 6 | $1/2 = 0.500$ | $2 \sin(\pi/18) = 0.347$ |
| diamond (\diamond) | 4 | 0.429 | 0.389 |
| sc | 6 | 0.311 | 0.249 |
| bcc | 8 | 0.246 | 0.180 |
| fcc | 12 | 0.199 | 0.120 |
| $d = 4$ hyper-cubic | 8 | 0.197 | 0.160 |
| $d = 5$ hyper-cubic | 10 | 0.141 | 0.118 |
| $d = 6$ hyper-cubic | 12 | 0.107 | 0.094 |
| $d = 7$ hyper-cubic | 14 | 0.089 | 0.079 |

Table 5.2: Critical percolation probabilities for some lattices [101, 100, 102]. $\bar{\Delta}$ denotes the average vertex degree. By increasing the dimension the critical probabilities decrease.

with the help of bond percolation. The lattice (we will use a hyper-cubic one in d dimensions) will be divided into blocks which will be reduced later by means of single qubit measurements to single qubits, thus “renormalising” the lattice (which will be square in 2 dimensions) in a physical sense, rather than a purely mathematical trick. Whenever a block contains a crossing open cluster that connects the block’s four faces (in the first two dimensions), we will refer to this crossing cluster as a *renormalised qubit*. If the crossing clusters of two neighbouring blocks actually touch each other (*i.e.*, there exists an open path between two vertices – one of the first and the second renormalised qubits each – that lies completely within the union of the two respective blocks), then the reduction of the blocks to single qubits will yield a *renormalised bond* between these qubits. How this reduction actually works is the scope of the following sections. Now, we are looking for the probability $P_p(\mathfrak{U}(N, k))$ of the event $\mathfrak{U}(N, k)$ to occur. $\mathfrak{U}(N, k)$ denotes the event that the renormalised square lattice of size $N \times N$ with hyper-cubic blocks of size $k^{\times d}$ is fully occupied and connected. Given a dependence of the block size on the lattice size, $k(N)$, we will use the abbreviation $P_p(N) := P_p(\mathfrak{U}(N, k(N)))$.

The result is more precisely phrased as follows:

Theorem 27 (Resource consumption without re-routing). *Let $p > p_d^{(c)}$ and $d \geq 2$ be the dimension of a hyper-cubic lattice. Then for any $\mu > 0$, the probability $P_p(N)$ of having an $N \times N$ renormalised cubic lattice fulfils*

$$\lim_{N \rightarrow \infty} P_p(N) = 1 \quad (5.116)$$

with an overall resource consumption of $R(N) = O(N^{2+\mu})$.

Here, $p_d^{(c)}$ denotes the percolation threshold of the d -dimensional hyper-cubic lattice. R is used to refer to the number of initial resources, so the constituents that are placed on each lattice site with the ability of “growing” connections to their neighbours in a probabilistic manner. Note, that the dimension d can always be chosen such that the bond generation processes at hand ($p > 0$) operate in a regime that allows for percolation⁵ ($p > p_d^{(c)}$).

Also, in the special case of the two-dimensional square lattice (and the gates succeeding with $p > 1/2$), algorithms can be found that allow for the optimal scaling of $R(N) = O(N^2)$ [99].

Proof. ($d \geq 3$) In the case of dimensions $d > 2$, in contrast to $d = 2$, crossing paths in different directions not necessarily intersect. Nevertheless, this approach is often favourable due to the higher percolation threshold in higher dimensions. Let us fix $N \in \mathbb{N}$ and take

$$U := [1, 2kN]^{\times 2} \times [1, 2k]^{\times d-2} \subset \mathbb{Z}^d \quad (5.117)$$

for some $k \in \mathbb{N}$. This slab can be divided into N^2 disjoint hypercubes with an edge length of $2k$. With $A_y(k)$, $y = (y_1, y_2) \in [2, 2N]^{\times 2}$ we denote the $(2k)^{\times d}$ hypercube starting at $((y_1 - 2)k + 1, (y_2 - 2)k + 1, 1, \dots, 1)$. For $y = 2x$ with

$$x \in M = [1, N]^{\times 2}, \quad (5.118)$$

these hypercubes $A_y(k)$ are the disjoint blocks mentioned earlier, and M plays the role of the renormalised square lattice.

Furthermore, we will use the overlap between adjacent blocks in the first direction,

$$B_y(k) = A_y(k) \cap A_{(y_1+1, y_2)}(k) \quad (5.119)$$

for $y_1 = 2, \dots, 2N - 1$, and the union of disjoint neighbouring blocks in the second direction, $C_z(k) = A_z(k) \cap A_{(z_1, z_2+2)}(k)$ for $z_2 = 2, 4, \dots, 2(N - 1)$.

On these blocks we will define a series of *events* as follows:

- $\mathfrak{A}_y(k)$: There exists an open crossing cluster in $A_y(k)$ in the first dimension (later referred to as open *left-to-right* crossing cluster), so an open path containing open vertices a and b with $a_1 = (y_1 - 2)k + 1$, $b_1 = y_1 k$. For $p > p_d^{(c)}$

⁵ This can be explained intuitively as follows: By increasing the dimension, the vertex degree of the lattice will increase as well. Having more connections, in turn, leads to an improved robustness against bond loss, so a lower critical probability.

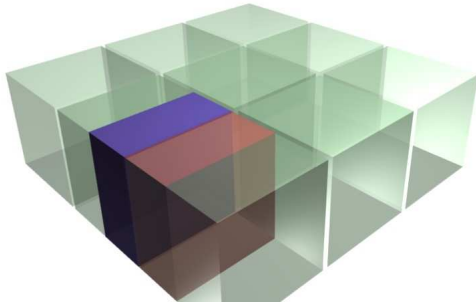


Figure 5.19: The blocks A_y for $y = 2x$. Together with the blue region the red one constitutes the block $A_{(3,2)}$, the blue region on its own being the overlap with a neighbouring block, so $B_{(3,2)}$.

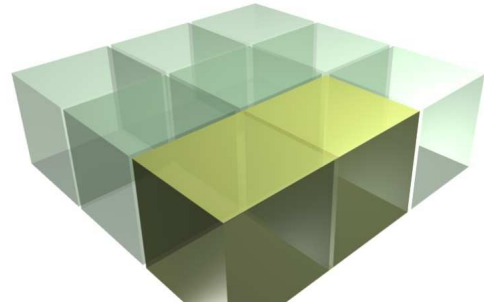


Figure 5.20: Again, the blocks A_y for $y = 2x$ are shown and the yellow region is the intersection $C_{(2,2)}$.

there exists a constant $g > 0$, which only depends on p , such that [100]

$$P_p(\mathfrak{A}_y(k)) \geq 1 - \exp(-gk^2). \quad (5.120)$$

- $\mathfrak{B}_y(k)$: The number of open left-to-right crossing clusters in $B_y(k)$ (see Fig. 5.19) does not exceed 1. It is shown in Ref. [103] that for $p > p_d^{(c)}$ there exist constants $a, c > 0$, only dependent on p , such that the probability of $\mathfrak{B}_y(k)$ occurring satisfies

$$P_p(\mathfrak{B}_y(k)) \geq 1 - (2k)^{2d} a \exp(-ck). \quad (5.121)$$

- $\mathfrak{D}_z(k)$ with $z_1 = 3, 5, \dots, 2N - 1$, $z_2 = 2, 4, \dots, 2N$ shall be the event that there exist open left-to-right crossing clusters in both blocks, $\mathfrak{A}_{(z_1-1, z_2)}(k)$ and $\mathfrak{A}_{(z_1+1, z_2)}(k)$, and that these two clusters are actually connected. Consequently, there exists an open left-to-right crossing cluster in $A_{(z_1-1, z_2)}(k) \cup A_{(z_1+1, z_2)}(k)$. In terms of the events defined above, $\mathfrak{D}_z(k)$ can be constructed in the following way: Given there exist open left-to-right crossing clusters in the three blocks $A_{(z_1+i, z_2)}(k)$ with $i = 0, \pm 1$ (the events $\mathfrak{A}_{(z_1-1, z_2)}(k)$ and $\mathfrak{A}_{(z_1+1, z_2)}(k)$ happen), but at most one in the overlaps $B_{(z_1-1, z_2)}(k)$ and $B_{(z_1, z_2)}(k)$ (occurrence of $\mathfrak{B}_{(z_1-1, z_2)}(k)$ and $\mathfrak{B}_{(z_1, z_2)}(k)$). Then, the crossing clusters in $A_{(z_1+i, z_2)}(k)$ have to be connected.

