Computational Perspectives on Bell Inequalities and Many-body Quantum Correlations

Matthew Joseph Hoban

A thesis submitted to

University College London

for the degree of

Doctor of Philosophy

Department of Physics and Astronomy
University College London
April 26, 2012

I, Matthew Joseph Hoban confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The predictions of quantum mechanics cannot be resolved with a completely classical view of the world. In particular, the statistics of space-like separated measurements on entangled quantum systems violate a Bell inequality [Bell1964].

We put forward a computational perspective on a broad class of Bell tests that study correlators, or the statistics of joint measurement outcomes. We associate particular maps, or functions to particular theories. The violation of a Bell inequality then implies the ability to perform some functions, or computations that classical, or more generally, local hidden variable (LHV) theories cannot.

We derive an infinite class of Bell inequalities that establish a link to so-called "non-local games" [Cleve2004]. We then make the connection between Raussendorf and Briegel's formulation of Measurement-based Quantum Computing (MBQC) [Raussendorf2001], and these non-local games. Not only can we show that a quantum violation implies a computational advantage in this model, we show that adaptive measurements are required to perform all quantum computations.

Finally, we explore post-selection of data in Bell tests from both a practical and conceptual point-of-view, with particular consideration to so-called "loopholes". Loopholes allow LHV theories to simulate quantum correlations through post-selection. We give a computational description of how loopholes can emerge in different post-selection scenarios. This motivates us to find a form of post-selection that does not lead to loopholes. Central again to this discussion is the description of LHV theories in terms of computations.

Interestingly, quantum correlators can be made more "non-classical" with this loophole-free post-selection. This method of post-selection also can simulate information processing tasks, such as MBQC, that have time-like separated components. This opens up new avenues for the study of time-like tasks studied within the space-like separated scenario of the Bell test.

Publications

The majority of the work in this thesis is based on the following publications:

M. J. Hoban, E. T. Campbell, K. Loukopoulos, and D. E. Browne, *Non-adaptive Measurement-based Quantum Computation and Multi-party Bell Inequalities*, New J. Phys. **13** 023014 (2011).

M. J. Hoban and D. E. Browne, *Stronger Quantum Correlations with Loophole-Free Postselection*, Phys. Rev. Lett. **107**, 120402 (2011).

M. J. Hoban, J. J. Wallman, and D. E. Browne, *Generalized Bell-inequality experiments and computation*, Phys. Rev. A **84**, 062107 (2011).

Acknowledgements

I am indebted to many different people for many things throughout the process of completing my PhD. First of all, I desperately need to thank my supervisor, Dan Browne. To paraphrase Winston Churchill, "Never was so much owed by one PhD student to one tireless supervisor." I thank him for his patience, enthusiasm, insight and for sharing his great ideas. Not only has he been a great supervisor, he has been a good friend.

My examiners Jon Barrett and Sougato Bose need to be thanked for their thorough reading of this thesis. Their experience and insight has only improved this document. It remains to be said that any remaining inaccuracies result from me.

A lot of the work contained in this tome would never have seen the light of day if I had not had such good collaborators. Earl Campbell was instrumental in the first two years of my PhD, sharing his great insights and discussing his new ideas. I am very proud of the paper we wrote together with Klearchos Loukopoulos, whose contribution to many discussions I am also grateful for. Joel Wallman has also been a great collaborator for discussing very many ideas in our time-zone defying email discourses.

I also thank Bob Coecke and Samson Abramsky for allowing me to join their group at Oxford. They have given me an independence that has allowed the continuation of some of the ideas resulting from the work contained herein.

I would also like to thank the Quantum Information group at UCL for producing a delightful environment in which to do research. In particular, I am indebted to Janet Anders for showing me around on my first day and always being around for good conversation, regardless of the topic. Also, my office mates at any one particular time including Sai-Yun Ye (for long discussions on short topics), Hussain Anwar, Brad Augstein, Tahir Sharaan, Hulya Yadsan-Appleby, and Peter Burns.

Outside of my office, along the corridor are the Public Engagement Unit. I would like to thank them for being just brilliant friends, especially Hilary Jackson

and Gemma Moore. They put up with my semi-coherent ramblings for far longer than they needed to. I will, and do already miss them.

Outside of my corridor, in big old London, I have to thank my friends for distracting me from everything. In particular James Millen (who became integrated into my office friends), Andy Sykes (who became integrated into my corridor friends), Philippa Stanger, Andy Webster, Keira Poland and countless others. They make it easy to miss London.

Outside of London, in the United Kingdom at large, I want to thank my family. My mum and dad, Christine and Chris Hoban, for everything they have done to allow me to get this far. I doubt anyone loves their parents as much I do and defy anyone to say otherwise. Along with my parents, I want to thank my brothers Kieran, Niall and Dominic for their support and love, and for testing the water before my arrival. Kieran especially needs to be thanked for the countless lunches, excellent conversation and emotional support.

Finally, and most importantly, I want to thank Francesca Richards for her kindness, intelligence, love and humour. Cesca (and Mwg) kept me sane throughout the whole of my PhD and I cannot begin to thank her enough.

Contents

To the memory of Margaret Ellen Hoban.

1 Introductions

"I tell you, we are here on Earth to fart around, and don't let anybody tell you different."

-Kurt Vonnegut

If this thesis has one central motivation it is this: to explore the interplay between the foundations and applications of quantum physics. The emergence of quantum information (the application of computer science ideas to quantum physics [Nielsen2000]) has motivated new insights into quantum mechanics. Indeed, new interpretations of quantum physics have been influenced by information theoretic concepts (e.g. [Caves2002]). In turn, ideas in quantum foundations have inspired new technological ideas and applications (e.g. [Ekert1991, Wootters1982]). The hope is that this work contributes to this fertile area of research by considering quantum mechanical correlations from a computational point-of-view.

In discussing the interplay between computation and correlations (in particular correlations of measurement statistics), we will discuss issues central to both computer science and quantum theory. Before we can address these issues we need to introduce basic concepts in quantum mechanics and quantum information. We will also mention how ideas in the foundations of quantum mechanics have inspired new applications of quantum theory, with a particular focus on the Bell inequality [Bell1964].

First, we introduce quantum mechanics and discuss the concept of entanglement [Schrödinger1936]. Einstein, Podolsky and Rosen used entanglement to argue that quantum mechanics is an incomplete theory [EPR1935]. This leads us to discuss Bell's argument that quantum mechanics is incompatible with "local realism" [Bell1964]. This incompatibility is epitomised by a violation of a Bell inequality [Bell1964, CHSH1969].

After the above discussion, we give a brief overview of quantum information

science. We indicate that entanglement has been shown to be a resource in quantum information [Nielsen2000]. The incompatibility of quantum mechanics with local realism is also a resource for certain tasks: device-independent quantum information protocols [Mayers98, Acín2007, Pironio2010]. We finish by discussing possible connections between Bell inequalities and Measurement-based Quantum Computing [Raussendorf2001]; the latter utilises entangled states to perform computational tasks. All of the work in this chapter is introductory and does not consist of new results produced by the author of this thesis.

1.1 Quantum Mechanics and Entanglement

In this section, we give a brief overview of the postulates of quantum mechanics. We also look at one of the consequences of these postulates: entanglement. There are very many clear and excellent pedagogical introductions to the quantum formalism (e.g. [Peres1993, Nielsen2000]). We base our introduction on that of Nielsen and Chuang [Nielsen2000]. The more relevant aspects of quantum theory will be emphasized, especially with regards to measurements.

Quantum mechanics is a mathematical framework for making predictions of outcomes of experiments. The problem of how this framework relates to a picture of physical reality is still open. An interesting research direction is to recover the quantum formalism from a set of axioms rooted in less mathematical, or more physical principles (e.g. [Hardy2001, Chiribella2011]). This subject will not be addressed in this thesis as it would be too much of a diversion from our discussion. Although, the issue of realism in a limited form will be encountered in section 2.1.

1.1.1 Postulates of Quantum Mechanics

In this subsection, we assume familiarity with linear algebra, complex vector spaces and Dirac notation ([Nielsen2000] is an excellent reference for these subjects). Physical systems described by quantum mechanics are associated with a complex inner product vector space, or Hilbert space $\mathcal{H}$. This idea can be formalised in the following postulate taken verbatim from [Nielsen2000].

Postulate 1 [Nielsen2000]. *Associated to any isolated physical system is a complex vector space with inner product (that is, a Hilbert space) known as the state space of the system. The system is completely described by its state vector, which*

is a unit vector in the system's state space.

A unit vector $|\psi\rangle$ in this Hilbert space $\mathcal{H}$ must satisfy $\langle\psi|\psi\rangle = 1$, where $\langle\psi|$ is the dual vector to $|\psi\rangle$ in the dual Hilbert space $\mathcal{H}^*$. For two-dimensional Hilbert spaces, all unit vectors are called "qubits" (quantum bits) and can be written as $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. By convention we choose the basis states in a d-dimensional Hilbert space $\mathcal{H}$ (where d is finite) to be $|j\rangle$ where $j \in \{0, 1, ..., (d-1)\}$[1].

More generally, quantum states can be associated with "density matrices" ρ, or an element of the space $L(\mathcal{H})$ of linear operators on $\mathcal{H}$. We may need to consider density matrices for physical systems that are not isolated or when an experimenter is not sure which state $|\psi\rangle$ a system is in; they assign probabilities to the possibilities. These density matrices represent statistical ensembles of the unit vectors described by Postulate 1. The unit vectors $|\psi\rangle$ are associated with "pure states" that are the density matrices $\rho = |\psi\rangle\langle\psi|$. If ρ is a statistical ensemble of pure states $|\psi_j\rangle$ then it can be represented as

$$\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j| \tag{1.1}$$

where j labels all possible pure states in an ensemble. The probabilities p_j are associated with each pure state $|\psi_j\rangle$ where $\sum_j p_j = 1$ and all $p_j \geq 0$.

For density matrices, the inner product is generalised to the operator trace $\mathrm{Tr}(...)$ such that $\mathrm{Tr}(\rho) = \sum_j p_j \mathrm{Tr}(|\psi_j\rangle\langle\psi_j|) = 1$, due to the cyclicity of trace. This is one of the conditions that a density matrix must satisfy along with the positivity condition $\rho \geq 0$. This second condition is satisfied for any arbitrary state $|\phi\rangle \in \mathcal{H}$ as $\langle\phi|\rho|\phi\rangle = \sum_j p_j |\langle\phi|\psi_j\rangle|^2 \geq 0$ due to $|\langle\phi|\psi_j\rangle|^2 = \langle\phi|\psi_j\rangle\langle\psi_j|\phi\rangle$.

The second postulate describes how quantum states can be transformed over time. Again this and all postulates are reproduced verbatim from [Nielsen2000].

Postulate 2 [Nielsen2000]. *The evolution of a closed system is described by a unitary transformation. That is, the state $|\psi\rangle$ of the system at time t_1 is related to the state $|\psi'\rangle$ of the system at time t_2 by a unitary operator U which depends only on the times t_1 and t_2,*

$$|\psi'\rangle = U|\psi\rangle. \tag{1.2}$$

[1]The set of integers $\{0, 1, ..., (d-1)\}$ can be described in terms of the cyclic group $\mathbb{Z}_d$.

We immediately see that a unitary operator preserves normalisation of a state as $\langle\psi'|\psi'\rangle = \langle\psi|U^\dagger U|\psi\rangle = \langle\psi|\psi\rangle = 1$ where $U^\dagger$ is the adjoint of U so that $U^\dagger U = \mathbb{I}$, the identity matrix. Unitary operators can also be applied to a density matrix as

$$U\rho U^\dagger = \sum_j p_j U|\psi_j\rangle\langle\psi_j|U^\dagger = \sum_j p_j|\psi'_j\rangle\langle\psi'_j|. \tag{1.3}$$

For open systems (i.e. systems that are not closed) we can generalise the unitary operator to a linear operator that must be completely positive and not increase the trace of ρ. The next postulate of quantum mechanics relates to measurements which are a form of completely positive and non-trace-increasing linear operator.

Postulate 3 [Nielsen2000]. *Quantum measurements are described by the collection $\{M_m\}$ of measurement operators. These are operators acting on the state space of the system being measured. The index m refers to the measurement outcomes that may occur in the experiment. If the state of the quantum system is $|\psi\rangle$ immediately before the measurement then the probability that result m occurs is given by*

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle, \tag{1.4}$$

and the state of the system after the measurement is

$$\frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{1.5}$$

The measurement operators satisfy the completeness equation,

$$\sum_m M_m^\dagger M_m = \mathbb{I}. \tag{1.6}$$

Again, the above postulate can be extended to density matrices ρ where $p(m)$ becomes $p(m) = \mathrm{Tr}(\rho M_m^\dagger M_m)$ and the state of the system after measurement is now

$$\frac{M_m \rho M_m^\dagger}{\mathrm{Tr}(\rho M_m^\dagger M_m)}. \tag{1.7}$$

Therefore measurement operators M_m act on density matrices in an analogous fashion to unitary operators. In calculating the probabilities of particular out-

comes and satisfying the completeness relation, $M_m^\dagger$ and M_m always appear together. For probabilities of measurement outcomes we rewrite $M_m^\dagger M_m$ as an operator E_m associated with a measurement outcome m. The operator E_m is called an "element" of a Positive Operator-Valued Measure (POVM) and is a positive operator[2] such that $\sum_m E_m = \mathbb{I}$ with probabilities $p(m) = \mathrm{Tr}(\rho E_m)$. The set of operators $\{E_m\}$ is then a POVM.

A special case of all possible measurements is the von Neumann projective measurement (PVM). This is the set $\{P_m\}$ where each element P_m is a projector associated with a measurement outcome m. These projectors satisfy an orthogonality constraint $P_m P_{m'} = \delta^m_{m'} P_m$ and for an arbitrary pure state $|\psi\rangle$, the state after a PVM is $|\nu_m\rangle \propto P_m|\psi\rangle$. Then, to satisfy the orthogonality constraint, we need N orthogonal vectors $|\nu_m\rangle$ to describe the projectors $P_m = |\nu_m\rangle\langle\nu_m|$, where N is the number of possible outcomes m of a measurement.

A PVM can be associated with an "observable" which matches each projector P_m of a PVM with a real eigenvalue λ_m. This observable $\hat{O}$ can be written as $\hat{O} = \sum_m \lambda_m P_m$ where λ_m is an observed outcome. The eigenvalue λ_m corresponds to a system being projected into the eigenstate $|\mu_m\rangle$ associated with P_m[3]. For example, for a two-dimensional Hilbert space, we can have observables with eigenvalues $\lambda_m = \pm 1$ associated with two-dimensional vectors $|\mu_m\rangle$ where m takes two possible values.

There is a beautiful result due to Naimark that shows that any POVM on a quantum state can be associated with a PVM [Paulsen2003]. That is, every POVM acting on a Hilbert space $\mathcal{H}$ can be implemented with a PVM on a larger Hilbert space $\mathcal{K}$. We can obtain some auxiliary (often referred to as an ancilla) system and take the composite of this system and our original Hilbert space $\mathcal{H}$ and perform a PVM on this new space. In order to consider composite systems we need to introduce the next postulate.

Postulate 4 [Nielsen2000]. *The state space of a composite physical system is the tensor product of the state spaces of the component physical systems. Moreover, if we have systems numbered* 1 *through* n*, and system number* i *is prepared in the state* $|\psi_i\rangle$*, then the joint state of the total system is* $|\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_n\rangle$*.*

[2]For all choices of states $|\phi\rangle$, $\langle\phi|E_m|\phi\rangle$ is a probability by definition, so E_m is a positive operator.

[3]These are eigenvalues and eigenstates as $\hat{O}|\mu_m\rangle = \lambda_m|\mu_m\rangle$.

We can replace pure states $|\psi_i\rangle$ in this postulate with density matrices ρ_i. Composite systems can be represented by density matrices as linear operators on a tensor product Hilbert space, i.e. $\rho \in L(\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_n)$ where $\mathcal{H}_i$ is the Hilbert space of each ith system. While in the postulate, we mention one pure state, $\bigotimes_{i=1}^n |\psi_i\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_n\rangle$, in particular, this is not the most general pure state in a composite Hilbert space $\bigotimes_{i=1}^n \mathcal{H}_i = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_n$. The state $\bigotimes_{i=1}^n |\psi_i\rangle$ is a "product state", but pure states that cannot be expressed in this form are said to be "entangled". This property will be discussed in the next subsection.

We have given a brief overview of the mathematical construction of quantum mechanics. In this thesis, we will be utilising the definition of a measurement and the description of composite systems. If a composite system consists of two space-like separated systems $\mathcal{H}_1$ and $\mathcal{H}_2$, then experimenters in each of these space-like separated systems can perform measurements on each of their respective subsystems. This way measurements can be written as a tensor product of these localised measurements, i.e. $\mathcal{M}_m^1 \otimes \mathcal{M}_m^2$ where $\mathcal{M}_m^i \in L(\mathcal{H}_i)$, a linear operator on $\mathcal{H}_i$. Assume that one can prepare all possible states (by whatever means) on the composite system, i.e. $\rho \in L(\mathcal{H}_1 \otimes \mathcal{H}_2)$. If the measurements are performed on *entangled* states then the statistics produced by this total system do not always factorise, i.e.

$$p(m, m') = \mathrm{Tr}(\rho \mathcal{M}_m^1 \otimes \mathcal{M}_{m'}^2) \neq \sum_j p_j \mathrm{Tr}(\rho_{1,j} \mathcal{M}_m^1) \mathrm{Tr}(\rho_{2,j} \mathcal{M}_{m'}^2) \tag{1.8}$$

as ρ is not necessarily equal to $\sum_j p_j \rho_{1,j} \otimes \rho_{2,j}$ where $\rho_{i,j}$ corresponds to a pure state of the ith system in the jth term of the decomposition of ρ.

This inability for the statistics of space-like separated measurements to be factorised will be central to the discussion of quantum correlations in this thesis. Entanglement is central to this subject. In the next subsection we will briefly discuss entanglement and how it can be quantified.

1.1.2 Entanglement

Schrödinger first introduced the term "entanglement" [Schrödinger1936]. This concept has become formalised for all possible density matrices ρ. First we describe systems in a bipartite scenario, that is where the Hilbert space of the system in question is the tensor product of two Hilbert spaces. An entangled

state represented by a density matrix ρ **cannot** be expressed as

$$\rho = \sum_i p_i |\psi_i^1\rangle\langle\psi_i^1| \otimes |\psi_i^2\rangle\langle\psi_i^2|. \tag{1.9}$$

The pure state $|\psi_i^j\rangle\langle\psi_i^j|$ is the jth party's state for the ith pure state in the probabilistic ensemble of ρ. There may be multiple, even infinite possible decompositions of ρ into a convex combination of pure states $|\psi_i^1\rangle\langle\psi_i^1| \otimes |\psi_i^2\rangle\langle\psi_i^2|$. For example, the density matrix $\rho = \frac{1}{4}\mathbb{I}$ in a composite Hilbert space of two, two-dimensional Hilbert spaces can be written as

$$\begin{aligned} \rho &= \frac{1}{4}\left(|00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10| + |11\rangle\langle 11|\right) \\ &= \frac{1}{4}\left(|++\rangle\langle ++| + |+-\rangle\langle +-| + |-+\rangle\langle -+| + |--\rangle\langle --|\right), \end{aligned} \tag{1.10}$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$[4]. This multiplicity of decomposition makes it difficult to ascertain whether an arbitrary density matrix is entangled or otherwise.

If a density matrix is a bipartite pure state, then there is a definite method to detect whether this state is entangled or not [Popescu1997, Plenio2007]. This method of detection also can quantify the *amount of entanglement*. For mixed states, this detection is a hard problem to compute [Gurvits2002].

The method of detecting entanglement for bipartite pure states involves finding the "Entropy of Entanglement" [Popescu1997]. To calculate this quantity, first one needs to find the reduced density matrix of ρ_1 and ρ_2 corresponding to party 1 and 2. The reduced density matrix is calculated from the partial trace of ρ, where we only take a trace over one party's system instead of the whole composite system. The partial trace of ρ over system 1 of two systems is written as $\mathrm{Tr}_1(\rho)$ and is calculated as

$$\mathrm{Tr}_1(\rho) = \sum_i \langle i_1|\rho|i_1\rangle, \tag{1.11}$$

where $|i_1\rangle$ are basis states on system 1. Without loss of generality, we assume that all subsystems have the same dimensional Hilbert space. We then calculate the von Neumann entropy $S(\rho_1)$ [vonNeumann1955] of this reduced density matrix[5]

[4] We make the standard abbreviation of omitting the tensor product for composite pure states, e.g. $|0\rangle \otimes |0\rangle$ becomes $|00\rangle$.

[5] The von Neumann entropy is the same for either sub-system [Nielsen2000].

ρ_1:

$$S(\rho_1) = -\mathrm{Tr}(\rho_1 \log_2(\rho_1)). \tag{1.12}$$

If $S(\rho_1) = 0$, then the reduced state ρ_1 is a pure state and so $\rho = \rho_1 \otimes \rho_2$ with both ρ_j being pure states. Importantly if $S(\rho_1) > 0$ then the pure state ρ is entangled. For $S(\rho_1) = 1$, then $\rho_1 = \frac{1}{2}\mathbb{I}$. The state ρ that results in $S(\rho_1) = 1$ is the "maximally entangled state" of two qubits, as it gives the maximum value of $S(\rho_1)$ for two qubits. The maximally entangled state $|\Psi\rangle$ of two d-dimensional systems can be written as

$$|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{(d-1)} |jj\rangle. \tag{1.13}$$

If we take the partial trace over system 1, then

$$\begin{aligned} \mathrm{Tr}_1(|\Psi\rangle\langle\Psi|) &= \sum_{i=0}^{(d-1)} \frac{1}{d} \left(\sum_{j=0}^{(d-1)} \langle i|jj\rangle \right) \left(\sum_{j=0}^{(d-1)} \langle jj|i\rangle \right) \\ &= \frac{1}{d} \sum_{j=0}^{(d-1)} |j\rangle\langle j| = \frac{1}{d}\mathbb{I}. \end{aligned} \tag{1.14}$$

In the case of two qubits we retrieve the value of entropy mentioned above, but in general, for these states $S(\rho_2) = S(\rho_1) = \log_2(d)$.

We have only discussed the bipartite case. In this thesis, we will also be interested in multipartite quantum systems. The definition of an entangled multipartite state is now where an entangled state cannot be written as (1.9) but with $|\psi_i^1\rangle\langle\psi_i^1| \otimes |\psi_i^2\rangle\langle\psi_i^2|$ now replaced with $\bigotimes_{j=1}^{n} |\psi_i^j\rangle\langle\psi_i^j|$. Entanglement of multipartite systems is relatively less well-studied but there do exist measures of entanglement in this scenario [Plenio2007]. There is also not one particular maximally entangled state for the multipartite setting like there is for the bipartite setting.

So far entanglement has been discussed as a mathematical construct and we have not discussed its physical consequences. In the next section we will discuss the impact of entanglement upon the foundations of quantum mechanics. That is, it causes a tension between quantum physics and a classical physics view of the world [Bell2004]. If quantum mechanics describes what is actually happening in the world then we need to accept some behaviour that is potentially incompatible with everyday intuition. We will make these issues more rigorous in the next

section.

1.2 EPR Paradox and Bell Inequalities

Albert Einstein played a crucial role in the development of quantum theory [Einstein1905]. However, upon being developed formally, he was famously dissatisfied with it. At its core, quantum mechanics predicts probabilities, and does not always make deterministic predictions[6]. It could be argued that this probabilistic feature convinced Einstein that quantum mechanics was a statistical theory akin to classical statistical, or Liouvillian mechanics [Liouville1838]. In Liouvillian mechanics objects have defined positions and momentum, but we may not have complete knowledge of these properties . Therefore, a probability distribution is assigned over a space of potential properties of a system. The state $|\psi\rangle$ could also resemble a probability distribution over some underlying reality describing a system. For more discussion of Einstein's potential view of quantum physics, see work by Harrigan and Spekkens [Harrigan2011].

A particular focus for Einstein's criticism of quantum mechanics became the issue of "locality". Locality has many different guises but we heuristically use it here in the sense that events in space-time can only "affect" each other if they are within each other's light-cone. It has been suggested by Bacciagaluppi and Valentini that Einstein had an argument against quantum theory based on a violation of locality at the 1927 Solvay Conference [Bacciagaluppi2009]. This discussion is beyond the scope of this thesis but we only mention it as a prelude to the argument presented by Einstein, Podolsky and Rosen (EPR) [EPR1935], often called the "EPR paradox"[7].

1.2.1 Realism and "Incompleteness" of Quantum Mechanics

In the original EPR paper, they argued that if one can predict a physical property, or quantity, with certainty then we associate that quantity with an "element of reality" [EPR1935]. If by the definition of EPR, a theory is "complete" then the properties that are found with certainty must be incorporated into the theory describing the system. Take two observables $\hat{O}_1$ and $\hat{O}_2$ that do not commute,

[6] Einstein's dissatisfaction can be summarised with one of his famous playful quotes: "... He[God] does not throw dice." [Einstein1971]

[7] The paradox being that if one accepts a particular picture of reality, then quantum mechanics contradicts this picture. It is not a paradox in the sense of demonstrating that quantum mechanics is inconsistent.

i.e. $[\hat{O}_1, \hat{O}_2] \neq 0$, and a state $|\psi\rangle$ being an eigenstate of $\hat{O}_1$ (with eigenvalue λ). If we make the constraint that the two observables do not share eigenstates nor are any of the eigenstates of one observable orthogonal to eigenstates of the other. We can predict the outcome λ of observable $\hat{O}_1$ with certainty, but cannot predict the outcome of $\hat{O}_2$ with certainty[8]. This means that we can only associate the observable $\hat{O}_1$ with an element of reality but not both observables. The following contradiction emerges if one asserts that elements of reality can only be associated with commuting observables. We follow Bohm's version of the EPR argument [Bohm1951].

Imagine that two parties share the entangled state (that is equivalent to the maximally entangled state[9]):

$$|\Psi\rangle_{\text{EPR}} = \frac{1}{\sqrt{2}} \left(|01\rangle - |10\rangle \right), \tag{1.15}$$

such that one party has access to one of the two-dimensional subsystems, or qubit, and the other party has access to the other qubit. We have put no constraint on the distance between the two parties, and in fact we make them space-like separated. The first party makes measurements of the observables $\hat{X} = (|+\rangle\langle+| - |-\rangle\langle-|)$ or $\hat{Z} = (|0\rangle\langle 0| - |1\rangle\langle 1|)$[10]. Each observable is associated with outcomes, or eigenvalues ± 1 and projectors P_m associated with this eigenvalue. According to EPR because the parties are space-like separated they can no longer "interact", and regardless of the observable performed by the first party, we must assign the same elements of reality to the second party [EPR1935].

If party 1 measures $\hat{X}$ and gets $+1$ or -1 then the second party's state will be $|-\rangle$ or $|+\rangle$ respectively with certainty (upto a global phase). Since we can predict the second party's state with certainty we must assign this property with an element of reality. If, on the other hand, party 1 measures $\hat{Z}$ then for outcomes $+1$ or -1 the second party's state will be $|1\rangle$ or $|0\rangle$ respectively (upto a global phase). Again, we can assign an element of reality since after the measurement, the first party knows the second party's state with certainty.

To summarise, if party 1 measures $\hat{X}$, then we can assign an element of reality with the second party's observable $\hat{X}$. When party 1 measures $\hat{Z}$, we assign an

[8] If we make a measurement of the observable $\hat{O}_1$ on $|\psi\rangle$, we obtain λ so that the projection $P_m = |\psi\rangle\langle\psi|$ has been performed on $|\psi\rangle$, giving the probability $p(m) = \langle\psi|\psi\rangle\langle\psi|\psi\rangle = 1$. However, $\hat{O}_2$ consists of projectors $P_m = |\phi\rangle\langle\phi|$ where $|\psi\rangle \neq |\phi\rangle$ so for $\hat{O}_2$ $p(m) = \langle\psi|\phi\rangle\langle\phi|\psi\rangle \neq 1$.

[9] One applies the unitary $\mathbb{I} \otimes U$ such that $U = |0\rangle\langle 1| - |1\rangle\langle 0|$ to both qubits.

[10] These are the Pauli-X and Pauli-Z measurements respectively.

element of reality with the second party's observable $\hat{Z}$. Since the measurements performed by party 1 are space-like separated from party 2, the elements of reality for party 2 should not be affected by the first party's measurements. This is the locality argument in the EPR paradox. However, $\hat{X}$ and $\hat{Z}$ do not commute, so we cannot assign an element of reality to each observable arriving at a contradiction. EPR reasoned that this contradiction means that quantum mechanics does not result in a complete picture of reality [EPR1935].

John Bell formalised the language of the EPR paradox away from the discussion of "incompleteness" and "elements of reality" into more mathematically rigorous concepts [Bell1964, Bell2004]. He showed that the assumption upon which the EPR paradox is based is that all physical systems obey "local realism" [Bell2004]. Local realism combines two separate assumptions invoking locality and realism and can be seen to limit the statistics of space-like separated measurements. In the following subsection we will briefly review local realism and show that it puts constraints on these statistics.

1.2.2 CHSH Inequality

We will describe local realism mathematically in section 2.1 of the next chapter but for now, we review the work of Clauser-Horne-Shimony-Holt (CHSH) [CHSH1969]. The seminal work of Bell [Bell1964] led to the formulation of the Bell inequality. This work was developed by CHSH into a mathematical expression that can be experimentally testable: the CHSH inequality.

We now describe the Bell-CHSH scenario, or "test" [CHSH1969]. There are two parties and each party chooses between two measurements. The choice of measurement is a completely random, free choice of the parties. This is a key assumption in the construction of Bell inequalities [Bell2004] (for consequences of dropping this assumption see [Barrett2011, Hall2011]). Each measurement has two possible outcomes ± 1. The measurements that the jth party chooses from are $\mathcal{M}_j^0$ and $\mathcal{M}_j^1$. These measurements can be described by an arbitrary theory and not just quantum theory. The statistics in this experiment that will be of interest to us are the correlations of the form

$$\mathbb{E}(\mathcal{M}_1^k \mathcal{M}_2^{k'}) = p(\mathcal{M}_1^k \mathcal{M}_2^{k'} = 1) - p(\mathcal{M}_1^k \mathcal{M}_2^{k'} = -1), \tag{1.16}$$

the expectation values of the joint outcome of both parties' measurements for choices k, $k' \in \{0, 1\}$ where $p(\mathcal{M}_1^k \mathcal{M}_2^{k'} = \pm 1)$ is the probability of getting the

joint measurement outcome ± 1.

There is actually a class of CHSH inequalities for this scenario [Fine1982], but we just pick out one particular expression

$$\mathbb{E}(\mathcal{M}_1^0\mathcal{M}_2^0) + \mathbb{E}(\mathcal{M}_1^0\mathcal{M}_2^1) + \mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^0) - \mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^1) \leq 2, \tag{1.17}$$

where the upper bound of 2 is satisfied for all physical systems that satisfy local realism [CHSH1969]. Local realism means outcomes of $\mathcal{M}_j^k$ are only dependent on some set of objective properties of each party's system. Secondly, these properties (which can be seen as elements of reality) are localised to each space-like separated region. Whilst they may have been shared properties when parties were not separated in the past, they are not affected by anything outside of their region. A locally realistic property for measurement M_j^k is then $\chi_j^k \in \{\pm 1\}$, so we can write the left-hand-side of (1.17) as

$$\mathbb{E}(\chi_1^0(\chi_2^0 + \chi_2^1) + \chi_1^1(\chi_2^0 - \chi_2^1)) = \sum_\chi p_\chi \left(\chi_1^0(\chi_2^0 + \chi_2^1) + \chi_1^1(\chi_2^0 - \chi_2^1)\right) \tag{1.18}$$

where p_χ is a probability distribution over all possible assignments of χ_j^k to measurements such that $\sum_\chi p_\chi = 1$ and $p_\chi \geq 0$. By convexity we can upper bound the right-hand-side of (1.18) by just considering the maximum value of $\chi_1^0(\chi_2^0 + \chi_2^1) + \chi_1^1(\chi_2^0 - \chi_2^1)$. If $(\chi_2^0 - \chi_2^1)$ is non-zero then $(\chi_2^0 + \chi_2^1)$ will be zero, resulting in

$$\sum_\chi p_\chi \left(\chi_1^0(\chi_2^0 + \chi_2^1) + \chi_1^1(\chi_2^0 - \chi_2^1)\right) \leq 2. \tag{1.19}$$

This expression then gives exactly the same right-hand-side of (1.17).

This result is interesting as we can derive a consequence of a theory with very few prior assumptions. More importantly though, in the following theorem, we can actually say something about quantum theory using the expression in (1.17).

Bell's Theorem [Bell1964]: *The predictions of quantum mechanics are not compatible with a locally realistic theory.*

Proof: To prove this theorem, we just need to show that the inequality (1.17) is not satisfied for all predicted values of $\mathbb{E}(\mathcal{M}_1^k\mathcal{M}_2^{k'})$ in quantum theory. We prove this by example. If two space-like separated parties share the state $|\Psi\rangle_{\text{EPR}}$ and make the measurements $\mathcal{M}_1^0 = \hat{X}$, $\mathcal{M}_1^1 = \hat{Z}$, $\mathcal{M}_2^0 = \frac{1}{\sqrt{2}}(-\hat{Z} - \hat{X})$, and

$\mathcal{M}_2^1 = \frac{1}{\sqrt{2}}(\hat{Z} - \hat{X})$, then the left-hand-side of (1.17) is

$$\begin{aligned} & \mathbb{E}(\mathcal{M}_1^0\mathcal{M}_2^0) + \mathbb{E}(\mathcal{M}_1^0\mathcal{M}_2^1) + \mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^0) - \mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^1) \\ = \; & \langle \left(\frac{\hat{X} \otimes (-\hat{Z} - \hat{X})}{\sqrt{2}} + \frac{\hat{X} \otimes (\hat{Z} - \hat{X})}{\sqrt{2}} + \frac{\hat{Z} \otimes (-\hat{Z} - \hat{X})}{\sqrt{2}} - \frac{\hat{Z} \otimes (\hat{Z} - \hat{X})}{\sqrt{2}} \right) \rangle. \end{aligned} \tag{1.20}$$

We have used the short-hand notation $\langle(...)\rangle = \langle\psi|(...)|\psi\rangle$ where $|\psi\rangle = |\Psi\rangle_{\text{EPR}}$. Calculation of the right-hand-side of (1.20) yields a value of $2\sqrt{2}$, which is greater than 2, thus violating the CHSH inequality[11]. Therefore quantum mechanics is incompatible with a theory satisfying local realism. □

The simplicity of the theorem and its proof has remarkable implications for the foundations of quantum mechanics. It means we must abandon the intuition of local realism, a constraint satisfied by classical physical systems. If the predictions of quantum theory are experimentally verified then if measurement outcomes result from elements of reality, then this reality does not satisfy locality. Or we could just abandon realism all together and not have to worry about locality.

1.2.3 Geometric Construction of Bell Inequalities

Beginning with the work of Froissart [Froissart1981], then developments by Fine [Fine1982], Pitowsky [Pitowsky1989] and Peres [Peres1999], the geometric picture of Bell inequalities has been well-developed. Correlations of space-like separated measurements are now elements of a vector in some real space. The space of correlations satisfying local realism is a convex polytope which can be described in terms of linear inequalities [Grünbaum2003]. These linear inequalities are examples of Bell inequalities. Finding these inequalities is then a problem in convex geometry.

This polytope approach to Bell inequalities is now an effective way of understanding the consequences of local realism. We will elaborate on and describe this approach in section 2.1.3 of the next chapter. Also we will comment on the hardness of finding the Bell inequalities that define the polytope of locally realistic correlations. The convex geometric approach has also been

[11] This value of $2\sqrt{2}$ is known as Tsirelson's bound [Tsirelson1980] as it is the largest possible quantum value of the left-hand-side of (1.20).

extended to the study of correlations that satisfy only a form of locality: space-like separated measurements that do not allow instantaneous communication [Barrett2005b, Pironio2011]. These issues will be discussed in section 2.3.

1.2.4 Experimental implementations for testing local realism

Testing whether quantum mechanics violates a Bell inequality in the laboratory is a difficult task. Firstly, measurements have to be space-like separated but transporting fragile quantum states over large distances can be hard. States may interact with the environment and become mixed states that are no longer entangled. Secondly, apparatus in the lab is not perfect and detectors may not always perfectly detect a measurement outcome. These difficulties can lead to "loopholes" (as we shall discuss in section 4.1 of chapter 4) whereby locally realistic theories are no longer constrained by the Bell inequality being tested [Pearle1970, Garg1987].

If we do not have space-like separated measurements then the local aspect of locally realistic theories is not constrained and we have the "locality loophole". For imperfect detection, the associated "detection loophole" is more subtle as it allows the possibility that the objective properties of a system can describe the statistics of detection [Pearle1970]. If we make the extra assumption that properties of the system we are observing are independent of the detection system, often called the "fair-sampling assumption" [Clauser1978], then violations of a Bell inequality have been observed in photonic systems [Aspect1981, Weihs1998]. Without this extra assumption, then ion-based systems have got around the detection loophole but suffer from the locality loophole [Rowe2001]. At the time of writing this thesis, completely loophole-free Bell inequality violations have not been observed. Although, there are promising avenues for future experimental work [Matsukevich2008, Vértesi2010]. In chapter 4, we will give a more thorough discussion of loopholes in Bell tests.

1.2.5 The GHZ Paradox

Bell's theorem can be proven using the now-famous Bell inequality. Did we need to construct this expression? There have been several arguments which have shown that quantum mechanics is incompatible with local realism but without use of a Bell inequality. For example, in 1983, Heywood and Redhead [Heywood1983] developed a proof that local realism cannot be compatible with

the statistics of two space-like separated, yet entangled spin-1 systems. This proof relied on an argument of determinism, in the spirit of the original EPR argument [EPR1935]. Later in 1989, Greenberger, Horne, and Zeilinger (GHZ) developed a proof of Bell's theorem without inequalities for three space-like separated parties [GHZ1989]. The GHZ argument has subsequently been developed by Mermin[12] [Mermin1990, Mermin1993]. Another notable example of Bell's theorem without the inequality is "Hardy's Paradox" which can be seen as a "possibilistic" proof, i.e. some things are possible in quantum mechanics that are not possible with locally realistic theories [Hardy1993]. We now present the GHZ argument, or "GHZ paradox" to which it is often referred, as a simple and beautiful proof of Bell's theorem.

We have three, space-like separated parties who (like in the CHSH construction) each have a completely free choice of measurement from a set of two measurements. We label the two measurements for the jth site $\mathcal{M}_j^0$ and $\mathcal{M}_j^1$ and each measurement takes one of two possible outcomes ± 1. As with the CHSH construction, each outcome is then a result of some objective property of each party's local system (which may have been shared in the past). Therefore, each measurement $\mathcal{M}_j^k$ again is assigned the value $\chi_j^k \in \{\pm 1\}$. Again we are interested in the correlations $\mathbb{E}(\mathcal{M}_1^k \mathcal{M}_2^l \mathcal{M}_3^m)$ where k, l, $m \in \{0, 1\}$. If we now obtain the following deterministic correlations for a particular set of measurements

$$\mathbb{E}(\mathcal{M}_1^0 \mathcal{M}_2^0 \mathcal{M}_3^0) = -1 \tag{1.21}$$

$$\mathbb{E}(\mathcal{M}_1^0 \mathcal{M}_2^1 \mathcal{M}_3^1) = -1 \tag{1.22}$$

$$\mathbb{E}(\mathcal{M}_1^1 \mathcal{M}_2^0 \mathcal{M}_3^1) = -1 \tag{1.23}$$

then we can assign values of χ_j^k deterministically to $\mathbb{E}(\mathcal{M}_1^k \mathcal{M}_2^l \mathcal{M}_3^m) = \chi_1^k \chi_2^l \chi_3^m$. If we multiply rows (1.21), (1.22), and (1.23) together after they have been assigned values of χ_j^k and observe that $(\chi_j^k)^2 = 1$, then in a locally realistic theory, we *must* obtain

$$\mathbb{E}(\mathcal{M}_1^1 \mathcal{M}_2^1 \mathcal{M}_3^0) = \chi_1^1 \chi_2^1 \chi_3^0 = -1. \tag{1.24}$$

However, measurements on an entangled quantum state can satisfy (1.21), (1.22), and (1.23) but contradict (1.24). The entangled state consists of three qubits

$$|\Psi\rangle_{\text{GHZ}} = \frac{1}{2}\left(-|00+\rangle + |01-\rangle + |10-\rangle + |11+\rangle\right), \tag{1.25}$$

[12]This argument was a development of a proof that quantum mechanics is "contextual" by Asher Peres developed into a proof of Bell's theorem.

where each jth site has one of these qubits and performs the measurements $\mathcal{M}_j^0 = \hat{X}$ or $\mathcal{M}_j^1 = \hat{Z}$. Calculating all expectation values, the statistics from these measurements on the state $|\Psi\rangle_{\text{GHZ}}$ satisfy correlations in (1.21), (1.22), and (1.23). However,

$$\mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^1\mathcal{M}_3^0) = \langle \hat{Z} \otimes \hat{Z} \otimes \hat{X} \rangle = +1, \tag{1.26}$$

thus contradicting (1.24). These quantum correlations have deterministically shown that local realism is inconsistent with quantum mechanics.

Mermin showed that we can still construct a Bell inequality from the correlations of the GHZ argument [Mermin1990]. We construct the following inequality

$$-\mathbb{E}(\mathcal{M}_1^0\mathcal{M}_2^0\mathcal{M}_3^0) - \mathbb{E}(\mathcal{M}_1^0\mathcal{M}_2^1\mathcal{M}_3^1) - \mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^0\mathcal{M}_3^1) + \mathbb{E}(\mathcal{M}_1^1\mathcal{M}_2^1\mathcal{M}_3^0) \leq 2, \tag{1.27}$$

which the correlations in (1.21), (1.22), (1.23) and (1.24) satisfy. Mermin showed that this inequality is satisfied for all locally realistic theories [Mermin1990], but the quantum mechanical correlations described above give a value of 4 for the left-hand-side of (1.27). We will show in section 1.4 that the GHZ-Mermin argument against local realism in quantum physics will be relevant to discussion about quantum information.

The CHSH inequality and the GHZ argument are ways of putting constraints on what is possible in a classical, or more generally, a locally realistic theory. The fact that quantum mechanics predicts contradictions to both constraints gives a remarkable departure from a classical view of the world. It indicates that when we are utilising the quantum mechanical formalism we can produce non-classical phenomena. One of the most enticing prospects for quantum mechanics is to use non-classical behaviour to perform some useful task that we could not achieve with classical resources. This motivation has led to the relatively nascent field of "quantum information science" [Nielsen2000]. One of the goals of this field is to process information via computation or communication and use quantum mechanical systems to do this "better" than with classical resources. In the next section we will give a broad overview of the field and how quantum systems could out-perform classical systems.

1.3 Quantum Information Processing

We have seen how quantum physics can be seen as non-classical in some concrete sense. Quantum information science has been developed to answer whether the non-classicality of quantum physics can be used to perform information processing tasks thought difficult or intractable with classical physical systems [Nielsen2000]. We now give a broad, and incomplete, overview of the field of quantum information in order to show that quantum resources can be useful for information processing.

The history of quantum information is itself an interesting topic for discussion. Stephen Wiesner developed the idea of "conjugate coding" circa 1970 but the result was not published until the 1980s [Wiesner1983]; this idea went on to influence the field of quantum cryptography. Alexander Holevo published his famous theorem in 1973 limiting the classical information in, say, a qubit to being at most one classical bit [Holevo1973]. Holevo's theorem is one of the most significant results in information theory applied to quantum systems and quantum "channels"[13]. The idea of a "quantum computer", or some quantum system capable of performing computations was first suggested by Richard Feynman in 1982 [Feynman1982]; the work of David Deutsch later formalised this concept [Deutsch1985]. Wootters and Zurek showed that unknown quantum information cannot be copied, called the "no-cloning" theorem [Wootters1982]. In the light of all of this work, we begin our discussion in the next subsection in 1984, with the seminal work by Bennett and Brassard (BB) on quantum cryptography [BB1984]. This work by BB brought together the ideas of the no-cloning theorem and conjugate cloning in a simple yet powerful way.

1.3.1 Quantum Cryptography

Two parties, referred to as Alice and Bob[14] want to communicate to each other without fear of eavesdroppers intercepting their messages. Alice encodes her message into another message or "ciphertext" with a "key" that Bob knows but no-one else does. Bob can use the "key" to unlock Alice's message from the

[13] Quantum channels consist of the positive linear operators on some "input" quantum state, mapping this state to another state. Perhaps this channel is a perfect communication channel for qubits and so would be the identity operator $\mathbb{I}$.

[14] These two characters have a long and auspicious career in computer science. Such is their success that the quantum information community talk often of Alice and Bob in quantum information procedures.

ciphertext. An eavesdropper can try and guess or calculate the key, but if it is random and Alice applies the "one-time pad", then a message can be made perfectly secure as defined by Shannon [Shannon1949]. The one-time pad consists of one bit of a message $x \in \{0,1\}$ being added (modulo 2) to a random bit $r \in \{0,1\}$ giving y, i.e. $y = x \oplus r$ where $\oplus$ represents modulo 2 addition. We need at least as many random bits as there are bits in the message, but as long as Bob knows every one of these random bits he can recover x by adding (modulo 2) r to y as $y \oplus r = x \oplus r \oplus r = x$. Shannon showed that this makes the ciphertext secure if an eavesdropper cannot obtain all values of r [Shannon1949].

How does Alice share the key consisting of the values of r to Bob? Since their goal was to communicate securely in the first place, they must find a secure way so that each party can communicate the random key. In 1984, BB showed that the combination of publicly communicating quantum states $|\psi\rangle$ from Alice to Bob and publicly communicating classical information about these states between Alice and Bob, secure values of r can be generated [BB1984]. An eavesdropper cannot perfectly copy the state that is publicly communicated by the no-cloning theorem, so must make a measurement to learn $|\psi\rangle$. The security is partly based on the fact that when an eavesdropper makes a measurement on the quantum state that is sent from Alice to Bob, they project the state into another state which may be different from $|\psi\rangle$. If the eavesdropper projects into a different state, Alice and Bob can compare measurement outcomes on the state to detect this. If Alice and Bob proceed with a particular protocol, with public quantum and classical communication, they can generate a secure random key. We then describe this as a method of "quantum key distribution" (QKD).

In 1991, Artur Ekert developed another method of QKD that utilised entanglement [Ekert1991]. This result alongside the discovery of "quantum teleportation" [Bennett1993] based upon sharing entanglement and classical communication led, in earnest, to entanglement being investigated as a resource for quantum information processing. Ekert based his protocol on a modified version of the CHSH Bell inequality test where Alice and Bob each receive one-half of the bipartite entangled state $|\Psi\rangle_{\mathrm{EPR}}$. The intuition behind the protocol is that a key is revealed by the act of space-like separated measurements on this entangled state; if a key existed before measurement it would be an "element of reality" and so incompatible with an entangled state.

For a given choice of measurements as discussed in the EPR paradox, outcomes are perfectly correlated generating a shared random bit. Alice and Bob randomly

choose measurements and announce the choice after receiving measurement outcomes. An eavesdropper can intercept the quantum state before it reaches either Alice or Bob and make a measurement, but this interception leaves the state in a separable state. They use the CHSH inequality to confirm that the state is entangled when they make measurements on it. Therefore, the protocol requires that the state is entangled and the CHSH inequality just confirms this, the security of the original 1991 protocol does not hinge directly on the incompatibility with local realism. Remarkably, in the spirit of Ekert's intuition, Barrett, Hardy and Kent designed a protocol whereby security was guaranteed by a Bell inequality violation [Barrett2005a]. Acín et al then made the connection to the original CHSH inequality that Ekert used (without assuming the quantum state shared), to confirm the security of a key [Acín2007].

1.3.2 Quantum Computing

If one does not use quantum cryptographic means to establish secure communication, then what means are there to establish a secure key? One of most commonly used tools is the Rivest-Shamir-Adleman (RSA) algorithm which is based upon a computational premise [RSA1978]. It is believed that it is hard for computers to find the prime factors of a large number. The RSA algorithm involves a public and a private key, where Alice makes the product of two large primes public and keeps these factors private. Bob receives the public key, encodes his message using it and sends his ciphertext to Alice in such a way that it can only be decrypted using Alice's private data. Therefore, if one can find the two prime factors of the public key efficiently, one can decode the message. However, as we mentioned, it is believed that this cannot be done efficiently with current computers and Alice receives the information from Bob securely. The RSA algorithm, as a result, is used quite successfully in many internet-based financial transactions.

Remarkably, if one could build a computer that works on quantum mechanics, a quantum computer[15], one could find the prime factors of a large number efficiently, thus breaking the RSA algorithm. The algorithm for finding these prime factors was invented by Peter Shor in 1994 [Shor1997] and became a key motivator for building a quantum computer.

In 1985, David Deutsch described a universal quantum computer which can

[15]Current desktop PCs rely on quantum theory to describe their workings. A quantum computer full exploits the quantum formalism and is based on the postulates of the theory.

perform any possible quantum computation [Deutsch1985]. A computation can be described in the "circuit model" of quantum computation where a quantum state consisting of n qubits is prepared in the product state $|\mathbf{0}\rangle = \bigotimes_{j=1}^{n} |0\rangle$ [Nielsen2000]. A computation then consists of a sequence of unitary operations performed on these qubits. Each unitary is considered 1 computational step, or "gate". After the requisite number of unitary operators is performed, some, or all of the n qubits can be measured.

Various algorithms have been designed for quantum computers indicating a potential improvement in computational time over classical computers [Deutsch1992, Shor1997, Grover1996, Harrow2009]. This improvement is conjectured in computational complexity terms as we currently do not even know the power of classical computers [Papadimitriou1994]. If quantum computers are more powerful than classical computers, then it would be of interest to know what aspect of quantum mechanics gives this improvement. It might even be the case that this property of quantum mechanics can assert the assumed separation between quantum and classical computers. Jozsa and Linden showed that in quantum computations on pure states, unbounded entanglement is necessary if there is to be a computational speed-up [Jozsa2003]. This does not mean that if there is entanglement in pure state quantum computation, the circuit cannot be simulated efficiently on a classical computer. A "Clifford circuit" is an example of a such a circuit that can be simulated efficiently with a classical computer [Nielsen2000, Aaronson2004]. It has also been shown by Vidal that if entanglement is bounded, then the quantum computation can be simulated efficiently classically [Vidal2003]. For quantum computations on mixed states, which will be those that are performed in the laboratory, the role of entanglement is unknown or possibly not even relevant [Jozsa2003, Datta2005].

1.3.3 Measurement-based Quantum Computing

There are several models of quantum computing that are equivalent to the circuit model of quantum computing[16] [Raussendorf2001, Zanardi1999, Aharonov2004, Kitaev2003, Leung2001]. In one particular class of models, the presence of entanglement is by construction a key ingredient in performing a computation. This is the class of models of Measurement-based Quantum Computing (MBQC) [Raussendorf2001, Raussendorf2003, Jozsa2003]. One of the origins of this model

[16] Equivalence means that every computation in one model can be efficiently simulated in another model.

can be seen in the teleportation-based quantum gate model developed by Gottesman and Chuang [Gottesman1999, Nielsen2003, Leung2001]. In teleportation-based quantum computing, n parties share bipartite maximally entangled states with their nearest neighbours, and make measurements at each site. Richard Jozsa has shown that this model is equivalent to a model proposed by Raussendorf and Briegel (RB) in 2001 [Jozsa2003, Raussendorf2001].