The events $\mathfrak{A}_y(k)$ and $\mathfrak{B}_y(k)$ are *increasing events*⁶, which allows for the application of

Lemma 28 (FKG inequality [104]). *Let \mathfrak{X} and \mathfrak{Y} be increasing events. Then*

$$P_p(\mathfrak{X} \cap \mathfrak{Y}) \geq P_p(\mathfrak{X})P(\mathfrak{Y}). \quad (5.122)$$

This results in an upper bound for $P_p(\mathfrak{D}_y(k))$, with

$$\mathfrak{D}_y(k) = \left(\bigcap_{a=0,\pm 1} \mathfrak{A}_{(y_1+a,y_2)}(k) \right) \cap \mathfrak{B}_{(y_1-1,y_2)}(k) \cap \mathfrak{B}_{(y_1,y_2)}. \quad (5.123)$$

- $\mathfrak{E}_z(k)$: The next event we need is the one that “connects” two blocks in the second dimension, similar to $\mathfrak{B}_y(k)$ in the first direction. Let $\mathfrak{E}_z(k)$ be the event that there exist at most one open left-to-right crossing cluster in $C_z(k)$. In order to apply the arguments of Ref. [103], we extend the blocks in the last $d-2$ dimensions by another $2k$. The probability of $\mathfrak{E}_z(k)$ occurring is bounded by

$$P_p(\mathfrak{E}_z(k)) \geq 1 - (4k)^{2d} a \exp(-2ck). \quad (5.124)$$

- The last event we will define here, $\mathfrak{F}_z(k)$, is that of having an open left-to-right crossing clusters in both, $A_{(z_1,z_2)}(k)$ and $A_{(z_1,z_2+2)}(k)$, but at most one in $C_z(k) = A_{(z_1,z_2)}(k) \cup A_{(z_1,z_2+2)}(k)$ (see Fig. 5.20). That means, again, that there is actually one left-to-right crossing cluster in $C_z(k)$, but it connects the left and right faces of $A_{(z_1,z_2)}(k)$ and $A_{(z_1,z_2+2)}(k)$ simultaneously. Similar to $P_p(\mathfrak{D}_z(k))$, by using the FKG inequality we can construct an upper bound to the probability of occurrence of $\mathfrak{F}_z(k) = \mathfrak{A}_{(z_1,z_2)}(k) \cap \mathfrak{A}_{(z_1,z_2+2)}(k) \cap \mathfrak{E}_z(k)$.

Having all these events at our disposal, the goal of realising a fully renormalised lattice can be formulated quite easily: we are looking for a simultaneous occurrence of $\mathfrak{D}_z(k)$ and $\mathfrak{F}_z(k)$ for a suitable set of $z = (z_1, z_2)$'s, so

$$\mathfrak{U}(N, k) = \left(\bigcap_{\substack{z_1=3,5,\dots,2N-1 \\ z_2=2,4,\dots,2N}} \mathfrak{D}_z(k) \right) \cap \left(\bigcap_{\substack{z_1=2,4,\dots,2N \\ z_2=2,4,\dots,2(N-1)}} \mathfrak{F}_z(k) \right). \quad (5.125)$$

By subsequent application of the FKG inequality as explained above, we can express an upper bound to the probability of $\mathfrak{U}(N, k)$ occurring in terms of the probabilities

⁶Let $\chi(\mathfrak{A}_p)$ denote the characteristic function of the event \mathfrak{A} for an elementary probability p . An increasing event \mathfrak{A} is one that satisfies $|\chi(\mathfrak{A}_p)| \leq |\chi(\mathfrak{A}_{p'})|$ if $p \leq p'$.

of the former events:

$$\begin{aligned}
P_p(\mathfrak{U}(N, k)) &\geq \prod_{\substack{y_1=2,3,\dots,2N \\ y_2=2,4,\dots,2N}} P_p(\mathfrak{A}_y(k)) \prod_{\substack{y_1=2,3,\dots,2N-1 \\ y_2=2,4,\dots,2N}} P_p(\mathfrak{B}_y(k)) \times \\
&\quad \prod_{\substack{y_1=2,4,\dots,2N \\ y_2=2,4,\dots,2(N-1)}} P_p(\mathfrak{C}_y(k)) \\
&\geq (1 - \exp(-gk^2))^{2N^2-N} \times \\
&\quad ((1 - (2k)^{2d} a \exp(-ck))^2 (1 - (4k)^{2d} a \exp(-c2k)))^{N(N-1)}.
\end{aligned} \tag{5.126}$$

Now, we will have to find the block size scaling $k(N)$ such that this probability is approaching unity for large N . Moreover, we are looking for a “good” scaling, in the sense that the overall resource scaling $N^2 k(N)^d$ does not differ too much from the optimal $O(N^2)$. In order to invert (5.126), so to find the best $k(N)$ consistent with this approach we, however, still need to relax the problem to some extent. By using the slowest increasing term in (5.126) we can bound the expression from above. There exists an integer k_0 such that

$$P_p(\mathfrak{U}(N, k)) \geq (1 - (2k)^{2d} a \exp(-ck))^{5N^2} \tag{5.127}$$

for all $k \geq k_0$. Let us now use the ansatz $k = \lceil N^\varepsilon \rceil$, for some $\varepsilon > 0$.

For any $x \in \mathbb{N}$ there exists a $N_0 \in \mathbb{N}$ such that for all $N \geq N_0$

$$1 - A(2N^\varepsilon)^{2d} \exp(-cN^\varepsilon) \geq 1 - 1/(xN^2). \tag{5.128}$$

Further,

$$\lim_{N \rightarrow \infty} (1 - 1/(xN^2))^{5N^2} = e^{-5/x} \tag{5.129}$$

and for every $\epsilon > 0$ we can find an x such that $1 - e^{-5/x} < 1 - \epsilon$.

Therefore, the chosen dependence of k on N is sufficient to achieve a success probability within a chosen ϵ around 1, getting arbitrary close in the limit of large N :

$$\lim_{N \rightarrow \infty} P_p(\mathfrak{U}(N, k(N))) = 1. \tag{5.130}$$

Combining this with the number of blocks used (N^2) induces a resource scaling of $R(N) = O(N^{2+d\varepsilon})$.

($d = 2$) In the two-dimensional case the connection between paths in the two directions is, of course, not an issue – whenever a block A_y is crossed in both directions,

these two crossing paths necessarily intersect. The events needed in this case are the following:

- $\mathfrak{G}_{i,r}$: The rectangle $C_{i,r} = \cup_{y:y_r=i} A_y$ is crossed in the r -th dimension. Here, $C_{i,r}$ is the i -th row or i -th column in case of $r = 1$ or $r = 2$, respectively. The probability for such an event to happen satisfies [100]

$$P_p(\mathfrak{G}_{i,r}) \geq 1 - skN \exp(-tk), \quad (5.131)$$

with $s, t > 0$.

Again, the $\mathfrak{G}_{i,r}$ are increasing events, so the probability of simultaneous crossings in all rows and columns,

$$\mathfrak{U} = \bigcap_{\substack{i=1,\dots,N \\ r=1,2}} \mathfrak{G}_{i,r} \quad (5.132)$$

satisfies

$$P_p(\mathfrak{U})(N, k) \geq (1 - skN \exp(-tk))^{2N}. \quad (5.133)$$

Now, the choice of $k = \lceil N^\varepsilon \rceil$ can be used again, together with the last steps for the case $d > 2$. Thus, in the two-dimensional case a resource scaling of $R(N) = O(N^{2+d\varepsilon})$ holds as well. \square

Although the proof was explicitly stated in terms of bond percolation, a reasoning along these lines will hold as well for site percolation or mixed site/bond percolation, as long as the probabilities in question are above the respective threshold.

5.3.3 Path identification

For the cluster state to be of any use in quantum computing, the number of resources required to simulate a given quantum circuit has to depend polynomially (such that efficient simulations of the circuit model are possible) on the size of the circuit. As the size N of the cluster state required to implement a given circuit has a polynomial dependence on the circuit's size [15], Theorem 27 already provides a suitable scaling in the number of qubits required.

Still, the amount of time and classical memory required to implement a given computation has to obey a well-tempered scaling as well. The “quantum part” (*i.e.*, the number of subsequent measurements in the preparation- and in the computing stage) only requires $O(1)$ time steps for preparation of the initial pieces and a single step for all the simultaneous entangling operations. Many of the measurements used

to reduce the percolated lattice to the renormalised one and perform the computation can be applied in parallel, but an upper bound is given by $R(N)$.

The classical amount of memory, of course, starts with $R(N)$ to store all gate outcomes (and therefore the percolated graph). In the following, the scaling of classical resources will be analysed in more detail.

Crossing clusters

For identification of the crossing clusters within the blocks, cluster finding algorithms such as the Hoshen-Kopelman-algorithm [105] can be employed. Out of the box this would require $O(k^{d-1})$ of classical memory and $O(kN^2)$ time steps. If there exists more than one crossing cluster (which is as of (5.124) highly improbable), only a single one (*e.g.*, the one with the largest surface) will be chosen for the subsequent procedure.

Connecting the blocks

A “mid-qubit” which is a member of a crossing cluster near the centre of the block will be chosen ($R(N)$ time-steps) in every block. Let us define an open path on $G = (V, E)$ between $a_1, a_{n+1} \in V$ by $\mathcal{P}(a_1, a_{n+1}) = \{(a_1, a_2), (a_2, a_3), \dots, (a_n, a_{n+1})\}$ with $(a_i, a_i + 1) \in E$, $i = 1, \dots, n$, and its length by $|\mathcal{P}(a_1, a_{n+1})| = n$. Because we are using undirected graphs, $(a, b) = (b, a)$ and there is a corresponding path from b to a for each path from a to b . Open paths between the mid-qubits of all pairs of neighbouring blocks are identified using a breadth-first-search (BFS) algorithm [106] ($R(N)$ time and memory complexity).

To prevent loops from being present in the paths in the first place, the following procedure is employed:

1. Using BFS on the crossing clusters starting from the mid-qubits and constrained to the respective block, each site is labelled with the length of the shortest path to the mid-qubit in this block. By going in the direction of decreasing length, the shortest path $\mathcal{P}(s, m(x))$ to a block’s x mid-qubit $m(x)$ can be found within its block, starting from any site $s \in x$.
2. The facing boundaries of all pairs of neighbouring blocks (x_1, x_2) are searched for the pair of sites (s_1, s_2) – one in each block – with the least sum of their distances $|\mathcal{P}(s_1, m_1)| + |\mathcal{P}(s_2, m_2)|$ from their respective mid-qubits m_1 and m_2 , and the bond (s_1, s_2) between them being open.

3. With the composite path $\mathcal{P}(m_1, s_1) \cup (s_1, s_2) \cup \mathcal{P}(s_2, m_2)$ a loop-free connection between m_1 and m_2 is found. Although paths from the mid-qubits to different neighbours might have sub-paths in common, there will be no loops inside a block due to starting always with the same site and the same algorithm (thus delivering a unique path to a common vertex) for all paths to the boundaries in a given block.

5.3.4 Reduction to a renormalised lattice

Instead of renormalising to the whole square lattice which would require to identify cross-like junctions within the blocks, the procedure will renormalise to a hexagonal lattice where only T-junctions are required. Of course, it will be embedded in the square lattice geometry in the obvious way. A square lattice would involve crosses, the construction of which is not obvious when T-junctions have been found and only local measurements are used for reduction. There are, however, easy ways to turn the whole lattice into a square one by using local measurements afterwards [107].

The procedure to cut out parts of the cluster and yanking paths straight involves single qubit measurements of the Pauli operators σ_z and σ_y , respectively. One can store the by-product for each remaining qubit (a memory requirement of $O(1)$ per qubit) and adjust the bases of the subsequent measurements accordingly. Because each qubit will be measured out by the end of the computation, it is sufficient to use this approach for the compensation of random measurement results.

The first application is to isolate the paths and eliminate the spare sites and dangling ends towards the mid-qubits. This is achieved by cutting out all unneeded sites using measurements of the respective qubits in the σ_z basis.

Now one is left with a hexagonal lattice where each edge possibly consists of a long path and each site might consist of a triangular structure in the worst case (depending on the type of lattice used, also the wanted single-qubit sites are possible). The triangles can be destroyed by suitable σ_y -measurements, as shown in Fig. 5.21. After that, only single-qubit junctions are left, the paths between which can be shrunk by subsequent applications of σ_y -measurements to a single edge.

When summing up all these contributions we realise that both – the amount of classical memory and the number of time-steps – are bounded by $R(N)$ as well.

It should be pointed out that one obstacle in one-way quantum computation is to keep the whole state in memory. Having fixed the algorithm in advance, the required state size is known and therefore also the block size for a fixed allowed overall error rate. Therefore, to grow individual blocks and reduce them to single

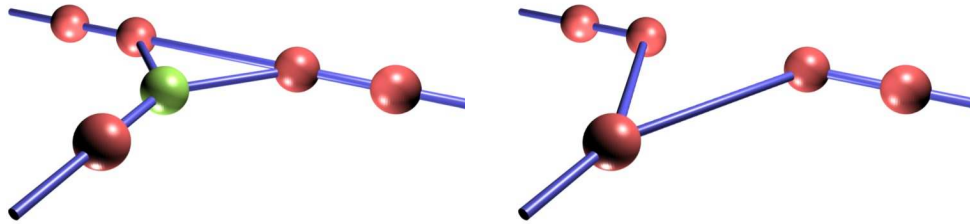


Figure 5.21: Effect of a σ_y measurement on a triangular junction of a cluster. If the three arms are, as shown, not immediately connected outside the triangle, a σ_y measurement on the green qubit has the effect of destroying the triangle in favour of a T-shaped junction. Due to symmetry, any of the three qubits in the triangle could have been chosen.

qubits, only the neighbouring blocks have to exist. This especially allows for growing the reduced lattice in the time direction while the computation moves on, requiring only $O(N^{1+\mu})$ qubits to be kept in memory at a given moment.

5.3.5 Practical considerations – decreasing the vertex degree

Choosing the appropriate lattice

To actually utilise the protocols based on percolation theory, the initial resources (*i.e.*, the stars sitting on the lattice sites) should be as small as possible. As it is more difficult to prepare larger states (this is the problem to be solved in the first place), the lattice with the lowest vertex degree for which $p_{\text{success}} > p^{(c)}$ is still fulfilled will be the most favourable one.

Again, results known in percolation theory but also specifics of the physical implementation can be used to decrease the vertex degree of the initial states. First, using star resources (states corresponding to star graphs – one central vertex with a couple of “arm” vertices, each of which is connected solely to the central one), one would have to look for the lattice with the smallest vertex degree that is still compatible with the bond probabilities at hand. It is not necessary to stay in two dimensions, as the blocking procedure can use higher dimensional lattices and renormalise them to two-dimensional square lattices.

For example, in the case mentioned above, the smallest vertex degree compatible with $p_{\text{success}} = 1/2$ is four, realised by the diamond lattice with $p_{\diamond}^{(c)} \approx 0.389$. That translates into five-qubit initial states. To see that the procedure still works

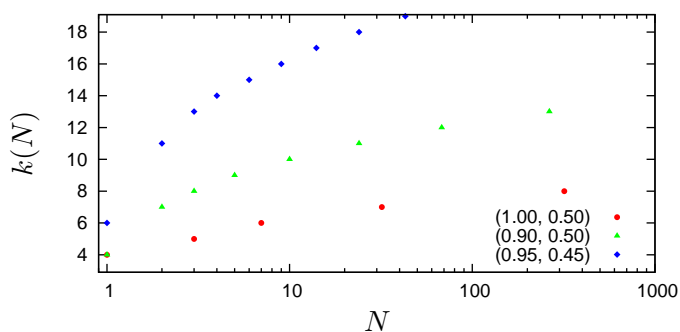


Figure 5.22: Results of Monte Carlo simulations to determine the scaling behavior of the renormalisation procedure on the diamond lattice. The dependence of the diamond lattice’s block size $k \times k \times k$ on the size $N \times N$ of the renormalised square lattice is shown for three different sets of site- and bond probabilities (p_s, p_b) . The threshold for the probability of \mathfrak{A} occurring was chosen to be $1/2$. 10^5 blocks of each size were created and used to randomly populate each lattice size 10^3 times.

for lattices different from cubic, especially the diamond lattice, see the results of numerical simulations in Fig. 5.22.