The model of MBQC proposed by RB consists of a multipartite entangled state, or "resource state" shared by n parties [Raussendorf2001]. This state is the "cluster state" consisting of n qubits [Raussendorf2001, Raussendorf2003]. However, in our discussion, we allow any possible state to be shared by these parties (see section 3.4) and do not restrict this aspect of MBQC. Also, from now on when we mention MBQC, we make it synonymous with the original RB construction but with any possible resource state [Anders2009]. The computation proceeds by each site performing a measurement with two outcomes ± 1 on their respective system. Measurements on, say, cluster states can have completely random outcomes, but a set of gates in the quantum circuit model corresponds to a set of unitary operators [Raussendorf2001]. The unitary evolution of a state is a deterministic operation. Remarkably, one can achieve determinism in MBQC by applying corrections at the end of the measurements and making measurements adaptive [Jozsa2003]. That is, the choice of measurement during the computation must be dependent on previous measurement outcomes.

To take into account this correction and adaptivity, a crucial component of MBQC is needed: the "classical control computer" [Raussendorf2003, Briegel2009, Anders2009]. This computer is a classical processor and processes bits corresponding to a choice of measurement and its respective outcome at each site. In the model of RB, one just needs a choice between two measurements at each site to get a universal quantum computer labelled by bit-values. The outcomes of a measurement are $\pm 1 = (-1)^x$ as described above where x is now a bit-value. In MBQC as formulated by RB, the control computer does not require all possible operations, or gates in classical computing. In fact, as pointed out by Anders and Browne, all that is needed is modulo 2 addition between classical data [Anders2009]. Using only these operations, a computer cannot perform all logical, or Boolean operations, and is therefore not *functionally complete* for all classical computations.

1.3.4 Entanglement as a Resource

In MBQC, entanglement can be seen as a resource that is "consumed" via single-qubit measurements [Briegel2009]. This model also provides a nice distinction between the quantum and classical parts of computation; the quantum part being the measurements on a quantum state and the classical control computer providing some, albeit limited processing to utilise this measurement data.

The idea of entanglement being a resource for information processing that is consumed can be seen in many aspects of quantum information [Horodecki2009]. Historically, beginning with entanglement as a resource for producing secure keys, then used as a channel for communicating quantum states via teleportation [Ekert1991, Bennett1993]. The interplay between quantum gates and teleportation as highlighted by Gottesman and Chuang, also highlights the role of entanglement with respect to computation [Gottesman1999, Jozsa2003]. Also relevant to quantum computation and communication, entanglement has been utilised as a resource for correcting errors [Brun2006].

Inspired by these information processing tasks, the resource theory of entanglement has been developed [Horodecki2009]. If parties are restricted to being only able to perform local operations on their respective subsystem and communicating classical information (LOCC), then they cannot produce an entangled quantum state [Horodecki2009]. Therefore, if parties have an entangled state, they can do tasks that they otherwise could not do with only LOCC. This theory has become well-developed and we refer the reader to [Horodecki2009] for a review of entanglement in quantum information.

1.4 Bell Inequalities and Quantum Information

Since entanglement is a resource for information processing, and entanglement was used to show an incompatibility of quantum physics with local realism, can this incompatibility also be used as a resource? In recent years, the answer to this question has been answered in the affirmative. The intuition behind Ekert's 1991 QKD protocol that if a key is some element of reality held by each party then an eavesdropper can threaten security and learn this data [Ekert1991]. As mentioned, Barrett, Hardy and Kent developed this intuition [Barrett2005a] and then Acín et al made the connection between security and Bell's theorem concrete [Acín2007]. They showed that if we put no constraint on the devices that Alice and Bob use (these devices can even be produced by the eavesdropper),

then the security of a key can be established *directly* by the violation of a Bell inequality. This is an example of "device-independent" quantum information processing [Mayers98, Acín2007, Pironio2010], and we now review this nascent field very briefly.

1.4.1 Device-independent Quantum Information

A violation of a Bell inequality indicates that if we assume our system is quantum mechanical, then the shared quantum state was entangled. Therefore, it is natural to say that a violation must detect entanglement without making any assumption on the system. Indeed this idea has been developed both in the bipartite and multipartite setting where a Bell inequality is used as a "witness" of entanglement [Liang2011, Rabelo2011]. In calculating entanglement of a state directly, one calculates this quantity directly from the state. However, if we do not know the state and we observe a violation, then it must be entangled[17].

There are two aspects of bipartite entanglement that make it useful for QKD. The fact that random, yet completely correlated outcomes can be generated for the shared key, and if measured, the system will no longer be entangled. The second fact ensures that the randomly generated key is securely generated. But randomness is in of itself a useful resource for many tasks [Knuth1981], including secure key distribution and cryptography in general [Shannon1949]. For example, in a Monte Carlo simulation of complicated systems, a random source is required to pick a data point at random on which to calculate something [Metropolis1949]. Also, randomly sampling from a probability distribution to perform statistical analysis is useful for ruling out statistical bias in this analysis.

Genuinely random processes are difficult to come by as classical physical systems are seemingly random due to lack of knowledge about all parameters of the systems. The underlying parameters of the system have deterministic properties but our inability to access all of them leads to the assignment of probabilities. This form of randomness can be seen as not true randomness due to the underlying determinism, but "pseudorandomness" [Knuth1981]. However, if we assume that locality must be respected then the random outcomes of observables on either side of a bipartite, space-like separated maximally entangled state cannot be due to some underlying real parameters. The randomness of the maximally entangled state is a good source of randomness.

[17] As well as this device-independent approach to entanglement, Bell inequalities can be used to gain information about the dimension of a quantum sytem [Gallego2010].

If we do not assume that we have a maximally entangled state shared between two parties, Pironio et al showed that true randomness can be generated from the violation of a Bell inequality [Pironio2010]. This randomness from a violation can then be used as a "seed" to generate something more random. Therefore, randomness can be generated without assuming anything about the underlying system that can possibly generate it, and so is device-independent. The generation of random numbers [Pironio2010, Colbeck2007] and cryptography [Mayers98, Acín2007, Pironio2009, Silman2011] are two main current implementations of device-independent protocols. The motivation behind both of these tasks comes from cryptography, but in the next section we give an example of computing based on a violation of a Bell inequality in the form of the GHZ paradox.

1.4.2 GHZ Paradox and Measurement-based Quantum Computing

Models of computing have been related to Bell inequalities. Communication complexity is a model where we have several parties and each party has unbounded computational power [Kushilevitz1996]. Each party has some data and the goal is to compute some function on all of this data. The question is whether all of this data needs to be sent between parties in order for the function to be computed? Communication complexity studies the minimum amount of communication needed to calculate a particular function. If a system violates a particular Bell inequality, then it can exhibit an advantage in a communication complexity task over a system that does not violate a Bell inequality [Brukner2004, Buhrman2010]. Another example of a computational model related to Bell inequalities is a "non-local game" [Cleve2004]. We will discuss these models in sub-section 3.3.2 of chapter 3 and so postpone discussion of the model until then.

In MBQC, the classical control computer can only perform addition modulo 2, or "XOR gates" as they are called in the Boolean circuit model of classical computing [Anders2009]. In order to have a full power classical computer, we require another gate: the "NAND gate" [Papadimitriou1994]. The XOR gate on two bits x_1 and x_2 is the function $f(x_1, x_2) = x_1 \oplus x_2$ but the NAND gate is $f(x_1, x_2) = 1 \oplus x_1 x_2$ where this function is 0 for $x_1 = x_2 = 1$ and 1 otherwise. Anders and Browne (AB) showed that in MBQC a NAND gate can be performed with three measurement sites and a single round of measurements [Anders2009]. This three-party system is also the minimal resource in MBQC that can produce

this function. AB used the GHZ paradox to demonstrate this result and we now review this result [Anders2009].

If we inspect the correlations in (1.21), (1.22), (1.23) and (1.26), the choice of measurements can be labeled by bit-values $s_j \in \{0, 1\}$ at each jth site. We relabel the values s_1 and s_2 to be some bit-values x_1 and x_2 respectively. Therefore, for the specific correlations in (1.21), (1.22), (1.23) and (1.26), the choice of third measurement s_3 is equal to $x_1 \oplus x_2$. This can be modeled as a computation in MBQC where the classical control computer sets the choice of measurement on the first two sites to be x_1 and x_2, and $x_1 \oplus x_2$ for the third site; the classical control computer calculates this third choice. Then if the measurements and quantum state are those in the GHZ paradox then we observe the correlations are

$$\mathbb{E}(\mathcal{M}_1^{x_1}\mathcal{M}_2^{x_2}\mathcal{M}_3^{x_1\oplus x_2}) = (-1)^{x_1x_2\oplus 1}. \tag{1.28}$$

The function corresponding to the NAND gate then appears on the right-hand-side. We obtain the measurement outcome from the jth site as $(-1)^{m_j}$ where $m_j \in \{0, 1\}$ and then the joint outcome of all three parties is $(-1)^{m_1\oplus m_2\oplus m_3}$. Then every instance of $\mathcal{M}_1^{x_1}\mathcal{M}_2^{x_2}\mathcal{M}_3^{x_1\oplus x_2}$ must deterministically produce an outcome $(-1)^{m_1\oplus m_2\oplus m_3}$ equal to $(-1)^{x_1x_2\oplus 1}$. If each site sends the value m_j to the classical control computer then it can calculate $m_1 \oplus m_2 \oplus m_3$ thus obtaining $x_1x_2 \oplus 1$ deterministically.

We have shown that classical correlations cannot reproduce the above quantum correlations. In order that we produce a NAND gate with classical correlations communication in the form of adaptivity is required in the measurement-based circuit. This does not minimise the resources required for a full classical computer and shows that correlations that are incompatible with local realism are useful in MBQC. These ideas will be developed further in section 3.4 of chapter 3.

1.5 Chapter Summary

We have introduced the quantum formalism and shown that it has an interesting mathematical consequence: entanglement. Not only is entanglement a mathematical curiosity, it has consequences for our understanding of quantum theory. In particular, it challenges the notion of local realism that is satisfied in classical physical systems [Bell2004]. The advent of quantum information placed entanglement in yet another context, that as an information theoretic resource [Horodecki2009]. It then has become an interesting research avenue to link the in-

compatibility with local realism with a potential information theoretic advantage. This has led to the development of device-independent quantum information.

Finally we have an indication of the possible applications of Bell inequalities to some computational models. When implementing a test of a Bell inequality, or Bell test, measurements must be space-like separated and classical communication is ruled out. Computation often involves time ordering of operations, or gates, and this time-ordering allows the possibility of communication between parties. However, when we process statistics from Bell tests, we are performing a computation on this data, and in the example of the GHZ paradox we can use this processing to obtain something "useful" from quantum correlations [Anders2009]. Correlations then are computations in this example after this processing. Throughout this thesis, this picture will become central to our understanding of correlations. That is, correlations can be used to compute particular functions on an "input" corresponding to the choice of measurement settings at all sites.

This computational insight on Bell tests will be used to give a new perspective on established ideas in Bell tests as well as new ideas for Bell tests. We will review the issue of loopholes in Bell tests [Pearle1970] and indicate that they have a computational interpretation that makes this subject more amenable pedagogically. Motivated by these issues, we describe a way of expanding Bell tests to include processing of statistical data but without introducing loopholes. We give the notion of a loophole a more technical grounding and present these results in chapter 4.

We have hinted at a connection between Bell tests and MBQC. We extend this connection and make it more concrete in chapter 3 by discussing MBQC without adaptivity. Using Bell tests we can actually say something about the power of MBQC without adaptivity as well as showing that quantum physics can do something that classical physics cannot. In chapter 4 we will discuss whether MBQC with adaptivity can be framed in terms of a Bell test. We give some indication that this is possible using a method of loophole-free data processing. Before we discuss applications of the Bell test to computation and vice versa, in the next chapter we introduce our framework for Bell tests.

2 Correlators and Bell Tests

In this chapter, we will lay the foundations for our study of Bell tests [Bell1964]. More precisely we will motivate the study of what we call correlators: the statistics of the joint outcome of many parties. We have already discussed (in section 1.2) correlators in terms of the expectation value of measurements made in the Bell-CHSH test [CHSH1969]. We now describe correlators in terms of conditional probabilities of a joint outcome given some measurement settings. We will make the connection to the Bell-CHSH test concrete and show that considering these conditional probabilities allows for greater scope when considering a broad class of Bell tests.

One prominent tool utilised in this chapter is to describe the correlators, or conditional probabilities as *stochastic maps* from a set of inputs (describing the measurement settings) to a set of outputs (describing corresponding measurement outcomes). These maps are then probabilistic maps from an input to an output. We will define particular classes of functions and show how they relate to correlators resulting from particular physical theories, more specifically locally realistic and quantum theories.

The famous Bell inequality emerges from a discussion on the geometry of stochastic maps. Correlators can be represented as vectors in a real vector space; every vector is a list of conditional probabilities for each joint outcome for every choice of measurement settings. In this real space, the space of LHV correlators can be defined as a *convex polytope*, an object which is the convex hull of a finite number of correlators (called extreme points) [Grünbaum2003]. The boundary, or surface of a convex polytope is made up of objects called faces. If the dimension of the space a polytope lives in is Δ, then a $(\Delta-1)$-dimensional face is called a *facet* and facets are defined by particular linear inequalities. These inequalities define half-spaces in the Δ-dimensional real space, and the intersection of these half-spaces also define a convex polytope [Grünbaum2003]. For the convex polytope of locally realistic correlators, the linear inequalities that define its facets are the facet Bell inequalities [Pitowsky1989, Froissart1981, Peres1999, Fine1982].

We combine this geometric picture of correlators with the discussion of stochastic maps or functions and show that we can describe Bell tests in terms of computations. This computational aspect allows us to capture locally realistic correlators in terms of computational expressiveness. Not only is this method used to describe correlators, it is used to say something about the full probability of distribution for all possible measurement outcomes and settings. In particular, correlators single out particular probability distributions that only satisfy special relativity (non-signalling) and no other physical constraints [Popescu1994]. Finally, we also characterise correlators, and correlations in general that appear in a model constructed originally by George Svetlichny [Svetlichny1987]. This model allows a sub-set of parties to share unconstrained correlations but satisfy local realism with respect to others.

This chapter in the main motivates the study of correlators as a simplification from studying the full statistics of a Bell test. Despite the simplification, the study of correlators yields significant insights into the study of the full probability distribution. We also establish the framework upon which results in later chapters are built. Section 2.1 consists of review material, and the work in section 2.2 introduces a new computational framework for correlators. Section 2.3 consists of new results describing non-signalling correlations and section 2.4 recasts Svetlichny correlations in terms of a computational description. The original work in sections 2.2, 2.3 and 2.4 was completed in collaboration with Joel Wallman and Dan Browne and published as [Hoban2011c].

2.1 A General Framework for Bell tests

Bell tests are carried out by space-like separated parties that each make a choice from a set of measurements and each measurement produces an outcome from a set of possible outcomes [Bell1964]. From this starting point, it has been insightful to think of Bell tests, and other physical processes from an operational point-of-view [Hardy2001, Hardy2011, Barrett2007]. In an operational framework, each measurement site is an abstract object, often referred to as a "box", that takes an "input" as the choice of measurement setting and returns an "output" in the form of a measurement outcome. Operationally then, we only concern ourselves with the statistics resulting from these boxes and not necessarily their "inner-workings". We only want to infer the properties of these boxes from their statistics making minimal assumptions.

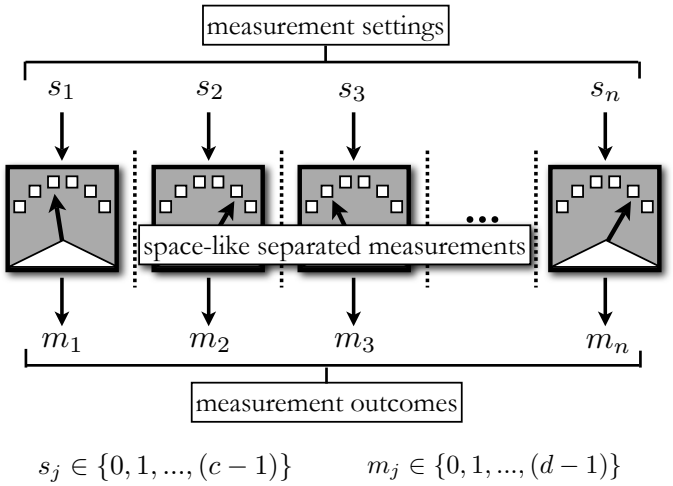

Figure 2.1: In a Bell test, n parties each make a measurement from c possible choices, where each measurement has d possible outcomes. Labelling the jth party's measurement choice and outcome by s_j and m_j respectively, we can describe each run of the experiment with n-digit strings $\mathbf{m}$ and $\mathbf{s}$. (Copyright: American Physical Society, 2011).

We then consider n space-like separated parties, or boxes. Each jth site for $j \in \{1, 2, ..., n\}$ makes a measurement $\mathcal{M}_{s_j}$ from a choice of c_j measurements where $s_j \in \{0, 1, ..., (c_j - 1)\}$ labels the choice of measurement and is expressed in terms of an integer, or digit in $\mathbb{Z}_{c_j}$, the cyclic group of c_j elements. Each measurement $\mathcal{M}_{s_j}$ has $d(s_j)$ possible outcomes $\mathcal{O}_{m(s_j)}$, where $m(s_j) \in \{0, 1, ..., (d(s_j)-1)\}$ is an element of $\mathbb{Z}_{d(s_j)}$, the cyclic group of $d(s_j)$ elements. Therefore, in operational terms, each jth box takes an input s_j and returns an output $m(s_j)$ for each input. From now on, we assume that $d(s_j) = d_j$ is constant for all measurements labelled by s_j. We include a schematic of the Bell test in Figure 2.1.

Inputs into all n boxes are represented by n-length digit-strings $\mathbf{s} \in \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j}$, the Cartesian product of all sites' inputs $\mathbf{s} = \{s_1, s_2, ..., s_n\}$ [1]. All digit-strings will be expressed in bold typeface, with jth elements not in bold, but with sub-script j. Outputs are then expressed as n-length digit-strings $\mathbf{m} = \{m_1, m_2, ..., m_n\}$ where we have not explicitly written the dependence on $\mathbf{s}$ for brevity, but this

[1] We are using this perhaps unconventional notation for the Cartesian product for the sake of brevity. Typically the Cartesian product between sets A and B is represented as $A \times B$ and $A_1 \times ... \times A_n$ for an n-fold Cartesian product of sets A_j. In this non-standard notation $A_1 \times ... \times A_n = \bigoplus_{j=1}^{n} A_j$.

dependence is implied. That is, every output is always a particular output $\mathbf{m}$ for a given input $\mathbf{s}$. Given non-empty sub-sets $\mathcal{J} \subseteq \{1, 2, ..., n\}$ of all n with $|\mathcal{J}|$ being the number of parties in the sub-set, the outputs of this sub-set is written as $\mathbf{m}^{\{j|j\in\mathcal{J}\}} = \{m_j | j \in \mathcal{J}\}$ such that singleton sub-sets are the elements m_j. The same notation is applied also to the inputs $\mathbf{s}$ with $\mathbf{s}^{\{j|j\in\mathcal{J}\}} = \{s_j | j \in \mathcal{J}\}$ being the inputs on a sub-set $\mathcal{J}$. When $\mathcal{J}$ includes all n parties then we recover $\mathbf{m}$ as the output again. There are then $(2^n - 1)$ of these non-empty sub-sets $\mathcal{J}$.

Central to the standard construction of Bell tests is that choice of measurement setting is independent of anything else in the experiment [Bell1977]. In other words, the choice of measurement is completely random, i.e. $p(\mathbf{s}) = \prod_{j=1}^{n} \frac{1}{c_j}$. The consequences of relaxing the constraint of measurement independence have been shown to be detrimental to Bell tests [Barrett2011, Hall2011].

As mentioned, in Bell tests, statistics are calculated from the data obtained from the boxes. The statistics are the conditional probabilities $p(\mathbf{m}^{\{j|j\in\mathcal{J}\}}|\mathbf{s})$, the probability of obtaining outputs $\mathbf{m}^{\{j|j\in\mathcal{J}\}}$ given the input $\mathbf{s}$ for all sub-sets $\mathcal{J}$ of n parties. Crucially though we can obtain every probability $p(\mathbf{m}^{\{j|j\in\mathcal{J}\}}|\mathbf{s})$ for a proper sub-set $\mathcal{J}$ from the full distribution $p(\mathbf{m}|\mathbf{s})$ by taking a sum of outcomes on the complement sub-set $\mathcal{J}^c = \{j|\{1, 2, ..., n\} \setminus \mathcal{J}\}$ to $\mathcal{J}$ of all n parties, i.e. $p(\mathbf{m}^{\{j|j\in\mathcal{J}\}}|\mathbf{s}) = \sum_{j\in\mathcal{J}^c} p(\mathbf{m}|\mathbf{s})$. Therefore from now on we only need to consider the full probability distribution $p(\mathbf{m}|\mathbf{s})$. In the following sub-section, we discuss the basic geometric objects that will dominate our discussion of correlations: the convex polytope.

2.1.1 Convex Polytopes and Stochastic Maps

The conditional probabilities $p(\mathbf{m}|\mathbf{s})$ are stochastic maps producing the map, or function $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \bigoplus_{j=1}^{n} \mathbb{Z}_{d_j}$ with some probability. Throughout this thesis, we will use a geometric picture to consider these (and other forms of) stochastic maps. These conditional probabilities $p(\mathbf{m}|\mathbf{s})$ are elements of a vector $\vec{p}$ in a real vector space. We can reduce the number of probabilities we need to consider by the normalisation condition that $\sum_{\mathbf{m}} p(\mathbf{m}|\mathbf{s}) = 1$ where $p(\mathbf{0}|\mathbf{s}) = 1 - \sum_{\mathbf{m}\neq\mathbf{0}} p(\mathbf{m}|\mathbf{s})$. Vectors $\vec{p}$ have length $D = (\prod_{j=1}^{n} d_j - 1) \prod_{j=1}^{n} c_j$ in $\mathbb{R}^D$ real space [2]. Since the elements of $\vec{p}$ are probabilities, they will live in a bounded sub-space in $\mathbb{R}^D$ satisfying the constraints that all $p(\mathbf{m}|\mathbf{s}) \geq 0$ and $\sum_{\mathbf{m}\neq\mathbf{0}} p(\mathbf{m}|\mathbf{s}) \leq 1$; the positivity and normalisation constraints respectively.

[2] This can be seen from the fact that we have $\prod_{j=1}^{n} c_j$ normalisation conditions (one for each input string) and $\prod_{j=1}^{n} d_j \prod_{j=1}^{n} c_j$ original probabilities $p(\mathbf{m}|\mathbf{s})$.

These inequalities essentially describe a *convex polytope* in $\mathbb{R}^D$ which we call $\mathcal{F}$. Convex polytopes will be a central part of this thesis for all manner of different real spaces and so we shall define them for all possible real spaces now.

A convex polytope $\mathcal{C}$ in a real space $\mathbb{R}^\Delta$ of dimension Δ can be defined in two ways: first is the *half-space representation* and the second is the *vertex representation* [Grünbaum2003]. We will now formally define $\mathcal{C}$ in terms of each representation:

Definition 1. *(Half-space representation)*: A convex polytope $\mathcal{C}$ in a real space $\mathbb{R}^\Delta$ of dimension Δ is the intersection of closed half-spaces. These closed half-spaces are defined by linear inequalities of the form $\sum_j^\Delta a_j v_j \leq b$ for real values a_j and elements v_j of a vector $\vec{v} \in \mathbb{R}^\Delta$.

This definition is general enough to encompass *unbounded* polytopes. We say a convex polytope is bounded if it can be contained in a ball of finite radius and unbounded otherwise. We impose the extra constraints that there are a finite number of inequalities that form a bounded polytope [Grünbaum2003].

The linear inequalities in the above definition are "facet-defining" which we define formally later on but can be informally seen as the boundary of the convex polytope $\mathcal{C}$. If we return to the example of $\mathcal{F}$ as the space of all possible conditional probabilities $p(\mathbf{m}|\mathbf{s})$ then the linear inequalities defining $\mathcal{F}$ are the positivity and normalisation constraints. As mentioned, dual to the half-space representation, the vertex representation of a convex polytope describes the polytope in terms of all points, or vectors in the polytope:

Definition 2. *(Vertex representation)*: **convex polytope** $\mathcal{C}$ in a real space $\mathbb{R}^\Delta$ of dimension Δ is the convex hull of E extreme points, or vectors $\vec{v}_e \in \mathbb{R}^\Delta$ for $e \in \{1, 2, ..., E\}$.

The convex polytope $\mathcal{C}$ then is the set of vectors that can be written as a convex combination of E vectors in the Δ-dimensional real space. For example, the polytope $\mathcal{F}$ can then be written in terms of the convex combination of E vectors $\vec{p}$ which we call $\vec{p}_e$ for $e \in \{1, 2, ..., E\}$ with probability distribution $p(E)$ over each $\vec{p}_e$. These vectors have the elements $\vec{p}_e$ that are the *deterministic* probabilities $p(\mathbf{m}|\mathbf{s}) \in \{0, 1\}$. For these deterministic probabilities we associate values $\{0, 1\}$ with each map $f : \mathbf{s} \to \mathbf{m}$. This way the probabilities can be written as $p(\mathbf{m}|\mathbf{s}) = \sum_E p(E)\vec{p}_E = \sum_f p_f \delta_{\mathbf{m}}^{f(\mathbf{s})}$ where $f(\mathbf{s})$ is the image of $\mathbf{s}$ under f and p_f is a probability distribution over all maps f.

2.1.2 The Non-signalling Polytope

One can make extra assumptions upon the statistical data obtained from space-like separated sites: each measurement site cannot communicate with each other outside each other's light-cone. This assumption is expressed in terms of the *no-signalling condition* which can be formally stated as:

$$\sum_{\mathbf{m}^{\{j|j\in\mathcal{J}\}}} p(\mathbf{m}|\mathbf{s}) = \sum_{\mathbf{m}^{\{j|j\in\mathcal{J}\}}} p(\mathbf{m}|\mathbf{s}') = p(\mathbf{m}^{\{j|j\in\mathcal{J}^c\}}|\mathbf{s}^{\{j|j\in\mathcal{J}^c\}}), \tag{2.1}$$

where $\mathcal{J}$ is any sub-set of all n parties and $\mathcal{J}^c = \{j \in \{1, 2, ..., n\} | j \notin \mathcal{J}\}$ is the complement of this sub-set and $\mathbf{s} \neq \mathbf{s}'$ such that the inputs differ in elements s_j with $j \in \mathcal{J}$ [Popescu1994]. Each of these conditions forms a hyperplane in $\mathbb{R}^D$ and then the intersection of hyperplanes is the space of correlations that satisfies the no-signalling condition. Or just as before, one can reduce the dimensionality of the space of statistics by imposing these equalities and then define inequalities on the reduced space.

Therefore one can construct another convex polytope called $\mathcal{NS}$ which is the intersection of half-spaces defined by inequalities resulting from the normalisation, positivity and no-signalling conditions [Barrett2005b]. We shall discuss the polytope $\mathcal{NS}$ in section 2.3 of this chapter. Now we consider the correlations that satisfy local realism, or local hidden variable theories.

2.1.3 Local Hidden Variable Theories and Bell Inequalities

A Bell test is an experiment that aims to test whether the statistics produced by boxes can be satisfied by a theory that obeys *local realism*. Systems that satisfy local realism satisfy two conditions (covered thoroughly in [Bell2004]):

1. Realism: There are objective properties of a system that are elements, or "hidden" variables $\lambda \in \Lambda$ in a (generally continuously defined) space of hidden variables Λ. These variables have a pre-existing value before the measurement is made and can influence measurement outcomes;

2. Locality: The variables λ possessed by a party at any site are not affected by events that occur outside of the light-cone of the measurement made at this site. These variables are called *Local Hidden Variables* (LHV).

Each party's measurement is influenced by λ and the measurement choice made at that party's site. In an LHV theory, space-like separated parties cannot com-

municate their measurement information to each other via the LHV, or any other means, due to locality. Measurement outcomes are then influenced by s_j and λ alone.

To be more precise, there is a probability distribution $p(\lambda)d\lambda$ over Λ such that $p(\lambda) \geq 0$ and $\int_\Lambda p(\lambda)d\lambda = 1$. This can occur, for example, if the parties have some shared source of randomness over the variables λ. Therefore each set of measurement outcomes conditioned upon measurement settings can be written in the following form [Bell1964],

$$p(\mathbf{m}|\mathbf{s}) = \int_\Lambda p(\lambda)d\lambda \prod_{j=1}^{n} p(m_j|s_j, \lambda). \tag{2.2}$$

This expression can be written in terms of a convex combination of deterministic maps $g_j : \mathbb{Z}_{c_j} \to \mathbb{Z}_{d_j}$ at each site. The single site probabilities are then written as a convex combination over all deterministic maps, i.e. $p(m_j|s_j, \lambda) = \sum_{g_j} p_{g_j} \delta_{g_j}^{m_j}$ where g_j are the single site maps with $p_{g_j} \geq 0$ and $\sum_{g_j} p_{g_j} = 1$. If one considers all n deterministic single-site maps, then we can deterministically obtain the resulting output digit-string from all parties $\mathbf{m} = \{g_1(s_1), g_2(s_2), g_3(s_3), ..., g_n(s_n)\}$ where $g_j(s_j)$ is the image of s_j under the single-site map g_j. As a result, equation (2.2) can be rewritten as:

$$p(\mathbf{m}|\mathbf{s}) = \sum_{g_1,g_2,...,g_n} p_{g_1,g_2,...,g_n} \prod_{j}^{n} \delta_{g_j(s_j)}^{m_j}, \tag{2.3}$$

taking a convex combination over all combination of single site maps g_j so that $p_{g_1,g_2,...,g_n} \geq 0$ and $\sum_{g_1,g_2,...,g_n} p_{g_1,g_2,...,g_n} = 1$ is satisfied. Note that the decomposition in (2.3) is not unique; uniqueness is only guaranteed when $p_{g_1,g_2,...,g_n} = 1$ for a particular choice of single site maps.

We see immediately from (2.3) that the space $\mathcal{L}_\mathcal{F} \subseteq \mathcal{F}$ of LHV correlations $p(\mathbf{m}|\mathbf{s})$ is also a convex polytope as defined in terms of a vertex representation. The vertices of $\mathcal{L}_\mathcal{F}$ are the deterministic probabilities $p(\mathbf{m}|\mathbf{s}) = \prod_j^n \delta_{g_j(s)}^{m_j}$ corresponding to each combination of single site maps g_j. There is also the facet representation of the polytope $\mathcal{L}_\mathcal{F}$ in terms of facet-defining linear inequalities. These linear inequalities are the facet-defining Bell inequalities, which we abbreviate to **facet Bell inequalities**, that constrain and define the consequences of LHV theories [Collins2004, Froissart1981, Pitowsky1989, Peres1999]. We now formally define what is means for a linear inequality to be facet-defining.

Definition 3. A linear inequality is **facet-defining** for a convex polytope $\mathcal{C}$ in a real space $\mathbb{R}^{\Delta}$ of dimension Δ when at least Δ affinely independent extreme points of $\mathcal{C}$ saturate the inequality (i.e. satisfy the equality of the linear inequality).

A set $\mathcal{S}$ of K vectors $\vec{p}_i$, $\mathcal{S} = \{\vec{p}_0, \vec{p}_1, ..., \vec{p}_{(K-1)}\}$ is *affinely independent* if for every $\vec{p}_k \in \mathcal{S}$, the $(K-1)$ vectors in the set $\{\vec{p}_i - \vec{p}_k | \vec{p}_i \neq \vec{p}_k\}$ are linearly independent. A linear inequality for the space of correlations is of the form:

$$\sum_{\mathbf{m},\mathbf{s}} \beta_{\mathbf{m},\mathbf{s}} p(\mathbf{m}|\mathbf{s}) \leq \gamma_{\mathcal{L}}, \tag{2.4}$$

where $\beta_{\mathbf{m},\mathbf{s}}$ are real pre-factors depending on $\mathbf{m}$ and $\mathbf{s}$ and $\gamma_{\mathcal{L}} \in \mathbb{R}$ as the upper bound resulting from LHV correlations in (2.3). All LHV correlations satisfy (2.4) whether the inequality is facet-defining or otherwise. For the inequalities to be facet Bell Inequalities the following conditions must be satisfied:

$$\sum_{\mathbf{m},\mathbf{s}} \beta_{\mathbf{m},\mathbf{s}} \prod_{j}^{n} \delta_{g_j(s)}^{m_j} = \gamma_{\mathcal{L}}, \tag{2.5}$$

for at least $(\prod_{j=1}^{n} d_j - 1) \prod_{j=1}^{n} c_j$ affinely independent vectors $\vec{p}$ such that elements are $p(\mathbf{m}|\mathbf{s}) = \prod_j^n \delta_{g_j(s)}^{m_j}$. We can demonstrate this schematically in Figure 2.2 where we show that a facet Bell inequality picks out the surface of the LHV polytope, whereas the inequalities in (2.4) might just bound the LHV polytope. We will make these ideas concrete in chapter 3.

The problem of finding the facets of a polytope given the vertices is known as the *facet enumeration problem* [Collins2004] and software does exist that performs this task (e.g. [Polymake2000]). However, it is currently in general both theoretically and practically hard to find these inequalities as we shall discuss in the subsequent sub-section. The hardness of this problem will motivate us to think about simplified Bell inequality settings, and then relate these simplified settings to a more general setting.

2.1.4 Facet Bell Inequalities and Computational Complexity

Given our abstract setting for n parties each with c_j possible inputs and d_j possible outputs, it is immediately natural to ask how hard is it to obtain Facet Bell Inequalities? Pitowsky notably studied this question by studying the intimate link between convex polytopes and propositional logic [Pitowsky1989]. The latter then has a deep connection to computational complexity, the branch

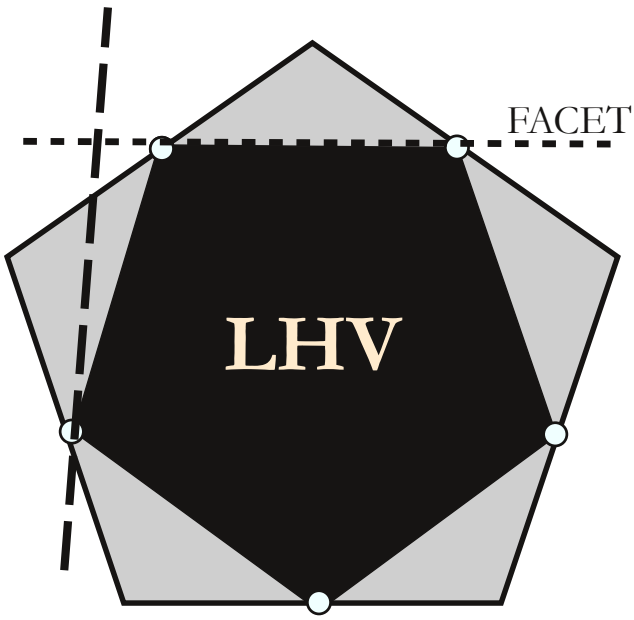

Figure 2.2: The LHV polytope can be defined in terms of the facet Bell inequalities. These inequalities intersect the surface of this polytope; whereas an arbitrary Bell inequality might only intersect one vertex as shown in this schematic.

of theoretical computer science devoted to the hardness of computational problems [Papadimitriou1994].

Whether a problem is easy or not can be defined in terms of time efficiency of finding a solution on a Turing Machine (an abstract computer that can simulate all other computers [Turing1937]) with respect to the size of the input. Time is defined in terms of computational time, or the number of computational steps in an algorithm. The computational time as a function of input size is then the indicator of computational efficiency, or hardness. If this function is a polynomial in the size of the input, then it is efficient, if super-polynomial (e.g. exponential) then it is inefficient [Papadimitriou1994].

The problems that are given to a Turing Machine are decision problems. That is, given an input the machine outputs 'yes', represented as the bit 0 or 'no', represented as bit 1; the decision problems are questions with binary potential outcomes. If the algorithm performed by the machine to make this decision operates in a number of steps that is polynomial in the size of the input, then it is in the complexity class called P. It is key to note that the algorithm must be polynomial in the input size for all possible inputs, as there may be inputs that are easier to compute than others. If problems in P are efficiently solvable, then there is another class of problems where solutions can be checked (for veracity) in an efficient amount of time. This complexity class is called NP.

The question of whether P is equal to NP is one of the greatest mathematical unsolved puzzles. Discussion of this problem is outside of the discussion of this thesis. However, if P = NP then in loose terms it would be as easy to solve a problem as to check the validity of the solution; this may seem intuitively incorrect to both a casual reader and a computer scientist. The hardest decision problems in NP are called NP-complete and so if P = NP then these problems have a polynomial time solution. At the current time, no polynomial time solution is known for NP-complete problems.

Pitowsky has shown that finding the facet Bell inequalities is at least as hard as any of the NP-complete problems, if not necessarily in NP [Pitowsky1989, Pitowsky1991]. In the terminology of computatational complexity, this problem is NP-hard. Heuristically, Pitowsky showed this by relating the problem of finding a facet Bell inequality to a Boolean satisfiability problem [Pitowsky1991]. These problems ask whether there exist variables that result in a Boolean function being 'true' and are NP-complete [Papadimitriou1994]. The vertices of the LHV polytope $\mathcal{L}_{\mathcal{F}}$ consist of vectors with elements being 0 or 1, which are truth

value assignments. This relationship between vertices of the polytope and truth assignments allowed Pitowsky to say that finding facet Bell inequalities is at least as hard as a Boolean satisfiability problem.

We have given an overview of the geometric construction of Bell inequalities and the space of correlations. The implications of local realism are connected to the study of convex polytopes. Polytopes have also been used to study the space of non-signalling correlations; we shall return to this subject in section 2.3 of this chapter. Finally we have shown that finding the Bell inequalities that define the LHV polytope is a computationally hard problem. This motivates our study of correlators, the probability of a joint outcome between all n parties instead of the full probability distribution. This simplifies a hard problem by decreasing its dimensionality. Correlators also have a computational perspective that will become crucial to our study of Bell inequalities.

2.2 Correlators

Finding facet Bell inequalities is difficult. This difficulty scales with the size of the problem such as the number of possible inputs and outputs for n parties. Firstly, we assume that $d_j = d_k = d$ (for $j \neq k$) is the same for all parties and is prime. We also simplify the type of probabilities that we need to consider from the correlations $p(\mathbf{m}|\mathbf{s})$ to a *correlator* which takes the form:

$$p(k|\mathbf{s}) = \sum_{\mathbf{m}} \delta_k^{\left[\sum_{j=1}^n m_j\right]_d} p(\mathbf{m}|\mathbf{s}) = p([\sum_{j=1}^n m_j]_d = k|\mathbf{s}), \tag{2.6}$$

where throughout this thesis (unless otherwise stated) we write all modulo x arithmetic in brackets with a sub-script $[...]_x$. From this perspective, the Bell test now consists of inputs $\mathbf{s} \in \bigoplus_{j=1}^n \mathbb{Z}_{c_j}$ and a single value output $k = \left[\sum_{j=1}^n m_j\right]_d$ is returned.

The correlator $p(k|\mathbf{s})$ is a stochastic map $f : \bigoplus_{j=1}^n \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ and due to the normalisation $\sum_{k=0}^{d-1} p(k|\mathbf{s}) = 1$ for all $\mathbf{s}$, we only need to consider $(d-1)\prod_{j=1}^n c_j$ correlators. We do not consider, then, the correlator $p(0|\mathbf{s})$ for all $\mathbf{s}$ as it can be recovered by normalisation. These correlators are now elements of a real vector $\vec{k} \in \mathbb{R}^{(d-1)\prod_{j=1}^n d_j}$ which we call a *correlator vector* [3].

[3] If there is no conflict in meaning, we may shorten correlator vector to just correlator. For example, if we refer to correlators being in some space, this means the resulting correlator vectors are in some space.

We can describe the space of all possible correlator vectors as another convex polytope $\mathcal{P}$. This object, analogous to $\mathcal{F}$, has a simple description in terms of vertices and linear inequalities. First, $\mathcal{P}$ has E vertices, or extreme points, $\vec{k}_e$ for $e \in \{1, 2, ..., E\}$ that have the elements $p_e(k|\mathbf{s}) \in \{0, 1\}$ for all k and $\mathbf{s}$. Therefore, these vectors correspond to deterministic maps where a single value of k is the output given the input $\mathbf{s}$ so elements are $p_e(k|\mathbf{s}) = \delta^k_{f(\mathbf{s})}$ where $f(\mathbf{s})$ is the image of $\mathbf{s}$ under the map $f : \bigoplus_{j=1}^n \mathbb{Z}_{c_j} \to \mathbb{Z}_d$. Any correlator vector $\vec{k} \in \mathbb{R}^{(d-1)\prod_{j=1}^n d_j}$ can be written as a convex combination of these extreme points:

$$\vec{k} = \sum_e p_e \vec{k}_e = \sum_f p_f \vec{k}_f \tag{2.7}$$

where every extreme point $\vec{k}_e$ corresponds one-to-one with a vector $\vec{k}_f$ resulting from a function $f : \bigoplus_{j=1}^n \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ and $p_e \geq 0$, $p_f \geq 0$ with $\sum_e p_e = 1$ and $\sum_f p_f = 1$ for all functions f.

Equivalently we can describe $\mathcal{P}$ in terms of the linear inequalities corresponding to positivity and normalisation: $p(k|\mathbf{s}) \geq 0$ for all k and $\mathbf{s}$ and $\sum_{k\neq 0} p(k|\mathbf{s}) \leq 1$. This is analogous to the way we defined $\mathcal{F}$ but interestingly, every vector in $\mathcal{P}$ can be produced by at least one probability distribution in $\mathcal{NS}$, the non-signalling polytope. If we allow all probability distributions that satisfy only the no-signalling condition we can completely saturate $\mathcal{P}$. As an example, every vertex of $\mathcal{P}$ corresponding to the map $f : \bigoplus_{j=1}^n \mathbb{Z}_{c_j} \to \mathbb{Z}_d$, we can always write this probability distribution:

$$p(\mathbf{m}|\mathbf{s}) = \begin{cases} d^{1-n} & \text{if } [\sum_{j=1}^n m_j]_d = f(\mathbf{s}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

for the function f as above. All reductions of this probability distribution are $p(\mathbf{m}^{\{j|j\in\mathcal{J}\}}|\mathbf{s}^{\{j|j\in\mathcal{J}\}}) = d^{-|\mathcal{J}|}$ if $|\mathcal{J}| \neq n$ for all $\mathbf{m}^{\{j|j\in\mathcal{J}\}}$ and $\mathbf{s}^{\{j|j\in\mathcal{J}\}}$. Since this distribution is uniformly random for all sub-sets of parties, it satisfies the no-signalling condition. We shall elaborate on the connections between $\mathcal{NS}$ and $\mathcal{P}$ in a subsequent section 2.3.

Another motivation for these correlators is that they are a generalisation of the well-studied CHSH Bell Inequality setting for many parties [CHSH1969, Werner2001, Żukowski2002]. This generalisation also coincides with the Collins-Gisin-Linden-Massar-Popescu (CGLMP) setting again generalised to many parties [CGLMP2002]. An example of work in a many-setting CGLMP framework

includes that by Acín et al [Acín2004].

Also, in the literature, correlators can be considered to result from the expectation value of the outcome of joint measurements if the outcomes of measurements are complex numbers of unit modulus [Lee2007, Son2006]. More specifically, every jth party's measurement $\mathcal{M}_{s_j}$ has the outcome values $e^{i2\pi\frac{k}{d}}$ for $k \in \mathbb{Z}_d$, then the expectation value of the joint measurement $\mathbb{E}(\mathbf{s}) = \mathbb{E}(\prod_{j=1}^{n} \mathcal{M}_{s_j})$ is:

$$\begin{aligned}
\mathbb{E}(\mathbf{s}) &= \sum_{\mathbf{m}} \prod_{j=1}^{n} e^{i2\pi\frac{m_j}{d}} p(\mathbf{m}|\mathbf{s}) \\
&= \sum_{k=0}^{(d-1)} e^{i2\pi\frac{k}{d}} p(k|\mathbf{s}) \\
&= 1 + \sum_{k=1}^{(d-1)} \left[e^{i2\pi\frac{k}{d}} - 1 \right] p(k|\mathbf{s}). \qquad (2.9)
\end{aligned}$$

These expectation values $\mathbb{E}(\mathbf{s})$ can be written in terms of correlators. Every measurement that has two possible outcomes $\{+1, -1\}$ results in expectation values of measurements being $\mathbb{E}(\mathbf{s}) = 1 - 2p(1|\mathbf{s})$; the expectation values are equivalent to a single correlator $p(1|\mathbf{s})$. This is the many-party generalisation of the CHSH setting for two parties. This equivalence has allowed research in the past to interchangeably use expectation values as well as conditional probabilities.

One can coarse-grain research into generalized Bell inequality setting as either obtaining statistics in terms of correlators (e.g. [CHSH1969, CGLMP2002, Acín2004]) or the full probability distribution (e.g. [CH1969, Collins2004]). The latter can be reduced to the former but much literature has been devoted to the study of correlators. As well as being able to infer structure of $\mathcal{NS}$ from $\mathcal{P}$ (see section 2.3), these correlators are at the centre of much research into Bell inequalities. We will now try and formalise the structure of the space of correlator vectors by considering the maps performed by all possible theories.

2.2.1 Correlators as Computations

Throughout this thesis we argue for a computational approach to Bell inequality experiments by considering in what sense correlations are computing a function $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ on the inputs $\mathbf{s}$. In this section we now want to introduce some of the tools associated with these functions so that we can be more specific about the computational power of correlations from physical (or non-physical)

theories.

Every function $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ can be written as a list (a single column table) with each row representing $f(\mathbf{s})$, the image of $\mathbf{s}$ under f. In turn this list is an element of the module $\mathbb{M}$ over the ring $\mathbb{Z}_d$. The module $\mathbb{M}$ consists of the abelian group $\mathbb{Z}_d^D$ for $D = \prod_{j=1}^{n} c_j$ with the group multiplication being modulo d addition of these elements, the module also has (left or right) scalar multiplication $\mathbb{Z} \times \mathbb{Z}_d^D \to \mathbb{Z}_d^D$ of elements in the group. In order to satisfy $\mathbb{M}$ being a module then for all $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{Z}_d^D$, and all a, b in $\mathbb{Z}_d$ then:

1. $1\mathbf{x} = \mathbf{x}$ (existence of the identity)
2. $a(b\mathbf{x}) = (ab)\mathbf{x}$ (associativity)
3. $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$ (distributivity over $\mathbb{Z}_d^D$)
4. $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$ (distributivity over $\mathbb{Z}_d$),

where we could have written the scalar multiplication in terms of left or right multiplication [Anderson1992]. All arithmetic is modulo d but we have suppressed the notation $[...]_d$ for clarity.

Every $f \in \mathbb{M}$ can be written in terms of Kronecker delta functions with elements $f(\mathbf{s}) = \delta_{\mathbf{y}}^{\mathbf{s}}$ which is 1 for only one input $\mathbf{s} = \mathbf{y} \in \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j}$ and 0 otherwise. Therefore every element $f(\mathbf{s})$ of any function f can be written as

$$f(\mathbf{s}) = \sum_{\mathbf{y} \in \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j}} f(\mathbf{y}) \delta_{\mathbf{y}}^{\mathbf{s}} \tag{2.10}$$

with $f(\mathbf{y}) \in \mathbb{Z}_d$. The delta functions then form something analogous to the basis vectors for a vector space and we can replace one of the delta functions with the constant, all-ones function with elements $f(\mathbf{s}) = 1$. The delta function we choose to replace is $f(\mathbf{s}) = \delta_{\mathbf{y}}^{\mathbf{s}}$ with $\mathbf{y} = \mathbf{0}$, the all-zeroes digit-string.

For every $s_j \in \mathbb{Z}_{c_j}$ we can choose to represent $\mathbb{Z}_{c_j}$ as a Cartesian product of cyclic groups of dimension being the prime factors of c_j. The set of prime factors of c_j are written as $\{{}^1c_j, {}^2c_j, ..., {}^{q_j}c_j\}$ for ${}^k c_j$ as the kth prime factor and q_j being the number of prime factors, therefore $s_j = \{{}^1s_j, {}^2s_j, ..., {}^{q_j}s_j\} \in \bigoplus_{k=1}^{q_j} \mathbb{Z}_{{}^k c_j}$ with ${}^k s_j \in \mathbb{Z}_{{}^k c_j}$. The delta functions $\delta_{\mathbf{y}}^{\mathbf{s}}$ can be written now in terms of inputs

$\mathbf{s} \in \bigoplus_{j=1}^{n} \left(\bigoplus_{k=1}^{q_j} \mathbb{Z}_{{}^k c_j} \right)$ and $\mathbf{y} \in \bigoplus_{j=1}^{n} \left(\bigoplus_{k=1}^{q_j} \mathbb{Z}_{{}^k c_j} \right)$ giving

$$\begin{aligned} \delta_{\mathbf{y}}^{\mathbf{s}} &= \prod_{j=1}^{n} \prod_{k=1}^{q_j} \delta_{{}^k y_j}^{{}^k s_j} \\ &= \prod_{j=1}^{n} \prod_{k=1}^{q_j} \left[1 - \left({}^k s_j - {}^k y_j \right)^{{}^k c_j - 1} \right]_{{}^k c_j} \\ &= \prod_{j=1}^{n} \prod_{k=1}^{q_j} \left[1 - \sum_{l=0}^{{}^k c_j - 1} (-1)^l \binom{{}^k c_j - 1}{l} ({}^k y_j)^l \left({}^k s_j \right)^{{}^k c_j - (l+1)} \right]_{{}^k c_j} \end{aligned} \tag{2.11}$$

where the second line is guaranteed by Fermat's little theorem. That is, the modular arithmetic expression is $\left[\left({}^k s_j - {}^k y_j \right)^{{}^k c_j - 1} \right]_{{}^k c_j} = 1$ for ${}^k s_j \neq {}^k y_j$ and coprime with ${}^k c_j$. The third line above just results from the binomial theorem.

In the instance where $c_j = c_k = d$ being prime for all $j \neq k$, the delta functions just simplify to being a polynomial over the field $\mathbb{Z}_d$ as indicated by the third line above. For example, for $d = 2$ the delta functions are Boolean functions $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ expressed as polynomials over $\mathbb{Z}_2$. We will use these properties more explicitly in subsequent chapters of this thesis.

We now introduce classes of functions that will be used to characterise the correlators resulting from particular theories. The first class of functions we now describe as "n-partite linear functions" have connections to LHV theories.