Besides via the possible vertex degrees, the elementary success probabilities also influences the resource consumption by its distance to the percolation threshold. For exact statements on the required block size, the trade-off between the size of the resource states and the pre-factor of the $N^{2+\mu}$ -term has to be considered. Fig. 5.23 shows the $k(p)$ dependence for the square lattice (vertex degree $\nu = 4$) with sites being occupied probabilistically and the cubic lattice ($\nu = 6$) with probabilistic edges. Especially increasing probabilities near the threshold results in a significant decrease of the block size. This decay happens to be more pronounced for lattices with higher vertex degree.

Covering lattice

If the entangling operations at hand have the property that one of the qubits survives (like the probabilistic parity check, the following property of bond percolation can further reduce the size of the initial pieces. So far the sites of the lattice were occupied by single qubits and the bonds were given by edges in the underlying entanglement graph. If one qubit is left by the entangling operation acting on the stars’ arms, we will not think of this one as being a site itself, but rather belonging to the bond between its neighbouring sites.

Having such a graph state (of which exact lattice type does not matter), we can measure the central qubit a of an initial star (the green one in Fig. 5.24) in the σ_y -basis. Given the specific structure we have at hand, this operation actually

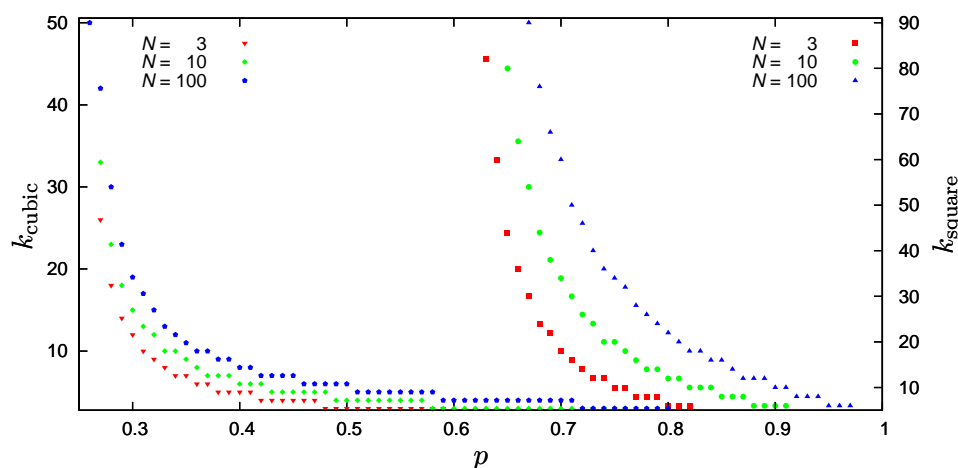


Figure 5.23: Dependence of the block size k on the site occupation probability p to achieve a fixed overall probability (here $P_p(N) \geq 1/2$) with a fixed lattice size ($N = 3, 10, 100$), obtained from Monte Carlo simulations.

For the data on the left side, bond percolation on the three-dimensional cubic lattice ($p^{(c)} \approx 0.249$) was simulated. Each block was realised 10^5 times and each lattice 10^3 times for every combination of k and p displayed. The right side shows the respective graph for site percolation on a square lattice ($p^{(c)} \approx 0.592$) with $5 \cdot 10^5$ block samples and $5 \cdot 10^3$ lattice samples.

Apart from a lower threshold for the lattice with higher vertex degree, the decay is steeper.

performs a transformation from the lattice type we had before to its covering lattice: now think of the qubits that were sitting on the bonds as proper sites. The old sites have disappeared (they have been measured out) and the new ones are connected to all the new sites that were in the same neighbourhood of an old site (local complementation).

Intuitively, the covering lattice has the same connectivity properties as the original lattice before. Paths through a star between two arms are existent iff the star was present and the two entangling operations involving these arms were successful. The same holds on the covering lattice. This property is reflected by the equation $p_{G,b}^{(c)} = p_{G_c,s}^{(c)}$, so bond percolation on the original lattice induces site percolation on the covering one.

As the local complementation inside the stars commute with the entanglement operation between them, the central qubit might be neglected from the very start (see Fig. 5.24). Therefore, one further qubit can be saved by starting with the fully connected graph state (locally equivalent to the GHZ state) that consists of one qubit less than the corresponding star. In case of the diamond lattice, the covering lattice (*pyrochlore lattice*) can be built using four-qubit GHZ states (tetrahedral states).

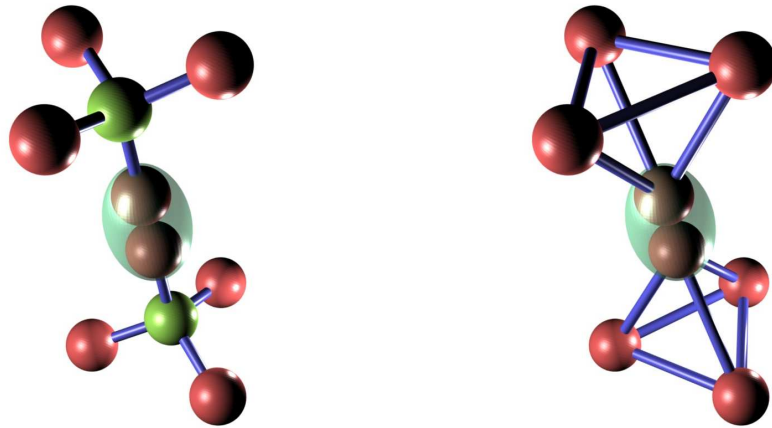


Figure 5.24: Pairs of initial resources for the diamond lattice and its covering one, the pyrochlore lattice. A probabilistic entangling gate is applied between pairs of qubits of all neighbouring initial states. The central qubit, here shown in green, is measured out in the σ_y -basis, resulting in the transformation from diamond to pyrochlore lattice.

When success probabilities in the regime $p_{\diamond}^{(c)} \approx 0.653 < p_{\text{success}} < 1$ are available, the minimum sized resource – a four-qubit state (which amounts to three qubits on the covering lattice) – can be utilised to build a hexagonal lattice.

Further methods to reduce the amount of conditional dynamics in linear optical quantum computing

Surely, if the scheme requires large initial stars, they can be prepared with the same tools probabilistically, starting from smaller stars. For a fair assessment, however, the constraints that led to the percolation scheme in the first place have to be imposed here as well, that is, the restriction to a static setup. Whether composite stars can be used in a static layout now depends on how the entangling gates work in detail, *i.e.*, what the failure outcomes are. That is of interest due to the fact that in general a failure in the star preparation step would require back-up steps that do not allow for subsequent application of further entangling gates without re-routing. A type of gates with suitable properties is the one that acts as a σ_z -measurement on both qubits on failure. For example, parity check based gates in linear optics offer this feature.