Definition 4. An **n-partite linear function** is a function $g : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ where the image of $\mathbf{s}$ under g can be written as

$$g(\mathbf{s}) = \left[\sum_{j=1}^{n} g_j(s_j) \right]_d \tag{2.12}$$

with $g_j(s_j)$ the image of s_j under the single-site map $g_j : \mathbb{Z}_{c_j} \to \mathbb{Z}_d$.

These functions are not strictly linear as the single-site maps g_j are not always linear in s_j, but for $c_j = d = 2$, then these maps are linear. We use the nomenclature of linearity *only* to highlight the fact that there is addition modulo d between single-site maps and not multiplication. If a map cannot be expressed as an n-partite linear function then we say it is a **non-n-partite linear function**.

Any function $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ can be described in terms of a sum of an n-partite linear function and non-n-partite linear function, i.e. $f(\mathbf{s}) = [g(\mathbf{s}) + h(\mathbf{s})]_d$

where $g(\mathbf{s})$ and $h(\mathbf{s})$ are n-partite and non-n-partite linear functions respectively. First we write an n-partite linear function in terms of the delta functions $\delta_{y_j}^{s_j}$ for single-site maps to obtain:

$$g(\mathbf{s}) = \left[\alpha + \sum_{j=1}^{n} \sum_{k=1}^{c_j - 1} \beta_{j,k} \delta_k^{s_j} \right]_d , \tag{2.13}$$

with α, $\beta_{j,k} \in \mathbb{Z}_d$. The constant α emerges from taking the sum modulo d of the constant function $\alpha_j \in \mathbb{Z}_d$ for each site, that replaces the delta function $\delta_0^{s_j}$, as discussed.

For inputs $\mathbf{s}$ with only one non-zero element, there is only a single delta function $\delta_k^{s_j}$ (for s_j being the non-zero element) that describes the value of $g(\mathbf{s}) = \left[\alpha + \beta_{j,k}\delta_k^{s_j}\right]_d$. For the all-zeroes digit-string $\mathbf{s} = \mathbf{0}$, then the only function describing $g(\mathbf{s})$ is the constant function α. We call the set of digit-strings $\mathbf{s}$ with at most one non-zero element $\mathcal{T}$.

Now we briefly consider a column list of the images of s under f for only these digit-strings $\mathbf{s} \in \mathcal{T}$. Then for this restricted list delta functions $\delta_k^{s_j}$ (and constant α), similar to before, form a basis for any function $f(\mathbf{s})$ with $\mathbf{s} \in \mathcal{T}$. This is because they are equivalent to the delta functions $\delta_{\mathbf{y}}^{\mathbf{s}} = \delta_k^{s_j} \prod_{l \neq j}^{n} \delta_0^{s_l}$ for these particular input digit-strings. A basis is formed in the sense that these functions are linearly independent over $\mathbb{Z}_d$.

In order to achieve any function $f(\mathbf{s})$ we need a basis for the functions for all possible input strings $\mathbf{s}$ including $\mathbf{s} \in \mathcal{T}$. We do this by supplementing the function $g(\mathbf{s})$ above with the delta functions $\delta_{\mathbf{y}}^{\mathbf{s}}$ with $\mathbf{y} \notin \mathcal{T}$. Therefore, any function can be written as a sum of an n-partite linear function $g(\mathbf{s})$ and a non-n-partite linear function

$$\begin{aligned} f(\mathbf{s}) &= \left[\alpha + \sum_{j=1}^{n} \sum_{k=1}^{c_j - 1} \beta_{j,k} \delta_k^{s_j} + \sum_{\mathbf{y} \notin \mathcal{T}} \gamma_{\mathbf{y}} \delta_{\mathbf{y}}^{\mathbf{s}} \right]_d \\ &= \left[g(\mathbf{s}) + h(\mathbf{s}) \right]_d , \end{aligned} \tag{2.14}$$

with $\gamma_{\mathbf{y}} \in \mathbb{Z}_d$ and $h(\mathbf{s}) = \sum_{\mathbf{y} \notin \mathcal{T}} \gamma_{\mathbf{y}} \delta_{\mathbf{y}}^{\mathbf{s}}$ as a non-n-partite linear function by construction. We describe this form of $f(\mathbf{s})$ as the *decomposition* of the function into n-partite linear and non-n-partite linear functions. If $h(\mathbf{s}) = 0$ for all $\mathbf{s}$, then $f(\mathbf{s})$ is necessarily an n-partite linear function, otherwise it is necessarily a non-n-partite linear function. If, on the other hand, $g(\mathbf{s}) = 0$ we have the following

result:

Lemma 1. *If a function* $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ *has no* n*-partite linear function in its decomposition, then* $f(\boldsymbol{s}) = 0$ *for all* $\boldsymbol{s} \in \mathcal{T}$.

Proof - If a function $f(\mathbf{s}) = [g(\mathbf{s}) + h(\mathbf{s})]_d$ has no n-partite linear function part, i.e. $g(\mathbf{s}) = 0$ for all $\mathbf{s}$ but some non-zero non-n-partite linear part, i.e. $h(\mathbf{s}) \neq 0$ for some inputs $\mathbf{s}$, then it can be written as,

$$f(\mathbf{s}) = \left[\sum_{\mathbf{y} \notin \mathcal{T}} \gamma_{\mathbf{y}} \delta_{\mathbf{y}}^{\mathbf{s}} \right]_d . \tag{2.15}$$

Then $f(\mathbf{s})$ must be zero for all $\mathbf{s} \in \mathcal{T}$. □

We will use this lemma in the proof of Lemma 3 in section 2.3.1 and is a useful consequence of choosing this decomposition of functions. We shall also show in the following subsection that this decomposition is physically motivated and not just mathematically convenient.

Another class of functions will now be introduced and shown to be useful in later sections 2.3 and 2.4 of this chapter. They can be seen to be a relaxation of the constraint of n-partite linear functions where instead of taking a sum of single-site maps, we take a sum of maps produced by sub-sets of all parties. In particular, we consider all the ways in which n parties can be partitioned into a non-empty sub-set $\mathcal{J}$ and its complement $\mathcal{J}^c$ as introduced in section 2.1. The class of functions called "bipartite linear functions" are then a generalization of n-partite linear functions defined for these partitions.

Definition 5. A **bipartite linear function** is a function $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ where the image of $\mathbf{s}$ under f can be written as

$$f(\mathbf{s}) = \left[f^1(\mathbf{s}^{\{j|j\in\mathcal{J}\}}) + f^2(\mathbf{s}^{\{j|j\in\mathcal{J}^c\}}) \right]_d \tag{2.16}$$

with $f^1 : \bigoplus_{j\in\mathcal{J}} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ and $f^2 : \bigoplus_{j\in\mathcal{J}^c} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ being functions mapping inputs for each partition into $\mathcal{J}$ and $\mathcal{J}^c$ to a single output.

These functions are equivalent to an n-partite linear function for $n = 2$ as the partition can be seen as a coarse-graining of n parties into two sub-sets, where each sub-set can be considered a party in its own right. Then the input string $\mathbf{s}^{\{j|j\in\mathcal{J}\}}$ is now a single input to one 'collective' party and $\mathbf{s}^{\{j|j\in\mathcal{J}^c\}}$ the input to

the other collective party. Each subset's collective output is just then the sum modulo d of all of their outputs m_j for $j \in \mathcal{J}$ and $j \in \mathcal{J}^c$ for each respective subset. As a result, for a given partition, any function $f(\mathbf{s})$ can be written as a sum of a bipartite linear function and a **non-bipartite linear function**.

We have described classes of functions, and every function describes a vertex of $\mathcal{P}$. To recapitulate, a vertex of $\mathcal{P}$ has the elements $p(k|\mathbf{s}) = \delta^k_{f(\mathbf{s})}$ for every $\mathbf{s}$. A correlator captures a computation whereby given some input, an output is produced with some probability. The region of $\mathcal{P}$ that is subsumed by a particular theory can then have a computational interpretation in terms of how 'close' the region of a particular theory gets to vertices of $\mathcal{P}$. In the following subsection we will discuss the region of correlators achievable in an LHV or quantum theory.

2.2.2 Correlators from Physical Theories

Bell tests aim to expose statistics that do not result from a particular class of theories viz. LHV theories. We will now describe the space of correlators $\mathcal{L} \subseteq \mathcal{P}$ resulting from LHV theories. This space can be defined in terms of the language of stochastic maps, and in particular, the functions defined in the subsection 2.2.1. We now present the following theorem which defines $\mathcal{L}$ in terms of a sub-class of all possible functions $f : \bigoplus_{j=1}^n \mathbb{Z}_{c_j} \to \mathbb{Z}_d$.

Theorem 2. *The space of LHV correlators $\mathcal{L}$ is the convex hull of deterministic correlators $p(k|\boldsymbol{s}) = \delta^k_{f(\boldsymbol{s})}$ for $f(\boldsymbol{s})$ being all of the n-partite linear functions, for all $\boldsymbol{s}$.*

Proof: The proof follows simply from how the probabilities $p(\mathbf{m}|\mathbf{s})$ are defined in (2.3) to obtain correlators:

$$\begin{aligned} p(k|\mathbf{s}) &= \sum_{\mathbf{m}} \delta_k^{[\sum_{j=1}^n m_j]_d} \sum_{g_1,g_2,\ldots,g_n} p_{g_1,g_2,\ldots,g_n} \prod_j^n \delta^{m_j}_{g_j(s_j)} \\ &= \sum_{g(\mathbf{s})} p_{g(\mathbf{s})} \delta^k_{g(\mathbf{s})}, \end{aligned} \tag{2.17}$$

where $g(\mathbf{s}) = [\sum_{j=1}^n g_j(s_j)]_d$ is an n-partite linear function by definition and $p_{g(\mathbf{s})} \geq 0$ and $\sum_{g(\mathbf{s})} p_{g(\mathbf{s})} = 1$. Therefore all LHV correlators are contained in the convex hull of n-partite linear functions. □

The consequence of this theorem then is that correlators resulting from LHV theories have a limited *computational expressiveness*. That is, no correlator re-

sulting from an LHV theory can deterministically perform a non-n-partite linear function. This is one of the main computational perspectives that we employ in this thesis, and we will return to this result throughout.

The CHSH inequality is a facet-defining Bell inequality for the LHV polytope $\mathcal{L}$ for $n = 2$ and $c_1 = c_2 = d = 2$ [Fine1982]. We have shown previously that this inequality can be violated by quantum correlators, therefore they cannot in general be confined to the polytope $\mathcal{L}$ for all possible values of n, c_j and d. Tsirelson showed that there is an equivalent CHSH inequality for quantum correlators denoted:

$$p(1|00) + p(1|01) + p(1|10) - p(1|11) \leq 1 + \sqrt{2} \approx 2.41, \tag{2.18}$$

whereas the upper bound for LHV correlators is 2 [Tsirelson1980]. However, the vertex of $\mathcal{P}$ described as $p(1|\mathbf{s}) = \delta^1_{f(\mathbf{s})}$ with $f(\mathbf{s}) = [s_1 s_2 + 1]_2$ gives a value of 3 for the CHSH inequality. Therefore, there is a hierarchy of spaces of correlators such that $\mathcal{L} \subseteq \mathcal{Q} \subseteq \mathcal{P}$ with $\mathcal{Q}$ as the space of quantum correlators.

Defining the space $\mathcal{Q}$ of quantum correlators (and correlations in general) is still a major open question but we can indicate some general properties of $\mathcal{Q}$. As Pitowsky has previously shown, $\mathcal{Q}$ is convex, but not a polytope [Pitowsky1989]. Quantum correlators can be written in terms of the probabilities $p(\mathbf{m}|\mathbf{s})$ which result from measurements on a quantum state ρ, i.e.

$$\begin{aligned} p(k|\mathbf{s}) &= \sum_{\mathbf{m}} \delta_k^{[\sum_{j=1}^n m_j]_d} p(\mathbf{m}|\mathbf{s}) \\ &= \sum_{\mathbf{m}} \delta_k^{[\sum_{j=1}^n m_j]_d} \mathrm{Tr}(\rho \bigotimes_{j=1}^{n} P_{m_j}^{s_j}) \end{aligned} \tag{2.19}$$

where $P_{m_j}^{s_j}$ is a single-site POVM corresponding to an outcome m_j given the choice of measurement s_j so that $\sum_{m_j} P_{m_j}^{s_j} = \mathbb{I}$, the identity matrix. If each measurement site has access to a Hilbert space $\mathcal{H}_j$, then the state ρ is in general, a density matrix acting over the tensor-product of these n Hilbert spaces $\bigotimes_{j=1}^n \mathcal{H}_j$. The dimension of each Hilbert space is arbitrary (and possibly infinite).

Naimark's theorem indicates that any POVM is equivalent to a PVM on an ancilla Hilbert space (it also applies for infinite dimensional systems) [Paulsen2003]. Therefore every correlator can be written in terms of a state ρ' and projectors $Q_{s_j}^{m_j}$ on the Hilbert space $\bigotimes_{j=1}^n \mathcal{H}'_j$ where $\mathcal{H}'_j$ is each jth site's enlarged Hilbert space. Projectors $Q_{s_j}^{m_j}$ are expressed in terms of each site's orthogonal basis $|m_j\rangle_{s_j}$ for

the s_j choice of basis, i.e. $Q_{s_j}^{m_j} = |m_j\rangle_{s_j}\langle m_j|_{s_j}$ such that $Q_{s_j}^{m_j} Q_{s_j}^{m'_j} = Q_{s_j}^{m_j}\delta_{m'_j}^{m_j}$. A density matrix can be constructed from a convex combination of pure states $\rho' = \sum_l p_l |\psi_l\rangle\langle\psi_l|$ so that,

$$p(k|\mathbf{s}) = \sum_l p_l \sum_{\mathbf{m}} \delta_k^{[\sum_{j=1}^n m_j]_d} \left| \langle\psi_l| \bigotimes_{j=1}^n |m_j\rangle_{s_j} \right|^2 . \tag{2.20}$$

Since $p(k|\mathbf{s}) = \sum_{\mathbf{m}} \delta_k^{[\sum_{j=1}^n m_j]_d} \left| \langle\psi_l| \bigotimes_{j=1}^n |m_j\rangle_{s_j} \right|^2$ is itself a quantum correlator, $\mathcal{Q}$ is convex and the extreme points of $\mathcal{Q}$ will be defined by particular measurements on particular pure states, i.e. $\rho' = |\psi_l\rangle\langle\psi_l|$. $\mathcal{Q}$ is not a polytope with a finite number of extreme points as the inner product $\left| \langle\psi_l| \bigotimes_{j=1}^n |m_j\rangle_{s_j} \right|^2$ is continuously defined over the reals for all pure states and bases $|m_j\rangle_{s_j}$. Heuristically, if an extreme point of $\mathcal{Q}$ is outside of $\mathcal{L}$ then there is a correlator that is arbitrarily close to this point resulting from a pure state that may also be extreme (see a far more rigorous analysis in [Pitowsky1989]).

As mentioned, actually finding the extreme points for all settings is a major open problem in current research. However, there do exist instances where the extreme points can be defined, particularly with n parties where each site has two inputs and two outputs (see section 3.2.3). Numerical methods exist for finding the boundary of $\mathcal{Q}$ using semi-definite programming [Navascués2008] and optimization over measurement bases given a particular state (e.g. the maximally entangled state for $n = 2$) [Durt2001, Acín2002]. We will elaborate on this point further on in section 3.2 of chapter 3.

There has recently been a different tack to defining $\mathcal{Q}$; is there some physical principle that captures the boundary of $\mathcal{Q}$? Instead of being a difficult calculation, is there is an underlying reason why the extreme points are the way they are? There is no definite answer to this, only indications of an answer (e.g. work presented in [Navascués2009] and [Oppenheim2010]). Interestingly, this approach has been extended to finding information theoretic principles that define extreme points. For example, if an extreme point were further from $\mathcal{L}$ then parties would be able to accumulate more information than is communicated to them [Pawłowski2009] or perform calculations with a "trivial" amount of communication [Brassard2006].

Popescu and Rohrlich began this exploration of finding what defines the quantum region [Popescu1994]. They originally asked whether it was special relativity

that limits the region of $\mathcal{Q}$ but the answer to this is negative. The vertex of $\mathcal{P}$ corresponding to $p(1|\mathbf{s}) = \delta^1_{f(\mathbf{s})}$ with $f(\mathbf{s}) = [s_1 s_2 + 1]_2$ can be produced by the following distribution:

$$p(m_1, m_2|s_1, s_2) = \begin{cases} \frac{1}{2} & \text{if } [m_1 + m_2]_2 = [s_1 s_2 + 1]_2, \\ 0 & \text{otherwise,} \end{cases} \tag{2.21}$$

which is an example of a non-signalling probability distribution, as described in section 2.2, more specifically it is a form of "Popescu-Rohrlich Non-local Box" (PR box) [Popescu1994, Barrett2005b]. However, it violates the Tsirelson-CHSH inequality above and so cannot result from quantum theory. On the other hand, it shows that there is a connection between the structure of $\mathcal{P}$ and $\mathcal{NS}$, the non-signalling polytope. In fact, this PR box is the *only* non-signalling distribution that can produce the corresponding vertex of $\mathcal{P}$. In the following section we make this unique connection more concrete.

2.3 Non-signalling Correlations

In this section, we will elaborate on the connections between the polytopes $\mathcal{P}$ and $\mathcal{NS}$. We have mentioned that the PR box in (2.21) is the only non-signalling correlation that can be associated with achieving a particular vertex of $\mathcal{P}$. If we assume our resources are non-signalling and we achieve a vertex of $\mathcal{P}$ associated with the function $[s_1 s_2 + 1]_2$ with only one possible probability distribution. This is no coincidence, but one example of an infinite number of non-signalling probability distributions of the form (2.8) that can be uniquely associated with a vertex of $\mathcal{P}$.

We suggest that a vertex of $\mathcal{NS}$ corresponding uniquely to a vertex of $\mathcal{P}$ is one possible way to generalise a PR box to more scenarios. We introduce another possible generalisation of a PR box in the next chapter in section 3.4.2. First we discuss the $n = 2$ situation and show that a vertex of $\mathcal{P}$ that is not in $\mathcal{L}$ can be uniquely associated with a vertex of $\mathcal{NS}$. We use the results we obtained for these bipartite PR boxes to consider the $n > 2$ case. In this multipartite scenario, again we can uniquely associate vertices of $\mathcal{P}$ with $\mathcal{NS}$. In all of the discussion in this section, we assume that d is prime.

2.3.1 Generalised bipartite PR boxes

The following lemma shows that there is a uniqueness relation between vertices of $\mathcal{P}$ and a distribution in $\mathcal{NS}$ for $n = 2$, or bipartite Bell tests. This result gives us many new ways of immediately generalising the PR box.

Lemma 3. *For every function $f : \mathbb{Z}_{c_1} \times \mathbb{Z}_{c_2} \to \mathbb{Z}_d$ that is non-n-partite linear for $n = 2$, the only non-signalling distribution compatible with the corresponding vertex $p(k|\boldsymbol{s}) = \delta^k_{f(\boldsymbol{s})}$ in $\mathcal{P}$ is*

$$p(m_1, m_2|s_1, s_2) = \begin{cases} d^{-1} & \text{if } [m_1 + m_2]_d = f(\boldsymbol{s}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.22}$$

Proof: The condition $p(k|\mathbf{s}) = \delta^k_{f(\mathbf{s})}$ for all $\mathbf{s} = \{s_1, s_2\}$ implies that for every value of m_1 in $p(m_1, m_2|s_1, s_2)$, there exists a unique value of $m_2 = [f(\mathbf{s}) - m_1]_d$. This immediately implies the equality for the following conditional distributions:

$$\begin{aligned} p(m_1 = x, m_2 = [f(\mathbf{s}) - x]_d|\mathbf{s}) &= \sum_{m_2} p(m_1 = x, m_2 = [f(\mathbf{s}) - x]_d|\mathbf{s}) \\ &= p(m_1 = x|\mathbf{s}) \\ &= \sum_{m_1} p(m_1 = x, m_2 = [f(\mathbf{s}) - x]_d|\mathbf{s}) \\ &= p(m_2 = [f(\mathbf{s}) - x]_d|\mathbf{s}), \end{aligned} \tag{2.23}$$

for all $x \in \mathbb{Z}_d$. The non-signalling condition further implies that $p(m_1 = x|\mathbf{s})$ is equal to $p(m_1 = x|s_1)$ and

$$p(m_1 = x|s_1) = p(m_2 = [f(\mathbf{s}) - x]_d|s_2), \tag{2.24}$$

which must be satisfied for all $\mathbf{s}$ and all x. We will show that repeated application of (2.24) for varying $\mathbf{s}$ allows us to prove that all non-marginal probabilities are equal provided that $f(\mathbf{s})$ has a non-n-partite linear element.

A function $f(\mathbf{s})$ can be decomposed into a non-n-partite linear and n-partite linear part, i.e.$f(\mathbf{s}) = [\mathrm{g}(\mathbf{s})+\mathrm{h}(\mathbf{s})]_d$ with $h(\mathbf{s})$ as a non-n-partite linear function. For every function, the n-partite linear part $g(\mathbf{s}) = [g_1(s_1) + g_2(s_2)]_d$ can be removed by local operations performed by each party; $g_j(s_j)$ is a single-site map that can be deleted from each party's outcome. Therefore, we only need to

consider functions $f(\mathbf{s})$ without an n-partite linear part. By lemma 1, we know for functions $f(\mathbf{s})$ without an n-partite linear part, $f(0,0) = f(0,s_2) = f(s_1,0) = 0$ for all s_1 and s_2. So repeatedly applying (2.24) gives

$$\begin{aligned} p(m_2 = [-x]_d|s_2) &= p(m_1 = x|s_1 = 0) \\ &= p(m_2 = [-x]_d|s_2 = 0) \\ &= p(m_1 = x|s_1) \\ &= p(m_2 = [f(s_1,s_2) - x]_d|s_2) \end{aligned} \tag{2.25}$$

for all x. Repeated iteration implies for the αth iteration,

$$p(m_2 = [-x]_d|s_2) = p(m_2 = [\alpha f(s_1,s_2) - x]_d|s_2) \tag{2.26}$$

for all $\alpha \in \mathbb{Z}_d$. The function $f(\mathbf{s})$ is non-n-partite linear so there must be at least one value of $\{s_1, s_2\}$ where $f(s_1,s_2)$ is non-zero. Since d is prime, $\alpha f(s_1,s_2)$ takes on all values in $\mathbb{Z}_d$, therefore the marginals are $p(m_2|s_2) = d^{-1}$ for all m_2. If the marginals are uniformly random for one particular input s_2, because $p(m_2 = [-x]_d|s_2) = p(m_1 = x|s_1 = 0) = p(m_2 = [-x]_d|s_2')$ for $s_2 \neq s_2'$, they will be uniformly random for all inputs.

Therefore, by the non-signalling conditions $p(m_1|s_1,s_2) = p(m_1|s_1,s_2') = d^{-1}$ and $p(m_2|s_1,s_2) = p(m_2|s_1',s_2) = d^{-1}$ implying $p(m_1|0,s_2) = p(m_1|0,0) = d^{-1}$ and $p(m_2|s_1,0) = p(m_2|0,0) = d^{-1}$; the marginals for all $\mathbf{s}$ must be completely random. Applying equation (2.24) implies that $p(m_1,m_2|s_1,s_2) = d^{-1}$ for all $\mathbf{m}$ such that $[m_1 + m_2]_d = f(\mathbf{s})$. □

2.3.2 Multipartite Generalisations of the PR box

Lemma 3 shows that for every vertex of $\mathcal{P}$ outside of $\mathcal{L}$ for $n = 2$ and d being prime, there is only one non-signalling probability distribution compatible with this vertex. As a corollary, $\mathcal{P}$ captures a lot of the structure of $\mathcal{NS}$ but with the space of statistics considered being smaller. We now go further and show that this one-to-one correspondence exists for $n > 2$.

Previous work has explicitly found the vertices of $\mathcal{NS}$ for $n = 3$, $c_j = 2$ for all j and $d = 2$ [Pironio2011]. This work revealed that multipartite non-signalling probability distributions can have an extremely complicated and unintuitive structure. For more general scenarios, very little is understood or been

investigated. Our approach, culminating in the following result, shows that correlators can give an insight into the multipartite structure of $\mathcal{NS}$.

Theorem 4. *For every function $f : \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ that is non-bipartite linear, the only non-signalling distribution compatible with the corresponding vertex $p(k|\boldsymbol{s}) = \delta^k_{f(\boldsymbol{s})}$ in $\mathcal{P}$ is*

$$p(\boldsymbol{m}|\boldsymbol{s}) = \begin{cases} d^{1-n} & \text{if } \left[\sum_{j=1}^{n} m_j\right]_d = f(\boldsymbol{s}), \\ 0 & \text{otherwise.} \end{cases} \tag{2.27}$$

Proof: As well as the above distribution of the form of (2.8) but for bipartite linear functions $f(\mathbf{s})$, we can explicitly construct another non-signalling probability distribution other than (2.8). This distribution can produce the corresponding vertex $p(k|\mathbf{s}) = \delta^k_{f(\mathbf{s})}$ of $\mathcal{P}$ for a bipartite linear function $f(\mathbf{s})$ and is

$$p(\mathbf{m}|\mathbf{s}) = \begin{cases} d^{2-|\mathcal{J}|-|\mathcal{J}^c|} = d^{2-n} & \text{if } [\sum_{j\in\mathcal{J}} m_j]_d = f_1(\mathbf{s}^{\{j|j\in\mathcal{J}\}}) \\ & \text{and } [\sum_{j\in\mathcal{J}^c} m_j]_d = f_2(\mathbf{s}^{\{j|j\in\mathcal{J}^c\}}), \\ 0 & \text{otherwise,} \end{cases} \tag{2.28}$$

since a bipartite linear function can be written as

$$f(\mathbf{s}) = [f_1(\mathbf{s}^{\{j|j\in\mathcal{J}\}}) + f_2(\mathbf{s}^{\{j|j\in\mathcal{J}^c\}})]_d \tag{2.29}$$

for all functions $f_1 : \bigoplus_{j\in\mathcal{J}} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ and $f_2 : \bigoplus_{j\in\mathcal{J}^c} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ for strict sub-set $\mathcal{J}$ and complement $\mathcal{J}^c$. The distribution is non-signalling across the partition as well as amongst the parties in the sub-set since in the sub-set it has the form (2.8).

This, therefore, leaves non-bipartite linear functions and their corresponding non-signalling probability distributions. As mentioned, every partition into $\mathcal{J}$ and $\mathcal{J}^c$ can be seen as a situation with two parties, where each side of the partition makes a choice from $c_{\mathcal{J}} = \prod_{j\in\mathcal{J}} c_j$ and $c_{\mathcal{J}^c} = \prod_{j\in\mathcal{J}^c} c_j$ inputs respectively; each partition also adds all their outputs togethers modulo d to obtain collective outputs $m_{\mathcal{J}} = [\sum_{j\in\mathcal{J}} m_j]_d$ and $m_{\mathcal{J}^c} = [\sum_{j\in\mathcal{J}^c} m_j]_d$ respectively. As a result, Lemma 3 now applies and if the resource produces $p(k|\mathbf{s}) = \delta^k_{f(\mathbf{s})}$ for $f(\mathbf{s})$ being a non-bipartite linear function, for all partitions into $\mathcal{J}$ and $\mathcal{J}^c$, then we obtain $p(m_{\mathcal{J}}|\mathbf{s}^{\{j|j\in\mathcal{J}\}}) = p(m_{\mathcal{J}^c}|\mathbf{s}^{\{j|j\in\mathcal{J}^c\}}) = d^{-1}$. This means all output strings $\mathbf{m}^{\{j|j\in\mathcal{J}\}}$ and $\mathbf{m}^{\{j|j\in\mathcal{J}^c\}}$ for all strict sub-sets occur with equal probability, unlike

the distribution in (2.28). This necessarily results in the distribution of the form (2.8) thus proving the theorem. □

The uniqueness relation between vertex of $\mathcal{P}$ and a distribution in $\mathcal{NS}$ says something about the vertices of $\mathcal{NS}$. This results from the extremality of vertices of $\mathcal{P}$ and the following result that says all non-signalling probability distributions that produce a vertex of $\mathcal{P}$ must form a face of $\mathcal{NS}$. The uniqueness result of Theorem 4 then collapses the face to a single vertex.

Proposition 5. *Every non-signalling probability distribution that produces a vertex of $\mathcal{P}$ forms a face of $\mathcal{NS}$.*

Proof: Every non-signalling probability distribution $p(\mathbf{m}|\mathbf{s})$ can be written as a convex combination of the set E of extreme points of $\mathcal{NS}$, i.e.

$$p(\mathbf{m}|\mathbf{s}) = \sum_E p(E) p_E(\mathbf{m}|\mathbf{s}) \tag{2.30}$$

where $p_E(\mathbf{m}|\mathbf{s})$ is a vertex distribution of $\mathcal{NS}$ and $p(E) \geq 0$ and $\sum_E p(E) = 1$. Of the set E, a sub-set of extreme points E' will each result in the same vertex of $\mathcal{P}$, and their convex combination will always result in a vertex $\vec{k}_E$ of $\mathcal{P}$. The region of distributions in $\mathcal{NS}$ which is formed by the convex hull of extreme points in E' is called $\mathcal{E}$.

First, we will point out that $\mathcal{E}$ has no points in the interior of $\mathcal{NS}$ and elements of $\mathcal{E}$ are only on the boundary (i.e. surface) of $\mathcal{NS}$. If we take the convex combination of an extreme point $p_{E'}(\mathbf{m}|\mathbf{s})$ in E' and an extreme point $p_{\not E'}(\mathbf{m}|\mathbf{s})$ in the set of extreme points not in E', then we have the convex line:

$$q p_{E'}(\mathbf{m}|\mathbf{s}) + (1-q) p_{\not E'}(\mathbf{m}|\mathbf{s}), \tag{2.31}$$

for $0 \leq q \leq 1$. If $\mathcal{E}$ has any elements in the interior of $\mathcal{NS}$ then $\mathcal{E}$ is intersected by at least one of the convex lines of (2.31) for $q \neq 0$ or $q \neq 1$. However, if $q \neq 1$ then this means that a probability distribution cannot result in the deterministic correlator $\vec{k}_E$ in $\mathcal{P}$, thereby leading to a contradiction. Therefore $\mathcal{E}$ must lie in at most a facet of $\mathcal{NS}$ because if it lies on one or more facets, then there will necessary be interior points of $\mathcal{NS}$ in $\mathcal{E}$.

Finally, we now show that if $\mathcal{E}$ is a Δ-dimensional sub-space of $\mathcal{NS}$, it does not lie in $\mathcal{X}$, a Δ'-dimensional sub-space (or Δ'-face) of $\mathcal{NS}$ where $\Delta' > \Delta$. As a result, $\mathcal{E}$ must be a face of $\mathcal{NS}$. If $\mathcal{E}$ lies in a larger space $\mathcal{X}$, then $\mathcal{X}$ has at least

one more extreme point than $\mathcal{E}$; this would mean that points in $\mathcal{E}$ can be written as a convex combination of extreme points in E' *and* not in E'. A contradiction again emerges as we would not obtain a deterministic correlator $\vec{k}_E$ in $\mathcal{P}$. □

Proposition 6. *A vertex $p(k|\boldsymbol{s}) = \delta^k_{f(\boldsymbol{s})}$ of $\mathcal{P}$ corresponding to $f(\boldsymbol{s})$ being a non-bipartite linear function results from a single vertex of $\mathcal{NS}$.*

Proof: Since there is only a single non-signalling probability distribution resulting in the vertex $p(k|\mathbf{s}) = \delta^k_{f(\mathbf{s})}$ of $\mathcal{P}$ for $f(\mathbf{s})$ being a non-bipartite linear function, the region $\mathcal{E}$ from the proof of Proposition 1 will necessarily consist of one extreme point. Therefore, $\mathcal{E}$ becomes a 1-face or vertex. □

The space of all possible correlators $\mathcal{P}$, uniquely captures properties of a full probability distribution that only satisfies special relativity. The study of $\mathcal{NS}$ has been motivated recently by foundational issues of what distinguishes quantum physics from something unphysical (e.g. [Pawłowski2009]). Vertices of $\mathcal{NS}$ have also been studied in the context of being an information theoretic resource [Barrett2005b]. Possession of particular resources that produce a vertex of $\mathcal{NS}$ not achievable with LHV or quantum resources (e.g. PR boxes) can lead to an information processing advantage in certain tasks (e.g. communication complexity [Brassard2006]). It has also been suggested that PR boxes can be seen as a unit of non-LHV correlations (often abbreviated as "non-locality"), though there is evidence both for and against this suggestion [Barrett2005c]. The fact that the space of correlators captures generalisations of the PR box (with respect to extremality of $\mathcal{NS}$) motivates the study of correlators as a smaller-dimensional problem revealing more general structures.

The bipartite linear functions are not only of relevance to Proposition 6 but also of relevance to the next section. In the next section, we discuss a generalisation of correlations discussed by George Svetlichny [Svetlichny1987]; these are correlations that exceed LHV correlations but do involve the space of all possible correlations. Interestingly, as Svetlichny has shown, these correlations do not fully capture all quantum correlations.

2.4 Svetlichny Correlations

George Svetlichny suggested an extension to the standard model of local hidden variables in the many-party scenario. More specifically, Svetlichny introduced

the scenario where there are three parties, and two parties are allowed to share whatever correlations they wish, but they are restricted to sharing only an LHV with the third party [Svetlichny1987]. Therefore if parties 1 and 2 can share whatever correlation they wish (it could even not respect special relativity), then party 3 only shares some local hidden variable $\lambda \in \Lambda$ with 1 and 2, to obtain the following distribution:

$$p(\mathbf{m}|\mathbf{s}) = \int_{\Lambda} p(\lambda) d\lambda p(m_1, m_2|s_1, s_2, \lambda) p(m_3|s_3, \lambda), \tag{2.32}$$

with the probability distribution $p(\lambda)d\lambda$ over Λ with $\int_{\Lambda} p(\lambda)d\lambda = 1$. There is no reason to privilege some parties over others and we allow permutations of parties so labels can be swapped, i.e. $\{1, 2, 3\} \to \sigma(\{1, 2, 3\})$ and σ is just a member of the permutation group.

In full generality, we can allow probabilistic combinations of distributions of the form (2.32) but with permutations of parties to give

$$\begin{aligned} p(\mathbf{m}|\mathbf{s}) \quad = \quad & p_{1,2} \int_{\Lambda} p_{1,2}(\lambda) d\lambda p(m_1, m_2|s_1, s_2, \lambda) p(m_3|s_3, \lambda) \\ & + p_{1,3} \int_{\Lambda} p_{1,3}(\lambda) d\lambda p(m_1, m_3|s_1, s_3, \lambda) p(m_2|s_2, \lambda) \\ & + p_{2,3} \int_{\Lambda} p_{2,3}(\lambda) d\lambda p(m_2, m_3|s_2, s_3, \lambda) p(m_1|s_1, \lambda), \end{aligned} \tag{2.33}$$

with $p_{i,j}$ and $p_{i,j}(\lambda)$ for $i,\ j \in \{1, 2, 3\}$ and $i \neq j$ being probabilities for a particular permutation such that $p_{1,2} + p_{1,3} + p_{2,3} = 1$. Therefore, all Svetlichny-type correlations in the form of (2.33) are in a sub-region of $\mathcal{F}$ that is a convex polytope $\mathcal{S}_{\mathcal{F}}$; the extreme points of $\mathcal{S}_{\mathcal{F}}$ are distributions of the form (2.32) but with both probabilities $p(m_j, m_k|s_j, s_k)$ and $p(m_l|s_l, \lambda)$ being deterministic for $j \neq k \neq l$.

2.4.1 Three-party Generalised Svetlichny Correlators

Since $\mathcal{S}_{\mathcal{F}}$ is a convex polytope, it will be defined as the intersection of half-spaces defined by a set of linear inequalities in analogy with the facet Bell inequalities. Svetlichny actually originally described his set of linear inequalities of correlators. We shall now take this original approach and describe Svetlichny correlations in terms of correlators where $\mathcal{S}$ is the space of Svetlichny correlators for three parties. The following result captures this space $\mathcal{S}$ in terms of the description of

functions that we have used in the last two sections.

Proposition 7. *The space $\mathcal{S}$ of Svetlichny correlators for three parties is the convex hull of vertices $p(k|\boldsymbol{s}) = \delta^k_{f(\boldsymbol{s})}$ of $\mathcal{P}$ corresponding to bipartite linear functions $f(\boldsymbol{s})$.*

Proof: If we take a probability distribution of the form in (2.32), then it can itself be written as a convex combination of deterministic probabilities of the form

$$p(\mathbf{m}|\mathbf{s}) = \delta^{\{m_j,m_k\}}_{g^1(s_j,s_k)}\delta^{m_l}_{g^2(s_l)}, \tag{2.34}$$

for the maps $g^1 : \mathbb{Z}_{c_j} \times \mathbb{Z}_{c_k} \to \mathbb{Z}_d \times \mathbb{Z}_d$ and $g^2 : \mathbb{Z}_{c_l} \to \mathbb{Z}_d$ with $j \neq k \neq l \in \{1,2,3\}$. The probability in (2.34) is defined for all possible maps g^1 and g^2, therefore we can rewrite (2.32) as a convex combination of these deterministic probabllities and permutations of $\{1,2,3\}$ to give

$$p(\mathbf{m}|\mathbf{s}) = \sum_{\{j,k,l\}\in\sigma\{1,2,3\}} \sum_{g^1,g^2} p_{g^1,g^2}\delta^{\{m_j,m_k\}}_{g^1(s_j,s_k)}\delta^{m_l}_{g^2(s_l)}, \tag{2.35}$$

where $p_{g^1,g^2} \geq 0$ is defined over all maps such that $\sum_{g^1,g^2} p_{g^1,g^2} = 1$. Therefore $\mathcal{S}_\mathcal{F}$ is the convex hull of extreme points defined by all possible maps of the form g^1 and g^2 for all different labellings of parties.

For correlators, we take the sum modulo d of all outcomes. Taking the sum $[m_j + m_k]_d$ results in all maps of the form g^1 now becoming all maps of the form $f^1 : \mathbb{Z}_{c_j} \times \mathbb{Z}_{c_k} \to \mathbb{Z}_d$. Finally, the sum of all outcomes is now $[m_j + m_k + m_l]_d = f(\mathbf{s}) = [f^1(\mathbf{s}) + f^2(\mathbf{s})]_d$ where $f^2 = g^2$. These functions $f(\mathbf{s})$ are by definition bipartite linear functions and so $\mathcal{S}$ is the convex hull of correlators resulting from bipartite linear functions. □

A facet Svetlichny inequality is a linear inequality that defines a facet of $\mathcal{S}$ in analogy with the facet Bell inequalities. One of the original facet Svetlichny inequalities for the setting with three parties, $c_j = d = 2$ for all j can be written in terms of correlators as [Svetlichny1987]

$$\begin{aligned} &p(1|000) + p(1|001) + p(1|010) - p(1|011) \\ &+p(1|100) - p(1|101) - p(1|110) - p(1|111) \quad \leq 2. \end{aligned} \tag{2.36}$$

Interestingly, despite the fact that we allow any possible correlation to be shared between two of the three parties, correlators in $\mathcal{Q}$ still violate (2.36) with the

quantum (Tsirelson-Svetlichny) upper bound $1+\sqrt{2}$ [Svetlichny1987]. Whilst quantum correlator vectors may be outside the space $\mathcal{S}$, this Svetlichny polytope is not strictly smaller than the space of quantum correlators, i.e. some vertices of $\mathcal{S}$ are not achievable with quantum correlators.

2.4.2 Multipartite Svetlichny Correlators

The above discussion has been restricted to Svetlichny's original work for three parties. It is natural to ask how this approach generalises to more than three parties. One could suggest a model where we allow only at most two out of n parties to share whatever correlation they wish and then share local hidden variables with the other $(n-2)$ parties. We will go further, and in line with other approaches (e.g. [Bancal2009, Bancal2011]), partition n parties into two sub-sets and parties in each of the two sub-sets is allowed to share whatever correlations they wish (signalling or otherwise). Then each partition only shares a local hidden variable $\lambda \in \Lambda$ (with probability distribution $p(\lambda)d\lambda$) with the other partition to obtain correlations of the form:

$$p(\mathbf{m}|\mathbf{s}) = \int_{\Lambda} p(\lambda)d\lambda p(\mathbf{m}^{\{j|j\in\mathcal{J}\}}|\mathbf{s}^{\{j|j\in\mathcal{J}\}},\lambda)p(\mathbf{m}^{\{j|j\in\mathcal{J}^c\}}|\mathbf{s}^{\{j|j\in\mathcal{J}^c\}},\lambda), \qquad (2.37)$$

where n parties are partitioned into sub-sets $\mathcal{J}$ and $\mathcal{J}^c$.

As with three parties, we allow convex combinations of distributions in (2.37) for all $(2^{n-1}-1)$ different partitions into strict sub-sets $\mathcal{J}$ and $\mathcal{J}^c$. Correlators resulting from this generalised Svetlichny model can again be expressed as a convex polytope as a generalisation of Proposition 7; the following result now captures this generalisation.

Theorem 8. *The space $\mathcal{S}$ of generalised Svetlichny correlators for n parties is the convex hull of vertices $p(k|\boldsymbol{s}) = \delta^k_{f(\boldsymbol{s})}$ of $\mathcal{P}$ corresponding to bipartite linear functions $f(\boldsymbol{s})$.*

Proof: The correlations in (2.37), as with the three-party case, can be written as a convex combination of deterministic probabilities resulting from deterministic maps labelled g^1 and g^2:

$$p(\mathbf{m}|\mathbf{s}) = \sum_{g^1,g^2} p_{g^1,g^2}\delta^{\mathbf{m}^{\{j|j\in\mathcal{J}\}}}_{g^1(\mathbf{s}^{\{j|j\in\mathcal{J}\}})}\delta^{\mathbf{m}^{\{j|j\in\mathcal{J}^c\}}}_{g^2(\mathbf{s}^{\{j|j\in\mathcal{J}^c\}})}, \qquad (2.38)$$

where $g^1 : \bigoplus_{j\in\mathcal{J}}\mathbb{Z}_{c_j} \to \mathbb{Z}_d^{|\mathcal{J}|}$ and $g^2 : \bigoplus_{j\in\mathcal{J}^c}\mathbb{Z}_{c_j} \to \mathbb{Z}_d^{|\mathcal{J}^c|}$ with $p_{g^1,g^2} \geq 0$ and

$\sum_{g^1,g^2} p_{g^1,g^2} = 1$. Now if we take the sum modulo d of all outcomes then we obtain the following correlators:

$$p(k|\mathbf{s}) = \sum_{f^1,f^2} p_{f^1,f^2} \delta^k_{[f^1(\mathbf{s})+f^2(\mathbf{s})]_d}, \tag{2.39}$$

with all possible maps of the form $f^1 : \bigoplus_{j\in\mathcal{J}} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ and $f^2 : \bigoplus_{j\in\mathcal{J}^c} \mathbb{Z}_{c_j} \to \mathbb{Z}_d$ and the distribution $p_{f^1,f^2} \geq 0$ such that $\sum_{f^1,f^2} p_{f^1,f^2} = 1$.

If we allow all possible correlators of the form (2.39) for all possible partitions into $\mathcal{J}$ and $\mathcal{J}^c$ then $\mathcal{S}$ is the convex hull of all deterministic correlators corresponding to functions $f(\mathbf{s}) = [f^1(\mathbf{s}) + f^2(\mathbf{s})]_d$. These are all of the bipartite linear functions by definition. □

The structure of bipartite linear functions gets translated from the three-party case to the n-party case. Despite the fact that we allowed signalling correlations within partitions of the n parties, we can impose the non-signalling conditions on all parties once again. This means that even within a sub-set of parties, the correlations they share must satisfy special relativity. Interestingly, even if we apply this restriction, the space of Svetlichny correlators for many parties is still $\mathcal{S}$ as defined by Theorem 8. This is simply because all deterministic correlators (or vertices of $\mathcal{P}$) can be achieved with non-signalling probability distributions $\mathcal{NS}$. All the deterministic correlators associated with bipartite linear functions can be achieved with probability distributions in $\mathcal{NS}$.

If one assumes that all correlations satisfy special relativity, then non-signalling correlations not achievable with Svetlichny-type correlations are said to be "truly n-partite non-local" [Bancal2009, Barrett2005b]. They are "non-local" in the sense that across all partitions of n parties, the correlations of the parties are not described by the parties sharing a local hidden variable. Therefore, the vertices of $\mathcal{P}$ that are not associated with bipartite linear functions can only result from truly n-partite non-local correlations. Of the non-signalling correlations in $\mathcal{NS}$, then for each of these vertices of $\mathcal{P}$ there is one truly n-partite non-local distribution, or vertex of $\mathcal{NS}$ as described by Theorem 4.

Instead of allowing all possible correlations within a sub-set of all parties or just allowing non-signalling correlations, one could allow correlations "in-between" that allow some, but not all forms of communication. Indeed, these issues have been investigated by Barrett and Pironio [Barrett2011a]. If one is only concerned with correlators, then whatever form of restricted, or unrestricted, communica-

tion within a partition of all parties, the space of Svetlichny-type correlators is $\mathcal{S}$ as described by Theorem 8. The space of correlators is conserved and we can always discuss the possibility of distinguishing between a model that permits, in part, an LHV description and something inconsistent with this model.

2.5 Chapter Summary

In this chapter we have motivated and presented the study of Bell correlators in a natural generalisation of the Bell-CHSH test. We have also discussed how finding Bell inequalities that define the space of LHV correlations/correlators is in general a hard problem. Motivated by this, studying correlators instead of a full probability distribution, we reduce the size of the problem, if not reducing the general hardness.

The language of stochastic maps and functions has been key to describing the correlators resulting from particular theories (both physical and non-physical). This description of correlators in terms of *computational expressiveness* is key to the central results of not only this chapter, but this entire thesis. To summarise, each potential theory has its own computational expressiveness and characterising this gains an insight into "which computations the theory is capable of performing". These ideas will be generalised in subsequent chapters to take into account data processing in Bell tests but the computational expressiveness insight will be key. Importantly, this computational point-of-view on correlators has allowed us to characterise the well-studied structures of LHV correlators in a new language.

This interpretation of correlators in terms of computation has also produced new results. We showed that vertices of the polytope of all correlators can correspond uniquely to vertices of the non-signalling polytope. As well as this, we have described the space of Svetlichny correlators in terms of computational expressiveness. Again, this description of Svetlichny correlations gives us a new insight into well-studied areas of research.

3 Constructing Bell Inequalities and Quantum Violations

In the previous chapter, we focussed mostly on the description of the local hidden variable (LHV) polytope in terms of its vertices. Now we shift to a facet representation of the LHV polytope in terms of the facet Bell inequalities: linear inequalities defining the facets of this polytope [Fine1982, Pitowsky1989]. If a correlator is outside of the polytope it must necessarily violate at least one of these inequalities. However, recall that finding them is a hard problem.

A Bell inequality is a linear inequality of the following form

$$\sum_{\mathbf{s}} \sum_{k=1}^{(d-1)} \beta_{k,\mathbf{s}} p(k|\mathbf{s}) \leq \gamma_{\mathcal{L}}, \tag{3.1}$$

for some real coefficients $\beta_{k,\mathbf{s}}$ where $\gamma_{\mathcal{L}}$ is the tight upper bound for all LHV correlators in $\mathcal{L}$ [1]. We introduce the vernacular that a "Bell expression" is the left-hand-side of (3.1). We make the distinction between Bell expression and Bell inequality as we can substitute correlators not in $\mathcal{L}$ into a Bell expression and they could violate a Bell inequality.

We optimize over values $\beta_{k,\mathbf{s}}$ and $\gamma_{\mathcal{L}}$ in (3.1) to find the facet Bell inequalities. But this optimization, in the worst case, is a hard computational task. In this chapter we look for these facet Bell inequalities but only manage to find them for a select number of scenarios on a desktop PC using Polymake [Polymake2000]. We give some indications of the possible connections between the violations of facet inequalities and the possibility of performing a non-n-partite linear function. However, this connection is not completely clear as the structure of $\mathcal{L}$ is in general, rather complicated. On the other hand, we review the results of Werner, Wolf, Żukowski and Brukner [Werner2001, Żukowski2002] in the n party, 2 input, 2

[1] In the literature, tight Bell inequalities are synonymous with facet Bell inequalities. Our use of the word tight reflects that the Bell inequality intersects the LHV polytope at (at least) one of its extreme points.

output scenario and relate the structure of $\mathcal{L}$ in this scenario to a particular class of Boolean functions.

In spite of the difficulty in understanding the structure of $\mathcal{L}$ and even finding the facet Bell inequalities, we find a general class of Bell inequalities that have a natural computational perspective. We call these inequalities non-trivial Bell inequalities. They are non-trivial in the sense that they provide a separation between all possible correlators in $\mathcal{L}$ and all possible correlators in $\mathcal{P}$. We go on to relate these inequalities to an information processing paradigm called a "non-local game" [Cleve2004]. We then use the construction of a non-local game to derive more of these non-trivial Bell inequalities.

Finally in this chapter, we make interesting connections between the discussion of Bell inequalities and Measurement-based Quantum Computing (MBQC) [Raussendorf2001, Raussendorf2003, Briegel2009, Jozsa2006]. In particular, we show that a sub-class of computations in Briegel and Raussendorf's construction of MBQC [Raussendorf2001, Raussendorf2003] can be cast as non-local games. Through the language of non-local games, we relate these quantum computations to non-trivial Bell inequalities. All of these connections truly highlight the rich interplay between the foundations of quantum mechanics and its applications.

The original material in sections 3.1 and 3.2 along with subsections 3.3.1 and 3.3.2 were completed in collaboration with Joel Wallman and Dan Browne and published in part as [Hoban2011c]. The subsections of 3.1.5 and 3.2.3 consist of rederivations of results in [Werner2001] with a focus on the computational description of correlators. The original work in subsection 3.3.3 and section 3.4 were done in collaboration with Earl Campbell, Klearchos Loukopoulos and Dan Browne and published as [Hoban2011a].

3.0.1 Notation

From now on, we simplify the scenarios of Bell tests that we consider by having the number of inputs at each site being the same, i.e. $c_j = c_{j'}$ for all $j \neq j'$. We introduce the notation (n, c, d) to describe Bell tests with n parties, c inputs and d outputs at each site. We also carry over the notation from chapter 2 of $\mathcal{L}$, $\mathcal{S}$, $\mathcal{Q}$ and $\mathcal{P}$ being the LHV polytope, the Svetlichny polytope, the space of quantum and all possible correlators respectively for each scenario (n, c, d).

The majority of the remainder of this thesis will be devoted to the study of the $(n, 2, 2)$ scenario. We privilege this scenario by assigning it a particular notation not shared by any others. Since the number of the inputs at each site is the same,

inputs are always $\mathbf{s} \in \bigoplus_{j=1}^{n} \mathbb{Z}_{c_j} = \mathbb{Z}_c^n$. As a result of this simplification, we will no longer use the notation $\bigoplus$ to describe the Cartesian product of groups $\mathbb{Z}_{c_j}$. We will use $\bigoplus$ to denote summation modulo 2, i.e. $\bigoplus = [\sum ...]_2$. This notation is used only in the $(n, 2, 2)$ scenario along with the notation $\oplus$ to describe addition modulo 2, i.e. $\oplus = [... + ...]_2$. Modulo 2 multiplication between elements in $\mathbb{Z}_2$ is exactly multiplication of these elements for standard arithmetic. Therefore, for the $(n, 2, 2)$ scenario and only this scenario we re-write expressions in modulo 2 arithmetic in terms of this notation. For example, the expression $[x_1x_2 + x_3 + 1]_2$ becomes $x_1x_2 \oplus x_3 \oplus 1$, and, $\left[\left(\sum_{j=1}^{4} x_j\right) + x_5 + 1\right]_2$ becomes $\left(\bigoplus_{j=1}^{4} x_j\right) \oplus x_5 \oplus 1$.

For scenarios other than $(n, 2, 2)$, we retain the notation from the previous chapter. That is, all arithmetic in $[...]_x$ is modulo x arithmetic. Even if either c or d is equal to 2 (but not both), we will use the notation $[...]_2$ for modulo 2 arithmetic.

3.1 Facet Bell Inequalities

In this section, we will discuss the facet Bell inequalities for particular (n, c, d) scenarios. We used the Polymake package of algorithms to find the facet Bell inequalities for a small number of cases [Polymake2000]. These are the (n, c, d) scenarios where finding the inequalities was computationally tractable on a desktop PC[2]. We will show that these inequalities can be grouped together into symmetries, or in group theoretical terms, orbits; these orbits are generated by operations that preserve the region $\mathcal{L}$ [Pitowsky1991]. For the number of (n, c, d) scenarios studied, we will describe elements in these orbits. Then we discuss the facet Bell inequalities for the $(n, 2, 2)$ scenario; there is a closed-form expression for these inequalities [Werner2001, Żukowski2002].

In Table 3.1 we have listed the number of facet Bell inequalities for a few scenarios that could be computed using Polymake. Included in the number of facet Bell inequalities are the c^n normalization and $(d-1)c^n$ positivity inequalities that define $\mathcal{P}$. Despite these dc^n inequalities, there are still a significant number of inequalities remaining. On the other hand, Pitowsky has shown that correlation polytopes have certain symmetries [Pitowsky1991]. These symmetries are generated by operations on the inputs and outputs as well as permutations of parties. The group of these symmetry operations generates *orbits* of facet Bell

[2]iMac with 2.4 GHz Intel Core 2 Duo (TM) Processor and 2 GB 800 MHz DDR2 SDRAM.

n	c	d	# Vertices	# Facet Bell inequalities
2	2	2	8	16
2	2	3	27	66
2	2	4	64	216
2	2	5	125	1020
3	2	2	16	256
3	2	3	81	125,412
2	3	2	32	90
2	4	2	128	27,968

Table 3.1: A table of number of facet Bell inequalities for each scenario (n, c, d) and the number of vertices for the LHV polytope.

inequalities[3]. Every facet Bell inequality in each orbit can be mapped to every other inequality in that orbit via these symmetry operations. Therefore, we do not need to consider every single facet Bell inequality for each (n, c, d) scenario but only one inequality in each orbit. In the following subsection we consider these symmetry operations.

3.1.1 Symmetries of the LHV Polytope

Pitowsky has shown that given a facet Bell inequality for an LHV correlation polytope, we may find more inequalities by some simple operations on data $\mathbf{m}$ and $\mathbf{s}$ [Pitowsky1991]. These operations G map from the set $\mathcal{E}$ of extreme points of $\mathcal{L}$ to themselves, i.e. $G : \mathcal{E} \to \mathcal{E}$. By convexity, we only need to consider the extreme points. The symmetry operations G that produce these maps are the following:

1. permutations of parties - $\{s_i, s_j, ..., s_n\} \to \{s_{i'}, s_{j'}, ..., s_{n'}\}$ where $k' = \sigma(k)$ is an element of the permutation group S^n of order n;

2. relabeling of measurement scenarios - $s_j \to s_j + a_j$ for some $a_j \in \mathbb{Z}_c$;

3. relabeling of measurement outcomes - $m_j \to m_j + b(s_j, j)$ where $b(s_j, j) \in \mathbb{Z}_d$.

The operations G and their products GG' (for either $G \neq G'$ or $G = G'$) form a group $\mathbb{G}$ such that $G \in \mathbb{G}$. There are $n!$ permutations of n parties and c^n ways of relabeling measurement scenarios. Since for each input s_j we add a value $b(s_j)$, for each input $\mathbf{s}$, $b(\mathbf{s}) = \sum_{j=1}^{n} b(s_j)$ is added to $\sum_{j=1}^{n} m_j$. There will be at most

[3]We are using the terminology used by Werner and Wolf [Werner2001].

n	c	d	# Facet Bell inequalities	# Orbits
2	2	2	16	2
2	2	3	66	2
2	2	4	216	4
2	2	5	1020	5
3	2	2	256	5
3	2	3	125,412	63
2	3	2	90	2
2	4	2	27,968	15

Table 3.2: The number of orbits for each scenario (n, c, d) under the symmetry operations described in the text. One of the orbits for each scenario is the orbit of normalization and positivity conditions.

d^{cn} values of $b(\mathbf{s})$. In total, there are at most $n!c^n d^{cn}$ elements of $\mathbb{G}$ in order for there to be closure[4].

The n-partite linear functions are closed under all of these operations. Using the facet-defining condition, the vertices of $\mathcal{L}$ that saturate a facet Bell inequality must be equivalent to another set of vertices in $\mathcal{L}$; this new set also saturates a facet Bell inequality. In group theoretic terms, if we have one facet Bell inequality and perform all possible sequences of operations G, then the set of facet Bell inequalities produced by these operations forms an *orbit* (see the use of terminology in [Werner2001]). In Table 3.2 we have listed the number of orbits for each of the scenarios in Table 3.1. These orbits were numerically found using a search algorithm on all of the facet Bell inequalities. For each instance of (n, c, d), it was found that one of the orbits consists of the normalisation and positivity inequalities; we call this orbit the "trivial orbit". Orbits which do not include the normalisation and positivity inequalities are called "non-trivial orbits".

For each of the (n, c, d) scenarios, we only need to consider one inequality from each orbit. For the $(2, 2, 2)$, $(2, 2, 3)$ and $(2, 3, 2)$ scenarios, there is only one non-trivial orbit. In each of these scenarios, we then only need to consider one inequality. If one of these inequalities in each orbit is violated by a quantum correlator, then the above symmetry operations can be applied to that quantum correlator so that it will violate every other inequality in said orbit. The possibility of violation of facet Bell inequalities with quantum correlators is, as a result,

[4]In principle, the number of operations could be smaller as the values of $b(\mathbf{s})$ may be overcomplete for all possible transformations. For example, in [Werner2001] the cardinality of $\mathbb{G}$ is $n!2^{2n+1}$.

rendered easier to study.

For the $(2, 2, 2)$ scenario, as Fine has also shown in [Fine1982], the only facet Bell inequality we need to consider is the CHSH inequality [CHSH1969]. In the following subsection we consider other facet Bell inequalities for $n = 2$. We show that the CHSH inequality and a generalisation in d (for $c = 2$) of this inequality (the CGLMP inequality [CGLMP2002]) between them generate a lot of the structure of $\mathcal{L}$. In later subsections 3.1.4 and 3.1.5 we will discuss the tripartite and multipartite scenario (i.e. for $n > 2$). First we briefly introduce some new notation.

3.1.2 Notation for Bell inequalities

We now introduce a piece of notation to describe all Bell inequalities. If we write vectors $\vec{k}$ of correlators that have elements $p(k|\mathbf{s})$, we can express an inequality as an inner product. The real pre-factors $\beta_{k,\mathbf{s}}$ of (3.1) are elements of an $(d-1)c^n$-length row vector $\vec{b} \in \mathbb{R}^{(d-1)c^n}$. Therefore, every inequality results from the Euclidean inner product $\vec{b} \cdot \vec{p} \leq \gamma_{\mathcal{L}}$ of these two vectors.

We adopt a convention to order the elements $\beta_{k,\mathbf{s}}$ of $\vec{b}$ from left-to-right starting with $\beta_{1,\mathbf{0}}$ and ending with $\beta_{(d-1),\mathbf{c}}$ with $\mathbf{c} = \{(c-1), (c-1), ..., (c-1)\}$, the digit-string of all inputs being $(c-1)$. To be explicit, each digit-string $\mathbf{s} \in \mathbb{Z}_c^n$ can be written as an integer in $\mathbb{Z}$, the set of positive integers. Digit-strings $\mathbf{s} \in \mathbb{Z}_c^n$ can be ordered in terms of these integers in $\mathbb{Z}$. For example for $c = 2$, the digit-string $\mathbf{s} = \{1, 0, 0\}$ corresponds to the integer 4 and for $c = 3$ the same digit-string is equal to 9. We order elements $\beta_{1,\mathbf{0}}$ from left-to-right for increasing values of $k \in \mathbb{Z}_d$ for each ordered value of $\mathbf{s}$.

To give a concrete example, the CHSH inequality [CHSH1969]

$$p(1|00) + p(1|01) + p(1|10) - p(1|11) \leq 2, \tag{3.2}$$

corresponds to the vector $\vec{b} = (\beta_{1,\{0,0\}}, \beta_{1,\{0,1\}}, \beta_{1,\{1,0\}}, \beta_{1,\{1,1\}}) = (1, 1, 1, -1)$. We will employ this notation for specific values n, c and d. For brevity, in more general expressions we may choose to write the inequality in terms of the sum in (3.1). In the next subsection we will write both in terms of the sum in (3.1) and the vector notation introduced above.

3.1.3 Bipartite facet Bell inequalities

In this subsection we will restrict ourselves to the $n = 2$ scenario for particular values of c and d. The CGLMP inequality [CGLMP2002] is a facet Bell inequality for all d in $(2, 2, d)$ scenarios, as shown by Masanes [Masanes2003]. For all d, this inequality can be written as

$$\mathcal{C}_{\text{CGLMP}} = d\times p(1|0,0) - \sum_{\mathbf{s}} (-1)^{s_1+s_2} p(1|\mathbf{s}) + \sum_{\mathbf{s}} (-1)^{s_1+s_2} \sum_{k=2}^{d-1} (d-k-1) p(k|\mathbf{s}) \leq d. \tag{3.3}$$

The CHSH inequality is exactly this inequality when $d = 2$. For $d = 2$, 3, the only non-trivial orbit is generated by the CGLMP inequality. Whilst for $d = 4$, 5 the CGLMP inequality generates one of $(d-1)$ non-trivial orbits. For all possible correlators in $\mathcal{P}$, the maximal value of the left-hand-side of the CGLMP inequality is $2d-1$, thus violating it. In fact, for all d, this maximal violation of the CGLMP is obtained by a vertex of $\mathcal{P}$ corresponding to the function $f(\mathbf{s}) = [s_1 s_2 + 1]_d$, i.e. the correlator $p(k|\mathbf{s}) = \delta^k_{[s_1 s_2+1]_d}$.

In the $(2, 2, 2)$ scenario there are $2^4 - 2^3 = 8$ non-n-partite linear functions and also 8 inequalities in the non-trivial orbit of the CHSH inequality. This is no coincidence as every Bell inequality in this orbit in maximally violated by a vertex of $\mathcal{P}$ corresponding to a non-n-partite linear function. This also occurs for the $(2, 2, 3)$ scenario where there are $3^4 - 3^3 = 54$ non-n-partite linear functions and 54 inequalities in the orbit of the CGLMP inequality. It can also be checked that every inequality in this orbit is violated by a different non-n-partite linear function.

For $(2, 2, 4)$, one of the orbits is generated by a generalisation of the CHSH inequality

$$\mathcal{C}^1_{d=4} = \sum_{\mathbf{s}} (-1)^{s_1 s_2} \left[p(1|\mathbf{s}) + p(3|\mathbf{s})\right] \leq 2. \tag{3.4}$$

This expression is essentially the CHSH inequality if each party groups their outcomes m_j into modulo 2 terms. Since 1 mod 2 is equal to 3 mod 2, each party just maps from modulo 4 arithmetic to modulo 2. For all possible correlators in $\mathcal{P}$, the Bell expression in inequality (3.4) achieves the value of 3. This value is achieved for two vertices of $\mathcal{P}$ corresponding to functions $f(\mathbf{s}) = [s_1 s_2 + 1]_2$ or $f(\mathbf{s}) = [s_1 s_2 + 1]_4$. Therefore the one-to-one relationship between inequality and maximal violation from a vertex of $\mathcal{P}$ breaks down for $d = 4$ (and also $d = 5$). This is confirmed by the number of facet Bell inequalities in non-trivial orbits

for $(2,2,4)$ being $216 - 26 = 200$ whereas the number of non-n-partite linear functions is $4^4 - 4^3 = 192$.

The third and final non-trivial orbit for $(2,2,4)$ is generated by the following inequality (expressed in the notation described earlier):

$$\mathcal{C}^2_{d=4} = (1,2,1,1,2,1,1,2,1,-1,-2,-1)\cdot\vec{k} \leq 4. \tag{3.5}$$

It is worth noting that this can be constructed by adding $\sum_{\mathbf{s}} 2(-1)^{s_1 s_2} p(2|\mathbf{s})$ to the left-hand-side of the previous inequality (3.4). The maximal value of 6 of the left-hand-side (i.e. Bell expression) results from the vertex of $\mathcal{P}$ corresponding to the function $f(\mathbf{s}) = [2s_1 s_2 + 2]_4$.

For $(2,2,5)$, there are 4 non-trivial orbits. One of these is generated by the CGLMP inequality and the other three are given by

$$\begin{aligned}
\mathcal{I}_1 &= \frac{1}{2}(6,2,3,4,4,-2,2,1,4,-2,2,1,-4,2,-2,-1)\cdot\vec{k} \leq 5,\\
\mathcal{I}_2 &= (3,1,-1,-3,2,-1,-4,-2,2,-1,-4,-2,-2,1,4,2)\cdot\vec{k} \leq 5,\\
\mathcal{I}_3 &= (2,-1,1,-2,3,1,-1,2,3,1,-1,2,-3,-1,1,-2)\cdot\vec{k} \leq 5.
\end{aligned} \tag{3.6}$$

The inequality for the Bell expression $\mathcal{I}_1$ and the CGLMP inequality are maximally violated by the vertex corresponding to $f(\mathbf{s}) = [s_1 s_2 + 1]_5$. The Bell expressions $\mathcal{I}_2$ and $\mathcal{I}_3$ are maximally violated by the vertex corresponding to $f(\mathbf{s}) = [2s_1 s_2 + 1]_5$. As we can seen there is a corresponding function for each of these inequalities that leads to a maximal violation.

We now consider scenarios with $c > 2$ but with $d = 2$. As can be seen from Table 3.2 for the $(2,3,2)$ scenario there is only one non-trivial orbit. The Bell inequality generating this orbit is another generalisation of the CHSH inequality:

$$\mathcal{C}_{c=3} = \sum_{\mathbf{s}} (-1)^{s_1 s_2} \prod_{j=1}^{2} (\delta_0^{s_j} + \delta_1^{s_j}) p(1|\mathbf{s}) \leq 2. \tag{3.7}$$

For the $(2,4,2)$ scenario, three of these non-trivial orbits are forms of the CHSH inequality embedded in the larger number of inputs. For completeness, we have listed all 14 Bell inequalities in Table 3.3. We now explicitly write out one of these inequalities:

$$\mathcal{C}^1_{c=4} = \sum_{\mathbf{s}} (-1)^{s_1 s_2} \prod_{j=1}^{2} (\delta_0^{s_j} + \delta_1^{s_j}) p(1|\mathbf{s}) \leq 2. \tag{3.8}$$

	$\vec{b}$															
$\mathcal{B}_1$	2	2	1	1	2	-1	-1	-2	1	-1	-2	2	1	-2	2	1
$\mathcal{B}_2$	2	2	1	1	2	-1	-1	-2	1	-2	2	1	1	-1	-2	2
$\mathcal{B}_3$	2	2	1	1	2	-1	-2	-1	1	-2	1	2	1	-1	2	-2
$\mathcal{B}_4$	2	2	1	1	1	-1	2	-2	1	-2	1	2	2	-1	-2	-1
$\mathcal{B}_5$	2	2	1	1	1	-2	2	1	1	-1	-2	2	2	-1	-1	-2
$\mathcal{B}_6$	2	1	1	0	1	-1	-1	1	1	-1	-1	-1	0	1	-1	0
$\mathcal{B}_7$	2	1	1	0	1	-1	-1	1	0	1	-1	0	1	-1	-1	-1
$\mathcal{B}_8$	2	1	1	0	0	1	-1	0	1	-1	-1	1	1	-1	-1	-1
$\mathcal{B}_9$	2	1	0	1	1	-1	1	-1	0	1	0	-1	1	-1	-1	-1
$\mathcal{B}_{10}$	2	1	0	1	0	1	0	-1	1	-1	1	-1	1	-1	-1	-1
$\mathcal{B}_{11}$	2	0	1	1	0	0	1	-1	1	1	-1	-1	1	-1	-1	-1
$\mathcal{C}^1_{c=4}$	1	1	0	0	1	-1	0	0	0	0	0	0	0	0	0	0
$\mathcal{C}^2_{c=4}$	1	1	0	0	0	0	0	0	1	-1	0	0	0	0	0	0
$\mathcal{C}^3_{c=4}$	1	0	1	0	0	0	0	0	1	0	-1	0	0	0	0	0

Table 3.3: The facet Bell inequality expressions that each belong to a particular non-trivial orbit for $(2, 4, 2)$. Each row corresponds to a particular inequality belonging to a different symmetry class. Each column of $\vec{b}$ is an element of this vector that forms an inner product with $\vec{p}$. The LHV upper bound for inequalities $\mathcal{B}_1$ to $\mathcal{B}_5$ is 8 and 4 for $\mathcal{B}_6$ to $\mathcal{B}_{11}$.

which is almost exactly the same as $\mathcal{C}^1_{c=3}$. The other two inequalities, $\mathcal{C}^2_{c=3}$ and $\mathcal{C}^3_{c=3}$ are similar to this inequality except with altered delta functions for $\mathcal{C}^2_{c=3}$ via the substitutions:

$$\prod_{j=1}^{2}(\delta_0^{s_j} + \delta_1^{s_j}) \to (\delta_0^{s_1} + \delta_2^{s_1})(\delta_0^{s_2} + \delta_1^{s_2}), \tag{3.9}$$

and for $\mathcal{C}^3_{c=3}$:

$$\prod_{j=1}^{2}(\delta_0^{s_j} + \delta_1^{s_j}) \to (\delta_0^{s_1} + \delta_2^{s_1})(\delta_0^{s_2} + \delta_2^{s_2}). \tag{3.10}$$

We can see that the CHSH inequality generates a lot of the structure of the LHV polytope in the bipartite scenario. In general though, we have given some insight into the richness of structure of $\mathcal{L}$. This might give some indication why finding the facet Bell inequalities is a complicated task. All of this discussion is even before we consider more than 2 parties. In the following subsection we discuss the $n = 3$ case. Despite not having as many results in this scenario due to the scaling of the size of $\mathbb{R}^{(d-1)c^n}$ in n, we show some of the structure of $\mathcal{L}$ can be

obtained from the $n = 2$ scenario.

3.1.4 Tripartite facet Bell inequalities

We have given an indication that facet Bell inequalities for $n = c = 2$ have a computational interpretation. Every facet Bell inequality we have found is maximally violated uniquely by a vertex of $\mathcal{P}$ when $d = 2$, 3, and 5. In this sense the violation of a facet Bell inequality can quantify how computationally powerful a theory is. For situations with $n > 2$, this becomes more complicated even for $n = 3$ and $c = d = 2$. The Mermin inequality [Mermin1990] which we introduced in the first chapter (see section 1.2) can be expressed as

$$p(1|000) + p(1|011) + p(1|101) - p(1|110) \leq 2, \tag{3.11}$$

and forms a non-trivial orbit [Werner2001]. This inequality is maximally violated by more than one vertex of $\mathcal{P}$. If expressed in terms of expectation values of measurements, it can be generated from the CHSH inequality by a form of substitution [Werner2001]. WW showed that all inequalities for $(n, 2, 2)$ can be generated by this substitution [Werner2001]. We now discuss a possible method of doing this for $(3, 2, 3)$.

Analogously to the Mermin inequality (3.11), we define a CGLMP inequality for three parties using the two party inequality. We have three parties but now we only consider non-zero terms in a Bell inequality when the third party's input is $s_3 = 0$. For LHV correlators $p(k|s_1, s_2, 0)$ the n-partite linear functions that can be achieved are $f(\mathbf{s}) = [\alpha_1 s_1 + \alpha_2 s_2 + \alpha_3]_3$ with $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{Z}_d$: the n-partite linear functions on two variables s_1 and s_2. Since the CGLMP inequality is facet-defining for the region of LHV correlators for two parties, or variables s_1 and s_2, it is facet-defining for this space of the $n = 3$ correlators for $s_3 = 0$. Then we can write the tripartite CGLMP inequality as

$$\begin{aligned} \mathcal{C}'_{\text{CGLMP}} = d \times p(1|0,0,0) - \sum_{\mathbf{s}} (-1)^{s_1+s_2} p(1|s_1, s_2, 0) \\ + \sum_{\mathbf{s}} (-1)^{s_1+s_2} \sum_{k=2}^{d-1} (d-k-1) p(k|s_1, s_2, 0) \quad \leq d. \end{aligned} \tag{3.12}$$

For the case of $(3, 2, 3)$, this tripartite CGLMP inequality is facet-defining and forms an orbit of 324 inequalities. There are 61 other non-trivial orbits for $(3, 2, 3)$. Inequalities from each of these orbits can be found in the supplementary

material in [Hoban2011c]. Interestingly though, the Mermin inequality (3.11) above which can be rewritten as:

$$p(1|000)+p(1|011)+p(1|101)-p(1|110)=\sum_{\mathbf{s}}\delta^{s_3}_{s_1\oplus s_2}(-1)^{s_1s_2}p(1|\mathbf{s})\leq 2, \quad (3.13)$$

does not generalize directly to the $(3,2,3)$ scenario. If we were to naively write the generalisation as

$$\begin{aligned}\mathcal{C}''_{\text{CGLMP}} = d\times p(1|0,0,0) &- \sum_{\mathbf{s}}\delta^{s_3}_{[s_1+s_2]_2}(-1)^{s_1+s_3}p(1|\mathbf{s}) \\ &+\sum_{\mathbf{s}}\delta^{s_3}_{[s_1+s_2]_2}(-1)^{s_1+s_3}\sum_{k=2}^{d-1}(d-k-1)p(k|\mathbf{s}) \quad \leq d, \end{aligned} \quad (3.14)$$

then the right-hand-side is not $d=3$ in the $(3,2,3)$ scenario but $2d-1=5$, the algebraic upper bound for all possible correlators and not just LHV correlators. This upper bound of 5 is achieved by vertices of $\mathcal{P}$ corresponding to the function $f(\mathbf{s})=[s_1s_2+1]_3$. However, if parties produce the n-partite linear function $f(\mathbf{s})=[2s_1+2s_2+s_3+1]_3$ and the only non-zero terms in the above inequality occur when $[s_1+s_2]_2=s_3$, then $f(\mathbf{s})=[2s_1+2s_2+[s_1+s_2]_2+1]_3=[s_1s_2+1]_3$.

Despite the fact that some of the facet Bell inequalities can be obtained from bipartite inequalities, understanding the full structure of $\mathcal{L}$ is a difficult task in general. For example, the straightforward substitution of the CGLMP inequality into expressions for $(3,2,3)$ still leaves a large number of orbits without characterisation. On the other hand, $\mathcal{L}$ in the $(n,2,2)$ scenario is well-understood as a hyperoctahedron [Werner2001, Żukowski2002]. The facet Bell inequalities can be described in terms of Boolean functions where each facet inequality results from each particular Boolean function. In the following subsection we review the insight obtained by Werner and Wolf [Werner2001] as well as Żukowski and Brukner [Żukowski2002].

3.1.5 Multipartite facet inequalities for $(n,2,2)$

So far we have found facet Bell inequalities numerically. The size and hardness of the problem means that as n gets larger, finding the facet inequalities quickly becomes intractable on a desktop PC. Convex polytopes are generalisations of the polyhedra and the geometry of these objects has been studied for thousands of years [Grünbaum2003]. A natural question to ask is whether there are an-

alytical tools in convex geometry that can help us define $\mathcal{L}$ in terms of linear inequalities? This is not immediately obvious in the case of general (n, c, d) but the case of $(n, 2, 2)$ has been amenable to this approach. Werner and Wolf (WW) independently with Żukowski and Brukner (ŻB) have shown that in this specific case, $\mathcal{L}$ is a hyperoctahedron [Werner2001, Żukowski2002].

Out of preference, we follow the WW construction of facet Bell inequalities [Werner2001]. Augmenting this approach we will use a central result from the previous chapter that $\mathcal{L}$ is the convex hull of n-partite linear functions. For the $(n, 2, 2)$ scenario, these n-partite linear functions are the *linear Boolean functions*. The linear Boolean functions are a class of functions that have existed in the study of computer science and propositional logic well before our usage here. For example, linear Boolean functions are generated in error correction such as with the Hamming code [MacWilliams1977]. The following result demonstrates yet another application of the study of linear Boolean functions.

Corollary 1. *The space $\mathcal{L}$ of LHV correlators in the $(n, 2, 2)$ scenario is the convex hull of linear Boolean functions.*

Proof: Since this corollary is a special case of Theorem 2 we just need to show that for the $(n, 2, 2)$ scenario, all the n-partite linear functions are the linear Boolean functions. Linear Boolean functions $f(\mathbf{s})$ for an n-length bit-string $\mathbf{s}$ can be written in terms of the Algebraic Normal Form (ANF) as:

$$f(\mathbf{s}) = \left(\bigoplus_{j=1}^{n} a_j s_j \right) \oplus b, \tag{3.15}$$

where a_j, $b \in \{0, 1\}$. Whereas, an n-partite linear function $g(\mathbf{s})$ in this scenario can be written as

$$g(\mathbf{s}) = \bigoplus_{j=1}^{n} g_j(s_j), \tag{3.16}$$

for single-site map $g_j : \mathbb{Z}_2 \to \mathbb{Z}_2$. Crucially, as a special case, all single-site Boolean functions of this form can be expressed as $g_j(s_j) = b_j \oplus a_j \delta_1^{s_j} = a_j s_j \oplus b_j$ since $\delta_1^{s_j} = s_j$ for a_j, $b_j \in \{0, 1\}$. Therefore, take the sum modulo 2 of all of these maps and setting $b = \bigoplus_{j=1}^{n} b_j$ returns the expression in (3.15). □

The above corollary is a rederivation of the LHV convex polytope that was derived by WW and ŻB [Werner2001, Żukowski2002]. However, this rederivation

is in terms of a language of *computational expressiveness* whereas the original derivation is in the language of expectation values of measurements with outcomes ± 1. The "linearity" (in the Boolean function sense of the word) is not explicit but buried in the mathematical derivation of $\mathcal{L}$. The language of computational expressiveness sheds a new light on an old result and this new perspective will become central to a lot of discussion in this chapter; the next chapter will also have Corollary 1 at its heart.

As mentioned above, both constructions due to WW and ŻB use the expectation values of $\mathbb{E}(\mathbf{s}) = p(0|\mathbf{s}) - p(1|\mathbf{s})$ rather than the correlators themselves. However, due to the "law of the excluded middle" giving $\mathbb{E}(\mathbf{s}) = 1 - 2p(1|\mathbf{s})$, expectation values and correlators are in one-to-one correspondence. For brevity of reproduction of results, we will work in terms of $\mathbb{E}(\mathbf{s})$ and then map back to correlators $p(1|\mathbf{s})$ at a final stage.

Taking on this notation, we construct all Bell inequalities in the $(n, 2, 2)$ scenario in the following way [Werner2001]:

$$\left| \sum_{\mathbf{s}} \beta_{\mathbf{s}} \mathbb{E}(\mathbf{s}) \right| \leq 1, \tag{3.17}$$

such that the real coefficients $\beta_{\mathbf{s}}$ always give 1 for LHV correlators. We are just choosing a normalisation convention without loss of generality. By convexity we only need to consider the extreme points of $\mathcal{L}$ which correspond to the linear Boolean functions. We rewrite these extreme points $\mathbb{E}(\mathbf{s})_E$ in terms of the expectation values, i.e. $\mathbb{E}(\mathbf{s})_E = \sum_k (-1)^k \delta^k_{g(\mathbf{s})} = (-1)^{g(\mathbf{s})}$ where $g(\mathbf{s})$ is a linear Boolean function. We are also only interested in extreme points that maximally saturate the upper bound as these extreme points will define a facet. Putting all of this information together, we can rewrite (3.17) as:

$$\sum_{\mathbf{s}} \beta_{\mathbf{s}} (-1)^{g(\mathbf{s})} = (-1)^{\gamma_{g(\mathbf{s})}}, \tag{3.18}$$

where $\gamma_{g(\mathbf{s})} \in \{0, 1\}$ depends on the linear Boolean function. The linear Boolean functions can be written as $g(\mathbf{s}) = (\bigoplus_{j=1}^{n} a_j s_j) \oplus b$ but the overall sign $(-1)^b$ leaves (3.17) unaffected. Therefore we only need to consider linear Boolean functions with $b = 0$, thus leaving 2^n such functions. In order to show that the inequalities in (3.18) are facet-defining, then we must form affinely independent 2^n-length vectors with elements $(-1)^{g(\mathbf{s})}$ for each $\mathbf{s}$. To demonstrate affine independence we utilise the following lemma.

Lemma 9. *For 2^n-length vectors $\vec{g} \in \mathbb{R}^{2^n}$ with elements $g(\boldsymbol{s})$ being the non-constant linear Boolean functions, a set of $(2^n - 1)$ vectors $\vec{g}$ are linearly independent as long as no two vectors, $\vec{g}^1$ and $\vec{g}^2$, corresponding to two linear Boolean functions $g^1(\boldsymbol{s})$ and $g^2(\boldsymbol{s})$ respectively, have all elements $g^1(\boldsymbol{s}) = g^2(\boldsymbol{s}) \oplus 1$.*

Proof: We demonstrate linear independence by mapping linear Boolean functions from $\mathbb{Z}_2$ to $\mathbb{R}$. Every linear Boolean function can always be expressed as $g(\mathbf{s}) = (\bigoplus_{j=1}^{n} a_j s_j) \oplus b$ and is non-constant as long as at least one value of a_j is non-zero. For two functions $g^1(\mathbf{s}) = g^2(\mathbf{s}) \oplus 1$, b is 1 for either of the functions and 0 for the other. Mapping from $\mathbb{Z}_2$ to $\mathbb{R}$, we can write $g^2(\mathbf{s}) \oplus 1$ as

$$g^2(\mathbf{s}) \oplus 1 = 1 - g^2(\mathbf{s}). \tag{3.19}$$

Therefore if we prove that the $(2^n - 1)$ non-constant linear Boolean functions $g(\mathbf{s}) = (\bigoplus_{j=1}^{n} a_j s_j)$ produce $(2^n - 1)$ linearly independent vectors $\vec{g}$, then this holds if functions are $g(\mathbf{s}) \oplus 1$. This is true if the set of linear Boolean functions does not include two functions $g^1(\mathbf{s})$ and $g^2(\mathbf{s})$ where $g^2(\mathbf{s}) = g^1(\mathbf{s}) \oplus 1$ for all $\mathbf{s}$.

First, all variables s_j will produce vectors $\vec{g}$ that are linearly independent from all $\vec{g}$ resulting from s_k by construction where $k \neq j$. As a shorthand, we say that a function is linearly independent from other functions if the associated vectors $\vec{g}$ are linearly independent. We show that linear Boolean functions dependent on more than one variable s_j are linearly independent. We start with the linear function $s_1 \oplus s_2$ which can be rewritten as

$$s_1 \oplus s_2 = s_1 + s_2 - 2s_1 s_2. \tag{3.20}$$

This expression is linearly independent from functions s_1 and s_2 due to the $s_1 s_2$ term being multiplicative. This function will also be linearly independent from all functions $s_j \oplus s_k \neq s_1 \oplus s_2$ due to $s_j s_k$ being linearly independent from $s_1 s_2$. Having shown that all linear Boolean functions dependent on 2 variables and 1 variable are all linearly independent from each other, we proceed inductively. For functions dependent on 3 variables s_j, eg. $g(\mathbf{s}) = s_1 \oplus s_2 \oplus s_3$, we can again map this function into standard arithmetic as

$$s_1 \oplus s_2 \oplus s_3 = s_1 + s_2 + s_3 - 2(s_1 s_2 + s_1 s_3 + s_2 s_3) + 4 s_1 s_2 s_3. \tag{3.21}$$

Again $s_1 s_2 s_3$ is linearly independent from all terms $s_j s_k$ for $j \neq k$ and single variable terms s_j, as well as all linear functions $s_j \oplus s_k \oplus s_l \neq s_1 \oplus s_2 \oplus s_3$.

Proceeding inductively for each function $g(\mathbf{s}) = (\bigoplus_{j=1}^{n} a_j s_j)$ with q non-zero values of a_j, writing $g(\mathbf{s})$ in standard arithmetic we have the product of these q elements of $\mathbf{s}$. This product of q elements of $\mathbf{s}$ is linearly independent from all other products of q and $q' < q$ elements of $\mathbf{s}$. The linear Boolean function $g(\mathbf{s}) = (\bigoplus_{j=1}^{n} s_j)$ can thus be written as

$$\bigoplus_{j=1}^{n} s_j = \frac{1}{2}\left[1 - \prod_{j}^{n}(1 - 2s_j)\right]. \tag{3.22}$$

This function is finally linearly independent from all other linear Boolean functions due to the term $\prod_{j=1}^{n} s_j$. Therefore all of the non-constant linear Boolean functions $g(\mathbf{s}) = \bigoplus_{j=1}^{n} a_j s_j$ produce vectors $\vec{g}$ that are linearly independent. □

As a result of this lemma, the extreme points $(-1)^{g(\mathbf{s})}$ for the linear Boolean functions $g(\mathbf{s}) = \bigoplus_{j=1}^{n} a_j s_j$ are affinely independent. The dimension of $\mathcal{P}$ for $(n, 2, 2)$ is 2^n, and so (3.18) is facet-defining if this expression is satisfied for all of these linear Boolean functions.

The key observation made by WW is that (3.18) is a discrete Fourier Transform and its inverse is

$$\beta_{\mathbf{s}} = \frac{1}{2^n}\sum_{g(\mathbf{s})}(-1)^{\gamma_{g(\mathbf{s})}}(-1)^{g(\mathbf{s})} \tag{3.23}$$

which is now a sum over all linear Boolean functions $g(\mathbf{s}) = (\bigoplus_{j=1}^{n} a_j s_j)$. Therefore, for each facet Bell inequality we now have some choice of the variables $\gamma_{g(\mathbf{s})} \in \{0, 1\}$ for all functions $g(\mathbf{s})$. There are then 2^{2^n} possible choices of these 2^n values of $\gamma_{g(\mathbf{s})}$. We now express (3.18) in terms of correlators $p(1|\mathbf{s})$ instead of expectation values $\mathbb{E}(\mathbf{s})$,

$$-\sum_{\mathbf{s}} \beta_{\mathbf{s}} p(1|\mathbf{s}) \leq \frac{1 - \sum_{\mathbf{s}} \beta_{\mathbf{s}}}{2} \in \{0, 1\}. \tag{3.24}$$

The sum of coefficients $\sum_{\mathbf{s}} \beta_{\mathbf{s}}$ is equal to ± 1 as it is equal to $(-1)^{\gamma_{g(\mathbf{s})}}$ when $g(\mathbf{s}) = 0$ for all $\mathbf{s}$. There are therefore 2^{2^n} facet Bell inequalities in the $(n, 2, 2)$ scenario of the form in (3.24).

In the $(n, 2, 2)$ scenario, we show that if we deal with expectation values we can derive all of the facet inequalities. As mentioned above, all of these inequalities can be obtained through substitution of the CHSH inequality in terms of expectation values [Werner2001]. The CHSH inequalities are expressed as a poly-

nomial in measurement operators on two sites, called "Bell polynomials". Every other inequality for $n > 2$ are multiples of these polynomials with measurement operators on other sites. This substitution of the CHSH inequality is clear in the expectation value scenario but not so clear in the correlator description. Despite this drawback, the insight we gain from Lemma 9 allows us to demonstrate that a particular inequality for each n is facet-defining as we now show.

We have utilised a form of substitution in constructing tripartite CGLMP inequalities by having non-zero terms in the inequality when the input satisfies a particular constraint, e.g. $s_1 = s_2$. But not all inequalities in the $(n, 2, 2)$ scenario can be constructed from the CHSH inequality by this simple method. For example, the following facet Bell inequality in the $(3, 2, 2)$ scenario as found by WW [Werner2001],

$$\begin{aligned} &\frac{1}{4}\left[p(1|000) + p(1|001) + p(1|010) + p(1|011)\right] \\ +&\frac{1}{4}\left[p(1|100) + p(1|101) + p(1|110) - 3p(1|111)\right] \quad \leq \quad 1 \end{aligned} \tag{3.25}$$

has non-zero coefficients for all inputs $\mathbf{s}$. However, we can generalise this inequality to n parties utilising the result from Lemma 1 (in a slightly modified form). A generalisation of this inequality is

$$\frac{1}{2^{n-1}}\left(-2^{n-1}p(1|\mathbf{1}) + \sum_{\mathbf{s}} p(1|\mathbf{s})\right) \leq 1. \tag{3.26}$$

It is worth noting that this inequality not only reduces to (3.25) for $n = 3$, but also the CHSH inequality for $n = 2$.

We observe that the upper bound on the right-hand-side of (3.26) is saturated for all $(2^n - 1)$ linear Boolean functions $g(\mathbf{s}) = \left(\bigoplus_{j=1}^{n} a_j s_j\right) \oplus b \neq 0$ where $g(\mathbf{1}) = 0$. The upper bound is also saturated when $g(\mathbf{s}) = 1$ for all $\mathbf{s}$. These 2^n linear Boolean functions are also affinely independent by the argument of Lemma 9. For a particular linear Boolean function $g^1(\mathbf{s})$, only one out of the two functions $g^1(\mathbf{s})$ and $g^1(\mathbf{s}) \oplus 1$ satisfy the condition that $g(\mathbf{1}) = 0$. Therefore the set of $(2^n - 1)$ linear Boolean functions $g(\mathbf{s}) = \left(\bigoplus_{j=1}^{n} a_j s_j\right)$ with some of these functions having 1 added mod 2, will be the set $g(\mathbf{s}) = \left(\bigoplus_{j=1}^{n} a_j s_j\right) \oplus b \neq 0$ where $g(\mathbf{1}) = 0$. By Lemma 9, the former set forms a linearly independent set of functions, and we can just add the constant function $g(\mathbf{s}) = 1$ for all $\mathbf{s}$ to make an affinely independent set.

We have used an insight from the computational perspective of LHV correlators in the $(n, 2, 2)$ scenario to define a facet Bell inequality for all n. Interestingly, for all correlators in $\mathcal{P}$, the inequality in (3.26) is *only* maximally violated by the correlator $p(1|\mathbf{s}) = \delta_1^{\prod_{j=1}^n s_j \oplus 1} = \prod_{j=1}^n s_j \oplus 1$ corresponding to the function $f(\mathbf{s}) = \prod_{j=1}^n s_j$ for all n. This is contrary to the Mermin inequality which is maximally violated by more than one correlator in $\mathcal{P}$.

So far in the discussion in this chapter, we have described facet Bell inequalities. They define the space of $\mathcal{L}$. They also guarantee that if a correlator is outside of $\mathcal{L}$, it must violate one of these facet Bell inequalities. We have shown throughout that this violation can be achieved (uniquely or otherwise) maximally by particular vertices of $\mathcal{P}$. Heuristically then, a violation of a Bell inequality can be associated with a *computational advantage*. The advantage being that non-LHV correlators can be associated with computations of non-n-partite linear functions. This insight will be utilised in section 3.3 where Bell inequalities may not be facet-defining, which can highlight the computational advantage of non-LHV theories.

Of the possible theories that can be associated with non-LHV correlators, quantum theory is currently the only working theory. Whether the predictions of quantum theory in the form of a violation of a Bell inequality can be verified in a laboratory will be discussed in chapter 4. In the next section, we will discuss quantum correlators, or the space $\mathcal{Q}$. We will explore methods used to find the maximal violations of Bell inequalities possible with quantum theory. This will give some indication of the extreme points of the space $\mathcal{Q}$.

3.2 Quantum Violations of Bell Inequalities

We have described the structure of $\mathcal{L}$ in terms of the facet Bell inequalities. We now give some indication of the structure of $\mathcal{Q}$. By giving an indication, we mean that we find the maximal violation of the facet Bell inequalities. It is still an open question of defining the extreme points of $\mathcal{Q}$ in general. In the specific $(n, 2, 2)$ scenario, WW have described the extreme points of the quantum region [Werner2001], but otherwise, we can only numerically find particular extreme points.

In this section, numerical methods [Navascués2008, Kaszlikowski2000] used to find the maximum quantum values of a Bell expression are reviewed. Using these methods we present numerical values for the bipartite facet Bell expressions we

found in subsection 3.1.3. In particular, we find the maximum quantum values for an expression in each orbit. Therefore, finding this value for an expression in an orbit also finds the quantum value for all expressions in that orbit. This is because the set of quantum correlators are also unaffected by the local operations on values **m** and **s** and permutations of parties.

We also comment on the relationship between entanglement and violation of bipartite facet Bell inequalities. We show that the maximal quantum violation may not be achieved by a maximally entangled quantum state. Although a violation is a "witness" of entanglement (see section 1.4), more entanglement may not mean more non-classicality.

We present the result of WW that all extreme points of $\mathcal{Q}$ have a closed form [Werner2001]. The maximum quantum value of all Bell expressions is an optimization over these points. What is more, these maximal expressions can be obtained from projective measurements on the n-party Greenberger-Horne-Zeilinger (GHZ) state [GHZ1989, Werner2001]. The GHZ state can be considered as a natural, if ambiguous [Plenio2007], multipartite generalization of the maximally entangled state.

3.2.1 Numerical Methods for finding Violations of Bell Inequalities

In the literature, there are two main methods of finding violations of Bell inequalities. The first approach which we call the "multiport beam-splitter" or MBS approach [Kaszlikowski2000, Durt2001]. This method fixes the quantum state shared by both parties as the maximally entangled state $|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{(d-1)} |jj\rangle$. We then optimize over projective measurements made by each party to find a lower bound of the maximum quantum violation of a Bell inequality, if a violation occurs.

A second, more general approach for finding a quantum violation of a Bell inequality involves semi-definite programming (SDP) [Boyd2004]. Therefore we call this approach the "SDP approach" as developed by Navascués, Pironio and Acín [Navascués2007, Navascués2008]. This approach involves constructing a positive semi-definite Gram matrix of (sequences of) correlations. The Bell expression is then a linear function on elements of this matrix and we maximize this linear, or "objective" function. This second approach produces an upper bound on the violation of a Bell inequality. However, if the Gram matrix satisfies a certain property (called a rank loop) then the maximized objective is equal to the maximal violation of a Bell inequality [Navascués2007]. On the other hand,

if we do not satisfy this property if the lower bound produced by the MBS approach is equal to the upper bound of the SDP approach then we have found the maximum quantum violation.

Both of these approaches have been developed in the bipartite scenario but can be extended to the multipartite scenario [Navascués2007, Żukowski1999]. Naturally though, with an increasing number of parties, the optimization for both approaches becomes harder for a desktop PC. In this subsection, we only use these two methods for finding bipartite quantum violations, so we will only describe them in these two scenarios. We now proceed to describe each approach in more detail.

The MBS approach is described as follows [Kaszlikowski2000, Durt2001]. The quantum state shared by two parties is first fixed as the d^2-dimensional maximally entangled state $|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |jj\rangle$ and both parties attain measurement outcomes associated with projectors $|\mu_j\rangle_{s_j}\langle\mu_j|_{s_j} = V_{s_j}|k\rangle\langle k|V_{s_j}^\dagger$, where $\{|k\rangle | k \in \mathbb{Z}_d\}$ is the standard basis of $\mathcal{H}^D$. The V_{s_j} is a unitary matrix and can be written as $V_{s_j} = FD_{s_j}$ where F is the d-by-d Quantum Fourier Transform matrix with elements for the jth row and kth column $F_{j,k} = \frac{1}{\sqrt{d}} e^{\frac{2\pi i}{d}(j-1)(k-1)}$. The d-by-d matrix D_{s_j} is a diagonal matrix $D_{s_j} = \text{diag}(e^{i\phi_1(s_j)}, e^{i\phi_2(s_j)}, ..., e^{i\phi_d(s_j)})$ with $\phi_j(s_j)$ as real phases. Therefore we optimise over these phases $\phi_j(s_j)$ to numerically maximize the quantum violation for the maximally entangled state.

This first approach can be modified further by altering the quantum state after optimization of the phases $\phi_j(s_j)$, as indicated by Acin et al [Acín2002]. We first obtain the optimal angles $\phi_j(s_j)$ found for the maximally entangled quantum state. We then substitute these optimal angles into the projectors $V_{s_j}|k\rangle\langle k|V_{s_j}^\dagger$. Then we construct the Bell expression in terms of these optimal projectors giving

$$\begin{aligned} \sum_{\mathbf{s}} \beta_{\mathbf{s}} p(k|\mathbf{s}) &= \langle\psi| \left(\sum_{\mathbf{s},\mathbf{m}} \beta_{\mathbf{s}} \delta^k_{[m_1+m_2]_d} |\mu_1\rangle_{s_1}\langle\mu_1|_{s_1} \otimes |\mu_2\rangle_{s_2}\langle\mu_2|_{s_2} \right) |\psi\rangle \\ &= \langle\psi|\mathcal{W}|\psi\rangle \end{aligned} \tag{3.27}$$

where $|\psi\rangle$ is not necessarily the maximally entangled state $|\Psi\rangle$. Finding the largest possible quantum value of the Bell expression is then a case of finding the largest eigenvalue of $\mathcal{W}$. Acin et al used this method to find a larger quantum violation of the CGLMP inequality for $(2, 2, 3)$ with a non-maximally entangled state [Acín2002]. We will discuss the connection between entanglement and Bell inequality violation in subsection 3.2.2.

We now briefly present the SDP approach. Central to the SDP approach is the construction of a positive semi-definite Gram matrix Γ. The elements Γ_{jk} of this matrix are $\Gamma_{jk} = \langle\psi|O_j^\dagger O_k|\psi\rangle$ where O_j is a linear combination of products of projectors E_{m_j,s_j} that depend on m_j and s_j at each jth site. These projectors correspond directly to probabilities of getting m_j given s_j, i.e. $p(m_j|s_j) = \langle E_{m_j,s_j}\rangle$. The projectors act on an arbitrary dimension Hilbert space which is shared by all parties. They also satisfy $\langle\psi|E_{m_j,s_j}|\psi\rangle \geq 0$ for all states $|\psi\rangle$, $E_{m_j,s_j} = E^\dagger_{m_j,s_j}$ (Hermiticity), $E_{m_j,s_j}E_{m'_j,s_j} = \mathbb{I}\delta^{m_j}_{m'_j}$ (orthogonality) and $\sum_{m_j} E_{m_j,s_j} = \mathbb{I}$.

We associate the degree of this product (i.e. the number of terms in the product of projectors) with a set of quantum operators, i.e. for degree of products being ν we have the set $\mathbb{Q}_\nu$. For example, the set $\mathbb{Q}_1$ can be associated with the identity matrix $\mathbb{I}$ and O_j consisting solely of linear combinations of single projectors E_{m_j,s_j}. $\mathbb{Q}_2$ is the set of 2-term products $E_{m_j,s_j}E_{m'_j,s'_j}$ for $s'_j \neq s_j$ and $E_{m_j,s_j}E_{m'_j,s_{j'}}$ for $j \neq j'$. Another set of interest is $\mathbb{Q}'_2$, an intermediate set[5] between $\mathbb{Q}_1$ and $\mathbb{Q}_2$, where we have all the operators which are the pairwise product of projectors between parties j and j' where $j' \neq j$.

The set $\mathbb{Q}_\infty$ of all products of projectors is then the set of all values of $\langle\psi|O_j^\dagger O_k|\psi\rangle$ possible with quantum mechanics. However, it is possible that Γ_ν, the Gram matrix of operators associated with $\mathbb{Q}_\nu$ may already contain all values $\langle\psi|O_j^\dagger O_k|\psi\rangle$ in $\mathbb{Q}_\infty$. If this occurs then the rank of Γ_ν is equal to the rank of $\Gamma_{\nu-1}$, resulting in a "rank loop". For more detail see [Navascués2008].

To find the quantum upper bound for a Bell expression we perform the following semi-definite program:

$$\begin{aligned} \text{maximize} \quad & \mathrm{tr}(B^{\mathrm{T}}\Gamma) \\ \text{subject to} \quad & \Gamma \geq 0 \\ & \mathrm{tr}(F_j^{\mathrm{T}}\Gamma) = 0,\ j \in \{0,1,...,x\}, \end{aligned} \tag{3.28}$$

where B is a matrix of the coefficients of the Bell expression for each probability $p(\mathbf{m}|\mathbf{s}) = \langle\prod_{j=1}^n E_{m_j,s_j}\rangle$. The matrices F_j are x linear constraints on elements of Γ. Of course, Γ_∞ will be infinitely large, so if we restrict at first to Γ_1, we obtain an upper bound on $\mathrm{tr}(B^{\mathrm{T}}\Gamma)$ for all quantum probabilities. It is an upper bound as there are fewer constraints on the elements of Γ, and so Γ might not be compatible with quantum physics. The bound can then be subsequently lowered

[5] In [Navascués2008], this set is written as $\mathbb{Q}_{1+AB}$ where A and B represent two parties, and the set includes pairwise products of the projectors for each party.

if we consider matrices Γ'_2 (corresponding to the set $\mathbb{Q}'_2$) which will impose more constraints on the products of projectors compatible with quantum physics.

The value of $\mathrm{tr}(B^{\mathrm{T}}\Gamma)$ will be the true quantum value (up to numerical error) if we have a rank loop as described above. Semi-definite programming forms part of the subject of convex optimization [Boyd2004]. There are algorithms for dealing with semi-definite programming such as those in the packages of YALMIP [Yalmip] and SeDuMi [SeDuMi]. We utilise these numerical methods to find quantum bounds of Bell expressions, and also to look for a rank loop. However, we do not need to look for a rank loop if the value of $\mathrm{tr}(B^{\mathrm{T}}\Gamma)$ is equal to the lower bound of the MBS method within numerical error. These two methods then give us an indication of the extreme points of $\mathcal{Q}$.

In the construction of the SDP approach we did not explicitly say that $n = 2$. Indeed this method can be utilised in the multipartite case but in order to have the correlations of n parties in Γ, one needs to go to at least $\mathbb{Q}_{\lceil \frac{n}{2} \rceil}$. The MBS approach can also be generalised to the multipartite scenario but again the problem becomes more complicated. In the following subsection we will utilise both the MBS and SDP approaches to find the maximal quantum violations of all bipartite facet Bell inequalities. Therefore, consideration of multipartite generalisations will not be relevant for our discussion.