That this behaviour might actually bring some benefit is shown by the following example (see Fig. 5.25). Two instances of such an entangling gate are applied to a pair of five-qubit stars, one to a pair of arms and one to the central qubits. On success of the “arm” fusion gates, the two stars are connected by a two-edge chain. The middle qubit of this chain (green) will be measured out in the σ_x -basis, leaving

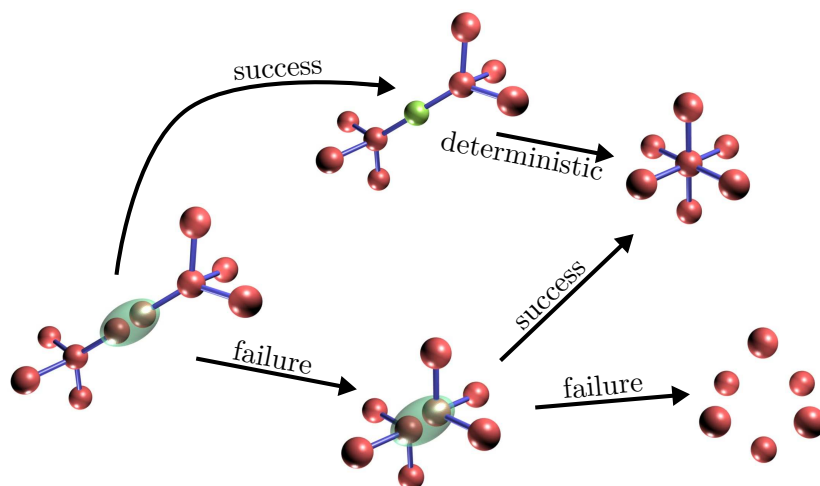


Figure 5.25: A pair of five-qubit stars (stars with four arms and one central qubit) can be used to create a single seven-qubit star with a probability of $p_{\text{success}} = 1 - (1 - p_{\text{gate}})^2$. With $p_{\text{failure}} = (1 - p_{\text{gate}})^2$ the six arm qubits are separated from each other.

the two centres merged in a redundantly encoded qubit which now constitutes the centre of a star with six arms. A second application of such a gate on the two qubits of the new centre will always succeed due to the fact that locally they sit in the symmetric subspace, and will reduce the level of redundancy encoding in the centre qubit by one.

If, however, the first gate operation failed, the two arms the gate was acting on will be cut off as a consequence of the σ_z failure outcome. The σ_x -measurement now acts on one part of a product state, leaving the other part – the two stars – unchanged. Now, a second attempt is possible by application of the entangling gate to the centre qubits. The success outcome is a six-arm star as well. On failure, however, the centres will be cut off, leaving the six qubits in the product state $|+\rangle^{\otimes 6}$. The failure outcome is the result of two consecutive failures of these entangling gates, which means for the success probability of site creation $p_{\text{success}} = 1 - (1 - p_{\text{gate}})^2$, which amounts to $p_{\text{success}} = 3/4$ in the case of the linear optics gate mentioned earlier.

All these operations do not need any classical post-processing or re-routing. Therefore, this scheme is suitable to be used in the static state production procedures introduced above. Because the centre qubits are simply cut off when a failure occurs, this procedure and the bond percolation involving the arms are completely independent. So, both processes together can be modelled as a mixed site/bond percolation with the site probability being $p_s = p_{\text{success}}$.

As long as the bond probability and the initial state preparation (site-) probability are above the percolation threshold of the respective lattice, this scheme might be useful to further reduce the size of the initial star shaped quantum states.

5.3.6 Loss tolerance

Although the primary concern of this chapter was to show how to effectively remove active feed-forward and to deal with probabilistic gates by using tools from percolation theory, we briefly address the role of further imperfections in such a setting. To start with, *heralded losses* – those losses for which one has a classical signal indicating its occurrence, without the need for destructively measuring them (*e.g.*, probabilistic entangling gates applied to atomic systems which might escape from the cavity) – can be accounted for by measuring out the qubits around failure sites in the σ_z -basis. This is effectively a new percolation model, albeit with correlated probabilities. Numerical investigations show that a similar reasoning as above is expected to hold, demonstrating that losses of 10% can easily be accounted for.

One type of heralded errors in linear optics is loss of the photons carrying the arm qubits. When using – instead of type-I – the type-II fusion gates (projective measurements which condition on detecting two photons), the post-selection on two detected photons during gate application allows to infer whether arm-photons were lost. Because both arm-qubits will be measured out, the covering lattice cannot be used in this setting, though.

More importantly, concerning *blind losses* such as photon loss, which are only detected by destructively measuring the central qubits, this scheme is no different than others, *i.e.*, standard techniques can make it loss tolerant, although, needless to say, with significantly more effort⁷. Our approach readily suggests two strategies to cope with such errors: On the one hand, by fixing the block size, one would fix the effective (non-heralded) site occupation probability. Then, schemes for fully-fledged fault-tolerant one-way computation can be used, once above the respective fault-tolerant threshold [110]. On the other hand, to suppress loss rates, specifically photon loss, it is legitimate to consider initial encodings like tree-structures [68]. They in turn can be grown probabilistically as well, still without any need for active switching. For example, a crude estimate is the following: To correct for these losses

⁷Apart from losses, other types of imperfections would have to be handled using standard error correction techniques. Not only detectors with their low efficiencies and optical networks with mode mismatching, but even the very process of photon generation leads to errors such as slightly distinguishable wave packets (first solutions to this problem can be found in Refs. [108, 109]).

occurring with a rate of 10%, trees with branching parameters $(6, 7, 7, 1)$ [111] can be used to suppress the loss to an effective rate of 10^{-5} . Together with blocks of the size of $6^{\times 3}$ elementary diamond cells (which fail to be crossed with a probability of $4.76 \cdot 10^{-6}$), the overall site loss rate on the renormalised lattice lies below the $3 \cdot 10^{-3}$ [110] limit.

5.4 Discussion

In this chapter we investigated the question how fusion-like probabilistic gates can be employed in the process of cluster state generation in an optimal way. The first part was concerned with a rather general approach with only fixed gates and a fixed type of resources. While a vast variety of possible decisions can arise during the process, we were able to capture the essential information by rigorously taking care of the averages occurring and finding the minimum number of steps needed in a drastically trimmed version of the problem. Focusing upon the linear optics parity checks (but still valid for arbitrary gate probabilities) we were able to identify strict upper and lower bounds for the optimal solution to both – the problem where a fixed number of resources is given as well the one where the objective length of the chain is fixed. These bounds tell that the best procedure based on gates with $p_s = 1/2$ requires a constant investment of between 6.47 and 6.53 EPR pairs per additional edge in the objective chain on average. Further, we showed how to use these tools to construct converging series of upper and lower bounds.

These results exceed the ones known so far as only finite size examples and indications to the possibility of linear growth in principle [82, 38, 86, 83] were discussed before.

For the case of a two-dimensional cluster state, we supplied the first rigorous proof of the possibility of the consumption of $O(N^2)$ EPR pairs during the construction of a cluster state of size $N \times N$. In contrast to results known before ($O(N^2 \log(N^2))$ [82, 83, 84, 85]), N^2 is the optimum because new edges cannot be generated in the process.

In the second part a rather more restricted setting was considered. Motivated by experimental constraints we introduced an experimentally more feasible, but more restricted, static scheme. We showed that by lifting the requirements for conditional dynamics quite substantially (no re-routing involved anymore and only mode-matching between neighbouring sites required), still a scaling of $O(N^{2+\mu})$ can be achieved for any strictly positive probability of the elementary gates. By analysing the amount of classical computation needed we prove the feasibility of the

scheme. Finally, the validity for a variety of lattices is exploited to minimise the initial resource states for the linear optics case (4-qubit GHZ states).

In this thesis we have discussed topics in the field of linear optics quantum computing. While the nature of problems arising from linear optics varies a lot depending on the specific question, our approach was two-pronged to cover different aspects of the area.

On the one hand we have derived some statements on the overall scalability of linear optics in principle. We were able to show that cluster states – a possible resource for universal quantum computing – of size $n \times n$ can be generated with $O(n^2)$ EPR pairs despite the probabilistic behaviour of the linear optics quantum gates at hand. Even when refraining from active switching of optical paths – a demanding task when it comes to interferometric stability – a procedure achieving roughly the same scaling was identified. Then, for any $\mu > 0$ the number of initial ingredients (which are now GHZ states of a certain size depending on the probabilities involved) scales as $O(n^{2+\mu})$.

Due to its resistance against errors induced by probabilistic gates, this static approach also shows an intrinsic robustness against certain types of losses. Further effort could be invested to analyse the impact of other types of errors including mode mismatching and photon loss. Perhaps the robustness of percolation itself can be exploited for more general error correction as well.

On the other hand, methods for analysing specific small-scale optical networks were discussed. Applied to small problems ranging from state-production and the implementation of quantum gates to measurements they proved to be reliable tools for network construction and assessment of their performance, and might help to gain intuition on the internals of linear optics. Optimal linear transformations of the creation operators, inducing post-selected controlled phase gates have been obtained and it was tried to cast them into a form suitable for experiments. An interesting feature that was observed is the following. When decreasing the phase of the optimal controlled phase gate (also applies to gates with more than one control qubit) below $\varphi = \pi$, the probability of success does not increase monotonously. Instead, it drops to a smaller value until a steep incline brings it to the trivial case $p_s(0) = 1$.