3.2.2 Bipartite Quantum Violations and Entanglement

We now describe the maximal quantum violations of facet Bell inequalities for $n = 2$. We used both methods described in the previous subsection first finding a lower bound using the MBS approach and then the SDP approach to confirm that this is the maximal value. We list all of the maximal quantum violations for $n = 2$ facet Bell inequalities numerically in Table 3.4. The numerical error in these values is of the order of $\pm 10^{-9}$ and maximal violations resulting from both the MBS and SDP approaches agree within this error. Also in Table 3.4 we have indicated which maximal violations result from the maximally entangled state $|\Psi\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{(d-1)} |jj\rangle$. For $d \neq 2$, there are instances where maximal violation is not a result of maximal entanglement.

While the construction of Bell inequalities was initially partly motivated by the issue of entanglement, the connection between entanglement and violation is not completely clear. A violation of a Bell inequality indicates that measurements are made on an entangled state, but entanglement does not necessarily result in a violation of a particular inequality [Werner1989]. For the CHSH inequality, the

n	c	d	Orbit	LHV bound	Quantum bound	Entanglement
2	2	2	$\mathcal{C}_{d=2}$	2	$2.4142^{\dagger}$	1.000
2	2	3	$\mathcal{C}_{CGLMP}$	3	3.9149	1.555
2	2	4	$\mathcal{C}_{CGLMP}$	4	5.4594	1.938
2	2	4	$\mathcal{C}^1_{d=4}$	2	2.4142	1.000
2	2	4	$\mathcal{C}^2_{d=4}$	4	$4.8284^{\dagger}$	2.000
2	2	5	$\mathcal{I}_1$	5	6.3145	2.310
2	2	5	$\mathcal{I}_2$	5	7.6290	2.310
2	2	5	$\mathcal{I}_3$	5	7.0314	2.230
2	2	5	$\mathcal{C}_{CGLMP}$	5	7.0314	2.230
2	3	2	$\mathcal{C}_{c=3}$	2	$2.4142^{\dagger}$	1.000
2	4	2	$\mathcal{B}_1$ to $\mathcal{B}_5$	8	$9.7570^{\dagger}$	1.000
2	4	2	$\mathcal{B}_6$ to $\mathcal{B}_{11}$	4	$5.0825^{\dagger}$	1.000
2	4	2	$\mathcal{C}^1_{c=4}$ to $\mathcal{C}^3_{c=4}$	2	$2.4142^{\dagger}$	1.000

Table 3.4: We list the bipartite maximal quantum violations for particular facet Bell inequalities for c and d. We have grouped the orbits of inequalities $\mathcal{B}_1$ to $\mathcal{B}_5$ (increasing numerically in the label of the inequality as they have the same LHV and quantum upper bounds. The same grouping also applies for inequalities $\mathcal{B}_6$ to $\mathcal{B}_{11}$. Those violations that are achieved with the bipartite maximally entangled state of d^2 dimension are labelled with a †. We also present the numerical calculation of entropy of entanglement for the pure state associated with each maximal violation. Recall that the maximally entangled state will have entanglement $\log_2(d)$.

maximal violation allowed by quantum mechanics is produced by the maximally entangled state [Tsirelson1980]. As we can see from Table 3.4, this is not true in general. Also in Table 3.4, we have calculated the entanglement of the pure state that maximally violates each inequality. The entanglement of bipartite pure states $|\psi\rangle\langle\psi| \in \mathcal{H}^d \otimes \mathcal{H}^d$ in d^2-dimensional Hilbert space is calculated from the entropy of entanglement $E(|\psi\rangle\langle\psi|)$ [Plenio2007]. Interestingly from Table 3.4, the entanglement of the state that maximally violates the CGLMP inequality decreases with d.

It has been established previously that a violation of a Bell inequality and entanglement are two different, but related issues [Vidick2011, Liang2011]. For example, statistics that violate a Bell inequality can be seen as a "resource" for demonstrating non-classicality, and entanglement can also be seen as a resource (see section 1.3). It has been shown that these two resources are different if one wants to use one resource to simulate the statistics of the other [Brunner2005].

In subsection 3.2.1, we mentioned that if one attains a rank loop between a Gram matrix Γ_ν and another Gram matrix $\Gamma_{\nu-1}$, then the quantum value of Bell expression has reached its maximum value for Γ_ν. For all of the examples in Table 3.4, there was a rank loop found between Γ'_2 and Γ_1. This observation is confirmed for the CGLMP inequalities by results obtained by Navascués, Pironio and Acín [Navascués2008]. This leads us to conjecture that the maximal quantum value resulting from $\mathcal{Q}$ for all bipartite Bell expressions for correlators is obtained from correlations in the set $\mathbb{Q}'_2$.

In this subsection we have indicated that all of the bipartite facet Bell inequalities found in this chapter are violated by quantum correlators. However, the maximum possible violation is not achieved by the relevant maximally entangled state. This implies it might not be favourable to use a maximally entangled state for the largest violation. This behaviour has also been observed when considering Bell inequalities expressed in terms of elements of the full probability distribution $p(\mathbf{m}|\mathbf{s})$ [Vidick2011, Liang2011]. In the next subsection, we describe the quantum region $\mathcal{Q}$ for the $(n, 2, 2)$ scenario. The connection between maximal violation and quantum state is also far clearer for all n; it results from the GHZ state [GHZ1989]. The GHZ state for $n = 2$ case is the maximally entangled state for $d = 2$.

3.2.3 Quantum Upper Bounds of $(n, 2, 2)$ Bell Inequalities

We now consider the maximal quantum violation of any Bell inequality in the $(n, 2, 2)$ scenario. We state the following result (as obtained by WW [Werner2001]) in terms of the maximal quantum value of a Bell expression.

Theorem 10. *The maximal quantum value of a Bell expression for the* $(n, 2, 2)$ *scenario is*

$$\sum_{\boldsymbol{s}} \beta_{\boldsymbol{s}} p(1|\boldsymbol{s}) = \underset{\{\theta_j\}}{sup} \left[\left(\sum_{\boldsymbol{s}} \frac{\beta_{\boldsymbol{s}}}{2} \right) + \left| \sum_{\boldsymbol{s}} \frac{\beta_{\boldsymbol{s}}}{2} e^{i(\sum_{j=1}^{n} s_j \theta_j)} \right| \right] \tag{3.29}$$

where θ_j *are* n *angles, or real parameters. These maximal quantum values result from von Neumann measurements on the GHZ state:*

$$|GHZ\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes n} + |1\rangle^{\otimes n} \right). \tag{3.30}$$

Proof: We map from the correlators $p(1|\mathbf{s})$ to the expectation values $\mathbb{E}(\mathbf{s})$ for measurements, or observables having outcomes ± 1. For quantum correlators, the measurements are Hermitian operators $\hat{M}_{s_j} = Q^0_{s_j} - Q^1_{s_j}$ where $Q^{m_j}_{s_j}$ are the projectors corresponding to outcome m_j. Therefore, $-\mathbb{I} \leq \hat{M}_{s_j} \leq \mathbb{I}$ and $\hat{M}^2_{s_j} = \mathbb{I}$ where $\mathbb{I}$ is the identity matrix. The expectation value is then for all pure states $|\psi\rangle$, $\mathbb{E}(\mathbf{s}) = \langle\psi| \bigotimes_{j=1}^{n} \hat{M}_{s_j} |\psi\rangle$, which can be substituted into a Bell expression to achieve the maximal quantum value

$$\sum_{\mathbf{s}} \beta_{\mathbf{s}} p(1|\mathbf{s}) = \sup_{\{|\psi\rangle, \hat{M}_{s_j}\}} \frac{1}{2} \left[\sum_{\mathbf{s}} \beta_{\mathbf{s}} - \sum_{\mathbf{s}} \beta_{\mathbf{s}} \langle\psi| \bigotimes_{j=1}^{n} \hat{M}_{s_j} |\psi\rangle \right]. \tag{3.31}$$

It remains then to minimize the expression $\sum_{\mathbf{s}} \beta_{\mathbf{s}} \langle\psi| \bigotimes_{j=1}^{n} \hat{M}_{s_j} |\psi\rangle$ over all states and choice of measurements. This equates to finding the minimum eigenvalue of the operator $\sum_{\mathbf{s}} \beta_{\mathbf{s}} \bigotimes_{j=1}^{n} \hat{M}_{s_j}$, or the operator norm $\|...\|$ of $-\sum_{\mathbf{s}} \beta_{\mathbf{s}} \bigotimes_{j=1}^{n} \hat{M}_{s_j}$. To find this operator norm, we need to diagonalize the operator and we can do

this in the following way since the identity commutes with all operators:

$$\begin{aligned}\sup_{\{|\psi\rangle,\hat{M}_{s_j}\}}\left[-\sum_{\mathbf{s}}\beta_{\mathbf{s}}\langle\psi|\bigotimes_{j=1}^{n}\hat{M}_{s_j}|\psi\rangle\right] &= \|-\bigotimes_{j=1}^{n}\hat{M}_0\|\|\sum_{\mathbf{s}}\beta_{\mathbf{s}}\bigotimes_{j=1}^{n}(\hat{M}_0\hat{M}_1)^{s_j}\| \\ &= \|\sum_{\mathbf{s}}\beta_{\mathbf{s}}\bigotimes_{j=1}^{n}(U_{s_j})^{s_j}\| \\ &= \sup_{\{\theta_j\}}\left|\sum_{\mathbf{s}}\beta_{\mathbf{s}}e^{i(\sum_{j=1}^{n}s_j\theta_j)}\right| \end{aligned} \tag{3.32}$$

We obtain this sum of complex terms as $\hat{M}_0\hat{M}_1 = U_j$ is a unitary matrix as $(\hat{M}_0\hat{M}_1)^\dagger = \hat{M}_1\hat{M}_0$ and $(\hat{M}_0\hat{M}_1)^\dagger\hat{M}_0\hat{M}_1 = \mathbb{I}$. The last line is then the norm of a linear combination of unitary matrices.

We now show that the value of (3.32) is attained by observables $\hat{M}_{s_j}$ on the GHZ state $|\text{GHZ}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$. We prove this by construction where each party's measurement is

$$\hat{M}_{s_j} = e^{i(\phi+s_j\theta_j)}|1\rangle\langle 0| + e^{-i(\phi+s_j\theta_j)}|0\rangle\langle 1|. \tag{3.33}$$

With these measurements, we obtain the following expectation values

$$-\sum_{\mathbf{s}}\beta_{\mathbf{s}}\langle GHZ|\bigotimes_{j=1}^{n}\hat{M}_{s_j}|GHZ\rangle = -\sum_{\mathbf{s}}\beta_{\mathbf{s}}\cos\left(n\phi+\sum_{j=1}^{n}s_j\theta_j\right). \tag{3.34}$$

We can write the expression over which we take the supremum in (3.32) as

$$\left|\sum_{\mathbf{s}}\beta_{\mathbf{s}}e^{i(\sum_{j=1}^{n}s_j\theta_j)}\right| = \sum_{\mathbf{s}}\beta_{\mathbf{s}}\text{Re}\left(e^{i(\psi+\sum_{j=1}^{n}s_j\theta_j)}\right) = \sum_{\mathbf{s}}\beta_{\mathbf{s}}\cos\left(\psi+\sum_{j=1}^{n}s_j\theta_j\right). \tag{3.35}$$

We choose $\phi = \frac{\psi+\pi}{n}$ and so the optimal values of θ_j in (3.32) can be substituted into the measurement in (3.33). Therefore, these measurements on a GHZ state attain the maximum quantum upper bound of a Bell expression. □

A corollary of this theorem is that since the quantum correlators

$$p(1|\mathbf{s}) = 1 - 2\cos\left(\psi+\sum_{j=1}^{n}s_j\theta_j\right) \tag{3.36}$$

can be optimized to maximally violate a Bell inequality, these correlators are extreme points of $\mathcal{Q}$ for all ψ, θ_j. The space $\mathcal{Q}$ must contain every one of these extreme points, and so is the convex hull of these correlators [Werner2001].

We now illustrate how the above theorem can be used to find the maximal quantum violation for the CHSH and Mermin inequality respectively. The phase values $\{\psi, \theta_j | j \in \{1, ..., n\}\}$ for the CHSH inequality are $\psi = -\frac{\pi}{4}$ and $\theta_1 = \theta_2 = \frac{\pi}{2}$. Substituting this into (3.32), we obtain Tsirelson's bound $1 + \sqrt{2}$ [Tsirelson1980]. For the Mermin inequality (3.13), $\psi = 0$ and $\theta_1 = \theta_2 = -\theta_3 = -\frac{\pi}{2}$ we have the maximal quantum (and algebraic) upper bound of 3. The quantum violations for facet Bell inequalities in the $(3, 2, 2)$ and $(4, 2, 2)$ cases are listed in [Werner2001].

In this section, we focussed on the quantum violation of facet Bell inequalities in various (n, c, d) scenarios. However, all of the methods described so far apply to any Bell inequality, facet-defining or otherwise. We have used the facet Bell inequalities to show that in all of the scenarios investigated, $\mathcal{Q}$ is strictly larger than $\mathcal{L}$. The facet Bell inequalities are associated with their own difficulty; we have only shown that $\mathcal{Q}$ is larger than $\mathcal{L}$ for a small number of scenarios where we could actually find the facet Bell inequalities. On the other hand, if we suspend the necessity for the facet-defining condition and demonstrate a violation of an arbitrary Bell inequality, then $\mathcal{Q}$ is still strictly larger than $\mathcal{L}$. In the next section, we will consider Bell inequalities that are not facet-defining and show that they are of importance for considering quantum correlations. These inequalities are also of relevance when considering information processing tasks.

3.3 Non-trivial Bell Inequalities

Bell inequalities were first constructed in order to show that the statistics resulting from LHV theories [Bell1964, CHSH1969] are constrained; this constraint does then not apply to quantum theory. The facet Bell inequalities go further and not only constrain LHV statistics but also *define* the space of LHV correlators. We have indicated that to find these region-defining inequalities is a difficult task. However, if we just want to find Bell inequalities that distinguish between LHV and non-LHV correlators, satisfying the facet-defining condition is not necessary. We say that Bell inequalities are "non-trivial" if there are correlators in $\mathcal{P}$ that

violate it, i.e.

$$\sum_{\mathbf{s}} \sum_{k=1}^{(d-1)} \beta_{k,\mathbf{s}} p(k|\mathbf{s}) \leq \gamma_{\mathcal{L}} < \gamma_{\mathcal{P}}, \tag{3.37}$$

with $\beta_{k,\mathbf{s}}$ as real pre-factors and $\gamma_{\mathcal{L}}$ as the upper bound resulting from all correlators in $\mathcal{L}$; $\gamma_{\mathcal{P}}$ is the upper bound of the inequality for all possible correlators in $\mathcal{P}$. As indicated above, for a non-trivial Bell inequality, there is the strict separation $\gamma_{\mathcal{L}} < \gamma_{\mathcal{P}}$.

We describe an explicit set of Bell inequalities that are non-trivial. We then employ a connection between these inequalities and an information processing task called a "non-local game" [Cleve2004] to derive an infinite number of non-trivial Bell inequalities. We begin our discussion in the simplest scenario by discussing the CHSH inequality and utilise its "computational nature" [vanDam2000]. We show that the intuition of the CHSH inequality as measuring the ability to perform a non-linear Boolean function with classical correlations can be applied to all scenarios. Again, central to our discussion is the computational perspective of LHV correlators. We utilise the limited computational expressiveness of LHV theories to derive consequences of this limitation.

3.3.1 Non-trivial Inequalities as Generalisations of the CHSH Inequality

When the CHSH inequality [CHSH1969] was originally derived, the characterisation of correlations in terms of convex polytopes had not yet been considered. It may be considered a happy coincidence that this inequality is facet-defining for the LHV polytope. Despite being placed in the context of convex polytopes, the CHSH inequality has been redefined in the context of non-local games [Cleve2004] as we shall discuss in the next subsection. Such a versatile inequality also has a computational perspective that helps understand why it puts a restriction on LHV correlators [vanDam2000]. We will exploit this perspective to derive a generalisation of the CHSH inequality for all (n, c, d) scenarios.

In order to describe this computational perspective we again write out the CHSH inequality

$$p(1|00) + p(1|01) + p(1|10) - p(1|11) \leq 2, \tag{3.38}$$

and make the substitution $p(1|11) = 1 - p(0|11)$, to obtain

$$\begin{aligned} p(1|00) + p(1|01) + p(1|10) + p(0|11) &= \\ \sum_{\mathbf{s}} \sum_{k=0}^{1} \delta^k_{s_1 s_2 \oplus 1} p(k|\mathbf{s}) &\leq 3. \end{aligned} \tag{3.39}$$

LHV correlators $p(k|\mathbf{s})$ are contained in the convex hull of linear Boolean functions $g(\mathbf{s})$ on $\mathbf{s}$. So, $p(k|\mathbf{s}) = \sum_{g(\mathbf{s})} p_{g(\mathbf{s})} \delta^k_{g(\mathbf{s})}$ with $p_{g(\mathbf{s})} \geq 0$ and $\sum_{g(\mathbf{s})} p_{g(\mathbf{s})} = 1$. Then, by convexity, the following expression must be satisfied for all linear Boolean functions $g(\mathbf{s})$ in the $(2, 2, 2)$ scenario:

$$\begin{aligned} \sum_{\mathbf{s}} \sum_{k=0}^{1} \delta^k_{s_1 s_2 \oplus 1} \delta^k_{g(\mathbf{s})} &= \\ \sum_{\mathbf{s}} \delta^{g(\mathbf{s})}_{s_1 s_2 \oplus 1} &\leq 3. \end{aligned} \tag{3.40}$$

By listing all possible functions $g(\mathbf{s})$ and seeing when they overlap with $s_1 s_2 \oplus 1$, we see that the maximum overlap is 3. We can then rewrite the original CHSH inequality in terms of correlators $p(1|\mathbf{s})$ and this derivation of the LHV upper bound $\gamma_{\mathcal{L}}$

$$p(1|00) + p(1|01) + p(1|10) - p(1|11) \leq \max_{g(\mathbf{s})} \left[\sum_{\mathbf{s}} \delta^{s_1 s_2 \oplus 1}_{g(\mathbf{s})} \right] - 1. \tag{3.41}$$

Essentially, this inequality "measures" the inability for LHV correlators to achieve the non-n-partite linear function $s_1 s_2 \oplus 1$ deterministically [vanDam2000]. If LHV theories could achieve this function deterministically then $\gamma_{\mathcal{L}} = 3$ as $\sum_{\mathbf{s}} \delta^{g(\mathbf{s})}_{f(\mathbf{s})} = 4$. This is, however, not possible and this is the upper bound $\gamma_{\mathcal{P}}$ for all correlators in $\mathcal{P}$ so $\gamma_{\mathcal{P}} > \gamma_{\mathcal{L}}$.

The CHSH inequality is not the only example of a well-studied Bell inequality that can be written in terms of the overlap between a non-n-partite linear and n-partite linear function. The Svetlichny inequality [Svetlichny1987] as mentioned in chapter 2, section 2.4,

$$\begin{aligned} p(1|000) + p(1|001) + p(1|010) - p(1|011) & \\ + p(1|100) - p(1|101) - p(1|110) - p(1|111) &\leq 2, \end{aligned} \tag{3.42}$$

can be rewritten as

$$\sum_{\mathbf{s}}\sum_{k=0}^{1}\delta^{k}_{s_1s_2\oplus s_1s_3\oplus s_2s_3\oplus 1}p(k|\mathbf{s})\leq 6, \tag{3.43}$$

after making the substitution of $p(1|\mathbf{s}) = 1 - p(0|\mathbf{s})$ when the prefactors in (3.42) are -1; this is the case when $s_1s_2 \oplus s_1s_3 \oplus s_2s_3 \oplus 1 = 0$. Again, by convexity the upper bound of this inequality just results in the maximum overlap $\sum_{\mathbf{s}}\delta^{g(\mathbf{s})}_{s_1s_2\oplus s_1s_3\oplus s_2s_3\oplus 1}$ for all linear Boolean functions $g(\mathbf{s})$ for $(3,2,2)$. The function $f(\mathbf{s}) = s_1s_2 \oplus s_1s_3 \oplus s_2s_3 \oplus 1$ is a non-linear Boolean function and so the overlap $\sum_{\mathbf{s}}\delta^{g(\mathbf{s})}_{f(\mathbf{s})}$ by definition will always be lower than $2^n = 8$. Again, we can rewrite the above Svetlichny inequality as

$$\begin{aligned} & p(1|000) + p(1|001) + p(1|010) - p(1|011) + p(1|100) - p(1|101) \\ & -p(1|110) - p(1|111) \\ \leq\ & \max_{g(\mathbf{s})}\left[\sum_{\mathbf{s}}\delta^{g(\mathbf{s})}_{s_1s_2\oplus s_1s_3\oplus s_2s_3\oplus 1}\right] - 4 = 2. \end{aligned} \tag{3.44}$$

For all possible correlators in $\mathcal{P}$, the upper bound is then 4 thus it is a non-trivial Bell inequality, as expected. However, it is not a facet Bell inequality for the region $\mathcal{L}$, but facet-defining for the Svetlichny region, $\mathcal{S}$. The region $\mathcal{S}$ is a sub-region of $\mathcal{P}$ but larger than $\mathcal{L}$, therefore bounds the region $\mathcal{L}$. Non-trivial Bell inequalities can then provide a useful tool to bound $\mathcal{L}$ away from the whole space $\mathcal{P}$.

The CHSH and Svetlichny inequalities above notably utilise the fact that linear Boolean functions cannot be equal to non-linear Boolean functions for all inputs $\mathbf{s}$. Given that LHV correlators are associated with the former and not the latter, we can write down inequalities of the following form for all scenarios (n, c, d):

$$\sum_{\mathbf{s}}\sum_{k=0}^{(d-1)}\delta^{k}_{f(\mathbf{s})}p(k|\mathbf{s})\leq\max_{g(\mathbf{s})}\sum_{\mathbf{s}}\delta^{f(\mathbf{s})}_{g(\mathbf{s})}, \tag{3.45}$$

for all non-n-partite linear functions $f(\mathbf{s})$ and n-partite linear functions $g(\mathbf{s})$. The above inequality in (3.45) is defined for all correlators $p(k|\mathbf{s})$ and not the normalised set of correlators for $k \in \{1, 2, ..., (d-1)\}$. Therefore in order to describe this inequality in terms of normalised correlators, i.e. vectors in $\mathcal{P}$, we impose the normalisation condition that $1 - \sum_{k=1}^{(d-1)} p(k|\mathbf{s}) = p(0|\mathbf{s})$. The

expression on the left-hand-side of (3.45) becomes

$$\sum_{\mathbf{s}} \left[\delta^0_{f(\mathbf{s})} \left(1 - \sum_{k=1}^{(d-1)} p(k|\mathbf{s}) \right) + \sum_{k=1}^{(d-1)} \delta^k_{f(\mathbf{s})} p(k|\mathbf{s}) \right] \leq \max_{g(\mathbf{s})} \sum_{\mathbf{s}} \delta^{f(\mathbf{s})}_{g(\mathbf{s})}, \tag{3.46}$$

which can be rewritten in a form similar to the CHSH inequality,

$$\sum_{\mathbf{s}} \left[\sum_{k=1}^{(d-1)} \left(\delta^k_{f(\mathbf{s})} - \delta^0_{f(\mathbf{s})} \right) p(k|\mathbf{s}) \right] \leq \max_{g(\mathbf{s})} \sum_{\mathbf{s}} \left(\delta^{f(\mathbf{s})}_{g(\mathbf{s})} - \delta^0_{f(\mathbf{s})} \right). \tag{3.47}$$

The upper bound $\gamma_{\mathcal{L}}$ then is strictly smaller than $\gamma_{\mathcal{P}} = c^n - \sum_{\mathbf{s}} \delta^0_{f(\mathbf{s})}$, so the inequality is non-trivial. As we have already demonstrated, the CHSH inequality and Svetlichny inequality are examples of these non-trivial inequalities. As with the Svetlichny inequality, they are not necessarily facet inequalities for $\mathcal{P}$, but necessarily bound the region $\mathcal{L}$. A non-trivial Bell inequality must also intersect $\mathcal{L}$ at, at least, one vertex otherwise the right-hand-side of the inequality in (3.47) is not tight.

Not only are these inequalities interesting because of their ability to bound $\mathcal{L}$, but they have a role in information processing tasks. The particular task of relevance is a non-local game [Cleve2004]. One can be successful at such a game if they violate a Bell inequality, hence the use of "non-local", as in non-LHV resources. One wants to achieve some task (expressed as a game) with as great a probability as possible. Games in general are of interest in computer science and in fields of applied mathematics such as economics [vonNeumann1944]. In some non-local games such as the "XOR games" [Cleve2004], the Bell inequality can quantify the probability of achieving a task and so have a natural role in these games. We use the language and structure of non-local game to describe an *infinite* number of non-trivial Bell inequalities for *each* scenario (n, c, d).

3.3.2 Non-local Games

We have discussed the operational perspective of Bell tests where we have many parties each with inputs and outputs. Many information processing tasks can be abstracted to a process with an input, and a transformation of the input to produce an output. We now focus on one particular task that has a natural connection to Bell tests, the non-local game (NLG) [Cleve2004]. In this section we discuss the set-up of an NLG and how it is relevant to the discussion of

constructing non-trivial Bell inequalities. In the next section, NLG will again be discussed and made relevant to the subject of MBQC. Therefore, these games are of relevance to a great deal of discussion to both Bell tests and this thesis in particular. They also give an interesting computational perspective on Bell tests that has been of interest to the quantum information science community.

We now describe a particular NLG with n parties, or "players" as they are often called. These n parties do not communicate with each other, and so for all intents and purposes, are space-like separated as in Bell tests. As well as these n parties, there is another party that is not a player, but a "referee". A referee can be seen as the experimenter in a Bell test who calculates the correlators $p(k|\mathbf{s})$. However, one distinct aspect in NLG from Bell tests is in the role of the referee as the person who distributes inputs to the n parties as well as retrieving their outputs. In the format of Bell tests that we have discussed so far, the inputs at each site are generated randomly by the parties themselves, in NLG this is not the case. To summarise, n parties each receive an input from the referee and then generate an output which they send to the referee. The referee finally calculates some function on the outputs; the objective of these parties is to maximize the mean probability (for all inputs) of this function being equal to some desired value. We now specify the particular NLG that is of relevance to our discussion:

1. A referee sends the input digit-string $\mathbf{s} = \mathbb{Z}_c^n$ to the n non-communicating parties. The inputs $\mathbf{s}$ are sent with probability distribution $\pi(\mathbf{s})$ such that $\sum_{\mathbf{s}} \pi(\mathbf{s}) = 1$ and all $\pi(\mathbf{s}) \geq 0$;

2. All n parties generate an output digit-string $\mathbf{m} = \mathbb{Z}_d^n$ which is sent to the referee;

3. The referee calculates the sum modulo d of all outcomes $[\sum_{j=1}^{n} m_j]_d = k$;

4. The goal of the game is for the players to maximise the average success probability of $\left[\sum_{j=1}^{n} m_j\right]_d = f(\mathbf{s})$ for some function $f : \mathbb{Z}_c^n \to \mathbb{Z}_d$.

Examples of these games include the well-studied multi-party "XOR games" where $c = d = 2$ [Cleve2004]. The average success probability $\bar{p}_{f(\mathbf{s})}$ of achieving

$k = \left[\sum_{j=1}^{n} m_j\right]_d = f(\mathbf{s})$ can be written in terms of the correlators $p(k|\mathbf{s})$:

$$\bar{p}_{f(\mathbf{s})} = \sum_{\mathbf{s}} \pi(\mathbf{s}) \sum_{k=0}^{(d-1)} \delta^k_{f(\mathbf{s})} p(k|\mathbf{s}). \tag{3.48}$$

In order to maximize this average success probability, we want to find the optimal correlators $p(k|\mathbf{s})$. We can then distinguish between the maximum average success probability $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})}$ and $\bar{p}^{\mathcal{Q}}_{f(\mathbf{s})}$ resulting from quantum and classical (or LHV) correlators respectively. If $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} < \bar{p}^{\mathcal{Q}}_{f(\mathbf{s})}$, it is optimal to use quantum resources instead of classical resources. Also, (3.48) produces a Bell inequality if the correlators result from LHV theories which is upper bounded by $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})}$; if $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} < \bar{p}^{\mathcal{Q}}_{f(\mathbf{s})}$, we have a violation of this Bell inequality. One way that we can possibly have a separation $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} < \bar{p}^{\mathcal{Q}}_{f(\mathbf{s})}$, is if the function $f(\mathbf{s})$ is a non-n-partite linear function. If $f(\mathbf{s})$ is an n-partite linear function, then $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} = 1$.

To make the connection to Bell inequalities explicit, if $\pi(\mathbf{s}) = \pi(\mathbf{s}') = \frac{1}{c^n}$ for all $\mathbf{s} \neq \mathbf{s}'$, then we obtain a modification of the non-trivial Bell inequalities (3.45) discussed in the previous subsection. We can rewrite the inequality in (3.47) in terms of $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})}$ for $f(\mathbf{s})$ being a non-n-partite linear function:

$$\frac{1}{c^n} \sum_{\mathbf{s}} \left[\sum_{k=1}^{(d-1)} \left(\delta^k_{f(\mathbf{s})} - \delta^0_{f(\mathbf{s})} \right) p(k|\mathbf{s}) \right] \leq \bar{p}^{\mathcal{L}}_{f(\mathbf{s})} - \frac{1}{c^n} \sum_{\mathbf{s}} \delta^0_{f(\mathbf{s})}, \tag{3.49}$$

where

$$\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} = \max_{g(\mathbf{s})} \frac{1}{c^n} \sum_{\mathbf{s}} \delta^{f(\mathbf{s})}_{g(\mathbf{s})} < 1. \tag{3.50}$$

For (3.48), if the correlators are all possible correlators in $\mathcal{P}$, then the maximum average success probability is $\bar{p}^{\mathcal{P}}_{f(\mathbf{s})} = \sum_{\mathbf{s}} \pi(\mathbf{s}) = 1$. Therefore, the fact that $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} < \bar{p}^{\mathcal{P}}_{f(\mathbf{s})}$ indicates that the inequality (3.49) is non-trivial.

In order to establish a non-trivial Bell inequality, we need to find a probability distribution $\pi(\mathbf{s})$ such that $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} < \bar{p}^{\mathcal{P}}_{f(\mathbf{s})}$. In the following result, we describe an infinite number of simple probability distributions such that we can generate a non-trivial Bell inequality.

Proposition 11. *All inequalities of the form*

$$\sum_{s} \pi(s) \left[\sum_{k=1}^{(d-1)} \left(\delta^k_{f(s)} - \delta^0_{f(s)} \right) p(k|s) \right] \leq \bar{p}^{\mathcal{L}}_{f(s)} - \sum_{s} \pi(s) \delta^0_{f(s)}, \tag{3.51}$$

are non-trivial Bell inequalities for all non-zero probabilities $\pi(\boldsymbol{s})$ if $f(\boldsymbol{s})$ is a non-n-partite linear function.

Proof: In order to prove this we just need to show that $\bar{p}^{\mathcal{L}}_{f(\mathbf{s})} < \bar{p}^{\mathcal{P}}_{f(\mathbf{s})} = 1$ for the probability distribution $\pi(\mathbf{s})$ being non-zero for all values of $\mathbf{s}$. That is,

$$\max_{g(\mathbf{s})} \sum_{\mathbf{s}} \pi(\mathbf{s}) \delta^{f(\mathbf{s})}_{g(\mathbf{s})} < 1, \tag{3.52}$$

which is true as $0 < \pi(\mathbf{s}) < 1$ for each $\mathbf{s}$ by definition and $\sum_{\mathbf{s}} \delta^{f(\mathbf{s})}_{g(\mathbf{s})} < 1$. □

For a distribution $\pi(\mathbf{s})$ satisfying $0 < \pi(\mathbf{s}) < 1$ for each $\mathbf{s}$ and $\sum_{\mathbf{s}} \pi(\mathbf{s}) = 1$, we can construct $\left(d^{c^n} - d^{n(c-1)+1}\right)$ non-trivial Bell inequalities: the number of non-n-partite linear functions $f(\mathbf{s})$. We are able to construct an infinite number of non-trivial Bell inequalities parametrized by $\pi(\mathbf{s})$ utilizing a computational perspective on Bell tests. Crucially though, the non-trivial Bell inequalities in (3.52) are not dependent on an NLG construction, they exist outside of NLG. More specifically, the probability distribution $\pi(\mathbf{s})$ is just a positive, non-zero weighting on correlators for a particular $\mathbf{s}$. In the context of NLG, the inequalities in (3.52) are related to the success probability of the game, but outside of this context we still have an infinite number of non-trivial Bell inequalities.

So far our discussion has applied to all possible scenarios (n, c, d) and we have stated general results for all these scenarios. In the discussion in the next subsection we will focus on a particular example of non-trivial Bell inequality in $(n, 2, 2)$ for all n of the form in (3.47). This example is a generalisation of the CHSH inequality to n parties. In contrast to the CHSH inequality, we will show that the upper bound of the quantum correlators is no better than the LHV upper bound.

3.3.3 $(n, 2, 2)$ scenario and the n-partite NAND function

It is natural at this point to question the motivation for finding non-trivial Bell inequalities when facet Bell inequalities are more useful for determining the consequences of LHV correlators. As well as the motivations from computational complexity that finding facet Bell inequalities is hard, the utility of bounding $\mathcal{L}$ and connection to information processing tasks, we have further motivation in the $(n, 2, 2)$ scenario. We show that in this scenario, facet Bell inequalities can be related to the non-trivial Bell inequalities described earlier.

As well as this general discussion, we will give an example of a non-trivial Bell inequality of the form (3.47) for the $(n, 2, 2)$ scenario. This non-trivial Bell inequality for all n is a direct generalisation of the CHSH inequality. Essentially it is associated with a non-n-partite linear function, itself a generalisation of the function $s_1 s_2 \oplus 1$ for n parties. We will use this function also to say something about MBQC in the section 3.4. We will show that for more than 2 parties this non-trivial inequality cannot be violated by quantum correlators.

Firstly, we rewrite the non-trivial Bell inequalities in (3.49) in the specific $(n, 2, 2)$ scenario as

$$\sum_{\mathbf{s}} \pi(\mathbf{s})(-1)^{f(\mathbf{s})+1} p(1|\mathbf{s}) \leq \max_{g(\mathbf{s})} \sum_{\mathbf{s}} \pi(\mathbf{s}) \left(\delta_{g(\mathbf{s})}^{f(\mathbf{s})} - \delta_{f(\mathbf{s})}^{0} \right), \tag{3.53}$$

but now we allow probabilities $0 \leq \pi(\mathbf{s}) \leq 1$. If we allow probabilities $\pi(\mathbf{s}) \in \{0, 1\}$ then we might not have non-trivial Bell inequalities as we can choose probabilities such that $\pi(\mathbf{s}) = 0$ when $f(\mathbf{s}) \neq g(\mathbf{s})$ and non-zero otherwise. In this instance, $\sum_{\mathbf{s}} \pi(\mathbf{s}) \delta_{g(\mathbf{s})}^{f(\mathbf{s})} = 1$, hence $\gamma_{\mathcal{L}} = \gamma_{\mathcal{P}}$ and we do not have a non-trivial Bell inequality. However, for a choice of function $f(\mathbf{s})$ and probability distribution $\pi(\mathbf{s})$ we can construct a facet Bell inequality. Since all prefactors $\beta_{\mathbf{s}}$ of a facet Bell inequality are real, then they can be rewritten as $\beta_{\mathbf{s}} = |\beta_{\mathbf{s}}| \text{sign}(\beta_{\mathbf{s}})$. We now fix values as $\pi(\mathbf{s}) = \frac{|\beta_{\mathbf{s}}|}{\sum_{\mathbf{s}} |\beta_{\mathbf{s}}|}$ and $\text{sign}(\beta_{\mathbf{s}}) = (-1)^{f(\mathbf{s})+1}$, then we multiply both sides of (3.53) with $\sum_{\mathbf{s}} |\beta_{\mathbf{s}}|$ to obtain the inequality of the form in (3.1). If $f(\mathbf{s})$ is an n-partite linear function in (3.53), then we cannot define a non-trivial Bell inequality and so cannot define a non-trivial, facet Bell inequality. Finding the facet Bell inequalities for $(n, 2, 2)$ is then a case of finding a probability distribution $\pi(\mathbf{s})$ where we satisfy the facet-defining condition.

For example, the inequality (3.25):

$$\begin{aligned} &\frac{1}{4} \left[p(1|000) + p(1|001) + p(1|010) + p(1|011) \right] \\ +&\frac{1}{4} \left[p(1|100) + p(1|101) + p(1|110) - 3p(1|111) \right] \quad \leq \quad 1, \end{aligned} \tag{3.54}$$

can be rewritten in the form of (3.53) with $\pi(\mathbf{s}) = \frac{1}{10}$ for all $\mathbf{s} \neq \{1, 1, 1\}$ and $\pi(\mathbf{s}) = \frac{3}{10}$ for $\mathbf{s} = \{1, 1, 1\}$ and $f(\mathbf{s}) = s_1 s_2 s_3 \oplus 1$. With these substitutions (3.54) can be retrieved from (3.53) but now both sides of the inequality in (3.54) are multiplied by $\frac{1}{10}$. Now, $\max_{g(\mathbf{s})} \sum_{\mathbf{s}} \pi(\mathbf{s}) \left(\delta_{g(\mathbf{s})}^{f(\mathbf{s})} - \delta_{f(\mathbf{s})}^{0} \right) = \frac{4}{10}$ so that $\bar{p}_{f(\mathbf{s})}^{\mathcal{L}} = \frac{7}{10}$. The latter result can be seen from the fact that the n-partite linear function

$g(\mathbf{s}) = 1$ overlaps with $f(\mathbf{s})$ for 7 values of $\mathbf{s}$; these are the inputs $\mathbf{s} \neq \mathbf{1}$. We have briefly shown that at least one of the non-trivial Bell inequalities as described in Proposition 12 is also a facet Bell inequality.

We have shown that for a function $f(\mathbf{s})$ we can find a probability distribution $\pi(\mathbf{s})$ where the resulting non-trivial Bell inequality in (3.53) is a facet Bell inequality. We now take a different approach and fix the probability distribution to be $\pi(\mathbf{s}) = \frac{1}{2^n}$ for all inputs $\mathbf{s}$. This can be seen as the probability distribution of the Bell test being an NLG with inputs chosen randomly.

Given this probability distribution, we consider a function $f(\mathbf{s})$ for all n in $(n, 2, 2)$. This function is a natural generalisation of the function $f(\mathbf{s}) = s_1 s_2 \oplus 1$ corresponding to the function defining the CHSH inequality. This function will be discussed later with reference to quantum computing and so we define it now.

Definition 6. The **n-partite NAND function** is $f_2(\mathbf{s}) = \prod_{j=1}^{n} s_j \oplus 1$ acting on bit-string $\mathbf{s} \in \mathbb{Z}_2^n$.

A NAND function is defined on two bits s_1 and s_2 as $f(\mathbf{s}) = s_1 s_2 \oplus 1$. This is exactly the function that we used when describing the CHSH inequality earlier in this chapter. The NAND function is the negation (or NOT) of the AND function $f(\mathbf{s}) = s_1 s_2$ [Papadimitriou1994]. The n-partite NAND function consists of the entire NOT of a number of AND functions between variables in $\mathbf{s}$. What is clear is that it is a non-linear Boolean function due to the multiplication between elements of $\mathbf{s}$. It is also the function describing the facet Bell inequality (3.54) above.

For the n-partite NAND function and uniform probability distribution we obtain a non-trivial Bell inequality of the form

$$\frac{1}{2^n} \sum_{\mathbf{s}} (-1)^{f_2(\mathbf{s})+1} p(1|\mathbf{s}) \leq \frac{2^n - 2}{2^n}. \tag{3.55}$$

The upper bound on the right-hand-side is due to the fact that $\sum_{\mathbf{s}} \delta_{g(\mathbf{s})}^{f_2(\mathbf{s})} = (2^n - 1)$ if $g(\mathbf{s}) = 1$, and all linear Boolean functions $g(\mathbf{s})$ are never always equal to $f_2(\mathbf{s})$. We have shown that this inequality for $n = 3$ is related to the facet Bell inequality (3.54), and this relation extends to all n in (3.26). We now show that this natural generalisation of the CHSH inequality has no quantum violation whatsoever for $n \geq 3$.

Proposition 12. *The non-trivial Bell inequality* (3.55) *for the n-partite NAND function for uniform probability distribution* $\pi(\boldsymbol{s}) = 2^{-n}$ *for all* $\boldsymbol{s}$ *is not violated*

by quantum correlators for $n \geq 3$.

Proof: The quantum upper bound for the inequality (3.55) can be calculated from (3.32) to obtain:

$$\begin{aligned} 2\sum_{\mathbf{s}}(-1)^{f_2(\mathbf{s})+1}p(1|\mathbf{s}) &= \sup_{\{\theta_j\}}\left[\left(\sum_{\mathbf{s}}(-1)^{f_2(\mathbf{s})+1}\right)+\left|\sum_{\mathbf{s}}(-1)^{f_2(\mathbf{s})+1}e^{i(\sum_{j=1}^{n}s_j\theta_j)}\right|\right] \\ &= 2^n-2+\sup_{\{\theta_j\}}\left|\sum_{\mathbf{s}}(-1)^{f_2(\mathbf{s})+1}e^{i(\sum_{j=1}^{n}s_j\theta_j)}\right|. \end{aligned} \tag{3.56}$$

If there is a quantum violation of the inequality in (3.55) then the following relationship must be satisfied:

$$\sup_{\{\theta_j\}}\left|\sum_{\mathbf{s}}(-1)^{f_2(\mathbf{s})+1}e^{i(\sum_{j=1}^{n}s_j\theta_j)}\right| > 2^n-2 \tag{3.57}$$

We may, without loss of generality, restrict θ_j to the range $\theta_j \in (-\pi,\pi)$. We simplify inequality (3.57), using the fact that $(-1)^{f_2(\mathbf{s})+1}=1$ for all bit strings $\mathbf{s}$ except when $\mathbf{s}=\mathbf{1}$, to write

$$\begin{aligned} \sup_{\{\theta_j\}}\left|\sum_{\mathbf{s}}\prod_{k=1}^{n}e^{is_j\theta_j}-2e^{i\sum_k^n\theta_k}\right| &= \\ \sup_{\{\theta_j\}}\left|2^n\prod_{j=1}^{n}\cos\left(\frac{\theta_j}{2}\right)-2e^{i\sum_k^n\theta_k}\right| &> 2^n-2. \end{aligned} \tag{3.58}$$

We now adopt a geometric argument. The goal is to maximize the modulus of a sum of two complex numbers. These numbers may be represented, on the plane, as two sides of a triangle. The first side has length $2^n\prod_{j=1}^{n}\cos(\frac{\theta_j}{2})$, the second is of length 2 and the angle between these sides is $\sum_k^n\frac{\theta_k}{2}$. We complete the proof by showing that when $n>7$, the length of the third side of the triangle can never exceed 2^n-2, and hence (3.58) is never satisfied.

We proceed by assuming the opposite of what we want to prove and demonstrating a contradiction. Via the triangle inequality, for this inequality to be satisfied, the length of the base of the triangle must be greater than 2^n-4, and thus

$$\prod_{j=1}^{n}\cos\left(\frac{\theta_j}{2}\right) > 1-2^{2-n}. \tag{3.59}$$

Since $\theta_j \in (-\pi,\pi)$ all terms in the product are non-negative, hence we can impose

the weaker condition for (3.59) that $\forall\theta_j$, $\cos\left(\frac{\theta_j}{2}\right) > 1 - 2^{2-n}$. This implies that $|\sum_{j=1}^{n} \frac{\theta_j}{2}| < n \arccos(1 - 2^{2-n})$.

Proceeding geometrically, we now use the cosine rule to express the third side of the triangle (representing the modulus in (3.57)), and this expression must satisfy

$$4 + 2^{2n} \prod_{j=1}^{n} \cos^2\left(\frac{\theta_j}{2}\right) - 2^{n+2} \cos\left(\sum_{k=1}^{n} \frac{\theta_k}{2}\right) \prod_{l=1}^{n} \cos\left(\frac{\theta_l}{2}\right) > 4 + 2^2 n - 2^{n+2}, \quad (3.60)$$

to obtain a quantum violation. Since $\theta_j \in (-\pi, \pi)$, $\prod_{l=1}^{n} \cos(\frac{\theta_l}{2})$ is non-negative, and hence a violation can only be achieved if $\cos(\sum_{k=1}^{n} \frac{\theta_k}{2})$ is negative which implies $|\sum_{k=1}^{n} \frac{\theta_k}{2}| > \frac{\pi}{2}$. Using this, we achieve

$$n \arccos(1 - 2^{2-n}) > \pi/2 \quad (3.61)$$

or equivalently

$$\cos\left(\frac{\pi}{2n}\right) > (1 - 2^{2-n}). \quad (3.62)$$

This inequality is only satisfied for integers $n \leq 7$, hence, due to the contradiction with our initial assumption, for $n > 7$ the quantum and classical bounds of the non-trivial Bell inequality (3.55). Direct numerical verification of the bounds, via equation (3.32) for $n < 7$ indicates that the bounds coincide for all integer values $3 \leq n \leq 7$, thus completing the proof. □

This proof demonstrates that, $\mathcal{Q}$ is smaller than $\mathcal{P}$ for this scenario. It also demonstrates that quantum correlators are not always useful in every non-local game. If we modify the probability distribution $\pi(\mathbf{s})$ by weighting the input $\mathbf{s} = \mathbf{1}$ more than other inputs, we can regain a quantum advantage as with the inequality in (3.26).

This example of a non-trivial Bell inequality not being violated by quantum mechanics is not isolated. For example, the following non-trivial Bell inequality for $(2, 3, 3)$ corresponding to the function $f(\mathbf{s}) = \left[s_1^2 s_2^2 + 1\right]_3$ with uniform probability distribution $\pi(\mathbf{s}) = \frac{1}{9}$:

$$\begin{aligned} &\frac{1}{9}\left(p(1|00) + p(1|01) + p(1|02) + p(110) + p(2|11)\right) \\ &\quad + \frac{1}{9}\left(p(2|12) + p(1|20) + p(2|21) + p(2|22)\right) \leq \ \frac{8}{9} \end{aligned} \quad (3.63)$$

is also not violated by quantum correlators $\mathcal{Q}$. This upper bound of $\frac{8}{9}$ was found using both the MBS and SDP approach. Just like the n-partite NAND function, the function $f(\mathbf{s}) = \left[s_1^2 s_2^2 + 1\right]_3$ differs from all possible n-partite linear functions for only one value of $\mathbf{s}$.

We have shown that quantum resources are not always better than classical resources when trying to maximize the mean probability of winning an NLG. This is not new as Linden et al [Linden2007] devised a model of "non-local computation" where quantum resources do no better than classical, or LHV resources. The resulting Bell inequality defining over probabilities $p(\mathbf{m}|\mathbf{s})$ from this model is also not facet-defining. Perhaps more interesting, Almeida et al found an NLG where quantum resources do no better than classical resources [Almeida2010]; and for $n \geq 3$, this game defines a facet Bell inequality of $\mathcal{L}_{\mathbf{F}}$ for probabilities $p(\mathbf{m}|\mathbf{s})$. Investigating the limitations of quantum correlations therefore seems to be just as interesting as finding its advantages.

We will use this function in the next subsection to say something about MBQC [Raussendorf2001, Raussendorf2003]. In particular, we look at a restricted class of computations in MBQC and map this class into the framework of Bell tests. We employ the Bell test as an NLG but in the language of games, there is a "promise" on the inputs [Cleve2004]. That is, the inputs are the result of some pre-processing on a bit-string [Anders2009]. This pre-processing has a well-defined role in MBQC and we use our NLG to show that our restricted class of MBQC is not equivalent to a universal Quantum Computer. The key to all of these insights is the computational perspective of the space $\mathcal{L}$.

3.4 Non-adaptive Measurement-based Quantum Computing

MBQC as formulated by Raussendorf and Briegel [Raussendorf2001] has been one of the great breakthroughs in quantum computing. Whereas the original circuit model of quantum computing requires the ability to perform unitary operators over the length of the computation [Nielsen2000], MBQC reduced this to state preparation and sequential single-site (single-qubit) measurements [Raussendorf2001]. The state that is prepared is a multipartite entangled state, e.g. the "cluster state" [Raussendorf2003]. We immediately see that MBQC is more in the vein of a Bell test, which (for quantum correlators) consists of the preparation of a potentially entangled state and then single-site measurements

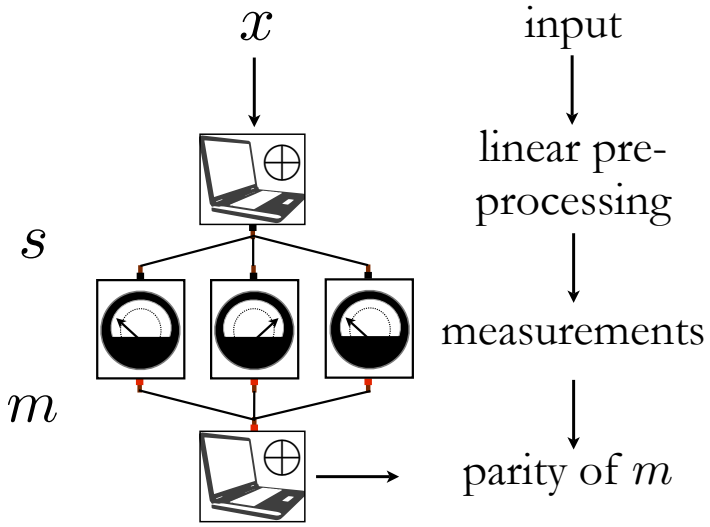

Figure 3.1: Non-adaptive MBQC consists of pre-processing on data, this data is then sent to the parties who make a single round of measurements. The classical control computer as well as performing pre-processing, processes the measurement outcomes. (Copyright: Institute of Physics, 2011).

on each part of this state. In this section, we show that the connection is more concrete than just this superficially shared language.

Briegel and Raussendorf showed that adaptivity is a key component of their formulation of MBQC [Raussendorf2001, Raussendorf2003], in order that all possible quantum circuits are implemented deterministically. A natural question is what happens when we remove adaptivity? If we do not have adaptivity, then all measurements can take place simultaneously. This also simplifies the technological implementation of an MBQC, where a state only needs to be prepared and then measured instantly. Adaptivity means that a state needs to be stored for a non-negligible amount of time between measurement rounds.

We now define the class of computations in MBQC without adaptivity or as we will call it, **nMBQC**. We do not place restrictions on the measurements, state prepared or number of sites. The element which remains the same in Raussendorf and Briegel's formulation of MBQC is the control computer [Raussendorf2003, Anders2009]. The control computer can only implement XOR gates, or addition modulo 2, on bits [Anders2009] (see section 1.3 and 1.4). However, the necessity for particular measurements and states is not as well-defined; the cluster state is

an example of a useful resource [Raussendorf2003], but there are other examples [Hein2005, VandenNest2006]. We now define nMBQC as an abstract model. See Figure 3.1 for an accompanying schematic of nMBQC.