Also, networks acting as generalised Toffoli gates, unambiguous Bell state discriminators, or photon number resolving detectors have been investigated with a range of counter-intuitive results. While the influence of (hard to implement) active re-routing turned out to be not that huge in that the scaling with and without does not differ much, other components of linear optics toolboxes have quite a considerable impact. The ability to change the parameters of local elements at random – which is a comparably easy experimental task – emerge as an indispensable good in state discrimination.

Further, the question of how to discriminate number states with linear optical means was settled. Until dichotomic detectors with very high efficiencies are available, cascading layouts will be the best choice.

Again, so far, only the perfect case was considered. In principle, the performance of such networks could exhibit a rather unstable behaviour when it comes to imperfections such as mode mismatching. For experimental realisations it would be helpful to have algorithms which generate networks which are “robust” and also “simple to implement”. For example, small deviations from the optimal probability of success do not carry so much weight if the experiments are simplified in turn, for example if the number of interferometers in the network is decreased.

Acknowledgements

The author would like to thank particularly Jens Eisert for his tutoring and support and David Gross for a lot of support and inspiration.

I am especially grateful to have received steady encouragement and support from my parents and my girlfriend Katrin in particular also during the three years abroad.

Also a special thanks goes to my office mates Alvaro Feito, Fernando Brandao, Doug Plato, and Oscar Dahlsten for many hours of discussions (not necessarily related to the research) and for providing such a fruitful environment.

During the course of this research I further benefitted a lot from insight- and joyful discussions with Terry Rudolph, Gunnar Pruessner, Pavel Lougovski, Witlef Wieczorek, Jonathan Mathews, Stefan Scheel, Jeremy O'Brien, Christian Schmid, Chris Dawson, and others. The hospitality of Jonathan Dowling made it possible to spend some time in his group and gain a lot of new inspiration.

Furthermore I want to acknowledge the help of Arwed Weigel, Tanja Gärtner and Katrin Lipke for spotting minor and major flaws in my use of the English language.

This work was supported by Microsoft Research through its European PhD Scholarship Programme.

Some of the symbols used in this thesis are not so well known in the quantum information community or differ from conventions. This list shall provide an overview of what the symbols shall mean.

| | |
|----------------------------------|--|
| $\delta_{i,j}$ | The Kronecker symbol: equals to 1 iff $i = j$ and 0 otherwise. |
| \mathbf{x} | Complex or real vector. |
| $\mathbb{K}[x_1, \dots, x_n]$ | Polynomial ring over the coefficient field \mathbb{K} with variables x_1, \dots, x_n (mainly used in Chapter 3). Elements are abstract polynomials in the x_i with coefficients from \mathbb{K} . |
| a_i, a_i^\dagger | Ladder operators (annihilation and creation operators) on the bosonic field mode i . See also page 16. |
| \mathbf{a}^\dagger | Formal vector of creation operators on all relevant modes. |
| A, U | Complex matrices describing linear transformations of the creation operators as introduced on page 16. U will be used to highlight when such a matrix is unitary (<i>i.e.</i> , $U^\dagger = U^{-1}$). |
| \mathcal{A}, \mathcal{U} | The operators on Hilbert space corresponding to A and U , respectively. |
| p_s | Probability of success of the network under consideration. |
| $\text{diag}\{x_1, \dots, x_n\}$ | $n \times n$ matrix with entries x_1, \dots, x_n on its main diagonal and 0 everywhere else. |
| $ \psi_{1,2,3,4}\rangle$ | The four <i>Bell-states</i> , defined on page 62. |

In Section 3.1.7, the general process of finding the optimal success probability of a linear optics gate is shown. Constraints to the optimisation problem are given by bounds to the singular values $1 \geq \sigma_1(A) \geq \dots \geq \sigma_n(A) \geq 0$ of the beam splitter matrix $A \in \mathbb{C}^{n \times n}$. Usually A is still subject to row-wise or column-wise rescaling, $A \mapsto XAY$ with diagonal matrices X and Y , and the constraints will become polynomial inequalities in these free parameters.

To obtain bounds to the success probability, bounds to the singular values of matrix products can be used. As supplementary material to the discussion in Section 3.1.7 we show the relations known from the literature:

- Let $A, B \in \mathbb{C}^{n \times n}$. For $r \in \mathbb{R}^+$ and any subset of singular values labelled by $1 \leq i_1 < \dots < i_k \leq n$, the relations

$$\sum_{t=1}^k \sigma_{i_t}^r(AB) \geq \sum_{t=1}^k \sigma_{i_t}^r(A) \sigma_{n-t+1}^r(B) \quad \text{and} \quad (\text{B.1})$$

$$\sum_{t=1}^k \sigma_t^r(AB) \geq \sum_{t=1}^k \sigma_{i_t}^r(A) \sigma_{n-i_t+1}^r(B) \quad (\text{B.2})$$

hold [112].

- For any $k = 1, \dots, n$ the following upper and lower bounds exist [113]:

$$\min_{1 \leq i \leq k} [\sigma_i(B) \sigma_{k+1-i}(A)] \geq \sigma_k(BA) \geq \max_{k \leq i \leq n} [\sigma_i(B) \sigma_{n+k-i}(A)]. \quad (\text{B.3})$$

- The singular values of the matrix product AB for any $r > 0$ is bounded by [114]

$$\sum_{j=n-k+1}^n \sigma_j^r(AB) \leq \sum_{j=n-k+1}^n \sigma_j^r(A) \sigma_j^r(B). \quad (\text{B.4})$$

Note, that only Eqns. (B.1) and (B.2) respect the non-commuting nature of the matrix product. The other approximations can only be expected to deliver tight bounds if the optimal X and Y are the same.

Publications

Results presented in this thesis are partly already published in refereed journals or as book chapters. The publications this thesis is based on are:

- K. Kieling, D. Gross, and J. Eisert, Minimal resources for linear optical one-way computing. *J. Opt. Soc. Am. B* **24** (184), 2006. 1
- D. Gross, K. Kieling, and J. Eisert, Potential and limits to cluster state quantum computing using probabilistic gates. *Phys. Rev. A* **74** (042343), 2006. 2
- K. Kieling, T. Rudolph, and J. Eisert, Percolation, renormalization, and quantum computing with non-deterministic gates. *Phys. Rev. Lett.* **99** (130501), 2007. 3
- N. M. VanMeter, P. Lougovski, D. B. Uskov, K. Kieling, J. Eisert, and J. P. Dowling, General linear-optical quantum state generation scheme: Applications to maximally path-entangled states. *Phys. Rev. A* **76** (063808), 2007. 4
- K. Kieling, D. Gross, and J. Eisert, Cluster state preparation using gates operating at arbitrary success probabilities. *New J. Phys.* **9** (200), 2007. 5
- K. Kieling and J. Eisert, Percolation in quantum computation and communication. In *Quantum and Semi-classical Percolation and Breakdown in Disordered Solids* (edited by A. K. Sen, K. K. Bardhan, and B. K. Chakrabarti), *Lecture Notes in Physics*, volume 762, Springer, 2008. 6

Bibliography

- C. E. Shannon, A mathematical theory of communication. *The Bell System Technical Journal* **27** (379–423, 623–656), 1948. 7
- A. M. Turing, On computable numbers, with an application to the entscheidungsproblem. *Proc. Lond. Math. Soc.* **s2-42** (230–265), 1937. 8
- R. Landauer, Information is physical. In *Proc. Workshop on Physics and Computation*, 1992. 9
- P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Statist. Comput.* **26** (1484), 1997. 10
- L. K. Grover, Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79** (325–328), 1997. 11
- C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing*, 1984. 12
- P. Kok, W. Munro, K. Nemoto, T. Ralph, J. P. Dowling, and G. Milburn, Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.* **79** (135–174), 2007. 13
- A. Politi, M. J. Cryan, J. G. Rarity, S. Yu, and J. L. O’Brien, Silica-on-silicon waveguide quantum circuits. *Science* **320** (646–649), 2008. 14
- R. Raussendorf and H. J. Briegel, A one-way quantum computer. *Phys. Rev. Lett.* **86** (5188–5191), 2001. 15
- M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge University Press, 2000. 16