Definition 7. The model of **Non-adaptive Measurement-Based Quantum Computing**, or nMBQC, involves the preparation of an n-partite quantum state $|\psi\rangle$ and a classical control computer $\mathcal{C}$. The computer $\mathcal{C}$ receives a bit-string $\mathbf{x}$ of length $|\mathbf{x}|$ with uniform probability $\frac{1}{2^{|\mathbf{x}|}}$. The control computer performs arbitrary XOR gates on a bit-string $\mathbf{x}$ and communicates the choice of measurement $s_j(\mathbf{x})$ to each jth site. There is a single round of measurements on all n non-communicating sites. The control computer receives the measurement outcomes from each site as bits m_j and computes the parity of $\mathbf{m}$: $\bigoplus_{j=1}^{n} m_j$.

The goal of this model is then to *deterministically* perform some Boolean function $f(\mathbf{x})$ *efficiently* on the original bit-string $\mathbf{x}$ for all $\mathbf{x}$. By efficient, we use the computational complexity convention that the amount of resources, in this case n, is polynomial in the size $|\mathbf{x}|$ of the input $\mathbf{x}$. We can ask what the *worst-case* number of resources, or measurements sites to perform this function so that the function can be performed for all instances of $\mathbf{x}$. In our definition above we add a uniform probability distribution on all inputs. This uniform probability distribution becomes relevant if we cannot perform a function deterministically, but want to maximize the probability of performing a function. The uniformity condition on all instances of $\mathbf{x}$ means there is no bias on any particular bit-string $\mathbf{x}$, since it may be easier to compute a function $f(\mathbf{x})$ for particular instances of $\mathbf{x}$. Some results in the following discussion (such as 13) do not require us to consider a distribution at all but we introduce it to cement a connection to Bell inequalities later on.

We have so far not mentioned any constraint on how our resource is constructed. Perhaps our resource can only be produced using exponential *quantum computations*. We place no constraint and just assume that the resource quantum state is "presented to us" and we make measurements on it. In fact, the optimal resource for nMBQC can be generated efficiently by a quantum computer. We will show that the optimal resource for all computations in nMBQC is the GHZ state, and this state can be generated efficiently [Hein2005].

The hope is that even in this model of nMBQC we might be able to perform (at least) all efficient classical computations (i.e. in the complexity class P [Papadimitriou1994]) efficiently. In this section, we show that this is not possible

and the model is quite limited. Even then, this model can simulate the statistics of Clifford circuits [Raussendorf2003, Jozsa2006], which are not believed even to be universal for classical computing [Aaronson2004] even though they produce entanglement. Computations in nMBQC are also in a recently studied class of limited quantum computations called "Instantaneous Quantum Polytime" (IQP) [Shepherd2009]. It is possible that IQP is not capable of simulating a full quantum computer, but IQP circuits are also not believed to be simulatable efficiently with a classical computer [Bremner2011].

We present an analogous result to that for IQP, but for nMBQC where there are functions that can be performed with greater mean success probability with quantum than classical resources. This result is in terms of computational expressiveness, rather than computational complexity. It also results from the fact that the model of nMBQC can be expressed as an NLG [Hoban2011a]. Non-trivial Bell inequalities can be derived from these games, and a violation of these Bell inequalities implies a computational advantage with quantum resources in nMBQC.

3.4.1 nMBQC, NLG and non-trivial Bell inequalities

We now formalise nMBQC and consider the tools required for our analysis. Firstly, the "goal" of nMBQC is to perform functions $f(\mathbf{x})$ both deterministically and efficiently in $|\mathbf{x}|$ for all instances of $\mathbf{x}$. This means that the mean success probability $\bar{p}(\bigoplus_{j=1}^{n} m_j = f(\mathbf{x}))$ of performing a function is

$$\bar{p}(\bigoplus_{j=1}^{n} m_j = f(\mathbf{x})) = \frac{1}{2^{|\mathbf{x}|}} \sum_{\mathbf{x}} p(\bigoplus_{j=1}^{n} m_j = f(\mathbf{x})|\mathbf{x}) = 1. \tag{3.64}$$

If this value is less than unity, a function $f(\mathbf{x})$ cannot be performed deterministically. We now relate this probability to the correlators $p(1|\mathbf{s})$, i.e. the statistics of obtaining $\bigoplus_{j=1}^{n} m_j = 1$ given inputs $\mathbf{s}$. In nMBQC, the inputs s_j are linear Boolean functions in $\mathbf{x}$, i.e. XOR gates performed on elements of $\mathbf{x}$. Also, without loss of generality, we consider inputs s_j being of the form $\bigoplus_{j=1}^{|\mathbf{x}|} a_j x_j \oplus b$ for $a_j \in \{0, 1\}$ and $b = 0$. If $b = 1$, each site can remove this constant from their input. We can then relate the bit-string $\mathbf{s}$ to $\mathbf{x}$ by an $|\mathbf{x}|$-by-n matrix $\mathbf{P}$ representing the linear transformations on $\mathbf{x}$ in mod 2. That is, every string $\mathbf{s}$ can be expressed as

$$\mathbf{s} = (\mathbf{P}\mathbf{x})_{\oplus} \tag{3.65}$$

where $(...)_\oplus$ represents matrix multiplication modulo 2. The strings $\mathbf{s}$ and $\mathbf{x}$ are then n-length and $|\mathbf{x}|$-length column vectors respectively. If we use the example from the introduction with 3 parties and have the input on the third party being $s_3 = s_1 \oplus s_2$, with the choice of measurement on site 1 and 2 being s_1 and s_2 respectively. If we fix $s_1 = x_1$ and $s_2 = x_2$ then $\mathbf{P} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{smallmatrix}\right]$. Therefore, for each computation in nMBQC we fix the matrix $\mathbf{P}$ that designates the computation performed by the control computer $\mathcal{C}$. The matrix $\mathbf{P}$ does not contain a row consisting of all-zeroes so there s_j is always dependent on elements of $\mathbf{x}$.

The figure of merit $\bar{p}(\bigoplus_{j=1}^n m_j = f(\mathbf{x}))$ can now be expressed in terms of correlators $p(1|\mathbf{s})$ to obtain the following:

$$\bar{p}(\bigoplus_{j=1}^{n} m_j = f(\mathbf{x})) = \frac{1}{2^{|\mathbf{x}|}} \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{P}\mathbf{x})_\oplus} \left(\delta_1^{f(\mathbf{x})\oplus 1} + (-1)^{f(\mathbf{x})\oplus 1} p(1|\mathbf{s}) \right). \quad (3.66)$$

We can immediately see that the right-hand-side is of the form of a non-trivial Bell inequality for $(n, 2, 2)$. The probability distribution $\pi(\mathbf{s})$ is $\sum_{\mathbf{x}} \frac{1}{2^{|\mathbf{x}|}} \delta^{\mathbf{s}}_{(\mathbf{P}\mathbf{x})_\oplus}$ but with the sum now over $\mathbf{x}$ instead of $\mathbf{s}$; $\mathbf{x}$ is however uniquely related to $\mathbf{s}$. The function $f(\mathbf{x})$ is then precisely the function in a non-trivial Bell inequality. This non-trivial Bell inequality is then

$$\begin{aligned} \frac{1}{2^{|\mathbf{x}|}} \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{P}\mathbf{x})_\oplus} \left((-1)^{f(\mathbf{x})\oplus 1} p(1|\mathbf{s}) \right) &\leq \frac{1}{2^{|\mathbf{x}|}} \max_{g(\mathbf{s})} \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{P}\mathbf{x})_\oplus} \left(\delta^{f(\mathbf{x})}_{g(\mathbf{s})} - \delta^0_{f(\mathbf{x})} \right) \\ &\leq \frac{1}{2^{|\mathbf{x}|}} \max_{g(\mathbf{x})} \sum_{\mathbf{x}} \left(\delta^{f(\mathbf{x})}_{g(\mathbf{x})} - \delta^0_{f(\mathbf{x})} \right) \quad (3.67) \end{aligned}$$

where $g(\mathbf{x})$ are all possible linear Boolean functions on $\mathbf{x}$. Since linear Boolean functions are $g(\mathbf{s}) = \bigoplus_{j=1}^n a_j s_j \oplus b$ for a_j, $b \in \{0, 1\}$ and $s_j = [(\mathbf{P}\mathbf{x})_\oplus]_j$ is a linear Boolean function on $\mathbf{x}$. Therefore $g(\mathbf{s})$ becomes an arbitrary linear Boolean function on $\mathbf{x}$ being $g(\mathbf{x})$.

Crucially the right-hand-side of (3.67) is independent of the number of sites n and is always strictly less than $1 - \frac{1}{2^{|\mathbf{x}|}} \sum_{\mathbf{x}} \delta^0_{f(\mathbf{x})}$ for $f(\mathbf{x})$ being a non-linear Boolean function. This latter fact also means that $\bar{p}(\bigoplus_{j=1}^n m_j = f(\mathbf{x})) < 1$ for classical resources. If $f(\mathbf{x})$ is linear then $\bar{p}(\bigoplus_{j=1}^n m_j = f(\mathbf{x})) = 1$ for classical resources; this is because the only deterministic correlators $p(1|\mathbf{s}) = \delta_1^{f(\mathbf{x})}$ possible in LHV theories are for the linear Boolean functions $f(\mathbf{x})$. Raussendorf has also shown that classical, or more generally, "noncontextual" resources can only perform linear Boolean functions deterministically in MBQC with a single round

of measurements [Raussendorf2009].

We now focus on using quantum resources in nMBQC. If quantum resources are more useful in nMBQC than classical resources then the inequality in (3.67) is violated by quantum correlators. Also if $\bar{p}(\bigoplus_{j=1}^{n} m_j = f(\mathbf{x})) = 1$, then from (3.29),

$$\sup_{\{\theta_j\}} \frac{1}{2^{|\mathbf{x}|+1}} \left[\left(\sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{Px})_\oplus} (-1)^{f(\mathbf{x})\oplus 1} \right) + \left| \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{Px})_\oplus} (-1)^{f(\mathbf{x})\oplus 1} e^{i(\sum_{j=1}^{n} s_j \theta_j)} \right| \right] = 1 - \frac{1}{2^{|\mathbf{x}|}} \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{Px})_\oplus} \delta^{0}_{f(\mathbf{x})} \tag{3.68}$$

We make the substitution of $(-1)^{f(\mathbf{x})\oplus 1} = 1 - 2\delta^0_{f(\mathbf{x})}$ to obtain

$$\sup_{\{\theta_j\}} \frac{1}{2^{|\mathbf{x}|}} \left| \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{Px})_\oplus} (-1)^{f(\mathbf{x})\oplus 1} e^{i(\sum_{j=1}^{n} s_j \theta_j)} \right| = 1.$$

This condition reduces to $(-1)^{f(\mathbf{x})\oplus 1} = e^{i(\sum_{j=1}^{n} [(\mathbf{Px})_\oplus]_j \theta_j)}$ where $s_j = [(\mathbf{Px})_\oplus]_j$. In the following result, we indicate that this condition can always be satisfied if n is at most equal to $2^{|\mathbf{x}|} - 1$. Therefore, all Boolean functions can be performed deterministically with quantum resources in nMBQC. The issue of efficiency will be discussed further into this section.

Theorem 13. *Every Boolean function $f(\boldsymbol{x})$ can be performed deterministically in nMBQC for at most $n = 2^{|\boldsymbol{x}|} - 1$ parties.*

Proof: As mentioned, every function $f(\mathbf{x})$ can be achieved deterministically if $(-1)^{f(\mathbf{x})\oplus 1}$ is equal to $e^{i(\sum_{j=1}^{n} [(\mathbf{Px})_\oplus]_j \theta_j)}$. This will be satisfied if each expression $\sum_{j=1}^{n} [(\mathbf{Px})_\oplus]_j \theta_j$ for $\mathbf{x} \neq \mathbf{0}$ is linearly independent from every other expression corresponding to each $\mathbf{s}$. To show this we just need to establish that all vectors $\mathbf{s} = (\mathbf{Px})_\oplus$ are linearly independent over $\mathbb{R}$. In other words, if we construct the $2^{|\mathbf{x}|} - 1$-by-n matrix $\mathbf{S}$ where rows are the vectors $\mathbf{s}^{\mathrm{T}}$ for $\mathbf{s} \neq \mathbf{0}$, then $\mathbf{S}$ must have rank $2^{|\mathbf{x}|} - 1$.

Every column of $\mathbf{S}$ has elements $g(\mathbf{x})$ where $g(\mathbf{x}) = \bigoplus_{j=1}^{n} a_j x_j$ is a linear Boolean function. We showed in Lemma 9 that $2^{|\mathbf{x}|} - 1$ vectors which have the elements being a different linear Boolean function of this form are linearly independent over $\mathbb{R}$. Therefore if each column of $\mathbf{S}$ corresponds to a different linear Boolean function $g(\mathbf{x}) = \bigoplus_{j=1}^{n} a_j x_j$ on $\mathbf{x}$, then the rank of $\mathbf{S}$ is $2^{|\mathbf{x}|} - 1$. □

Therefore, with quantum resources a function $f(\mathbf{x})$ can be computed. However, this result has only upper-bounded the resources n required to compute functions and this upper bound is inefficient. We now show that this upper bound is tight for all possible functions by using the example of the n-partite NAND function. We then use the following result to show that adaptivity is a crucial ingredient in MBQC.

Theorem 14. *The n-partite NAND function can only be performed deterministically in nMBQC for $n = 2^{|\boldsymbol{x}|} - 1$ parties.*

Proof: For ease of calculation, we prove this theorem for the n-partite NAND function with a NOT on each element of $\mathbf{x}$, i.e. $f(\mathbf{x}) = \prod_{j=1}^{|\mathbf{x}|}(x_j \oplus 1) \oplus 1$. However, the two functions are equivalent in our model as the control computer $\mathcal{C}$ can perform a NOT operation on each element of $\mathbf{x}$. We prove this theorem by assuming that this function can be performed deterministically with $2^{|\mathbf{x}|} - 2$ parties, and then obtain a contradiction. We first simplify the proof, instead of considering all possible $2^{|\mathbf{x}|} - 2$-by-$|\mathbf{x}|$ $\mathbf{P}$ matrices, we only need to consider one particular matrix $\mathbf{Q}$. This matrix $\mathbf{Q}$ is the matrix with all rows being bit-strings $\mathbb{Z}_2^{|\mathbf{x}|}$ not equal to either $\mathbf{0}$ or $\mathbf{1}$. Any matrix $\mathbf{P}$ not equal to $\mathbf{Q}$ can be turned into $\mathbf{Q}$ in the following way $(\Pi\mathbf{PM})_\oplus$ where Π is a $2^{|\mathbf{x}|}-2$-by-$2^{|\mathbf{x}|}-2$ permutation matrix and $\mathbf{M}$ is any binary, invertible $|\mathbf{x}|$-by-$|\mathbf{x}|$ matrix. This is because $\mathbf{P} \neq \mathbf{Q}$ then $\mathbf{P}$ contains a row equal to $\mathbf{1}$ and does not contain a bit-string $\mathbf{y} \in \mathbb{Z}_2^{|\mathbf{x}|}$. Therefore we use $\mathbf{M}$ to map $\mathbf{y}$ to $\mathbf{1}$ by right multiplication $(\mathbf{My})_\oplus = \mathbf{1}$ and $(\mathbf{PM})_\oplus$ contains all the same rows as $\mathbf{Q}$ but not necessarily in the same ordering. To establish the same ordering we left-multiply $(\mathbf{PM})_\oplus$ by the permutation matrix Π so that $(\Pi\mathbf{PM})_\oplus = \Pi(\mathbf{PM})_\oplus$.

The permutation matrix Π just is equivalent to permuting all n parties and so will leave the probability of performing a function invariant. If the function $f(\mathbf{x}) = \prod_{j=1}^{|\mathbf{x}|}(x_j \oplus 1) \oplus 1$ is performed deterministically with quantum resources then $e^{i(\sum_{j=1}^{n}[(\mathbf{Px})_\oplus]_j\theta_j)} = 1$ for $\mathbf{x} = \mathbf{0}$ and

$$e^{i(\sum_{j=1}^{n}[(\mathbf{Px})_\oplus]_j\theta_j)} = -1 \tag{3.69}$$

for all $\mathbf{x} \neq \mathbf{0}$. Since $(\mathbf{PM0})_\oplus = (\mathbf{P0})_\oplus$ then $e^{i(\sum_{j=1}^{n}[(\mathbf{PMx})_\oplus]_j\theta_j)} = 1$ for $\mathbf{x} = \mathbf{0}$. The set of strings $\{\mathbf{x}|\mathbf{x} \neq \mathbf{0}\}$ is equal to the set $\{(\mathbf{Mx})_\oplus|\mathbf{x} \neq \mathbf{0}\}$, then (3.69) is satisfied for both $\mathbf{P}$ and $(\mathbf{PM})_\oplus$. Therefore satisfying determinism for one matrix such as $\mathbf{Q}$ is equivalent to satisfying determinism for all $\mathbf{P}$ matrices.

We now show that deterministically performing $f(\mathbf{x}) = \prod_{j=1}^{|\mathbf{x}|}(x_j \oplus 1) \oplus 1$ is impossible for the matrix $\mathbf{Q}$. First we observe that if determinism is satisfied then from (3.69) we must satisfy

$$\sum_{j=1}^{n}[(\mathbf{Q}\mathbf{x})_{\oplus}]_j\theta_j = \pi(2t_{\mathbf{x}} + 1), \tag{3.70}$$

for all $\mathbf{x} \neq \mathbf{0}$ where $t_{\mathbf{x}}$ is an integer. We can construct a sum over all bit-strings $\mathbf{x}$ (including $\mathbf{0}$) which alternates in sign:

$$\sum_{j=1}^{2^{|\mathbf{x}|}-2}\left[\sum_{\mathbf{x}}(-1)^{W(\mathbf{x})}[(\mathbf{Q}\mathbf{x})_{\oplus}]_j\theta_j\right] = \sum_{\mathbf{x}\neq\mathbf{0}}(-1)^{W(\mathbf{x})}[\pi(2t_{\mathbf{x}} + 1)], \tag{3.71}$$

where $W(\mathbf{x})$ is the Hamming weight [MacWilliams1977] of $\mathbf{x}$, i.e. the number of non-zero elements of $\mathbf{x}$. We collect terms that have the same Hamming weight $W(\mathbf{x})$ on the right-hand-side of (3.71) and set $y = W(\mathbf{x})$. Defining $t_y = \sum_{\mathbf{x};W(\mathbf{x})=y} t_{\mathbf{x}}$ then the right-hand-side of (3.71) is equal to:

$$\sum_{y=1}^{|\mathbf{x}|}(-1)^y\pi\left(\frac{|\mathbf{x}|!}{y!(|\mathbf{x}|-y)!} + 2t_y\right) = \pi(2t-1), \tag{3.72}$$

where $t = \sum_{y=1}^{|\mathbf{x}|}(-1)^y t_y$ is some integer.

We now show that the left-hand-side of (3.71) is actually equal to zero, thus leading to the contradiction that $\pi(2t-1) = 0$ indicating that (3.69) is not true for all $\mathbf{x} \neq \mathbf{0}$ and determinism is not achieved. We can express the jth element of the sum in (3.69) as

$$\sum_{\mathbf{x}}(-1)^{W(\mathbf{x})}[(\mathbf{Q}\mathbf{x})_{\oplus}]_j = \sum_{\mathbf{x}}(-1)^{W(\mathbf{x})}\left(\bigoplus_{k=1}^{|\mathbf{x}|}Q_{j,k}x_k\right) \tag{3.73}$$

where $Q_{j,k}$ the element of $\mathbf{Q}$ corresponding to the jth row and kth column. Since every row of $\mathbf{Q}$ has at least one element $Q_{j,k} = 0$ for a particular value of k, then for two bit-strings $\mathbf{x}'$ and $\mathbf{x}''$ that only differ in their kth element, $\left(\bigoplus_{k=1}^{|\mathbf{x}|}Q_{j,k}x'_k\right) = \left(\bigoplus_{k=1}^{|\mathbf{x}|}Q_{j,k}x''_k\right)$. However, the two Hamming weights $W(\mathbf{x}')$ and $W(\mathbf{x}'')$ corresponding respectively to these bit-strings differ by 1 resulting in

$$(-1)^{W(\mathbf{x}')}\left(\bigoplus_{k=1}^{|\mathbf{x}|}Q_{j,k}x'_k\right)+(-1)^{W(\mathbf{x}'')}\left(\bigoplus_{k=1}^{|\mathbf{x}|}Q_{j,k}x''_k\right)=0. \tag{3.74}$$

Since all $2^{|\mathbf{x}|}$ bit-strings $\mathbf{x}$ can be paired into bit-strings that differ by one element, then (3.73) must be equal to zero. Therefore, we have reached a contradiction and deterministic computation of $f(\mathbf{x})=\prod_{j=1}^{|\mathbf{x}|}(x_j\oplus 1)\oplus 1$ cannot be achieved with less than $2^{|\mathbf{x}|}-1$ parties. □

This result says that determinism in nMBQC comes potentially at the price of an exponential overhead in resources. Contrast the above result with the fact that the n-partite NAND function can be implemented deterministically and efficiently by a classical computer. In fact, we do not need the full computing power of P, but a smaller complexity class called NC^1 which is contained in P [Papadimitriou1994]. Since a quantum computer can implement all computations in P, we have the following corollary.

Corollary 2. *It is impossible to efficiently achieve universal quantum computation deterministically in nMQBC.*

It is interesting that Bell tests, and in particular, non-trivial Bell inequalities have something to say about quantum computers. We know that Bell tests have a role in quantum cryptography and communication complexity; they now have some role to play in quantum computation. This relationship between foundations and applications of quantum physics is not unidirectional. We now discuss in the following subsection, how these results for nMBQC say something about Bell tests and correlations. In particular, they convey generalisations of the GHZ paradox [GHZ1989] mentioned earlier and indicate that there exist generalisations of the PR box [Popescu1994] that may not defined on all inputs $\mathbf{s}$.

3.4.2 Generalized GHZ Paradoxes and PR boxes

The original GHZ paradox [GHZ1989] was constructed as a way to demonstrate the incompatability of quantum physics with a LHV theory, but without the use of a Bell inequality. In the original paradox as discussed in section 1.2, the

following outcomes, translated into correlators:

$$\begin{aligned} p(1|000) &= 1 \\ p(1|011) &= 1 \\ p(1|101) &= 1 \end{aligned} \tag{3.75}$$

in an LHV theory deterministically predict that $p(1|110) = 1$. These statistics belong to the statistics of an extreme point of $\mathcal{L}$. However, measurements on a GHZ state lead to a contradiction where the expressions in (3.75) are satisfied but $p(1|110) = 0$. The LHV statistics in (3.75) result in a value of 2 for the Mermin inequality (3.13), but the quantum statistics result in a maximal algebraic violation of 3. Therefore our result in Theorem 13 can result in a GHZ paradox for $2^{|\mathbf{x}|} - 1$ parties. We can assign the following statistics in an LHV theory:

$$p(1|\mathbf{s}; \mathbf{s} = (\mathbf{P}\mathbf{x})_{\oplus}) = 1, \tag{3.76}$$

for all $\mathbf{x} \neq \mathbf{1}$ and $\mathbf{P}$ is the $2^{|\mathbf{x}|} - 1$-by-$|\mathbf{x}|$ matrix with rows consisting of all $2^{|\mathbf{x}|} - 1$ bit-strings $\mathbf{y} \in \mathbb{Z}_2^{|\mathbf{x}|}$ not equal to $\mathbf{0}$. If we put these statistics into the non-trivial Bell inequality in (3.67) corresponding to the n-partite NAND function, we obtain:

$$\begin{aligned} \frac{1}{2^{|\mathbf{x}|}} \sum_{\mathbf{x}} \sum_{\mathbf{s}} \delta^{\mathbf{s}}_{(\mathbf{P}\mathbf{x})_{\oplus}} (-1)^{f_2(\mathbf{x})+1} p(1|\mathbf{s}) &= \frac{1}{2^{|\mathbf{x}|}} \left(2^{|\mathbf{x}|} - 1 - p(1|\mathbf{s}; \mathbf{s} = (\mathbf{P}\mathbf{1})_{\oplus}) \right) \\ &\leq \frac{2^{|\mathbf{x}|} - 2}{2^{|\mathbf{x}|}}. \end{aligned} \tag{3.77}$$

then for LHV theories we can only assign the probability $p(1|\mathbf{s}; \mathbf{s} = (\mathbf{P}\mathbf{1})_{\oplus}) = 1$ deterministically. However, since with $2^{|\mathbf{x}|} - 1$ parties, we can perform the NAND function deterministically with quantum mechanics, we can satisfy both the probabilities in (3.76) and $p(1|\mathbf{s}; \mathbf{s} = (\mathbf{P}\mathbf{1})_{\oplus}) = 0$, leading to a contradiction.

We did not need to make this argument utilising a Bell inequality as we could have just used the statistics of the LHV correlator producing the linear Boolean function $f(\mathbf{x}) = 1$ deterministically. This deterministic correlator is the only correlator that satisfies all assignments in (3.76). In this sense then, we have a GHZ paradox for all choices of $|\mathbf{x}|$.

Finally, when we introduce the pre-processing on inputs $\mathbf{s} = (\mathbf{P}\mathbf{x})_{\oplus}$ and construct a non-trivial Bell inequality of the form in (3.67) then we do not consider all possible correlators $p(1|\mathbf{s})$ but only those correlators where $\mathbf{s}$ is defined by

$\mathbf{x}$ and $\mathbf{P}$. As a result we only consider probabilities $p(\mathbf{m}|\mathbf{s})$ that also satisfy this relationship between $\mathbf{s}$ and $\mathbf{x}$. We can consider non-signalling probability distributions $p(\mathbf{m}|\mathbf{s})$ that are of the following form

$$p(\mathbf{m}|\mathbf{s};\mathbf{s}=(\mathbf{P}\mathbf{x})_{\oplus}) = \begin{cases} \frac{1}{2^{n-1}} & \text{if } \bigoplus_{j=1}^{n} m_j = f(\mathbf{x}), \\ 0 & \text{otherwise,} \end{cases} \tag{3.78}$$

for any non-linear Boolean function $f(\mathbf{x})$. We are not concerned with inputs $\mathbf{s}$ that do not satisfy $\mathbf{s}=(\mathbf{P}\mathbf{x})_{\oplus}$, therefore, these distributions are not necessarily extreme points of $\mathcal{NS}$. The distributions may even be in the interior of $\mathcal{NS}$ but can be perceived as a generalisation of the PR box [Popescu1994], due to the fact that they maximally violate a Bell inequality for all correlators.

For example, the Mermin inequality is maximally violated by correlators resulting from a GHZ state, but we can also achieve the same maximal violation with vertices of $\mathcal{NS}$. The correlations $p(\mathbf{m}|\mathbf{s})$ that result from the GHZ state do not form a vertex of $\mathcal{NS}$. For $\mathbf{s} \notin \{\{000\},\{011\},\{101\},\{110\}\}$, the correlations $p(\mathbf{m}|\mathbf{s})$ resulting from the GHZ state do not resemble those of extreme points in $\mathcal{NS}$.

What Theorem 14 implies, is that even though (3.78) is defined on a subset of inputs $\mathbf{s}$, there exist non-signalling probability distributions for $n \leq 2^{|\mathbf{x}|}-2$ that cannot be achieved by quantum mechanics. More specifically, if $f(\mathbf{x})$ in (3.78) is the n-partite NAND function, since quantum physics cannot achieve this distribution for these values of n, they are as "unphysical" as the PR box.

Theorem 14 also implies that there are generalised PR boxes that can efficiently perform the n-partite NAND function in our nMBQC model. The fact that these unphysical resources can efficiently perform tasks unthinkable with physical resources has been analogously investigated in the field of communication complexity. An argument put forward first by Van Dam [vanDam2000] and then developed by Brassard et al [Brassard2006], is that if these unphysical, bipartite PR boxes exist then tasks in communication complexity are rendered "trivial". By trivial, we mean that only one bit of communication is required between two parties to achieve all Boolean functions. These ideas were also extended to the multipartite scenario [Marcovitch2008]. It could be argued that the result of Theorem 14 complements the idea that quantum mechanics cannot simulate all non-signalling probability distributions because information processing would be rendered "too easy".

In this section we have discussed the interplay between the computational perspectives on Bell tests and computation itself. In particular, we looked at a restricted class of computations in MBQC, itself a promising avenue for quantum computing. We have used Bell tests to show that adaptivity is crucial in Briegel and Raussendorf's MBQC scheme [Raussendorf2001]. With adaptivity comes the possibility for parties to communicate to each other and the connection between computation and Bell tests can break down. In the next chapter, we hint at a method to re-establish this connection.

3.5 Chapter Summary

When Bell first formulated his inequality he wanted to say something concrete about the interpretation of the wavefunction [Bell2004]. He established that if quantum mechanics is to be re-imagined as a local hidden variable theory, then a great deal of the theory's predictions would have to be "thrown out". Classical physics can be conceived as a local hidden theory, so there is an incompatibility between classical physics and quantum physics. This incompatibility is "witnessed" by a Bell inequality: a violation indicates incompatibility. It immediately tells us that quantum systems can do something that classical systems cannot.

It could be argued that it was inevitable that this tool for disambiguation between classical and non-classical would be used to show that quantum correlations can perform some tasks that classical correlations cannot. With the development of quantum information theory, Bell tests were approached with a new motivation: to find a quantum advantage for some quantum information processing tasks. For example, the application of Bell tests to cryptography [Acín2007] and random number generation [Pironio2010] has been successful.

Quantum computation could produce an advantage over classical computers [Shor1997]. The proof that quantum computers are more powerful than classical computers would have an immense impact on the study of classical computational complexity as it would provide a separation in a conjectured hierarchy of computational models [Papadimitriou1994]. Since the Bell test produces a clear cut distinction between quantum and classical, it could be considered a useful tool for proving this separation in computational models. The difficulty lies in communication, a resource not allowed in Bell tests, but not prohibited in most models of computation.

Immediately one can suggest that we study models of computation that do not require or even limit communication. Communication complexity is a model of computation that limits communication [vanDam2000], and non-local games do not allow communication between players but to the referee [Cleve2004]. Connections have been made to the latter with multi-prover interative proof systems, a model of computing based on the exchange of messages between parties in order to ascertain whether a potential solution to a problem is correct [Cleve2004]. Interactive proof systems have been shown to be extremely powerful, potentially far more powerful than computations in NP depending on the model [Jain2010]. If we want to say something about classical and quantum computers, then in these "simpler" models we will still want to place restrictions on communication. This motivates our study of MBQC circuits where the only communication allowed is between a classical computer and measurement sties, sites cannot communicate with each other and there is a single-round of measurements.

In this chapter, we began by discussing the space of LHV correlators in terms of the facet Bell inequalities. Finding facet Bell inequalities is hard and in practice we could only find them for a limited number of (n, c, d) scenarios. This motivated us to find a set of non-trivial Bell inequalities. These non-trivial inequalities were motivated by our computational insight into the space of LHV correlators, and were shown to be relevant for the study of non-local games (NLG). Finally, our restricted class of MBQC computations was shown to be cast as an NLG, and again made relevant to non-trivial Bell inequalities. Using the tools from the study of Bell inequalities, we showed that this restricted class of MBQC computations is not universal for quantum computing. However, in this model, due to the very nature of the Bell inequality, we showed that quantum resources can do something that classical resources cannot.

We have shown that there are concrete connections between Bell tests and some models of computing. On the other hand, we have also shown that communication in the form of adaptivity is vital for MBQC. In the next chapter, we will indicate how to simulate communication in computations within the framework of a Bell test. Perhaps surprisingly, this communication simulation still allows the possibility for disambiguating quantum and classical resources.

4 Data Post-selection in Bell Tests

The Bell test has been around formally for decades. A natural question is 'can we go beyond this formulation?' Of course, situations altering the number of parties, inputs and outputs have been studied. Despite these generalizations, the core of the gedankenexperiment still involves space-like separated parties making their measurements and then sending their data to be turned into statistics. However, in reality, data does not always emerge perfectly from experiments, and often it needs to be discarded. CHSH took this imperfection into account and added an extra assumption to the construction of Bell tests beyond Bell's formulation: the "fair-sampling assumption" [CHSH1969, Clauser1978, Berry2010]. This assumption essentially states that the experimental errors in performing a Bell test are independent of the choice of measurement at each site. In the history of experimental tests of Bell inequalities, this assumption has featured strongly, especially in optical tests [Freedman1972, Shih1988, Ou1988, Rarity1990, Tittel1998, Weihs1998].

Whilst the fair-sampling assumption may be rooted in common sense, we cannot assume, in general, that it is true. However, if we relax it then the discarding of data can be problematic. In particular, it can lead to the "detection loophole" [Pearle1970, Garg1987] as it is now often referred. A "loophole" emerges when some imperfection in the experiment can allow LHV correlations to simulate quantum correlations. There a several sources of loopholes in experimental Bell tests, some more subtle that others.

Two central constraints on the construction of Bell tests are measurement choice independence and space-like separation. If the latter is not respected in an experiment, then parties can communicate and from this communication, simulate whichever correlations they wish. Bell has emphasized himself how important that choice of measurement be completely random and independent of the parties' systems [Bell2004]. Barrett and Gisin have directly related the lack of measurement choice independence to simulating communication between parties. These central stipulations of the Bell test must be upheld if we want to

restrict what is possible with LHV correlations.

Modern, photonic-based Bell tests allow for space-like separated measurements [Tittel1998, Weihs1998]. The issue of freedom of measurement choice can tend towards philosophy, and the concept of "free will". These discussions are well beyond the scope of this thesis. It could be argued though that photon Bell tests can also address the need for random choice of measurements [Weihs1998]. As discussed in section 1.4 of chapter 1, random numbers can be generated by quantum processes, potentially in a device independent manner [Colbeck2007, Pironio2010]. Experimental groups have exploited this source of randomness to produce random measurements [Weihs1998].

The issues raised by more systematic failures to implement Bell tests are problematic. The detection loophole is a more subtle source of problems. It can be seen to result from a form of "post-selection". Here we use the term post-selection as a means of accepting measurement data if it satisfies particular criteria[1]. In the case of imperfect detection where our measurement devices (detectors) may or may not receive a measurement outcome (detection event), we can only calculate correlations for all parties if all parties have made a successful measurement. Therefore, we accept or post-select on measurement data if all sites successfully detected a measurement outcome. In the first section of this chapter, we will formalise these ideas in the $(n, 2, 2)$ scenario[2].

This chapter concerns itself more generally with data post-selection in Bell tests. In particular, we introduce two forms of post-selection and associate a loophole with each form of post-selection. In section 4.1 of this chapter, we discuss the form of post-selection in the presence of imperfect detection, whereas in section 4.2 we consider post-selection in *perfect* Bell tests. By the latter, we mean that we have perfect detection, space-like separation and freedom of measurement choice (the original gedankenexperiment) but introduce a form of post-selection on accepting measurement data. Whilst the post-selection in section 4.1 is experimentally motivated, the post-selection in section 4.2 is very much conceptually motivated. Despite their differing motivation there is an overlap in the language we use to describe the loopholes. This language is rooted in our *computational* insight into LHV correlators.

[1] Post-selection in quantum information can often mean the acceptance of a quantum state after measurement, if a particular measurement outcome is achieved. Otherwise the quantum state is discarded.

[2] These ideas can be extended to different scenarios, but for pedagogical clarity and the ease of producing new results we make this restriction.

Interestingly, whilst post-selection on successful detection can lead to the detection loophole (as we shall show), the post-selection in section 4.2 can be described as "loophole-free". As well as the latter constraining LHV correlators in the presence of post-selection, it can also *enlarge* the space of quantum correlators. We also indicate that connections can be made between MBQC and our new form of data post-selection. Finally in section 4.3, we give some indications that generalising the results of section 4.2 to different (n, c, d) scenarios may become problematic, and no longer loophole-free.

The original work in this chapter was developed in collaboration with Dan Browne. Section 4.1 (except subsection 4.1.3) is a rederivation of the work of Garg and Mermin in [Garg1987], but now in our computational description of Bell tests. Subsection 4.1.3 consists of a new result generalising the work of Garg and Mermin to n parties. In section 4.3 all of the work was completed also in collaboration with Joel Wallman. Results in section 4.2 have been published as [Hoban2011b] and some of the results in section 4.3 have been published in [Hoban2011c].

4.0.1 Notation

In this chapter, we will carry over the notation convention for modular arithmetic introduced in the last chapter. The first two sections of this chapter solely consider the $(n, 2, 2)$ scenario and so we use $\oplus$ and $\bigoplus$ to denote addition and summation modulo 2 for only the $(n, 2, 2)$ case. In section 4.3 we consider the (n, c, d) cases, and we enclose modulo x arithmetic in brackets, i.e. $[...]_x$. For further clarification see section 3.0.1 of chapter 3.

4.1 Post-selection and the Detection Loophole

We know that the space of quantum correlators is larger than $\mathcal{L}$ by Bell's theorem. This is a mathematical statement and testing it in the laboratory has been a major endeavour and challenge in the past few decades [CHSH1969, Freedman1972, Shih1988, Ou1988, Rarity1990, Tittel1998, Weihs1998, Rowe2001]. However, of these experiments, the majority have suffered from the detection loophole. Experiments such as [Rowe2001] that manage to overcome the detection loophole suffer from not having space-like separated measurements [Rowe2001]. There are currently no loophole-free Bell tests but there are promising routes for overcoming the detection loophole [Matsukevich2008, Vértesi2010, Sangouard2011].

The issue of imperfect detection, culminating in the detection loophole is a subtle issue [Pearle1970, Garg1987]. In a full treatment of a Bell test, a non-detection of an event is in itself an event. That is, if a measurement is the result of a detection and there are d possible outcomes, a non-detection must be another outcome. We cannot rule out the possibility that an LHV theory can produce all $(d+1)$ outcomes. The fair-sampling assumption aims to exclude this possibility by saying that the non-detection event is independent of our choice of measurement [CHSH1969, Clauser1978, Berry2010]. This assumption cannot itself be tested. For example, we construct an explicit LHV model that violates the fair-sampling assumption but the statistics of detection are random at each site. We cannot extract the dependence on $\mathbf{s}$ from the statistics alone. We do not therefore impose the fair-sampling assumption in our discussion.

Having imperfect detectors does not necessarily mean that LHV correlators can completely simulate quantum correlators. Recall that this simulation is how we describe a loophole, but we shall make this notion more rigorous in subsequent discussion. Work by Pearle [Pearle1970] which was then developed by Garg and Mermin [Garg1987] showed that if the detector efficiency (the ratio of successful detection to all incoming events) at each site is above some threshold, then a loophole can be ruled out. This detection efficiency threshold has been subsequently lowered by further research [Eberhard1993, Vértesi2010].

A final, somewhat more applied, motivation for considering the detection loophole comes from quantum key distribution [Ekert1991]. We discussed device-independent quantum key distribution [Acín2007, Pironio2009] in section 1.4 of the first chapter. Recall that the security of device-independent quantum key distribution can be ensured by the violation of a Bell inequality. The intuition is as follows: an adversary trying to learn the generated secret key (thus able to decode any secret message) can learn it if the key is described by an LHV. The secret information is contained in some "local" information at each site which can be "extracted" by said adversary. If the secret information is generated by some correlations incompatible with an LHV theory, then an adversary cannot localise it and obtain it. The detection loophole allows an adversary to learn a secret key that can be generated by LHV resources via the loophole [Acín2007].

We structure this section so that we introduce and describe the detection loophole. Our novel insight into this loophole is to use the language of computational expressiveness to describe what LHV correlators can do in the presence of imperfect detection. We show that the post-selection of accepting measurement

data based on successful detection induces a relationship between each party's shared hidden variables and inputs s_j. We use this discussion to derive the GM threshold detector efficiency, but also to generalise their result to n parties. We show that this threshold can be lowered by going from 2 to n parties. A previous reduction in the threshold detector efficiency for $(2, 2, 2)$ have resulted from considering the full probability distribution and not correlators [Eberhard1993].

In this section and the next, we will restrict ourselves to the study of Bell tests in the $(n, 2, 2)$ scenario. Therefore we will use Corollary 1 of Theorem 2 where $\mathcal{L}$ is the convex hull of the linear Boolean functions. If a correlator cannot be written as a convex combination of linear Boolean functions for all possible decompositions it must lie outside of $\mathcal{L}$.

4.1.1 The Detection loophole

The action of discarding data means that the person carrying out a Bell test is playing an active role[3]. Because of this active role, throughout this chapter, we will refer to an "experimenter" who does something non-trivial with the experimental data. We will describe the role of the experimenter in different contexts in more detail throughout this chapter. That is, what the experimenter can and cannot do will be prescribed.

How do we incorporate the issue of a non-detection event into an $(n, 2, 2)$ Bell test? Since the number of outputs of a successful measurement is binary, then the total number of outcomes is ternary, i.e. $m_j \in \mathbb{Z}_3$. What is an appropriate joint outcome, the sum modulo 2 of all outcomes, or the sum modulo 3? If we take the sum modulo 2 then a non-detection will necessarily get mapped to an event with a successful detection. Can we still talk in terms of Boolean functions if the number of outcomes at each site is ternary? Is a loophole is caused by de facto moving out of the scope of Boolean functions?

We can resolve this discussion by redescribing the scenario only in terms of bit-strings. Now instead of each jth site outputting a single digit m_j, they output two bits $\{t_j, m_j\} \in \mathbb{Z}_2 \times \mathbb{Z}_2$. Here t_j is a bit that indicates whether an event is successfully detected (represented by 1) or not detected (represented by 0). If $t_j = 1$ for all j, then the experimenter takes the sum modulo 2 of all outcomes m_j, if $t_j = 0$ for at least one site j, we throw away all data. The elements t_j

[3]With perfect detection, the experimenter only calculated the sum modulo d of outcomes. This can be seen as an active role, however, we take active to mean that they can do something non-trivial with the data.

make up an n-length bit-string $\mathbf{t}$ and we accept $\mathbf{m}$ if $\mathbf{t} = \mathbf{1}$, the string of all-ones. This discarding of data is a form of post-selection; we call this method of post-selection when $\mathbf{t} = \mathbf{1}$ "detection post-selection".

Definition 8. When the experimenter accepts, or post-selects on data $\mathbf{m}$ and $\mathbf{s}$ when $\mathbf{t} = \mathbf{1}$, this is **detection post-selection**. This data after post-selection is then used to calculate $\bigoplus_{j=1}^{n} m_j$.

This action of post-selection as we shall show can be a way of introducing loopholes. Before we define a loophole we need to introduce the mathematical construction we need to define them.

The convex polytope $\mathcal{P}$ is the space of correlators $p(1|\mathbf{s})$ that are perfectly detected, i.e. $\mathbf{t} = \mathbf{1}$ for all runs of an experiment. For imperfect detection, we need a new, more general space of correlators that are calculated *after* post-selecting on $\mathbf{m}$ and $\mathbf{s}$ when $\mathbf{t} = \mathbf{1}$. We call this more general space $\tilde{\mathcal{P}}$ and if every run of an experiment produces $\mathbf{t} = \mathbf{1}$, then $\tilde{\mathcal{P}} = \mathcal{P}$. However, more generally, correlators are now defined in the following way

$$\tilde{p}(1|\mathbf{s}) = p(\bigoplus_{j=1}^{n} m_j = 1|\mathbf{s}, \mathbf{t} = \mathbf{1}). \tag{4.1}$$

$\tilde{\mathcal{P}}$ is now the space of correlators of the form (4.1). However, the space $\tilde{\mathcal{P}}$ for the $(n, 2, 2)$ setting can be defined in an analogous way to $\mathcal{P}$. That is, $\tilde{\mathcal{P}}$ is the convex hull of all correlators $\tilde{p}(1|\mathbf{s}) = \delta^1_{f(\mathbf{s})}$ for any Boolean function $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$. We have put no restriction on the probability of detection $p(\mathbf{t} = \mathbf{1})$, only that $\mathbf{t} = \mathbf{1}$.

In the case for perfect detection, the space of LHV correlators is $\mathcal{L}$ as defined by Corollary 1. We define $\tilde{\mathcal{L}}$ as the space of correlators $\tilde{p}(1|\mathbf{s})$ resulting from LHV correlators, computed after detection post-selection. Is the space $\tilde{\mathcal{L}}$ always the convex hull of linear Boolean functions on $\mathbf{s}$? For perfect detectors where $\mathbf{t} = \mathbf{1}$ is always satisfied, then $\tilde{\mathcal{L}} = \mathcal{L}$. Another way of asking this is to write the CHSH inequalities in terms correlators $\tilde{p}(1|\mathbf{s})$,

$$\tilde{p}(1|00) + \tilde{p}(1|00) + \tilde{p}(1|00) - \tilde{p}(1|00) \leq 2. \tag{4.2}$$

If this inequality can be violated by correlators in $\tilde{\mathcal{L}}$ then the space $\tilde{\mathcal{L}}$ is no longer the convex hull of linear Boolean functions. We then associate this violation by LHV correlators (in the presence of imperfect detection) with a loophole in a Bell

test. We now define this loophole.

Definition 9. A **loophole** is introduced by an experimenter into a Bell test if after detection post-selection, the space $\tilde{\mathcal{L}}$ is larger than the convex hull of linear Boolean functions.

The intuition behind this being a loophole is that if we have a quantum correlator $\vec{q}$ (obtained with perfect detection) being outside of $\mathcal{L}$, then it will violate a facet Bell inequality. However, if the detectors which obtained this quantum correlator become imperfect, then after detection post-selection, the resulting quantum correlator in $\tilde{\mathcal{P}}$ will again[4] be equal to $\vec{q}$. The loophole means that $\vec{q}$ could now be in the space $\tilde{\mathcal{L}}$. It is possible that $\vec{q}$ could be outside of $\tilde{\mathcal{L}}$, but the facet Bell inequalities for $\mathcal{L}$ are possibly no longer relevant for informing us either way. We now show that loopholes are achievable with detection post-selection. In the following result we show that it is possible that $\tilde{\mathcal{L}}$ can no longer be confined to the convex hull of linear Boolean functions.

Proposition 15. *For all LHV theories, $\tilde{\mathcal{L}}$ is larger than the convex hull of linear Boolean functions on $\boldsymbol{s}$.*

Proof: We construct the following specific model with n sites. The $(n-1)$ sites for $j \in \{2, 3, ..., n\}$ have perfect detectors whereas the first site has an imperfect detector. The first detector outputs the detection bit as a function of an LHV λ and its input, $t_1 = s_1 \oplus b(\lambda) \oplus 1$ so that $b(\lambda) \in \{0, 1\}$, whereas $t_2 = 1$. When we post-select so that $t_1 = 1$ then $s_1 = b(\lambda)$. The variable $b(\lambda)$ is shared by all parties, and the second party's measurement outcome $m_2 = b(\lambda)s_2$. If for all $(n-1)$ sites where $j \neq 2$, the parties' measurement outcomes upon successful detection are $m_j = 0$, then when $\mathbf{t} = \mathbf{1}$, $\bigoplus_{j=1}^{n} m_j = s_1 s_2$. The resulting correlator is then $\tilde{p}(1|\mathbf{s}) = \delta_1^{s_1 s_2} = s_1 s_2$, which is a vertex outside of the convex hull of linear Boolean functions. □

This demonstrates how post-selection can be problematic in Bell tests. A drawback of the proof of the above result is the asymmetry in the detectors between the first detector and the rest. If we were to switch the detectors in the experiment and still got the same imperfect detection at site 1 then the rate of detection must be independent of the detector. The measure of detection is the

[4]We assume that the detection device is independent of the quantum state or choice of measurement made.

detection efficiency η which is the quotient of number of successful detections to the number of events incoming to the detector. We can obtain the efficiency of a detector if two sites each make measurements, and then condition the statistics of the detector upon the other detecting an event so that

$$\eta = \frac{p(t_1 = 1|t_2 = 1)}{p(t_1 = 0|t_2 = 1) + p(t_1 = 1|t_2 = 1)} = \frac{p(\mathbf{t} = \{1, 1\})}{p(\mathbf{t} = \{0, 1\}) + p(\mathbf{t} = \{1, 1\})}. \quad (4.3)$$

We assume that the detector efficiency η is the same for all n sites. Situations with non-uniform η amongst parties have been investigated (e.g. [Vértesi2010]) but is beyond the scope of our discussion here.

In the early literature discussing the detection loophole (e.g. [Pearle1970, Garg1987]), the probabilities in (4.3) are calculated from the number $N_{\mathbf{t}}$ of events where $\mathbf{t}$ occured. It is assumed that the number of events where $\mathbf{t} = \mathbf{0}$ is unobservable as they are non-events. Probabilities then become normalised relative to this inability to detect when $\mathbf{t} = \mathbf{0}$ and

$$p(\mathbf{t}) = \frac{N_{\mathbf{t}}}{\sum_{\mathbf{t} \neq \mathbf{0}} N_{\mathbf{t}}}. \quad (4.4)$$

Then for the above discussion about η for 2 parties, $N_{\mathbf{t}}$ being the number of events where $\mathbf{t} \in \{0, 1\}^2$ occurs, the total number of events is $N = N_{\mathbf{1}} + N_{\{0,1\}} + N_{\{1,0\}}$. The probabilities in (4.3) then are obtained in the limit where $N \to \infty$ giving the efficiency

$$\eta = \frac{N_{\mathbf{1}}}{N_{\mathbf{1}} + N_{\{0,1\}}}. \quad (4.5)$$

If we want η to be the same for all sites then $N_{\{0,1\}} = N_{\{1,0\}}$. We also now impose that the statistics $p(t_j = 1)$ should be independent of s_j. This is not as strong as the fair-sampling assumption and we can experimentally test whether single-site detection statistics are independent of s_j [Garg1987]. This reinforces the intuition that the properties of a detector such as η should be independent of whatever measurement we make.

In line with previous research such as in [Garg1987], we now weigh the correlation statistics with the statistics of detection. Therefore correlators now take the form

$$\bar{p}(1|\mathbf{s}) = p(\mathbf{t} = \mathbf{1})\tilde{p}(1|\mathbf{s}). \quad (4.6)$$

These correlators $\bar{p}(1|\mathbf{s})$ are not necessarily normalised so $\bar{p}(0|\mathbf{s}) \neq 1 - \bar{p}(1|\mathbf{s})$ in general. On the other hand, $p(\mathbf{t} = \mathbf{1})\tilde{p}(0|\mathbf{s}) = p(\mathbf{t} = \mathbf{1})(1 - \tilde{p}(1|\mathbf{s}))$. Expectation

values of outcomes in the space $\tilde{\mathcal{P}}$ can be defined in exact analogy with the expectation values $\mathbb{E}(\mathbf{s})$ over correlators in $\mathcal{P}$, giving

$$\tilde{\mathbb{E}}(\mathbf{s}) = 1 - 2\tilde{p}(1|\mathbf{s}). \tag{4.7}$$

Expectation values $\bar{\mathbb{E}}(\mathbf{s})$ for the correlators $\bar{p}(0|\mathbf{s})$ and $\bar{p}(1|\mathbf{s})$ can then be related to $\tilde{\mathbb{E}}(\mathbf{s})$ to obtain

$$\bar{\mathbb{E}}(\mathbf{s}) = \bar{p}(0|\mathbf{s}) - \bar{p}(1|\mathbf{s}) = p(\mathbf{t} = \mathbf{1})(1 - 2\tilde{p}(1|\mathbf{s})) = p(\mathbf{t} = \mathbf{1})\tilde{\mathbb{E}}(\mathbf{s}). \tag{4.8}$$

This relationship between expectation values will be utilised in the following section. In fact, because the correlators $\bar{p}(1|\mathbf{s})$ are not normalised, it will be more useful to work in terms of the expectation values $\bar{\mathbb{E}}(\mathbf{s})$. This means we only need to consider one number instead of both $\bar{p}(0|\mathbf{s})$ and $\bar{p}(1|\mathbf{s})$.