- J. Eisert and M. M. Wolf, Quantum computing. In *Handbook of nature-inspired and innovative computing* (edited by A. Y. Zomaya), chapter 8, 253–286, Springer, New York, 2004. 17
- S. Scheel, K. Nemoto, W. J. Munro, and P. L. Knight, Measurement-induced nonlinearity in linear optics. *Phys. Rev. A* **68** (032310), 2003. 18
- K. Kieling, *Linear optics methods in quantum information processing*. Diploma thesis, University of Potsdam, 2005. 19
- M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **73** (58–61), 1994. 20
- H. Minc, *Permanents*. Addison-Wesley, 1978. 21
- S. Scheel, Permanents in linear optical networks, 2004, [quant-ph/0406127](#). 22
- N. Kiesel, C. Schmid, U. Weber, R. Ursin, and H. Weinfurter, Linear optics c-phase gate made simple. *Phys. Rev. Lett.* **95** (210505), 2005. 23
- N. K. Langford, T. J. Weinhold, R. Prevedel, K. J. Resch, A. Gilchrist, J. L. O’Brien, G. J. Pryde, and A. G. White, Demonstration of a simple entangling optical gate and its use in bell-state analysis. *Phys. Rev. Lett.* **95** (210504), 2005. 24
- R. Okamoto, H. F. Hofmann, S. Takeuchi, and K. Sasaki, Demonstration of an optical quantum controlled-not gate without path interference. *Phys. Rev. Lett.* **95** (210506), 2005. 25
- D. A. Cox, J. Little, and D. O’Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Undergraduate Texts in Mathematics, Springer, 1991. 26
- D. A. Cox, J. Little, and D. O’Shea, *Using algebraic geometry*. Graduate Texts in Mathematics, Springer, 2004. 27
- J. von zur Gathen and J. Gerhard, *Modern computer algebra*. Cambridge University Press, 1999. 28
- K. Kühnle and E. W. Mayr, Exponential space computation of gröbner bases. In *Proceedings of the 1996 international symposium on Symbolic and algebraic computation*, 63–71, 1996. 29

- G. Matera and J. M. T. Torres, The space complexity of elimination theory: Upper bounds. In *Foundations of Computational Mathematics, Springer, 1997* (edited by F. Cucker and M. Shub), Springer, 1997. 30
- E. Knill, Quantum gates using linear optics and postselection. *Phys. Rev. A* **66** (052306), 2002. 31
- R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge University Press, 1985. 32
- L. Vandenberghe and S. Boyd, Semidefinite programming. *SIAM Review* **38** (49), 1996. 33
- S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004. 34
- N. Lütkenhaus, J. Calsamiglia, and K.-A. Suominen, Bell measurements for teleportation. *Phys. Rev. A* **59** (3295–3300), 1999. 35
- J. Calsamiglia, Generalized measurements by linear elements. *Phys. Rev. A* **65** (030301), 2002. 36
- E. Knill, R. Laflamme, and G. J. Milburn, A scheme for efficient quantum computation with linear optics. *Nature* **409** (46–52), 2001. 37
- D. E. Browne and T. Rudolph, Resource-efficient linear optical quantum computation. *Phys. Rev. Lett.* **95** (010501), 2005. 38
- A. Acín, A. Andrianov, L. Costa, E. Jan, J. I. Latorre, and R. Tarrach, Generalized schmidt decomposition and classification of three-quantum-bit states. *Phys. Rev. Lett.* **85** (1560–1563), 2000. 39
- H. A. Carteret, A. Higuchi, and A. Sudbery, Multipartite generalization of the schmidt decomposition. *J. Math. Phys.* **41** (7932–7939), 2000. 40
- J. D. Carroll and J.-J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition. *Psychometrika* **35** (283–319), 1970. 41
- R. Harshman, Parafac. *UCLA Working Papers in Phonetics* **16** (1–84), 1970. 42
- L. R. Tucker, Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** (279–311), 1966. 43

- L. D. Lathauwer, B. D. Moor, and J. Vandewalle, A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21** (1253–1278), 2000. 44
- W. Ruppert, Reducibility of polynomials $f(x, y)$ modulo p . *Journal of Number Theory* **77** (62–70), 1999. 45
- E. Kaltofen, Effective noether irreducibility forms and applications. *Journal of Computer and System Sciences* **50** (274–295), 1995. 46
- H. Lee, P. Kok, and J. P. Dowling, A quantum rosetta stone for interferometry. *J. Mod. Opt.* **49** (2325–2338), 2002. 47
- J. Fiurášek, Conditional generation of n -photon entangled states of light. *Phys. Rev. A* **65** (053818), 2002. 48
- X. Zou, K. Pahlke, and W. Mathis, Generation of entangled states of two traveling modes for fixed number of photon. *Phys. Rev. A* **66** (014102), 2002. 49
- P. Kok, H. Lee, and J. P. Dowling, Creation of large-photon-number path entanglement conditioned on photodetection. *Phys. Rev. A* **65** (052104), 2002. 50
- B. P. Lanyon, T. J. Weinhold, N. K. Langford, J. L. O’Brien, K. J. Resch, A. Gilchrist, and A. G. White, Manipulating biphotonic qutrits. *Phys. Rev. Lett.* **100** (060504), 2008. 51
- J. B. Kruskal, Rank, decomposition, and uniqueness for 3-way and n -way arrays. In *Multiway Data Analysis* (edited by R. Coppi and S. Bolasco), Elsevier, 1989. 52
- H. F. Hofmann and S. Takeuchi, Quantum phase gate for photonic qubits using only beam splitters and postselection. *Phys. Rev. A* **66** (024308), 2002. 53
- T. C. Ralph, N. K. Langford, T. B. Bell, and A. G. White, Linear optical controlled-not gate in the coincidence basis. *Phys. Rev. A* **65** (062324), 2002. 54
- J. Eisert, Optimizing linear optics quantum gates. *Phys. Rev. Lett.* **95** (040502), 2005. 55
- D. Gross and J. Eisert, Novel schemes for measurement-based quantum computation. *Phys. Rev. Lett.* **98** (220503), 2007. 56
- J. Eisert, K. Jacobs, P. Papadopoulos, and M. B. Plenio, Optimal local implementation of nonlocal quantum gates. *Phys. Rev. A* **62** (052317), 2000. 57
- LOQC – construction of small networks and asymptotic scaling 148

- J. I. Cirac, W. Dr, B. Kraus, and M. Lewenstein, Entangling operations and their implementation using a small amount of entanglement. *Phys. Rev. Lett.* **86** (544–547), 2001. 58
- D. W. Berry, Implementation of multipartite unitary operations with limited resources. *Phys. Rev. A* **75** (032349), 2007. 59
- Q. Zhang, X.-H. Bao, C.-Y. Lu, X.-Q. Zhou, T. Yang, T. Rudolph, and J.-W. Pan, Demonstration of a efficient scheme for generation of "event ready" entangled photon pairs from single photon source. *Phys. Rev. A* **77** (062316), 2008. 60
- T. C. Ralph, K. J. Resch, and A. Gilchrist, Efficient toffoli gates using qudits. *Phys. Rev. A* **75** (022313), 2007. 61
- J. Fiurášek, Linear optics quantum toffoli and fredkin gates. *Phys. Rev. A* **73** (062313), 2006. 62
- J. Calsamiglia and N. Lütkenhaus, Maximum efficiency of a linear-optical bell-state analyzer. *Appl. Phys. B* **72** (67–71), 2001. 63
- P. van Loock and N. Lütkenhaus, Simple criteria for the implementation of projective measurements with linear optics. *Phys. Rev. A* **69** (012302), 2004. 64
- S. Massar and S. Popescu, Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74** (1259–1263), 1995. 65
- H. Weinfurter, Experimental bell-state analysis. *Europhys. Lett.* **25** (559–564), 1994. 66
- S. L. Braunstein and A. Mann, Measurement of the bell operator and quantum teleportation. *Phys. Rev. A* **51** (R1727–R1730), 1995. 67
- M. Varnava, D. E. Browne, and T. Rudolph, Loss tolerant one-way quantum computation – a horticultural approach. *Phys. Rev. Lett.* **97** (120501), 2006. 68
- M. M. Wilde, F. Spedalieri, J. P. Dowling, and H. Lee, Optical cluster-state generation without number-resolving photon detectors. *Int. J. Quant. Inf.* **5** (617–626), 2007. 69
- P. Kok, H. Lee, and J. P. Dowling, Single-photon quantum-nondemolition detectors constructed with linear optics and projective measurements. *Phys. Rev. A* **66** (063814), 2002. 70