In the following two subsections we will work in the new space $\bar{\mathcal{L}}_{\mathbb{E}}$ of the expectation values $\bar{\mathbb{E}}(\mathbf{s})$ for LHV theories. In line with previous discussion, this space is a 2^n dimensional real space of vectors having the elements $\bar{\mathbb{E}}(\mathbf{s})$. These elements can now be negative but their magnitudes are bounded by unity. The space $\bar{\mathcal{L}}_{\mathbb{E}}$ is a sub-space of $\bar{\mathcal{P}}_{\mathbb{E}}$ which is now the space of all possible vectors of expectation values $\bar{\mathbb{E}}(\mathbf{s})$.

4.1.2 Rederivation of the GM detection efficiency

We now address the $(2, 2, 2)$ scenario and use it to give an upper bound on the detection efficiency η required in order to demonstrate a violation of the CHSH inequality in the presence of imperfect detectors. This upper bound was derived by GM [Garg1987] and has since been improved upon by Eberhard [Eberhard1993] in the Clauser-Horne inequality setting [CH1969]. As an aside, it has been suggested that if we consider different Bell test settings, we can lower the detection efficiency required to violate any Bell inequality [Vértesi2010].

We will use our computational interpretation of correlators to rederive the GM upper bound on the threshold detection efficiency η. In order to do this, we first describe the space $\bar{\mathcal{L}}_{\mathbb{E}}$ of expectation values $\bar{\mathbb{E}}(\mathbf{s})$. The following result now captures this space in terms of a vertex description.

Proposition 16. *The space $\bar{\mathcal{L}}_{\mathbb{E}}$ is the convex hull of all expectation values $\bar{\mathbb{E}}(\boldsymbol{s}) = (-1)^{g(\boldsymbol{s})}$ for $g(\boldsymbol{s})$ being a linear Boolean function on $\boldsymbol{s}$.*

Proof: First, just like measurement outcomes m_j resulting from LHV theories,

the detection values t_j can be, in general, written as $t_j = x_j(\lambda)s_j \oplus y_j(\lambda)$ for bits $x_j(\lambda) \in \{0,1\}$ depending on the local hidden variable λ. Therefore, if $y_j(\lambda) = 1$ and $x_j(\lambda) = 0$, then $p(\mathbf{t} = \mathbf{1}) = 1$ for all j, λ, otherwise if $y_j(\lambda) = x_j(\lambda) = 0$, then $p(\mathbf{t} = \mathbf{1}) = 0$ again for all j, λ. So then for all LHV maps where $x_j(\lambda) = 0$ for all j, λ, the probability of detection for a single-site is $p(t_j = 1) = p(y_j(\lambda) = 1)$.

Since s_j is randomly generated, for $x_j(\lambda) = 1$ for all j, λ, then $p(\mathbf{t} = \mathbf{1}) \neq 1$ for all $y_j(\lambda)$. Finally, the map $t_j = s_j$ or $t_j = s_j \oplus 1$ is forbidden as this means there is a direct dependence in the statistics of detection with the choice of input. We then instead have maps $t_j = s_j \oplus y_j(\lambda)$ where $y_j(\lambda)$ is shared by both parties and generated randomly so that $p(t_j|s_j) = p(t_j|s'_j)$ for $s_j \neq s'_j$. As a result of t_j being random, $p(\mathbf{t} = \mathbf{1})$ is at most equal to $\frac{1}{2}$.

If one party employs the strategy of $t_j = s_j \oplus y_j(\lambda)$ and the other site produces the deterministic map $t_{j'} = 1$ then $p(\mathbf{t} = \mathbf{1}) = \frac{1}{2}$. However, the detection efficiency η is not the same for both sides. We can maintain the same probability $p(\mathbf{t} = \mathbf{1}) = \frac{1}{2}$ while making the detection efficiency the same for both sides if both parties share a random bit $z(\lambda) \in \{0,1\}$. When $z(\lambda) = 0$, $t_1 = s_1 \oplus y_1(\lambda)$ and $t_2 = 1$, and when $z(\lambda) = 1$, $t_1 = 1$ and $t_2 = s_1 \oplus y_1(\lambda)$. As $z(\lambda)$ is randomly generated then $p(\mathbf{t} = \mathbf{1}) = \frac{1}{2}(\frac{1}{2} + \frac{1}{2}) = \frac{1}{2}$. If $z(\lambda)$ were not random then we bias one of the strategies and $N_{\{0,1\}} \neq N_{\{1,0\}}$ for $N \to \infty$, which is forbidden.

In this strategy where parties share $z(\lambda)$, one of the parties learns the other party's input s_j as it is equal to a variable $y_j(\lambda) \oplus 1$ when $t_j = 1$. If one party learns the other party's variable then they can compute the non-linear Boolean functions $f(\mathbf{s}) = (s_1 \oplus a)(s_2 \oplus b) \oplus c$ for $a, b, c \in \{0,1\}$ deterministically. Therefore, the parties can achieve the post-selected expectation value:

$$\tilde{\mathbb{E}}(\mathbf{s}) = (-1)^{f(\mathbf{s})}, \tag{4.9}$$

with $f(\mathbf{s})$ being the above non-linear Boolean function. This gives a value of

$$\bar{\mathbb{E}}(\mathbf{s}) = p(\mathbf{t} = \mathbf{1})\tilde{\mathbb{E}}(\mathbf{s}) = \frac{1}{2}(-1)^{f(\mathbf{s})}. \tag{4.10}$$

We take the convex combination of LHV strategies producing all allowed deterministic maps $t_j = x_j(\lambda)s_j \oplus y_j(\lambda)$ and then the possible deterministic values of $\tilde{\mathbb{E}}(\mathbf{s})$ for each strategy. This then produces the expectation values:

$$\bar{\mathbb{E}}(\mathbf{s}) = \sum_{g(\mathbf{s})} p_{g(\mathbf{s})}(-1)^{g(\mathbf{s})} + \sum_{f(\mathbf{s})} \frac{p_{f(\mathbf{s})}}{2}(-1)^{f(\mathbf{s})}. \tag{4.11}$$

with $f(\mathbf{s})$ and $g(\mathbf{s})$ being all of the non-linear and linear Boolean functions respectively. We have taken the convex combination with positive coefficients $p_{g(\mathbf{s})}$, $p_{f(\mathbf{s})}$ such that $\sum_{g(\mathbf{s})} p_{g(\mathbf{s})} + \sum_{f(\mathbf{s})} p_{f(\mathbf{s})} = 1$. Thus $\bar{\mathcal{L}}_{\mathbb{E}}$ is at least as large as the convex hull of $\bar{\mathbb{E}}(\mathbf{s}) = (-1)^{g(\mathbf{s})}$. If the expectation values in (4.10) are outside of this space then they will violate one of the CHSH inequalities

$$\left| \sum_{\mathbf{s}} (-1)^{f(\mathbf{s})} \bar{\mathbb{E}}(\mathbf{s}) \right| \leq 2, \tag{4.12}$$

where $f(\mathbf{s})$ can one of the non-linear Boolean functions $f(\mathbf{s}) = (s_1 \oplus a)(s_2 \oplus b) \oplus c$ for a, b, $c \in \{0, 1\}$. If we use the strategy of allowing the maps $t_j = s_j \oplus y_j(\lambda)$, then $p(\mathbf{t} = \mathbf{1}) = \frac{1}{2}$ even though $\left| \sum_{\mathbf{s}} (-1)^{f(\mathbf{s})} \tilde{\mathbb{E}}(\mathbf{s}) \right| \leq 4$. This lack of violation for the CHSH inequalities therefore concludes the proof. □

This result will give an upper bound on the efficiency η required of detectors in order to establish that certain values of $\bar{\mathbb{E}}(\mathbf{s})$ are not in $\bar{\mathcal{L}}_{\mathbb{E}}$. The result indicates the structure of $\bar{\mathcal{L}}_{\mathbb{E}}$ is the same as $\mathcal{L}$, and the CHSH inequalities are exactly the same, i.e.

$$\begin{aligned} \bar{\mathbb{E}}(00) + \bar{\mathbb{E}}(01) + \bar{\mathbb{E}}(10) - \bar{\mathbb{E}}(11) &= \\ p(\mathbf{t} = \mathbf{1}) \left(\tilde{\mathbb{E}}(00) + \tilde{\mathbb{E}}(01) + \tilde{\mathbb{E}}(10) - \tilde{\mathbb{E}}(11) \right) &\leq 2. \end{aligned} \tag{4.13}$$

If we assume that the values $\tilde{E}(\mathbf{s})$ are obtained from measurements on quantum systems, then the maximum quantum value of $\tilde{E}(00) + \tilde{E}(01) + \tilde{E}(10) - \tilde{E}(11)$ is Tsirelson's bound, $2\sqrt{2}$. In order to demonstrate a violation of the inequality (4.13), we must then satisfy $p(\mathbf{t} = \mathbf{1}) > \frac{1}{\sqrt{2}}$.

We now relate the value of $p(\mathbf{t} = \mathbf{1})$ to the detection efficiency η with the following expression:

$$p(\mathbf{t} = \mathbf{1}) = \lim_{N \to \infty} \frac{N_{\mathbf{1}}}{N_{\mathbf{1}} + N_{\{0,1\}} + N_{\{1,0\}}} = \lim_{N \to \infty} \frac{N_{\mathbf{1}}}{N_{\mathbf{1}} + 2N_{\{0,1\}}} = \frac{\eta}{2 - \eta}, \tag{4.14}$$

since $N_{\{0,1\}} = N_{\{1,0\}}$. A value of $p(\mathbf{t} = \mathbf{1}) > \frac{1}{\sqrt{2}}$ thus gives $\eta > \frac{2}{\sqrt{2}+1} \approx 0.8284$. This is exactly the detection efficiency derived by GM [Garg1987].

In GM's result of $\eta \approx 0.8284$, they use a Bell inequality derived for spin-0 particles [Mermin1982]. In this original work, it is perhaps not clear, in general, how a loophole is avoided or created. We have explicitly shown the mechanism of how loopholes are formed and this is due to the emergence of non-linear Boolean

functions in the event of post-selection. The beauty of our approach, as we shall show in the next subsection is that it can be generalised to n parties; something not immediately attainable in the GM approach[5]. In the following subsection, we describe this generalisation to $(n, 2, 2)$ scenarios.

4.1.3 Generalisation of the GM bound to Many Parties

In the previous subsection, we showed that as long as detection efficiency is above some threshold then quantum physics can violate a Bell inequality. The threshold we derived was already attained by GM. Our rederivation makes the mechanism of loopholes very clear and also establishes the framework for generalising to more than two parties. In this subsection we now present a new result.

We have shown that the GM threshold for detection efficiency is reached when the quantum systems achieve Tsirelson's bound. If the quantum systems do not achieve this bound then the detection efficiency needed to rule out an LHV description needs to be higher. That is if $\tilde{E}(00) + \tilde{E}(01) + \tilde{E}(10) - \tilde{E}(11) = 2 + \epsilon$ where $0 < \epsilon \leq 2(\sqrt{2} - 1)$ results from quantum correlators then the detection efficiency must satisfy $\eta > \frac{4}{4+\epsilon}$. A natural extension of this result is to find Bell inequalities in other Bell tests where the detection efficiency required is lower. Then a bigger range of quantum values of a Bell expression can be tolerated and rule out an LHV description.

This has also been investigated in the full probability distribution Bell setting (e.g. the Clauser-Horne Bell setting[CH1969]). For example, Eberhard showed that for the CH inequality the minimum detection efficiency is given by $\eta > \frac{2}{3}$ [Eberhard1993]. This value has been subsequently lowered if one increases the number of measurement settings that one can choose from [Vértesi2010]. However, we are focussing on the n-party setting with 2 inputs and 2 outputs; we will explore a generalisation of the derivation of the GM bound to the $(n, 2, 2)$ setting and show that the threshold for η decreases from $\eta \approx 0.8284$.

The intuition then is to find inequalities where the maximal quantum violation is larger than for the $(2, 2, 2)$ case. For the $(n, 2, 2)$ case, WW have shown [Werner2001] that the quantum violation of the Mermin-Klyshko inequalities [Mermin1990, Belinskii1993, Gisin1998] (and inequalities in its orbit) is the largest violation for any $(n, 2, 2)$ inequality. There is only one vertex of $\mathcal{P}$ that maximally violates this inequality (for n being even), as shown by Marcovitch

[5] One would need to find the facet Bell inequalities for 3 or more spin-0 particles.

and Reznik [Marcovitch2008]. We will restrict ourselves to the cases $(n, 2, 2)$ for n being even. We shall describe the odd n case as an extension of the even case.

The vertex of $\mathcal{P}$ that maximally violates the Mermin inequality for n being even is $p(1|\mathbf{s}) = \delta^1_{f(\mathbf{s})}$ where $f(\mathbf{s}) = \bigoplus_{j=1}^{n-1} s_j(\bigoplus_{k=j+1}^{n} s_k)$ [Marcovitch2008]. When we refer to $f(\mathbf{s})$ in this subsection we mean this function in particular. If we were to allow communication then this function could be performed deterministically. One method would be if each jth party received the inputs s_k for all kth parties where $(j+1) \leq k \leq n$ and $1 \leq j \leq (n-1)$. Each party did not even need to learn every other party's input. This protocol also works if we cyclically permute the parties as the function $f(\mathbf{s})$ is invariant under all permutations of parties. We now show that this communication protocol can be "simulated" if we perform detection post-selection.

We now describe how we can achieve the vertex $\tilde{p}(1|\mathbf{s}) = \delta^1_{f(\mathbf{s})}$ of $\tilde{\mathcal{P}}$ corresponding to the function $f(\mathbf{s}) = \bigoplus_{j=1}^{n-1} s_j(\bigoplus_{k=j+1}^{n} s_k)$ with LHV correlators. We do this by simulating the above communication protocol using detection post-selection. We call this post-selection protocol the "Mermin-Klyshko post-selection" (MKP) protocol: each kth party for $2 \leq k \leq n$ produces the map $t_k = s_k \oplus y_k(\lambda) \oplus 1$ where all n parties share the $(n-1)$ bit-values $y_k(\lambda)$. Party 1 produces the map $t_1 = 1$. As before, the variables $y_k(\lambda)$ are randomly generated. Therefore after detection post-selection, all parties have mapped the inputs s_k for $j \neq 1$ onto the shared variables $y_k(\lambda)$. Then each jth party for $1 \leq j \leq (n-1)$ outputs the value $m_j = s_j(\bigoplus_{k=j+1}^{n} y_k(\lambda)) = s_j(\bigoplus_{k=j+1}^{n} s_k)$ and the nth party outputs $m_n = 0$. As a result, we obtain the correlator $\tilde{p}(1|\mathbf{s}) = \delta^1_{f(\mathbf{s})}$. It is worth noting that we need all $(n-1)$ maps $t_k = s_k \oplus y_k(\lambda) \oplus 1$ so that the first party can obtain all other inputs.

As in the previous subsection, in order to consider the detection efficiency we need to consider the space $\bar{\mathcal{L}}_{\mathbb{E}}$. We need to consider the probabilities $p(\mathbf{t} = \mathbf{1})$ for the LHV maps t_j. The MKP protocol produces $p(\mathbf{t} = \mathbf{1}) = \frac{1}{2^{(n-1)}}$. However, $p(t_1 = 1) = 1$ in this protocol. To counter this the n parties share the variable $z(\lambda) \in \{1, 2, ..., n\}$ which corresponds to each cyclic permutation of the n parties. This variable is randomly generated and then the n parties produce the MKP protocol but for a particular cyclic permutation. As a result, $p(\mathbf{t} = \mathbf{1}) = \frac{1}{n}\frac{1}{2^{(n-1)}}n = \frac{1}{2^{(n-1)}}$.

Therefore, LHV theories can produce a convex combination of expectation values $\bar{\mathbb{E}}(\mathbf{s}) = (-1)^{g(\mathbf{s})}$ and $\bar{\mathbb{E}}(\mathbf{s}) = \frac{1}{2^{(n-1)}}(-1)^{f(\mathbf{s})}$ where $f(\mathbf{s})$. We can substitute

these expectation values in the Mermin-Klyshko inequality for even n

$$\frac{1}{2^{\frac{n}{2}-1}}\sum_{\mathbf{s}}(-1)^{f(\mathbf{s})}\bar{\mathbb{E}}(\mathbf{s})\leq 2, \tag{4.15}$$

For the expectation value $\bar{\mathbb{E}}(\mathbf{s})=\frac{1}{2^{(n-1)}}(-1)^{f(\mathbf{s})}$, the Bell expression takes the value $2^{2-\frac{n}{2}}$. This inequality is therefore not violated. For odd n, the Mermin-Klyshko inequality can be rewritten as [Marcovitch2008]

$$\frac{1}{2^{\frac{n-1}{2}-1}}\sum_{\mathbf{s}}\delta^{s_n}_{\bigoplus_{j=1}^{(n-1)}s_j}(-1)^{f(\mathbf{s})}\bar{\mathbb{E}}(\mathbf{s})\leq 2. \tag{4.16}$$

We can use the same argument for even n to show that this inequality is not violated for any vector of expectation values in $\bar{\mathcal{L}}_{\mathbb{E}}$. First, one can use the MKP protocol, as for even n, to give $\bar{\mathbb{E}}(\mathbf{s})=\frac{1}{2^{(n-1)}}(-1)^{f(\mathbf{s})}$. This gives a value of $2^{\frac{3-n}{2}}$ and so does not lead to a violation. On the other hand, due to the delta function $\delta^{s_n}_{\bigoplus_{j=1}^{(n-1)}s_j}$, the function $f(\mathbf{s})$ is now independent of s_n and becomes $f'(\mathbf{s})=\bigoplus_{j=1}^{n-2}s_j(\bigoplus_{k=j+1}^{n-1}s_k\oplus 1)$. This function can be achieved by $(n-1)$ parties carrying out the MKP protocol, thus producing a value of $2^{\frac{5-n}{2}}$ for the Bell expression. In summary then, the Mermin-Klyshko inequality is not violated for all expectation values in $\bar{\mathcal{L}}_{\mathbb{E}}$.

It now remains to express $p(\mathbf{t}=\mathbf{1})$ in terms of detector efficiency η. Again we assume that η is the same for all sites and so can be calculated from the number counts $N_{\mathbf{t}}$ (for $\mathbf{t}\neq\mathbf{0}$). Therefore taking the limit of $N=\sum_{\mathbf{t}\neq\mathbf{0}}N_{\mathbf{t}}\rightarrow\infty$, the efficiency is

$$\eta=\frac{N_{\{1,\mathbf{t}'\}}}{N_{\{1,\mathbf{t}'\}}+N_{\{0,\mathbf{t}'\}}}, \tag{4.17}$$

where $\mathbf{t}'\neq\mathbf{0}$ is any of the bit-strings for all of the 2-party sub-sets of all 3 parties. The notation $\{0,\mathbf{t}'\}$ ($\{1,\mathbf{t}'\}$) then says that the other bit not in the sub-set $\mathbf{t}'$ is 0 (1). We can obtain values of $N_{\{0,\mathbf{t}'\}}$ in terms of η and $N_{\{1,\mathbf{t}'\}}$ and substitute them into an expression for $p(\mathbf{t}=\mathbf{1})$ (using recursion) to obtain

$$p(\mathbf{t}=\mathbf{1})=\frac{\eta^n}{1-(1-\eta)^n}. \tag{4.18}$$

If we substitute the maximal quantum violation of the Mermin-Klyshko inequality $2^{\frac{n+1}{2}}$ for the expectation values $\bar{\mathbb{E}}(\mathbf{s})$ then we have the following expressions $p(\mathbf{t}=\mathbf{1})=2^{\frac{1-n}{2}}$. Therefore, for $n=3$, detection efficiency must satisfy $\eta>\frac{1}{2}(\sqrt{21}-3)\approx 0.7913$ in order to demonstrate a loophole-free violation of

a Bell inequality. Whilst this is a decrease from the GM bound, this value of η does not decrease dramatically; for example for $n = 25$, $\eta \gtrsim 0.7170$ but for $n = 75$, $\eta \gtrsim 0.7104$. The bound of $\eta > \frac{2}{3}$ found by Eberhard (and subsequently improved) is more effective for a loophole-free Bell test [Eberhard1993].

While these generalisations of the GM bound on η may not be impressive compared to the current literature, our discussion has been motivated by a qualitative description of loopholes. We have also connected the detection loophole to communication protocols (cf. [Barrett2011]). Detection post-selection can simulate communication between parties by correlating input data to shared hidden variables. We used this simulation of a communication protocol to derive these generalisations of the GM bound. We have also used our computational description of all possible LHV maps to make this loophole-producing mechanism clear.

4.1.4 Summary of Loopholes

We have discussed how experimental imperfections in Bell tests can lead to loopholes. We have briefly covered how loss of measurement freedom and no space-like separation can lead to loopholes. In more detail, we have discussed how the subtleties of the detection loophole can be made clearer with the language of Boolean functions. Our language in terms of computational expressiveness allowed us to redrive the GM bound and generalised it to n parties.

Beyond the loopholes we have discussed already, we will now briefly mention another: the memory loophole [Barrett2002]. The memory loophole emerges if parties retain their choice of input and subsequent output in a "memory" that can be communicated between parties in-between tests. From this memory, parties can make "educated guesses" about which measurement outputs to give for a particular input. This problem occurs from a finite number N of Bell tests from which we produce correlation statistics. However, the loophole does not become an issue as $N \to \infty$ [Barrett2002], heuristically, the region of the LHV polytope outside of the linear Boolean functions disappears exponentially in N. Since we have assumed that all statistics from experiments are obtained in this limit, the memory loophole is not a conceptual, problematic issue.

In the next section, we look again at post-selection but not from an experimental point-of-view. We will assume that Bell tests are perfectly implemented in the laboratory. The post-selection introduced establishes a relationship between measurement data in a non-trivial fashion. We have shown that with detection

post-selection, relationships are induced between hidden variables and measurement settings, thus leading to loopholes. In this new setting we will define a loophole in analogy to the definition in this section. Given this definition, we show that this new form of post-selection is free of loopholes. This new method is a way of conceptually modifying Bell tests but not modifying the implications of LHV theories.

4.2 Loophole-free Post-selection and Quantum Correlators

In the previous section, post-selection was a necessity in order to calculate correlators. For non-detection events, measurement outcomes are not defined so the sum modulo 2 of outcomes could not be calculated. We now explore the use of post-selection utilised by the experimenter out of choice rather than necessity. We assume that the Bell test has perfect detectors and the experimenter does not need to use detection post-selection. Therefore, data is perfectly obtained by the experimenter but they still choose to discard some of this data. We will construct a new model to reflect this choice and discuss the possibility of loopholes in this model.

For all of the discussion so far in this thesis, the variable $\mathbf{s}$ for each correlator $p(k|\mathbf{s})$ has two functions: 1) it labels the inputs to all sites corresponding to the choice of measurement settings; 2) $\mathbf{s}$ acts as a conditioning variable for the probability measure on all maps $\mathbb{Z}_2^n \to \mathbb{Z}_2$. In this section we will distinguish between these two roles by using post-selection on measurement data. This is done by relating measurement data to data that is independent of measurement settings or outcomes. We motivate this discussion by returning to the Mermin inequality [Mermin1990]:

$$p(1|000) + p(1|011) + p(1|101) - p(1|110) \leq 2 \tag{4.19}$$

and recall that as in the GHZ paradox, we are interested in correlators when $s_3 = s_1 \oplus s_2$. In the language of computer science, this is called a *promise* on the inputs that they satisfy a particular relation [Cleve2004]. This inequality is also superficially similar to the CHSH inequality but now with a third party whose inputs are related to the other two sites.

To make the connection to the CHSH inequality clearer, we notice that the

linear Boolean functions that LHV theories can achieve if $s_3 = s_1 \oplus s_2$ are written as $m_1 \oplus m_2 \oplus m_3 = \alpha s_1 \oplus \beta s_2 \oplus \gamma s_3 \oplus \delta = \alpha' s_1 \oplus \beta' s_2 \oplus \delta$ for α, β, α', β', γ, $\delta \in \{0, 1\}$. These functions are exactly the linear Boolean functions for the $(2, 2, 2)$ CHSH setting. Therefore, the linear Boolean functions that satisfy the CHSH inequality also satisfy the Mermin inequality. The Mermin inequality can be seen as a manifestation of the CHSH inequality.

If we reconsider experimental implementations of Bell tests, then how do parties obtain the input $s_3 = s_1 \oplus s_2$ if they are space-like separated from the other two parties? A possible solution is through data post-selection; the third party makes a completely random choice of s_3. After receiving all data $\mathbf{m}$ and $\mathbf{s}$ from all parties the experimenter only accepts data from all parties and calculate $p(1|\mathbf{s})$ if $s_3 = s_1 \oplus s_2$; otherwise data is discarded. Since s_1 and s_2 are also randomly generated, the rate at which the experimenter discards the data will tend to $\frac{1}{2}$ for $N \to \infty$ runs of the experiment.

We will proceed to generalise this method of post-selection utilised in the GHZ paradox. Central to this approach will be the linear Boolean functions. In the example of the GHZ paradox, the experimenter post-selects on one input being a linear Boolean function. This keeps the computational power of the LHV correlators confined to these linear functions. We showed in the previous section that loopholes can lead to LHV correlators performing non-linear Boolean functions. In analogy with the detection loophole, the post-selection in the GHZ paradox can be seen to avoid a loophole. This is the central insight in this section and we will now develop these ideas rigorously.

4.2.1 Post-selection, Linearity and Loopholes

We now introduce some more general structure beyond the GHZ paradox. The experimenter now has some bit-string $\mathbf{x}$ of length $|\mathbf{x}| \leq n$. Referring back to the two roles of $\mathbf{s}$ described above, $\mathbf{x}$ now plays the role of conditioning variable (role 2). That is, instead of the stochastic maps $p(1|\mathbf{s})$ being conditioned upon $\mathbf{s}$, they are now conditioned upon $\mathbf{x}$, i.e. the experimenter calculates $p(1|\mathbf{x}) = p(\bigoplus_{j=1}^{n} m_j = 1|\mathbf{x})$. The experimenter then relates their data $\mathbf{x}$ to the experimental data $\mathbf{m}$ and $\mathbf{s}$.

If we return to the GHZ paradox, we have three parties but the bit-string $\mathbf{x} \in \{0, 1\}^2$. The experiment now accepts, or post-selects on data $\mathbf{m}$ and $\mathbf{s}$ when $s_1 = x_1$, $s_2 = x_2$ and $s_3 = x_1 \oplus x_2$ is satisfied. Then the experimenter calculates

the correlator

$$p(1|\mathbf{x}) = p(m_1 \oplus m_2 \oplus m_3 = 1 | s_1 = x_1, s_2 = x_2, s_3 = x_1 \oplus x_2). \tag{4.20}$$

A relationship between data $\mathbf{s}$ and $\mathbf{x}$ is established by the experimenter's post-selection. We now generalise this approach of relating $\mathbf{s}$ to $\mathbf{x}$ with the following form of post-selection.

Definition 10. If an experimenter accepts, or post-selects on data $\mathbf{m}$ and $\mathbf{s}$ for every input s_j satisfying $s_j = g_j(\mathbf{x})$ where $g_j(\mathbf{x})$ is some Boolean function on $\mathbf{x}$, this is **input post-selection**. The experimenter fixes this relationship between $\mathbf{x}$ and $\mathbf{s}$ for all runs of the experiment. After post-selection, the experimenter calculates $p(1|\mathbf{x})$ for their value of $\mathbf{x}$.

After the post-selection there are now $2^{|\mathbf{x}|}$ correlators $p(1|\mathbf{x})$ for all values of $\mathbf{x}$. Just as with the correlators $p(1|\mathbf{s})$, the correlators $p(1|\mathbf{x})$ are elements of vectors $\vec{k}_{\mathbf{x}} \in \mathbb{R}^{2^{|\mathbf{x}|}}$. The deterministic correlators are $p(1|\mathbf{x}) = \delta^1_{f(\mathbf{x})}$ for all Boolean functions $f(\mathbf{x})$. Therefore, vectors $\vec{k}_{\mathbf{x}}$ are contained in a convex polytope $\mathcal{P}_{\mathbf{x}}$ with these extreme points being these deterministic correlators. There will also be the space of LHV correlators $\mathcal{L}_{\mathbf{x}}$ in analogy to $\mathcal{L}$. If $n = |\mathbf{x}|$ and the functions $g_j(\mathbf{x})$ in input post-selection are $s_j = x_j$, then we recover the original $(n, 2, 2)$ Bell test. For this example, $\mathcal{L}_{\mathbf{x}}$ is the convex hull of linear Boolean functions on $\mathbf{x}$. In analogy with the detection loophole defined in the previous section, we now define a loophole for input post-selection.

Definition 11. A **loophole** is introduced by an experimenter into a Bell test if after input post-selection, the space $\mathcal{L}_{\mathbf{x}}$ is larger than the convex hull of linear Boolean functions on $\mathbf{x}$.

In the next subsection we will show how loopholes are avoided if the experimenter utilises input post-selection. We will then develop input post-selection in subsection 4.2.3, now to encompass a relationship between $\mathbf{x}$ and *both* $\mathbf{m}$ and $\mathbf{s}$. This new form of post-selection will be called "output-input post-selection". In this case, we can still find a way to avoid loopholes in the sense that $\mathcal{L}_{\mathbf{x}}$ remains the convex hull of linear Boolean functions. We now address loopholes in input post-selection.

4.2.2 Linear Input Post-selection in $(n, 2, 2)$ tests

We begin our discussion with a key result for setting post-selection. This result informs us of how to avoid loopholes and will lead to us describing a particular class of input post-selections.

Theorem 17. *The space $\mathcal{L}_{\boldsymbol{x}}$ of LHV correlators is the convex hull of linear Boolean functions on $\boldsymbol{x}$ for input post-selections with $s_j = g_j(\boldsymbol{x})$ if and only if every $g_j(\boldsymbol{x})$ is a linear Boolean function on $\boldsymbol{x}$.*

Proof: First we prove the if statement. We only need to consider the extreme points of $\mathcal{L}$ corresponding to the deterministic linear Boolean functions $f(\mathbf{s})$ on $\mathbf{s}$, i.e. $f(\mathbf{s}) = \bigoplus_{j=1}^{n} a_j s_j \oplus b$ with a_j, $b \in \{0, 1\}$. If we post-select on $s_j = g_j(\mathbf{x})$ being a linear Boolean function, then $f(\mathbf{s}) \to f(\mathbf{x}) = \bigoplus_{j=1}^{n} a_j g_j(\mathbf{x}) \oplus b$, which is again a linear Boolean function now on $\mathbf{x}$. To prove the only if statement, if $g_j(\mathbf{x})$ is a non-linear Boolean function, then extreme points of $\mathcal{L}$ producing $f(\mathbf{x}) = \bigoplus_{j=1}^{n} a_j g_j(\mathbf{x}) \oplus b$ will in general be a non-linear Boolean function for all a_j and b. □

From this result, $\mathcal{L}_{\mathbf{x}}$ will be defined by the facet Bell inequalities for the $(|x|, 2, 2)$ setting replacing $p(1|\mathbf{s})$ with $p(1|\mathbf{x})$. Returning to our example, the Mermin inequality (with replacing $\mathbf{s}$ with $\mathbf{x}$) is the CHSH-like inequality defining a facet of $\mathcal{L}_{\mathbf{x}}$ with $|\mathbf{x}| = 2$. This all occurs only if the input post-selection consists of $g_j(\mathbf{x})$ being a linear Boolean function. We now formally define this particular class of post-selections:

Definition 12. Linear Input Post-selection (LI) is input post-selection but where all of the functions $g_j(\mathbf{x})$ are linear Boolean functions on $\mathbf{x}$.

This post-selection can be seen to simulate nMBQC as described in section 3.4 of chapter 3. Instead of pre-processing on inputs which are then distributed to n parties, we post-select on inputs satisfying the expressions that are described by the $\mathbf{P}$ matrices. Since LHV resources can only produce linear Boolean functions in nMBQC, then our post-selection simulates a model with the same computational power. Crucially both the pre-processing and post-selection is restricted to the linear Boolean functions.

The connection to MBQC can now be extended by considering adaptivity. In adaptive MBQC, inputs, or measurement settings at each site are influenced by previous measurement outcomes. Translating this into a Bell test, the input

s_j is now a function $h(m_{j'}|\forall j' \neq j)$ of measurements outcomes $m_{j'}$ from other sites. Directly, this would assume communication between sites. However, if we post-select on inputs s_j satisfying this function $h(m_{j'}|\forall j' \neq j)$, then we can simulate this communication. We now discuss this form of post-selection and show, remarkably, that we can avoid loopholes.

4.2.3 Linear Output-Input Post-selection in $(n, 2, 2)$ tests

We now extend input post-selection to consider functional relationships induced between the experimenter's variable $\mathbf{x}$ and $\mathbf{m}$ and $\mathbf{s}$. In particular, for the jth site, s_j can be related to outcomes $m_{j'}$ for $j \neq j'$. We introduce the notation $\mathbf{m}^{\setminus j}$ to describe a $(n-1)$-length bit-string which is $\mathbf{m}$ but *without* the bit-value m_j. For example, if $j = 1$, then $\mathbf{m}^{\setminus j} = \{m_2, m_3, ..., m_n\}$. With this new piece of notation we now introduce a new form of post-selection, first studied by Hoban and Browne [Hoban2011b]. "Output-input post-selection" is now the same as input post-selection but the experimenter now accepts data when $s_j = g_j(\mathbf{x}, \mathbf{m}^{\setminus j})$ instead of $s_j = g_j(\mathbf{x})$. Again, after the post-selection, the experimenter again calculates $p(1|\mathbf{x})$ for each $\mathbf{x}$.

For this output-input post-selection, the space of all possible correlators is $\mathcal{P}_{\mathbf{x}}$, the same as input post-selection. For LHV correlators, we describe the space of correlators after output-input post-selection as $\mathcal{L}_{\mathbf{x},\mathbf{m}^{\setminus j}}$. As an extension of the definition of a loophole for input post-selection, a loophole emerges if $\mathcal{L}_{\mathbf{x},\mathbf{m}^{\setminus j}}$ is larger than the convex hull of linear Boolean functions on $\mathbf{x}$. We now show when loopholes in output-input post-selection can be avoided.

Theorem 18. *The space $\mathcal{L}_{\boldsymbol{x},\boldsymbol{m}^{\setminus j}}$ of LHV correlators is the convex hull of linear Boolean functions on $\boldsymbol{x}$ for output-input post-selections $s_j = g_j(\boldsymbol{x}, \boldsymbol{m}^{\setminus j})$ if and only if every $g_j(\boldsymbol{x}, \boldsymbol{m}^{\setminus j})$ is a linear Boolean function on $\boldsymbol{x}$ and $\boldsymbol{m}^{\setminus j}$.*

Proof: First we prove the if statement. We recall that all deterministic LHV single-site maps can be written as $m_j = \alpha_j s_j \oplus \beta_j$ and we can take their convex combination. We assume that α_j and β_j is dependent on an LHV λ but these variables are in no way correlated with the inputs $\mathbf{s}$. Therefore, all extreme points of $\mathcal{L}$ from these deterministic maps result in $\bigoplus_{j=1}^{n} m_j$ being a linear Boolean function on $\mathbf{x}$ and $\mathbf{m}^{\setminus j}$.

If we do not assume that the values α_j and β_j for all j are not correlated to $\mathbf{s}$, there is a way in which this post-selection can allow correlations between bits from the LHV, α_j, β_j and inputs $\mathbf{s}$. We now demonstrate this method. We can

decompose a linear function $g_j(\mathbf{x}, \mathbf{m}^{\setminus j})$ as $g_j^{(1)}(\mathbf{m}^{\setminus j}) \oplus g_j^{(2)}(\mathbf{x})$, i.e. in terms of the linear functions $g_j^{(1)}(\mathbf{m}^{\setminus j})$ and $g_j^{(2)}(\mathbf{x})$ on $\mathbf{m}^{\setminus j}$ and $\mathbf{x}$ respectively. The outcomes in $\mathbf{m}^{\setminus j}$ contain information about λ, but s_j is random and uncorrelated to λ, $\mathbf{m}$ and $\mathbf{x}$. Therefore $g_j^{(1)}(\mathbf{m}^{\setminus j}) = g_j^{(2)}(\mathbf{x}) \oplus s_j$ means that $g_j^{(1)}(\mathbf{m}^{\setminus j})$ is random and uncorrelated to $g_j^{(2)}(\mathbf{x})$[6]. These random bits s_j play the role of the pad-bit in one-time pad cryptography which Shannon [Shannon1949] proved is perfectly secure for encrypting messages.

We finally prove the only if statement. If $g_j(\mathbf{m}^{\setminus j}, \mathbf{x})$ becomes non-linear then we can always produce this function $f(\mathbf{x}) = g_j(\mathbf{m}^{\setminus j}, \mathbf{x})$ as an output. Since values of $\mathbf{m}^{\setminus j}$ can be made to be equal to values of $\mathbf{x}$, there always exists a non-linear function in $\mathbf{x}$ if $g_j(\mathbf{m}^{\setminus j}, \mathbf{x})$ is non-linear. □

We now call output-input post-selection where $g_j(\mathbf{x}, \mathbf{m}^{\setminus j})$ is a linear Boolean function on $\mathbf{x}$ and $\mathbf{m}^{\setminus j}$, **Linear Output-Input Post-selection** (LOI). With LOI, we can simulate signalling processes by making inputs dependent on outputs at other sites. But, we can also keep the space of correlators confined to the linear Boolean functions on $\mathbf{x}$. This means that for all $n \geq |\mathbf{x}|$, the space $\mathcal{L}_{\mathbf{x},\mathbf{m}^{\setminus j}}$ for LOI is $\mathcal{L}_{\mathbf{x}}$, the convex hull of linear Boolean functions. The n-independence in the space of correlators is unusual given that in traditional Bell tests, the role of the number of parties is important. In some way, by considering $|\mathbf{x}|$, we unify all possible multi-party Bell settings for $n \geq |\mathbf{x}|$.

With regards to quantum correlators, we have already indicated that there is an n-dependence in the example of the Mermin inequality. For $n = |\mathbf{x}| = 2$, the maximal violation of the CHSH inequality is $2\sqrt{2}$. After LI, for $n = 3$ and $|\mathbf{x}| = 2$, the same CHSH inequality in terms of $p(1|\mathbf{x})$ has the maximal violation of 4. In the next subsection we will discuss the effect of LI and LOI upon the space of quantum correlators.

4.2.4 Bipartite Quantum correlators under post-selection

The space $\mathcal{Q}_{\mathbf{x}}$ of quantum correlators under LI needs to be specified for a particular value of n, i.e. $\mathcal{Q}^n_{\mathbf{x}}$. For LOI, the corresponding space of quantum correlators is $\mathcal{Q}^n_{\mathbf{x},\mathbf{m}^{\setminus j}}$ for n number of parties. Since LOI includes all possible post-selections in LI, then necessarily $\mathcal{Q}^n_{\mathbf{x}} \subseteq \mathcal{Q}^n_{\mathbf{x},\mathbf{m}^{\setminus j}}$.

[6] If $g_j^{(2)}(\mathbf{x}) = 0$, the bit s_j does become correlated with other sites' measurements m_k and hence λ but s_j will be uncorrelated to $\mathbf{x}$. If $g_j^{(1)}(\mathbf{m}^{\setminus j}) = 0$, we recover LI post-selection.

We now focus on $|\mathbf{x}| = 2$ as the smallest example of non-trivial behaviour of $\mathcal{Q}^n_{\mathbf{x}}$ and $\mathcal{Q}^n_{\mathbf{x},\mathbf{m}\setminus j}$. Since $\mathcal{Q}^2_{\mathbf{x}}$ is strictly smaller than $\mathcal{P}_{\mathbf{x}}$ for $|\mathbf{x}| = 2$, we can initially ask whether $\mathcal{Q}^2_{\mathbf{x},\mathbf{m}\setminus j}$ is larger than $\mathcal{Q}^2_{\mathbf{x}}$? This turns out not to be the case as we now demonstrate. For this situation, the most general LOI possible involves post-selecting on the following relations being satisfied: $s_1 = x_1 \oplus \alpha m_2$ and $s_2 = x_2 \oplus \beta m_1$ with $\alpha, \beta \in \{0, 1\}$. When $\alpha = \beta = 0$, we retrieve the standard, well-studied scenario. The two scenarios where $\alpha \neq \beta$ are equivalent up to changing of labels. If we consider the scenario where $\{\alpha, \beta\} = \{0, 1\}$ then the probabilities $p(1|\mathbf{x})$ can be rewritten in terms of probabilities $p(m_1, m_2|s_1, s_2)$:

$$\begin{aligned} p(1|\mathbf{x}) &= \sum_{m_1,m_2} \delta_1^{m_1 \oplus m_2} p(m_1, m_2|s_1 = x_1, s_2 = x_2 \oplus m_1) \\ &= p(0, 1|s_1 = x_1, s_2 = x_2) + p(1, 0|s_1 = x_1, s_2 = x_2 \oplus 1). \end{aligned} \quad (4.21)$$

The correlator can be written in this way as $p(\mathbf{m}|\mathbf{s})$ is a non-signalling distribution. Any non-signalling probability distribution can be written as a convex combination of the vertices of $\mathcal{NS}$. For the bipartite scenario there are two types of vertices: 1) local vertices where $p(m_1, m_2|s_1, s_2) = \prod_{j=1}^2 \delta^{m_j}_{a_j s_j \oplus b_j}$ for a_j, $b_j \in \{0, 1\}$; and 2) "non-local" vertices for $p(m_1, m_2|s_1, s_2) = \frac{1}{2}$ for $m_1 \oplus m_2 = (s_1 \oplus a)(s_2 \oplus b) \oplus c$, and 0 otherwise where a, b, $c \in \{0, 1\}$. In the former case, when $s_1 = x_1$, m_1 is either 0 or 1 deterministically and so (4.21) must be 1 for a local vertex. For a non-local vertex, (4.21) takes any of the values $\{0, \frac{1}{2}, 1\}$, so for non-signalling distributions, every correlator of the form (4.21) is at most 1.

To see if any correlators are outside of $\mathcal{L}_{\mathbf{x}}$, we put the correlators in (4.21) into the CHSH inequality (and any in its symmetry group) to obtain

$$\begin{aligned} &p(0, 1|0, 0) + p(1, 0|0, 0) + p(0, 1|0, 1) + p(1, 0|0, 1) + \\ &\quad p(0, 1|1, 0) - p(1, 0|1, 0) - p(0, 1|1, 1) + p(1, 0|1, 1) \quad \leq \quad 2. \end{aligned} \quad (4.22)$$

All bipartite non-local vertices satisfy

$$p(0, 1|1, 0) - p(1, 0|1, 0) = -p(0, 1|1, 1) + p(1, 0|1, 1) = 0. \quad (4.23)$$

Therefore, all non-signalling probability distributions do not violate the CHSH inequality with LOI for $\alpha \neq \beta$. As a corollary, quantum correlators satisfy the

CHSH inequality[7].

Finally, for the scenario of LOI with $\alpha = \beta = 1$, then $m_1 \oplus m_2 = s_1 \oplus s_2 \oplus x_1 \oplus x_2$. The correlators $p(1|\mathbf{x})$ are calculated when $s_1 \oplus s_2 = x_1 \oplus x_2 \oplus 1$, so $p(1|\mathbf{x}) = p(1|\mathbf{x}')$ for $\mathbf{x} = \mathbf{x}'$ if $x_1 \oplus x_2 = x_1' \oplus x_2'$. Substituting these values of the correlators into all of the CHSH inequalities never yields a violation, as two of the correlators will cancel[8]. To summarise then, for LOI for $\alpha \neq \beta$ or $\alpha = \beta = 1$, quantum correlators do not exceed the LHV polytope. Therefore the space of quantum correlators $\mathcal{Q}^2_{\mathbf{x},\mathbf{m}\setminus j} = \mathcal{Q}^2_{\mathbf{x}}$, with $\mathcal{Q}^2_{\mathbf{x}}$ being $\mathcal{Q}$ for $(2, 2, 2)$.

4.2.5 Multipartite quantum correlators

We have looked at the scenario when $n = 2$, we will now consider the space of quantum correlators $\mathcal{Q}^n_{\mathbf{x},\mathbf{m}\setminus j}$ for general n and $|\mathbf{x}|$. Having shown that LIO has no impact on quantum correlators, we now show the opposite in the multipartite setting. That is, the space of quantum correlators under LIO can be *larger* than the space of quantum correlators under LI. We begin by considering the $n = 3$ scenario and then use it to consider larger n for a particular $|\mathbf{x}|$.

Firstly, we observe that $\mathcal{Q}^3_{\mathbf{x},\mathbf{m}\setminus j} = \mathcal{P}_{\mathbf{x}}$ for $|\mathbf{x}| = 2$. From the GHZ paradox, $p(m_1 \oplus m_2 \oplus m_3 = 1|\mathbf{x}) = \delta_1^{x_1 x_2 \oplus 1} = x_1 x_2 \oplus 1$. We can map from the function $x_1 x_2 \oplus 1$, to the other non-linear Boolean functions $(x_1 \oplus a)(x_2 \oplus b) \oplus c$ (with a, b, $c \in \{0, 1\}$) with relabelling of bit-values x_j. All vertices of $\mathcal{P}_{\mathbf{x}}$ can be achieved by quantum correlators for $n = 3$ and $|\mathbf{x}| = 2$.

We might ask whether quantum correlators can saturate the whole of $\mathcal{P}_{\mathbf{x}}$ for particular values of n and $|\mathbf{x}|$? In the following lemma, quantum correlators for a given n and $|\mathbf{x}|$ can saturate the whole space $\mathcal{P}_{\mathbf{x}}$. In particular, any vertex of $\mathcal{P}_{\mathbf{x}}$ corresponding to non-linear Boolean functions can be attained with quantum correlators for a particular n. We use the $n = 3$, $|\mathbf{x}| = 2$ case to demonstrate this fact.

Lemma 19. *For all $|\boldsymbol{x}|$, $\mathcal{Q}^n_{\boldsymbol{x},\boldsymbol{m}\setminus j}$ contains the vertex $p(1|\boldsymbol{x}) = \delta^1_{f(\boldsymbol{x})}$ for $f(\boldsymbol{x}) = \prod_{j=1}^{|\boldsymbol{x}|} x_j$ if $n = 3(|\boldsymbol{x}| - 1)$.*

Proof: We prove this with an explicit LIO protocol. If we have $n = 3y$ parties for y as some non-zero positive integer, and divide them into y sets of three neigh-

[7] The inequalities in the CHSH inequality symmetry group are also not violated as we can map to all inequalities in this group via local re-labellings or an overall sign change, and we can also map from every non-local vertex of $\mathcal{NS}$ via the same operations

[8] After the terms that are equal but have opposite sign pre-factors in the inequality cancel, the inequalities reduce to either $2p(1|\mathbf{x}) \leq 2$ or $-2p(1|\mathbf{x}) \leq 0$ for a particular value of $\mathbf{x}$

bouring parties in the following way $\{\{1,2,3\},...,\{3y-2,3y-1,3y\}\}$. For each of these y sets $\{j,(j+1),(j+2)\}$, if inputs are s_j, $s_{(j+1)}$ and $s_{(j+2)} = s_j \oplus s_{(j+1)}$, then we can have $m_j \oplus m_{(j+1)} \oplus m_{(j+2)} = s_j s_{(j+1)}$ with quantum correlators[9]. For $j = 1$, if the experimenter post-selects upon $s_1 = x_1$, $s_2 = x_2$ and $s_3 = x_1 \oplus x_2$, then we have the situation for $n = 3$ and $|\mathbf{x}| = 2$ discussed above.

For $j = 3k+1$ with $k \in \{1,2,...,(y-1)\}$, the experimenter post-selects data if $s_j = \bigoplus_{l=1}^{j-1} m_l$, $s_{(j+1)} = x_{k+2} \oplus 1$ and $s_{(j+2)} = \bigoplus_{l=1}^{j-1} m_l \oplus x_{k+2} \oplus 1$. For $j = 4$, this results in $s_4 = x_1 x_2$, $s_5 = x_3 \oplus 1$ and $s_6 = x_1 x_2 \oplus x_3 \oplus 1$, and so $m_4 \oplus m_5 \oplus m_6 = x_1 x_2 (x_3 \oplus 1)$, resulting in $\bigoplus_{l=1}^{6} m_l = x_1 x_2 x_3$. Then by iteration, for $k \geq 2$, the above protocol results in $\bigoplus_{l=1}^{3y} m_l = \prod_{l=1}^{y+1} x_l$. Therefore, if $y = |\mathbf{x}| - 1$, the function $f(\mathbf{x}) = \prod_{l=1}^{|\mathbf{x}|} x_l$ can be achieved deterministically with $n = 3(|\mathbf{x}| - 1)$. □

If we consider LI, we know that for $n = 2^{|\mathbf{x}|} - 1$, every Boolean function can be achieved deterministically. This is because LI can simulate nMBQC directly, and in nMBQC we need at most this number of parties to achieve all Boolean functions. Translated into the language of post-selection, $\mathcal{Q}^n_{\mathbf{x},\mathbf{m}^{\setminus j}} = \mathcal{P}_{\mathbf{x}}$ for $n = 2^{|\mathbf{x}|} - 1$. What is more, we showed in Theorem 14, that to achieve $p(1|\mathbf{x}) = \delta^1_{f(\mathbf{x})}$ for $f(\mathbf{x}) = \prod_{j=1}^{|\mathbf{x}|} x_j$ with nMBQC, we require no fewer than $n = 2^{|\mathbf{x}|} - 1$ parties. For $|\mathbf{x}| \geq 3$, $2^{|\mathbf{x}|} - 1 > 3(|\mathbf{x}| - 1)$. This then gives us the following result.

Theorem 20. *$\mathcal{Q}^n_{\boldsymbol{x},\boldsymbol{m}^{\setminus j}}$ can be be larger than $\mathcal{Q}^n_{\boldsymbol{x}}$ for a fixed n and $|\boldsymbol{x}|$.*

If we utilise LOI for a particular number n of parties, then we can get a larger violation of a Bell inequality with this LOI than with LI. Quantum correlators can be perceived to be "more non-local" if we process our measurement data in a particular way. Then the action of discarding data can not only allow classical, or LHV correlators to simulate quantum correlators (as in the detection loophole), but used to emphasize the non-classical aspect of quantum physics.