- P. Kok, Limitations on building single-photon-resolution detection devices. *IEEE: Sel. Top. Quantum Electronics* **9** (1498), 2003. 71
- D. Achilles, C. Silberhorn, C. Sliwa, K. Banaszek, and I. A. Walmsley, Fiber-assisted detection with photon number resolution. *Opt. Lett.* **28** (2387), 2003. 72
- D. Achilles, C. Silberhorn, C. Sliwa, K. Banaszek, I. A. Walmsley, M. J. Fitch, B. C. Jacobs, T. B. Pittman, and J. D. Franson, Photon number resolving detection using time-multiplexing. *J. Mod. Opt.* **51** (1499), 2004. 73
- E. Waks, K. Inoue, E. Diamanti, and Y. Yamamoto, High efficiency photon number detection for quantum information processing. *IEEE: Sel. Top. Quantum Electronics* **9** (1502–1511), 2003. 74
- D. Schlingemann and R. F. Werner, Quantum error-correcting codes associated with graphs. *Phys. Rev. A* **65** (012308), 2002. 75
- M. Hein, J. Eisert, and H. J. Briegel, Multiparty entanglement in graph states. *Phys. Rev. A* **69** (062311), 2004. 76
- M. V. den Nest, J. Dehaene, and B. D. Moor, Graphical description of the action of local clifford transformations on graph states. *Phys. Rev. A* **69** (022316), 2004. 77
- J. A. Bondy and U. S. R. Murty, *Graph theory with applications*. Macmillan, 1976. 78
- O. Mandel, M. Greiner, A. Widera, T. Rom, T. W. Hänsch, and I. Bloch, Controlled collisions for multi-particle entanglement of optically trapped atoms. *Nature* **425** (937–940), 2003. 79
- H. J. Briegel and R. Raussendorf, Persistent entanglement in arrays of interacting particles. *Phys. Rev. Lett.* **86** (910–913), 2001. 80
- R. Raussendorf, D. E. Browne, and H. Briegel, Measurement-based quantum computation with cluster states. *Phys. Rev. A* **68** (022312), 2003. 81
- M. A. Nielsen, Optical quantum computation using cluster states. *Phys. Rev. Lett.* **93** (040503), 2004. 82
- L.-M. Duan and R. Raussendorf, Efficient quantum computation with probabilistic quantum gates. *Phys. Rev. Lett.* **95** (080503), 2005. 83

- Q. Chen, J. Cheng, K.-L. Wang, and J. Du, Efficient construction of 2-d cluster states with probabilistic quantum gates. *Phys. Rev. A* **73** (012303), 2006. 84
- D.-S. Diao, Y.-S. Zhang, X.-F. Zhou, and G.-C. Guo, Efficient construction of high-dimensional cluster state. *Chinese Physics Letters* **25** (3555–3557), 2008. 85
- S. D. Barrett and P. Kok, Efficient high-fidelity quantum computation using matter qubits and linear optics. *Phys. Rev. A* **71** (060310(R)), 2005. 86
- T. B. Pittman, B. C. Jacobs, and J. D. Franson, Probabilistic quantum logic operations using polarizing beam splitters. *Phys. Rev. A* **64** (062311), 2001. 87
- C. Cabrillo, J. I. Cirac, P. Garca-Fernndez, and P. Zoller, Creation of entangled states of distant atoms by interference. *Phys. Rev. A* **59** (1025–1033), 1999. 88
- L.-M. Duan and H. J. Kimble, Efficient engineering of multiatom entanglement through single-photon detections. *Phys. Rev. Lett.* **90** (253601), 2003. 89
- S. G. R. Louis, K. Nemoto, W. J. Munro, and T. P. Spiller, The efficiencies of generating cluster states with weak non-linearities. *New J. Phys.* **9** (193), 2007. 90
- X.-H. Bao, T.-Y. Chen, Q. Zhang, J. Yang, H. Zhang, T. Yang, and J.-W. Pan, Optical nondestructive controlled-not gate without using entangled photons. *Phys. Rev. Lett.* **98** (170502), 2007. 91
- L.-M. Duan, M. J. Madsen, D. L. Moehring, P. Maunz, R. N. K. Jr., and C. Monroe, Probabilistic quantum gates between remote atoms through interference of optical frequency qubits. *Phys. Rev. A* **73** (062324), 2006. 92
- N. J. A. Sloane, The on-line encyclopedia of integer sequences, sequence A000070. 2003. 93
- R. P. Stanley, *Enumerative combinatorics*. Wadsworth, 1986. 94
- W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Am. Stat. Ass.* **58** (13–30), 1963. 95
- G. Gilbert, M. Hamrick, and Y. S. Weinstein, Efficient construction of photonic quantum computational clusters. *Phys. Rev. A* **73** (064303), 2006. 96

- T. B. Pittman, B. C. Jacobs, and J. D. Franson, Demonstration of feed-forward control for linear optics quantum computation. *Phys. Rev. A* **66** (052305), 2002. 97
- R. Prevedel, P. Walther, F. Tiefenbacher, P. Bhi, R. Kaltenbaek, T. Jennewein, and A. Zeilinger, High-speed linear optics quantum computing using active feed-forward. *Nature* **445** (65–69), 2007. 98
- D. E. Browne, M. B. Elliott, S. T. Flammia, S. T. Merkel, A. Miyake, and A. J. Short, Phase transition of computational power in the resource states for one-way quantum computation. *New J. Phys.* **10** (023010), 2008. 99
- G. Grimmett, *Percolation*. Springer, 2nd edition, 1999. 100
- D. Stauffer and A. Aharony, *Introduction to percolation theory*. Taylor & Francis, 2nd edition, 1994. 101
- S. C. van der Marck, Percolation thresholds and universal formulas. *Phys. Rev. E* **55** (001514), 1997. 102
- M. Aizenman, On the number of incipient spanning clusters. *Nuclear Physics B* **485** (551–582), 1997. 103
- C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre, Correlation inequalities on some partially ordered sets. *Commun. Math. Phys.* **22** (89–103), 1971. 104
- J. Hoshen and R. Kopelman, Percolation and cluster distribution. I. cluster multiple labeling technique and critical concentration algorithm. *Phys. Rev. B* **14** (3438–3445), 1976. 105
- R. Sedgewick, *Algorithms*. Addison-Wesley, 1983. 106
- M. V. den Nest, A. Miyake, W. Dr, and H. J. Briegel, Universal resources for measurement-based quantum computation. *Phys. Rev. Lett.* **97** (150504), 2006. 107
- E. T. Campbell, J. Fitzsimons, S. C. Benjamin, and P. Kok, Adaptive strategies for graph state growth in the presence of monitored errors. *Phys. Rev. A* **75** (042303), 2007. 108
- E. T. Campbell, J. Fitzsimons, S. C. Benjamin, and P. Kok, Efficient growth of complex graph states via imperfect path erasure. *New J. Phys.* **9** (196), 2007. 109
- LOQC – construction of small networks and asymptotic scaling 152

- C. M. Dawson, H. L. Haselgrove, and M. A. Nielsen, Noise thresholds for optical quantum computers. *Phys. Rev. Lett.* **96** (020501), 2006. 110
- M. Varnava, 2007, private communication. 111
- B.-Y. Wang and B.-Y. Xi, Some inequalities for singular values of matrix products. *Linear Algebra and its Applications* **264** (109–115), 1997. 112
- L.-Z. Lu and C. Pearce, Some new bounds for singular values and eigenvalues of matrix products. *Annals of Operations Research* **98** (141–148), 2000. 113
- S. K. Godunov, *Modern aspects of linear algebra, Translations of Mathematical Monographs*, volume 175. American Mathematical Society, 1998. 114