The LOI can simulate a circuit where some outputs can affect some inputs. Traditionally, Boolean circuits have sequential gates so there is a temporal order of processes. In our post-selection, Boolean functions result from resources that are without temporal order or space-like separated. There has been a great deal of research into a field called Boolean circuit complexity and there is a natural

[9]Corresponding to the maximal quantum violation 1 of the Mermin inequality $-p(1|0,0,0) - p(1|0,1,1) - p(1|1,0,1) + p(1|1,1,0) \leq 0$, equivalent to the original Mermin inequality with the symmetry operation being adding 1 (modulo 2) to the joint outcome $m_1 \oplus m_2 \oplus m_3$.

overlap with discussion of this field to our discussion of LOI. Boolean circuit complexity asks how many fundamental operations or *gates* are required to perform any Boolean function (e.g. the AND and NOT gates) [Papadimitriou1994]. The application of Boolean circuit complexity results to LOI would be an interesting avenue of research.

In this section we have shown that post-selection can be used to conceptually change Bell tests. The post-selection described also establishes a link between Bell tests and the full MBQC model described by Briegel and Raussendorf [Raussendorf2001]. Whilst LI simulates nMBQC, the adaptivity in MBQC can be simulated by LOI. With this post-selection, processing on measurement data utilises addition modulo 2, as with the classical computer in MBQC. Heuristically, the Bell test with LOI is akin to a single round of measurements in MBQC, but we only accept the circuit if it corresponds to an adaptive circuit in MBQC; we discard the circuit otherwise. This is analogous to post-selected quantum state teleportation where we accept, post-select our system on the "correct" measurement outcome resulting in teleportation [Lloyd2011].

Central to our discussion in this section has been the computational description of correlators. We then used this computational description to consider input and output-input post-selection. We can limit the computational implications of this post-selection if we restrict ourselves to measurement data being related by linear Boolean functions. Linear Boolean functions are associated with LHV correlators. If all processing on data consists of linear Boolean functions, the computational power of LHV correlators remains linear. However, for general (n, c, d) scenarios, LHV correlators are associated with n-partite linear functions. We show in the next section, that generalising LI and LOI to these more general scenarios can be very problematic.

4.3 General settings and Input Post-selection

The discussion in the previous two sections 4.1 and 4.2 of this chapter have been in the $(n, 2, 2)$ scenario. We now consider generalisations of input post-selection to (n, d, d) scenarios where d is prime. We have demonstrated that we can avoid loopholes in the $(n, 2, 2)$ scenario, is this true for all d? In order to address this question we need to generalise the approach developed in the previous section. We will introduce two natural generalisations of the post-selection in LI, and we show that it is not loophole free. Despite this, we will again give an indication

that the region of quantum correlators can be *enlarged* by post-selection. We now proceed to introduce the framework for input post-selection in the (n, d, d) scenario.

As before, the experimenter has some $|\mathbf{x}|$-length digit string $\mathbf{x} \in \mathbb{Z}_d^{|\mathbf{x}|}$ which they have chosen. He receives data $\mathbf{m}$ and $\mathbf{s}$ from all n parties and then accepts this data if inputs s_j are equal to some function $g_j(\mathbf{x})$ on $\mathbf{x}$. If this function is not satisfied by all s_j then the experimenter discards this data. Once the data has been accepted by the experimenter they calculate $k = \left[\sum_{j=1}^{n} m_j\right]_d$ and produce the correlator $p(k|\mathbf{x})$.

The space of all possible correlators is $\mathcal{P}_{\mathbf{x}}$, the convex polytope of correlators $p(k|\mathbf{x}) = \delta^k_{f(\mathbf{x})}$ for any function $f(\mathbf{x}) : \mathbb{Z}_d^n \to \mathbb{Z}_d$. For the trivial post-selection $s_j = x_j$ where $n = |\mathbf{x}|$, then the space of LHV correlators is $\mathcal{L}_{\mathbf{x}}$: the convex hull of n-partite linear functions on $\mathbf{x}$. This space might be dependent on n, but this is implicitly assumed in our notation. As in the $(n, 2, 2)$ case, we define a **loophole** as a form of input post-selection that results in $\mathcal{L}_{\mathbf{x}}$ being larger than the convex hull of n-partite linear functions.

So far, this framework for all (n, d, d) scenarios for prime d is almost identical to the $(n, 2, 2)$ case. What is the generalisation of LI for this more general case? For $(n, 2, 2)$, the linear Boolean functions on $\mathbf{x}$ are both n-partite linear functions *and* the addition modulo 2 of variables x_j (upto some additional constant). In the (n, d, d) scenario, functions consisting of sums of elements x_j modulo d are a subclass of all n-partite linear functions; we call these functions **affine functions** on $\mathbf{x}$. In the next subsection we will consider input post-selection for affine functions $g_j(\mathbf{x})$. We will then consider the case where $g_j(\mathbf{x})$ is any n-partite linear function. In both cases, loopholes are introduced by the input post-selection. For all n-partite linear functions $g_j(\mathbf{x})$, the space of LHV correlators is equal to $\mathcal{P}_{\mathbf{x}}$ for some n; this is not possible for the affine functions $g_j(\mathbf{x})$. Finally we will briefly discuss the space of quantum correlators under input post-selection.

4.3.1 Input Post-selection with Affine Functions

We now consider input post-selection where $g_j(\mathbf{x})$ are the affine functions. The affine functions can be written as $h(\mathbf{x}) = \left[b + \sum_{j=1}^{|\mathbf{x}|} a_j x_j\right]_d$ for $a_j, b \in \mathbb{Z}_d$. It can be readily seen that for $d = 2$, these functions are the linear Boolean functions[10]. We now define the class of input post-selections for the affine functions.

[10]Linear Boolean functions are often referred to as affine Boolean functions.

Definition 13. Affine Input Post-selection (AI) is input post-selection where the experimenter accepts data when all s_j satisfy $s_j = h(\mathbf{x})$ where $h(\mathbf{x})$ is an affine function on $\mathbf{x}$.

The space of LHV correlators under AI is written as $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$. We are now in a position to present the following result that shows that AI introduces loopholes. Whilst loopholes are introduced, $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$ is still smaller than the space of all possible correlators. This fact will be utilised in subsection 4.3.3 to highlight the space of quantum correlators for AI.

Proposition 21. *The space $\mathcal{L}_{\boldsymbol{x}}^{AI}$ is larger than the convex hull of n-partite linear functions but smaller than $\mathcal{P}_{\boldsymbol{x}}$ for $n \geq |\boldsymbol{x}|$.*

Proof: We first use the results from chapter 2 to describe n-partite linear functions for $c = d$ being prime:

$$\begin{aligned} f(\mathbf{x}) &= \left[\alpha + \sum_{j=1}^{n}\sum_{k=1}^{(d-1)} \beta_{j,k}\left(1 - \sum_{l=0}^{(d-1)}(-1)^l \binom{(d-1)}{l} k^l (s_j)^{d-(l+1)}\right)\right]_d \\ &= \left[\alpha' + \sum_{j=1}^{n}\sum_{q=1}^{(d-1)} \beta'_{j,q}(s_j)^q\right]_d, \end{aligned} \tag{4.24}$$

with $\alpha \in \mathbb{Z}_d$ and $\beta_{j,k} \in \mathbb{Z}_d$ where

$$\begin{aligned} \alpha' &= [\alpha + \sum_{j=1}^{n}\sum_{k=1}^{(d-1)} \beta_{j,k}(1 - (-k)^{(d-1)})]_d \\ &= \alpha \end{aligned} \tag{4.25}$$

and

$$\beta'_{j,q} = \sum_{j=1}^{n}\sum_{k=1}^{(d-1)} \beta_{j,k}(-1)^{d-q}\binom{(d-1)}{d-(q+1)} k^{d-(q+1)}. \tag{4.26}$$

Thus n-partite linear functions are the sum modulo d of powers of s_j. In AI we calculate correlators $p(k|\mathbf{x})$ after post-selecting on $s_j = \left[b + \sum_{j=1}^{|\mathbf{x}|} a_j x_j\right]_d$ for a_j, $b \in \mathbb{Z}_d$. Therefore the extreme points of $\mathcal{L}$ corresponding to the n-partite linear functions get mapped to extreme points of $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$ with extreme points $p(k|\mathbf{x}) = \delta^1_{f(\mathbf{x})}$

corresponding to the functions

$$f(\mathbf{x}) = \left[\alpha' + \sum_{j=1}^{n} \sum_{q=1}^{(d-1)} \beta'_{j,q} (b + \sum_{j=1}^{|\mathbf{x}|} a_j x_j)^q \right]_d , \tag{4.27}$$

which is a function consisting of multiplication between elements of $\mathbf{x}$. This function is not an n-partite linear function on $\mathbf{x}$. The space $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$ of LHV correlators under AI is then not confined to the convex hull of n-partite linear functions on $\mathbf{x}$.

However, $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$ does not contain all vertices of $\mathcal{P}_{\mathbf{x}}$. In other words, the function in (4.27) is not equal to all functions $f : \mathbb{Z}_d^{|\mathbf{x}|} \to \mathbb{Z}_d$. We can demonstrate this by the example of the function $f(\mathbf{x}) = \left[\prod_{j=1}^{|\mathbf{x}|} (x_j)^{(d-1)} \right]_d$ that cannot be produced by powers of $\left[b + \sum_{j=1}^{|\mathbf{x}|} a_j x_j \right]_d$. Therefore $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}} \neq \mathcal{P}_{\mathbf{x}}$. □

We have shown that AI is *not* a loophole-free form of post-selection but LHV correlators cannot saturate the whole space $\mathcal{P}_{\mathbf{x}}$. This is somewhat analogous to discussion of the detection loophole, where for detection efficiency above some threshold, LHV correlators do not saturate the space of all possible correlators. In the subsequent subsection we will consider a more general class of input post-selections where $g_j(\mathbf{x})$ is now an n-partite linear function on $\mathbf{x}$. As a corollary of the above result, these input post-selections are also not loophole-free. However, in this new class of input post-selections, LHV correlators have greater computational expressiveness.

4.3.2 Input Post-selection with n-Partite Linear Functions

We now define input post-selection for n-partite linear functions $g_j(\mathbf{x})$. As can be seen from this definition, this post-selection includes AI, and therefore is not loophole-free.

Definition 14. n-Partite Linear Input Post-selection (PI) is input post-selection where the experimenter accepts data when all s_j satisfy $s_j = h(\mathbf{x})$ where $h(\mathbf{x})$ is an n-partite linear function on $\mathbf{x}$.

Again, we can define the space of LHV correlators under PI as $\mathcal{L}_{\mathbf{x}}^{\mathrm{PI}}$. This space is thus larger than the convex hull of n-partite functions on $\mathbf{x}$. In the following result we show that the space $\mathcal{L}_{\mathbf{x}}^{\mathrm{PI}}$ can be equal to $\mathcal{P}_{\mathbf{x}}$ for particular instances of $|\mathbf{x}|$ and n.

Proposition 22. *The space $\mathcal{L}_{\boldsymbol{x}}^{PI}$ is $\mathcal{P}_{\boldsymbol{x}}$ for a large enough n if $|\boldsymbol{x}| \leq (d-1)$.*

Proof: First we point out that for d being prime, any function $f(\mathbf{x})$ can be written as a polynomial of elements x_j in the following way:

$$f(\mathbf{x}) = \left[\sum_{\mathbf{z} \in \mathbb{Z}_d^{|\mathbf{x}|}} a_{\mathbf{z}} \prod_{j=1}^{|\mathbf{x}|} (x_j)^{z_j} \right]_d , \tag{4.28}$$

with $a_{\mathbf{z}} \in \mathbb{Z}_d$ where $\mathbf{z} \in \mathbb{Z}_d^{|\mathbf{x}|}$ are digit-strings. We now demonstrate that there are values of $n = n'$ when we can achieve any of the polynomials $\left[\prod_{j=1}^{|\mathbf{x}|}(x_j)^{z_j}\right]_d$, and then we can take $d^{|\mathbf{x}|}$ sets of these n' parties; each set outputs $\left[a_{\mathbf{z}} \prod_{j=1}^{|\mathbf{x}|}(x_j)^{z_j}\right]_d$ and we take the sum modulo d of all the sets outputs and as a result produce $f(\mathbf{x})$.

Now we demonstrate that for $n = n'$ parties we can produce the outcome $\left[\sum_{j=1}^{n'} m_j\right]_d = \left[\prod_{j=1}^{|\mathbf{x}|}(x_j)^{z_j}\right]_d$ deterministically. First, we show that all polynomial terms of length 2, i.e. $\left[\prod_{j=1}^{|\mathbf{x}|}(x_j)^{z_j}\right]_d$ with only 2 non-zero terms in $\mathbf{y}$, can be produced and proceed by induction. The length 2 polynomials can be achieved if a party outputs $m_j = \left[(s_j)^2\right]_d$ which is an n-partite linear function on $\mathbf{s}$. We then post-select on s_j satisfying the n-partite linear function on $\mathbf{x}$ in the following way $s_j = [(x_l)^{y_l} + (x_{l'})^{y_{l'}}]_d$ where l and l' labels the 2 elements of $\mathbf{y}$ which are non-zero. After this post-selection $m_j = \left[(x_l)^{2y_l} + (x_{l'})^{2y_{l'}} + 2(x_l)^{y_l}(x_{l'})^{y_{l'}}\right]_d$ and if we have two other parties that each outputs $m_{j+1} = [-s_{j+1}]_d$ and $m_{j+2} = [-s_{j+2}]_d$ and post-select on $s_{j+1} = (x_l)^{2y_l}$ and $s_{j+2} = (x_{l'})^{2y_{l'}}$. Then if we take the sum modulo d of these three outcomes we obtain $m_j \oplus m_{j+1} \oplus m_{j+2} = [2(x_l)^{y_l}(x_{l'})^{y_{l'}}]_d$, which is a length 2 polynomial. We can repeat this process with q sets of three parties and take the sum modulo d of the joint outcomes of all sets to obtain $[2q(x_l)^{y_l}(x_{l'})^{y_{l'}}]_d = [(x_l)^{y_l}(x_{l'})^{y_{l'}}]_d$ such that $[2q]_d = 1$ as d is prime.

For $|\mathbf{x}| = 3$, we have another party outputting $m_{j'} = [(s_j)^3]_d$ and post-selecting on the n-partite linear function on $\mathbf{x}$, $s_{j'} = [(x_1)^{y_1} + (x_2)^{y_2} + (x_3)^{y_3}]_d$. Thus we produce $m_{j'} = \left[((x_1)^{y_1} + (x_2)^{y_2} + (x_3)^{y_3})^3\right]_d = [3!(x_1)^{y_1}(x_2)^{y_2}(x_3)^{y_3} + ...]_d$ where "..." represents length 2 polynomials of $\mathbf{x}$. The length 2 and 1 polynomials can be subtracted from this output from the j'th site as they can be produced by other parties as shown above, so that the joint outcome can produce $[3!(x_1)^{y_1}(x_2)^{y_2}(x_3)^{y_3}]_d$. Again by taking q sets of parties that output this

in total and taking the joint outcome of all q sets produces

$$[q3!(x_1)^{y_1}(x_2)^{y_2}(x_3)^{y_3}]_d = [(x_1)^{y_1}(x_2)^{y_2}(x_3)^{y_3}]_d \tag{4.29}$$

for $[6q]_d = 1$ as d is prime.

We can repeat this process for $|\mathbf{x}| > 3$, where a party outputs $m_{j''} = \left[(s_{j''})^{|\mathbf{x}|}\right]_d$ and we post-select upon $s_{j''} = \left[\sum_{k=1}^{|\mathbf{x}|}(x_k)^{y_k}\right]_d$. This results in

$$m_{j''} = \left[(\sum_{k=1}^{|\mathbf{x}|}(x_k)^{y_k})^{|\mathbf{x}|}\right]_d = \left[|\mathbf{x}|!\prod_{k=1}^{|\mathbf{x}|}(x_k)^{y_k} + ...\right]_d \tag{4.30}$$

where "..." represents length $(|\mathbf{x}| - 1)$ polynomials of $\mathbf{x}$ which can be subtracted. Finally, again we can taking an arbitrary number of parties and the sum modulo d of the parties outputs will be $\left[\prod_{k=1}^{|\mathbf{x}|}(x_k)^{y_k}\right]_d$. This all applies when $|\mathbf{x}| \leq (d-1)$, and so when this is satisfied, all functions on $\mathbf{x}$ can be achieved with large enough n. □

Therefore in the presence of data post-selection that is a natural generalisation of LI post-selection, not only do we avoid loopholes, but we can completely saturate the space of all possible correlators $\mathcal{P}_{\mathbf{x}}$. This truly highlights the uniqueness of the scenario with binary inputs and outputs at each site. We now discuss the effect of input post-selection upon quantum correlators.

4.3.3 Quantum Correlators and Input Post-selection

For LI and LIO, the space of LHV correlators was unaffected, but the space of quantum correlators was n-dependent and could completely saturate $\mathcal{P}_{\mathbf{x}}$. Since LHV correlators can also saturate the whole correlator space with PI, we briefly consider the effect of AI on quantum correlators. The space of quantum correlators under AI post-selection is $\mathcal{Q}_{\mathbf{x}}^{\mathrm{AI}}$. As with $\mathcal{L}_{\mathbf{x}}$, there may be an n-dependence on the size of $\mathcal{Q}_{\mathbf{x}}^{\mathrm{AI}}$, but for brevity we will not make this explicit in our notation. The main result of this subsection is that for $|\mathbf{x}| = 2$ and $n = 3$, $\mathcal{Q}_{\mathbf{x}}^{\mathrm{AI}}$ is larger than $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$. We demonstrate this by an example for $d = 3$.

We have already shown in Proposition 21 that the vertex of $\mathcal{P}_{\mathbf{x}}$ corresponding to the function $f(\mathbf{x}) = \left[(x_1x_2)^2 + 1\right]_3$ is not in $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$. Therefore, we can adapt the non-trivial Bell inequality (3.63) from subsection 3.3.3 in chapter 3 for correlators

$p(1|\mathbf{x})$:

$$\frac{1}{9}\left(p(1|00)+p(1|01)+p(1|02)+p(110)+p(2|11)\right) \\ +\frac{1}{9}\left(p(2|12)+p(1|20)+p(2|21)+p(2|22)\right) \leq \ \tfrac{8}{9}. \tag{4.31}$$

The right-hand-side is exactly the same as (3.63), as all of the n-partite linear functions coincide with $f(\mathbf{x}) = \left[(x_1x_2)^2+1\right]_3$ for 8 out of 9 values of $\mathbf{x}$. Therefore, for all functions not equal to $f(\mathbf{x})$, this is the maximum overlap between functions. $\mathcal{L}_{\mathbf{x}}^{\mathrm{AI}}$ will be a convex polytope of functions not including $f(\mathbf{x}) = \left[(x_1x_2)^2+1\right]_3$, thus giving at most $\frac{8}{9}$ (for the Bell expression) for each of its extreme points. As discussed in chapter 3, this inequality is not violated by quantum correlators for $n = 2$. However, this inequality can be violated by quantum correlators for $n = 3$ with AI if $s_1 = x_1$, $s_2 = x_2$ and $s_3 = [x_1+x_2]_3$. We used the MBS approach to find a lower bound of $\approx 0.9314 > \frac{8}{9}$ on the quantum violation of (4.31).

Even in the presence of post-selection that introduces loopholes, the space of quantum correlators can be larger than the space of LHV correlators. Whilst not as dramatic as the effect that LI and LIO has on the quantum region, it is never-the-less interesting how "tactile" quantum correlators can be. That is, even if we imbue LHV correlators with more computational power (as with AI), quantum correlators can still have more computational expressiveness. It would be an interesting avenue of research to consider how quantum correlators are affected by non-loophole-free post-selection and whether their power can always be "boosted" by this post-selection.

4.4 Chapter Summary

The practical motivations of implementing Bell tests in the laboratory have motivated the study of loopholes and how they emerge when we have to reject "imperfect" measurement data [Pearle1970]. In this chapter we have used the insight from considering Bell tests from a computational point-of-view to say how and why loopholes emerge. By post-selecting on measurement data only when we have successful detection, we establish a relationship between the inputs and local hidden variables. This relationship allows other parties to indirectly learn the inputs of other sites via this shared data. By modelling this behaviour we retrieved the GM [Garg1987] bound on the necessary detection efficiency required

to establish a loophole-free violation of a Bell inequality. We then subsequently improved upon their bound by considering more parties.

Our improvement on the GM bound is not as impressive as the improvement attained by Eberhard [Eberhard1993] in the Clauser-Horne inequality setting. Eberhard's bound of $\frac{2}{3}$ has been improved upon further [Vértesi2010], this was a result of considering more measurement settings at each site. It would be interesting to consider the Bell inequalities on the full probability distribution for $(n, 2, 2)$ for $n > 2$, and whether the detection efficiency can be lowered further in analogy to our results.

Despite the issues associated with post-selection, there is a scenario where if we have perfect detections but the experimenter post-selects on measurement data by choice, we do not introduce loopholes. We associate LHV correlators with a limited computational expressiveness in the $(n, 2, 2)$ scenario: only linear Boolean functions can be achieved. If we post-select on data but only in a way that does not introduce non-linear Boolean functions, we avoid loopholes and still allow the possibility for a violation of a Bell inequality. In fact, we can increase the amount of quantum violation for particular Bell inequalities if we utilise post-selection.

However, we have also shown that the $(n, 2, 2)$ scenario is unique in the respect of not introducing loopholes; if we allow a greater number of inputs and outputs at each site, loopholes can again emerge. For $c = d > 2$, LHV correlators can produce powers of its input, and this inherent multiplication can be used to simulate all possible correlators. The ability to produce addition and multiplication modulo d for d being prime can be enough to produce any function $f : \mathbb{Z}_d^n \to \mathbb{Z}_d$. This has highlighted both how fragile Bell tests are in establishing a distinction between quantum and LHV correlators, and also how much descriptive power is accumulated by considering Bell tests from the computational point-of-view. Since we have shown the intimate link between correlators and functions on digit-strings, discussing functions has allowed to capture part of the picture of loopholes in Bell tests.

Interestingly, the models of post-selection we have discussed for $(n, 2, 2)$ involve the same level of data processing involved in MBQC. With LIO post-selection, we can simulate time-like separated processes such as adaptive MBQC circuits without introducing loopholes. Modelling signalling processes within Bell tests could lead to an insight into why we obtain improvements in information processing for quantum resources. We will summarise and consider some of these

ideas in the final chapter.

5 Summary and Outlook

The Bell inequalities have dictated and continue to dictate much of the discussion about the nature of quantum mechanics. In this thesis we have suggested that a general framework for Bell tests has a computational aspect. This both allows us to use methods and ideas in computer science to say something about Bell tests and methods developed in Bell tests to say something about computation. This collaboration between applied and fundamental science is what drives a large part of quantum information science [Hardy2010]. The diversity of connections addressed in this thesis have been made between the CGLMP-type Bell tests and basic number theory (in the form of functions on cyclic groups); loopholes, post-selection and quantum computing; we also connected quantum computing to non-local games and WW Bell tests.

In chapter 2, we outlined our approach to Bell tests, in particular looking at correlators: the expectation value of joint measurements. We showed that correlators can be associated with a notion of computation, that is functions on inputs. The calculation of a correlator maps raw statistical data into a stochastic map from inputs to a single output. This operational description allows to then think about information processing. This framework and description also has something to say about non-signalling theories and Svetlichny's model of correlations.

The discussion of correlators in chapter 2 was mostly in terms of the vertex description of convex polytopes. In chapter 3, we shifted to discussing the Bell inequality as defining the convex polytope of LHV correlators. We used the vertex description from chapter 2 to numerically calculate the linear inequalities, or facet Bell inequalities that define this polytope. However, it was only computationally feasible to find these inequalities for a relatively small number of settings. Given the hardness of the computational problem, we then just discussed non-trivial Bell inequalities, relaxing the need for the inequality to be facet-defining, but still potentially be violated. These non-trivial expressions then necessarily bound the space of LHV correlators to be smaller than the space of all possible correlators.

Non-trivial Bell inequalities are not only useful for bounding classical correlations, they have a natural interpretation in terms of non-local games. We looked at these non-local games in the many party, two-input, two-output scenario and showed that they have a concrete connection to Measurement-based Quantum Computing (MBQC). In particular, nMBQC, the class of non-adaptive circuits in the Raussendorf and Briegel model of MBQC can be shown to be inequivalent to a full quantum computer. However, within this nMBQC structure we still obtain natural generalisations of both the GHZ paradox and PR non-local box.

An interesting aspect of the nMBQC model is that data processing by a classical computer does not imbue LHV theories with any more computational power. In chapter 4, we applied this insight to data post-selection in Bell tests. We showed that post-selection in Bell tests, such as post-selecting on detecting outcomes in imperfect experiments, is problematic and introduces "loopholes". We used the computational insight from the rest of the thesis to show how the detection loophole can emerge and then rederived the Garg-Mermin bound on detection efficiency [Garg1987]. Throughout this discussion our computational perspective drove the understanding of loopholes.

After showing how in imperfect experimental Bell tests, loopholes can emerge, we turned to a different framework for data post-selection. We assume that we have perfect detection and data collection, but we post-select on inputs satisfying certain constraints. We showed that LHV correlators are unaffected in their computational expressiveness by this post-selection. We associate this conservation of computational power with the post-selection being "loophole-free". The notion of a loophole in both frameworks for post-selection is heuristically connected as allowing LHV correlators to have more computational expressiveness than just the linear Boolean functions.

The post-selection in the second framework is loophole-free if we constrain its form. These constraints however can still allow the post-selection to simulate the processing a classical computer imposes on data sent to measurement sites in MBQC. Also we can simulate signalling processes with this post-selection. This offers a potentially fruitful way of viewing quantum protocols and processes that have time-like separated elements into a framework where processes are now in the context of space-like separated parties. All of these results were developed in the $(n, 2, 2)$ scenario, and we showed that generalisations of these methods to other scenarios is problematic, thus highlighting the uniqueness of LHV expressed in terms of linear Boolean functions.

The work in this thesis is by no means a complete analysis of the role of computation in Bell tests, but perhaps strengthens the study of the relationship between the two. There is much work to be done still in understanding quantum correlations and whilst we have discovered new phenomena, the characterisation of quantum correlations remains broadly ill-understood. We have conjectured that all quantum correlators for the bipartite scenario can be captured by a particular set of quantum operators in the Navacués-Pironio-Acín hierarchy. It would be of great interest if this were true and if a similar behaviour occurred in the multipartite setting. This is an immediate problem raised by work in this thesis and worth pursuing as a continuation.

In recent years, a significant amount of effort into classifying the geometric nature of non-signalling correlations (see e.g. [Pironio2011]). We have shown that some of the extremal structure of the polytope of non-signalling theories can be revealed by the extremal structure of correlators. It would be interesting to see if there is a connection between correlators and the rest of the vertices of the non-signalling polytope. This picture is not clear as some of the non-LHV vertices of the non-signalling polytope for $(3, 2, 2)$ do not violate any of the facet Bell inequalities for correlators in this setting. However, the generality of the correlator description in terms of computations could give a handle on some of these ideas.

Continuing with the theme of characterising the full probability distribution instead of correlators, it would be interesting to study the effect of data post-selection on non-signalling resources. The difficulty in relating the inputs of parties to each other as we have done can allow LHV resources to achieve correlations that violate locality. Since in the correlator framework all single-site maps get mapped to a single output, this violation of locality has little or no effect. It would be interesting to allow resources that exploit "non-locality" in this way but still cannot produce something that quantum mechanics can produce. This is akin to the detector loophole where the LHV region is enlarged by post-selection, but below a threshold detection efficiency, still is not large enough to simulate quantum correlators.

Can our approach to correlators in terms of functions to applied to other issues in the study of Bell tests? An interesting potential avenue for further research could be the "monogamy of Bell correlations" [Pawłowski2003, Toner2006]. This is similar to the "monogamy of entanglement"[1] where we have three parties and

[1]This expression is thought to originate with Charlie Bennett [Toner2006].

if two parties are maximally entangled then the third party cannot be entangled with either of these two parties. It has been shown that Bell correlations behave in an analogous fashion where if two parties out of three violate a bipartite Bell inequality, then the correlations between either of these two parties and a third party cannot achieve a violation of the same inequality. This has been generalised to many parties with these parties divided into two overlapping sets [Pawłowski2003]. Can the language of functions, or computations explain that if one set of parties is trying to perform a computation, then by a satisfiability argument, the other set cannot produce this same function?

Finally, since we have established a connection between MBQC and Bell tests, it would be interesting to simulate quantum computations such as, say, Shor's algorithm [Shor1997] and see if it violates a Bell inequality. In some sense then it could be seen that this computation cannot be resolved with a classical picture of the world, or it would highlight the non-classical aspects of this algorithm. Post-selection and quantum computation have been studied before by Scott Aaronson [Aaronson2005], in a different format to our own framework. It was shown by Aaronson that quantum computation with post-selection of a different kind to ours is incredibly powerful. Speculatively, there may be some connection between our work and ideas in computational complexity. We have already made the connection to the class IQP [Shepherd2009], this class may be amenable to the study of our Bell tests with post-selection.

We hope to address the issues raised by this thesis in further research. We also hope that the work presented has produced the motivation to consider "device-independent" computing. This would be the ability to confirm that we have built something that uses quantum mechanics to compute but without knowing anything about the device. We have shown that the Bell inequality is a useful metric for quantum behaviour, in particular with regards to computation. More importantly, we hope that the work in this thesis can lead to new approaches of thinking about Bell tests, perhaps motivated by computation.

N. David Mermin once quoted a "distinguished Princeton physicist" as saying [Mermin1985],"Anybody who's not bothered by Bell's theorem has to have rocks in his head." The Bell inequality has been a profound addition to science and we hope that the work in this thesis contributes to new aspects of its study.

Bibliography

[Aaronson2004] S. Aaronson and D. Gottesman, *Improved Simulation of Stabilizer Circuits*, Phys. Rev. A **70**, 052328 (2004).

[Aaronson2005] S. Aaronson, *Quantum computing, postselection, and probabilistic polynomial-time*, Proc. R. Soc. A **461**, 3473-3482, (2005).

[Acín2002] A. Acín, T. Durt, N. Gisin, and J. I. Latorre, *Quantum nonlocality in two three-level systems*, Phys. Rev. A **65**, 052325 (2002).

[Acín2004] A. Acín, J. L. Chen, N. Gisin, D. Kaszlikowski, L. C. Kwek, C. H. Oh, and M. Żukowski, *Coincidence Bell Inequality for Three Three-Dimensional Systems*, Phys. Rev. Lett. **92**, 250404 (2004) .

[Acín2007] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, *Device-Independent Security of Quantum Cryptography against Collective Attacks*, Phys. Rev. Lett. **98**, 230501 (2007) .

[Aharonov2004] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, *Adiabatic Quantum Computation is Equivalent to Standard Quantum Computation*, 45th Ann. Symp. on the Found. of Comp. Science, 42-51 (2004).

[Almeida2010] M. L. Almeida, J. -D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, *Guess Your Neighbor's Input: A multipartite Nonlocal Game with No Quantum Advantage*, Phys. Rev. Lett. **104**, 230404 (2010).

[Anders2009] J. Anders and D. E. Browne, *Computational Power of Correlations*, Phys. Rev. Lett. **102**, 050502 (2009).

[Anderson1992] F. W. Anderson and K. R. Fuller, *Rings and Categories of Modules*, 2nd edition, Graduate Texts in Mathematics, **13**, Springer-Verlag (Berlin) (1992).

[Aspect1981] A. Aspect, P. Grangier, and G. Roger, *Experimental tests of realistic local theories via bell's theorem*, Phys. Rev. Lett. **47**, 460 (1981).

[Bacciagaluppi2009] G. Bacciagaluppi and A. Valentini, *Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference*, Cambridge University Press (2009).

[Bancal2009] J.-D. Bancal, C. Branciard, N. Gisin, and S. Pironio, *Quantifying multipartite nonlocality*, Phys. Rev. Lett. **103**, 090503 (2009).

[Bancal2011] J.-D. Bancal, N. Brunner, N. Gisin, and Y.-C. Liang, *Detecting Genuine multipartite Quantum Nonlocality: A Simple Approach and Generalization to Arbitrary Dimensions*, Phys. Rev. Lett. **106**, 020405 (2011).

[Barrett2002] J. Barrett, D. Collins, L. Hardy, A. Kent, and S. Popescu, *Quantum nonlocality, Bell inequalities, and the memory loophole*, Phys. Rev. A **66**, 042111 (2002)

[Barrett2005a] J. Barrett, L. Hardy, and A. Kent, *No Signalling and Quantum Key Distribution*, Phys. Rev. Lett. **95**, 010503 (2005).

[Barrett2005b] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu and D. Roberts, *Nonlocal correlations as an information-theoretic resource*, Phys. Rev. A **71**, 022101 (2005).

[Barrett2005c] J. Barrett, and S. Pironio, *Popescu-Rohrlich Correlations as a Unit of Nonlocality*, Phys. Rev. Lett. **95**, 140401 (2005).

[Barrett2007] J. Barrett, *Information processing in generalized probabilistic theories*, Phys. Rev. A **75**, 032304 (2007).

[Barrett2011] J. Barrett and N. Gisin, *How Much Measurement Independence Is Needed to Demonstrate Nonlocality?*, Phys. Rev. Lett. **106**, 100406 (2011).

[Barrett2011a] J. Barrett, S. Pironio, J.-D. Bancal, and N. Gisin *The definition of multipartite nonlocality*, arXiv:quant-ph/1112.2626v1 (2011).

[Belinskii1993] A.V. Belinskii and D.N. Klyshko, *Interference of light and Bell's theorem*, Sov. Phys. Usp. **36**, 653 (1993).

[Bell1964] J. S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics, **1**, 195 (1964).

[Bell1977] J. S. Bell, *Free variables and local causality*, Epistemological Letters, February 1977.

[Bell2004] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2nd edition, Cambridge University Press (1964).

[BB1984] C. H. Bennett and G. Brassard, *Quantum Cryptography: Public key distribution and coin tossing*, Proc. IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, 175 (1984).

[Bennett1993] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels*, Phys. Rev. Lett. **70**, 1895-1899 (1993).

[Berry2010] D. W. Berry, H. Jeong, M. Stobińska, and T. C. Ralph, *Fair-sampling assumption is not necessary for testing local realism*, Phys. Rev. A **81**, 012109 (2010).

[Bohm1951] D. Bohm, *Quantum Physics*, Constable, London (1951).

[Boyd2004] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004).

[Brassard2006] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, *Limit on Nonlocality in Any World in Which Communication Complexity Is Not Trivial*, Phys. Rev. Lett. **96**, 250401, (2006).

[Bremner2011] M. J. Bremner, R. Jozsa, and D. J. Shepherd, *Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy*, Proc. R. Soc. A **467**, 459-472, (2011).

[Briegel2009] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, M. Van den Nest, *Measurement-based quantum computation*, Nature Physics **5** 1, 19-26 (2009).

[Brukner2004] Č. Brukner, M. Żukowski, J. -W. Pan, and A. Zeilinger, *Bell's Inequalities and Quantum Communication Complexity*, Phys. Rev. Lett. **92**, 127901 (2004).

[Brun2006] T. A. Brun, I. Devetak, and M. -H. Hsieh, *Correcting Quantum Errors with Entanglement*, Science **314**, 436 (2006).

[Brunner2005] N. Brunner, N. Gisin, and V. Scarani, *Entanglement and nonlocality are different resources*, New J. Phys. **7**, 88 (2005).

[Buhrman2010] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, *Nonlocality and communication complexity*, Rev. Mod. Phys. **82**, 665-698 (2010).

[Caves2002] C. M. Caves, C. A. Fuchs, and R. Schack, *Quantum Probabilities as Bayesian Probabilities*, Phys. Rev. A **65**, 022305 (2002).

[Chiribella2011] G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Informational derivation of Quantum Theory*, Phys. Rev. A **84**, 012311 (2011).

[Collins2004] D. Collins and N. Gisin, *A relevant two qubit Bell inequality inequivalent to the CHSH inequality*, J. Phys. A **37**, 1775 (2004).

[CGLMP2002] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, *Bell Inequalities for Arbitrarily High-Dimensional Systems*, Phys. Rev. Lett. **88**, 040404 (2002).

[CH1969] J. F. Clauser, and M. A. Horne, *Experimental consequences of objective local theories*, Phys. Rev. Lett. **23**, 880 (1969).

[CHSH1969] J. F. Clauser, M. A. Horne, A. Shimony and R. Holt, *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett. **23**, 880 (1969).

[Tsirelson1980] B. S. Tsirelson, *Quantum Generalizations of Bell's Inequality*, Lett. Math. Phys. **4**, 93 (1980).

[Clauser1978] J. F. Clauser, and A. Shimony, *Bell's theorem. Experimental tests and implications*, Rep. Prog. Phys. **41**, 1881 (1978).

[Cleve2004] R. Cleve, P. Hoyer, B. Toner, and J. Watrous, *Consequences and limits of nonlocal strategies*, Proc. 19th Ann. IEEE Conf. Comput. Complexity, 236 - 249 (2004).

[Colbeck2007] R. Colbeck, *Quantum and Relativistic Protocols for Secure Multi-Party Computation*, PhD thesis, University of Cambridge (2007).

[Datta2005] A. Datta, S. T. Flammia, and C. M. Caves, *Entanglement and the Power of One Qubit*, Phys. Rev. A **72**, 042316 (2005).

[Deutsch1985] D. Deutsch, *Quantum theory, the Church-Turing principle and the universal quantum computer*, Proc. R. Soc. A **400**, 1818, 97-117 (1985).

[Deutsch1992] D. Deutsch and R. Jozsa, *Rapid solutions of problems by quantum computation*, Proc. R. Soc. A **439**, 1907, 553-558 (1992).

[Durt2001] T. Durt, D. Kaszlikowski, and M. Żukowski, *Violations of local realism with quantum systems described by N-dimensional Hilbert spaces up to N=16*, Phys. Rev. A **64**, 024101 (2001).

[Eberhard1993] P. H. Eberhard, *Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment*, Phys. Rev. A **47**, (R)747-750 (1993).

[Einstein1905] A. Einstein, *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt*, Annalen der Physik **17**, 6, 132-148 (1905).

[Einstein1971] A. Einstein, Letter to Max Born (4 December 1926); *The Born-Einstein Letters*, translated by Irene Born, Walker and Company, New York (1971).

[EPR1935] A. Einstein, B. Podolsky, and N. Rosen, *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?*, Physical Review **47**, 10, 777-780 (1935).

[Ekert1991] A. K. Ekert, *Quantum cryptography based on Bell's theorem*, Phys. Rev. Lett. **67**, 661-663 (1991).

[Feynman1982] R. P. Feynman, *Simulating Physics with Computers*, International Journal of Theoretical Physics **21**, 6-7, 467-488 (1982).

[Fine1982] A. Fine, *Hidden Variables, Joint Probability, and the Bell Inequalities*, Phys. Rev. Lett. **48**, 291 (1982).

[Freedman1972] S. J. Freedman and J. F. Clauser, *Experimental Test of Local Hidden-Variable Theories*, Phys. Rev. Lett. **28**, 938 (1972).

[Froissart1981] M. Froissart, *Constructive generalization of Bell's inequalities*, Nuovo Cimento **64** B, 241 (1981).

[Gallego2010] R. Gallego, N. Brunner, C. Hadley, A. Acín, *Device-independent tests of classical and quantum dimensions*, Phys. Rev. Lett. **105**, 230501 (2010).

[Garg1987] A. Garg and N.D. Mermin, *Detector inefficiencies in the Einstein-Podolsky-Rosen experiment*, Phys. Rev. D **35**, 3831 (1987).

[GHZ1989] D. M. Greenberger, M. A. Horne, and A. Zeilinger, *Going Beyond Bell's Theorem*, in Bell's Theorem, Quantum Theory, and Conceptions (edited by M. Kafatos), Kluwer, Dordrecht, 69-72 (1989).

[Gisin1998] N. Gisin and H. Bechmann-Pasquinucci, *Bell inequality, Bell states and maximally entangled states for n qubits*, Phys. Lett. A, **246**, 1-6 (1998).

[Gottesman1999] D. Gottesman and I. L. Chuang, *Quantum Teleportation is a Universal Computational Primitive*, Nature **402**, 390-393 (1999).

[Grover1996] L. K. Grover, *A fast quantum mechanical algorithm for database search*, Proc. 28th Annual ACM Symp. Theory of Comp., 212 (1996).

[Grünbaum2003] B. Grünbaum, *Convex Polytopes*, 2nd edition, Graduate Texts in Mathematics, **221**, Springer-Verlag (Berlin) (2003).

[Gurvits2002] L. Gurvits, *Quantum matching theory*, arXiv:quant-ph/0201022 (2002).

[Hall2011] M. J. W. Hall, *Relaxed Bell inequalities and Kochen-Specker theorems*, Phys. Rev. A **84**, 022102 (2011).

[Hardy1993] L. Hardy, *Nonlocality for two particles without inequalities for almost all entangled states*, Phys. Rev. Lett. **71**, 11, 1665-1668 (1993).

[Hardy2001] L. Hardy, *Quantum theory from five reasonable axioms*, arXiv:quant-ph/0101012v4 (2001).

[Hardy2010] L. Hardy and R. Spekkens, *Why Physics Needs Quantum Foundations*, Physics in Canada **66**, 2, 73-76 (2010).

[Hardy2011] L. Hardy, *Reformulating and reconstructing quantum theory*, arXiv:quant-ph/1104.2066v3 (2011).

[Harrigan2011] N. Harrigan and R. W. Spekkens, *Einstein, incompleteness, and the epistemic view of quantum states*, Found. Phys. **40**, 125 (2010).

[Harrow2009] A. W. Harrow, A. Hassidim, and S. Lloyd, *Quantum algorithm for solving linear systems of equations*, Phys. Rev. Lett. **103**, 150502 (2009).

[Hein2005] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, and H. J. Briegel, *Entanglement in Graph States and its Applications*, Proc. of the International School of Physics "Enrico Fermi" on "Quantum Computers, Algorithms and Chaos", Varenna, Italy, (2005).

[Heywood1983] P. Heywood and M. L. G. Redhead, *Nonlocality and the Kochen-Specker paradox*, Found. Phys. **13**, 481 (1983).

[Hoban2011a] M. J. Hoban, E. T. Campbell, K. Loukopoulos, and D. E. Browne, *Non-adaptive Measurement-based Quantum Computation and Multi-party Bell Inequalities*, New J. Phys. **13** 023014 (2011).

[Hoban2011b] M. J. Hoban and D. E. Browne, *Stronger Quantum Correlations with Loophole-Free Postselection*, Phys. Rev. Lett. **107**, 120402 (2011).

[Hoban2011c] M. J. Hoban, J. J. Wallman, and D. E. Browne, *Generalized Bell-inequality experiments and computation*, Phys. Rev. A **84**, 062107 (2011).

[Holevo1973] A. S. Holevo, *Bounds for the quantity of information transmitted by a quantum communication channel*, Problems of Information Transmission **9**, 177-183 (1973).

[Horodecki2009] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Quantum entanglement*, Rev. Mod. Phys. **81**, 865-942, (2009).

[Jain2010] R. Jain, Z. Ji, S. Upadhyay, and J. Watrous, *QIP = PSPACE*, Proc. 42nd ACM Symp. Theory of Comp., **53**, 12, (2010).

[Jozsa2003] R. Jozsa and N. Linden, *On the role of entanglement in quantum-computational speed-up*, Proc. R. Soc. A **459**, 2011-2032, (2011).

[Jozsa2006] R. Jozsa, *An Introduction to Measurement Based Quantum Computation*, NATO Science Series, III: Computer and Systems Sciences **199**: *Quantum Information Processing - From Theory to Experiment*, 137-158, (2006).

[Kaszlikowski2000] D. Kaszlikowski, P. Gnaciński, M. Żukowski, W. Miklaszewski, and A. Zeilinger, *Violations of Local Realism by Two Entangled N-Dimensional Systems Are Stronger than for Two Qubits*, Phys. Rev. Lett. **85**, 4418-4421 (2000).

[Kitaev2003] A. Yu. Kitaev, *Fault-tolerant quantum computation by anyons*, Ann. of Phys. **303**, 2 (2003).

[Knuth1981] D. Knuth, *The Art of Computer Programming Vol. 2, Seminumerical Algorithms*, Addison-Wesley, Boston (1981).

[Kushilevitz1996] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, Cambridge (1996).

[Lee2007] S.-W. Lee, Y. W. Cheong, and J. Lee, *Generalized structure of Bell inequalities for bipartite arbitrary-dimensional systems*, Phys. Rev. A **76**, 032108 (2007).

[Leung2001] D. W. Leung, *Two-qubit Projective Measurements are Universal for Quantum Computation*, Technical Report, NSF-ITP-01-174 (2001).

[Liang2011] Y.-C. Liang, T. Vértesi, and N. Brunner, *Semi-device-independent bounds on entanglement*, Phys. Rev. A, **83**, 022108 (2011).

[Linden2007] N. Linden, S. Popescu, A. J. Short, and A. Winter, *Quantum Nonlocality and Beyond: Limits from Nonlocal Computation*, Phys. Rev. Lett. **99**, 180502 (2007).

[Liouville1838] J. Liouville, *Note sur la Théorie de la Variation des constantes arbitraires*, Journ. de Math. **3**, 349 (1838).

[Lloyd2011] S. Lloyd, L. Maccone, R. Garcia-Patron, V. Giovannetti, and Y. Shikano, *The quantum mechanics of time travel through post-selected teleportation*, Phys. Rev. D **84**, 025007 (2011).

[MacWilliams1977] F. J. MacWilliams, and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam (1977).

[Marcovitch2008] S. Marcovitch and B. Reznik, *Is Communication Complexity Physical?*, Phys. Rev. A **77**, 032120 (2008).

[Masanes2003] Ll. Masanes, *Tight Bell inequality for d-outcome measurements correlations*, Quant. Inf. Comput. **3**, 345 (2003).

[Matsukevich2008] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, *Bell Inequality Violation with Two Remote Atomic Qubits*, Phys. Rev. Lett. **100**, 150404 (2008).

[Mayers98] D. Mayers and A. Yao, *Quantum Cryptography with Imperfect Apparatus*, Proc. 39th Ann. Symp. on Found. of Comp. Science, 503-509 (1998).

[Mermin1982] N. D. Mermin and G. M. Schwarz, *Joint distributions and local realism in the higher-spin Einstein-Podolsky-Rosen experiment*, Found. Phys. **12**, 2 (1982).

[Mermin1990] N. D. Mermin, *Extreme quantum entanglement in a superposition of macroscopically distinct states*, Phys. Rev. Lett. **65**, 1838 (1990).

[Mermin1993] N. D. Mermin, *Hidden variables and the two theorerns of John Bell*, Rev. Mod. Phys. **66**, 803 (1993).

[Mermin1985] N. D. Mermin, *Is the Moon there when nobody looks? Reality and the quantum theory*, Physics Today (American Institute of Physics), April (1985).

[Metropolis1949] N. Metropolis and S. Ulam, *The Monte Carlo Method*, Journal of the American Statistical Association **44**, 247, 335-341 (1949).

[Navascués2007] M. Navascués, S. Pironio, and A. Acín, *Bounding the Set of Quantum Correlations*, Phys. Rev. Lett. **98**, 010401 (2007).

[Navascués2008] M. Navascués, S. Pironio, and A. Acín, *A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations*, New J. Phys. **10**, 073013 (2008).

[Navascués2009] M. Navascués, and H. Wunderlich, *A glance beyond the quantum model*, Proc. Roy. Soc. Lond. A **466**, 881-890 (2009).

[Nielsen2000] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000).

[Nielsen2003] M. A. Nielsen, *Universal quantum computation using only projective measurement, quantum memory, and preparation of the* 0 *state*, Phys. Lett. A. **308**, 2-3, 96-100 (2003).

[Oppenheim2010] J. Oppenheim, and S. Wehner, *The Uncertainty Principle Determines the Nonlocality of Quantum Mechanics*, Science **330**, 6007, 1072-1074 (2010).

[Ou1988] Z. Y. Ou and L. Mandel, *Violation of Bell's Inequality and Classical Probability in a Two-Photon Correlation Experiment*, Phys. Rev. Lett. **61**, 50 (1988)

[Papadimitriou1994] C. Papadimitriou, *Computational Complexity*, Addison Wesley, (1994).

[Paulsen2003] Lucien Hardy, *Completely Bounded Maps and Operator Algebras*, Cambridge University Press, (2003).

[Pawłowski2003] M. Pawłowski and Č. Brukner, *Monogamy of Bell's inequality Violations in Nonsignaling Theories*, Phys. Rev. Lett. **102**, 030403 (2009).

[Pawłowski2009] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, *Information causality as a physical principle*, Nature **461**, 1101 (2009).

[Pearle1970] P Pearle, *Hidden-Variable Example Based upon Data Rejection*, Phys. Rev. D **2**, 1418-25 (1970).

[Peres1993] A. Peres, *Quantum Theory, Concepts and Methods*, Kluwer, Alphen aan den Rijn, Netherlands (1993).

[Peres1999] A. Peres, *All the Bell Inequalities*, Foundations of Physics **29**, 589-614 (1999).

[Pironio2009] S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani, *Device-independent quantum key distribution secure against collective attacks*, New J. Phys. **11**, 045021 (2009).

[Pironio2010] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, *Random Numbers Certified by Bell's Theorem*, Nature **464**, 1021 (2010).

[Pironio2011] S. Pironio, J. -D. Bancal, and V. Scarani, *Extremal correlations of the tripartite no-signaling polytope*, J. Phys. A: Math. Theor. **44**, 065303 (2011).

[Pitowsky1989] I. Pitowsky, *Quantum Probability - Quantum Logic*, Lecture Notes in Physics **321**, Springer (Berlin) (1989).

[Pitowsky1991] I. Pitowsky, *Correlation Polytopes: Their Geometry and Complexity*, Mathematical Programming **A50**, 395-414 (1991).

[Plenio2007] M. B. Plenio and S. Virmani, *An introduction to entanglement measures*, Quant. Inf. Comput. **7**, 1-51, (2007).

[Polymake2000] E. Gawrilow, and M. Joswig, *Polymake: a framework for analyzing convex polytopes*, Polytopes - combinatorics and computation (edited by G. Kalai and G. M. Ziegler), 43-73, Birkhäuser (2000).

[Popescu1994] S. Popescu and D. Rohrlich, *Quantum nonlocality as an axiom*, Foundations of Physics **24**, 379-385 (1994).

[Popescu1997] S. Popescu and D. Rohrlich, *Thermodynamics and the measure of entanglement*, Phys. Rev. A **56**, (R)3319 (1997).

[Rabelo2011] R. Rabelo, M. Ho, D. Cavalcanti, N. Brunner, and V. Scarani, *Device-independent certification of entangled measurements*, Phys. Rev. Lett. **107**, 050502 (2011).

[Rarity1990] J. G. Rarity and P. R. Tapster, *Experimental violation of Bell's inequality based on phase and momentum*, Phys. Rev. Lett. **64**, 2495 (1990).

[Raussendorf2001] R. Raussendorf and H. J. Briegel, *A One-Way Quantum Computer*, Phys. Rev. Lett. **86**, 5188 (2001).

[Raussendorf2003] R. Raussendorf, D. E. Browne, and H. J. Briegel, *Measurement-based quantum computation on cluster states*, Phys. Rev. A **68**, 022312 (2003).

[Raussendorf2009] R. Raussendorf, *Quantum computation, discreteness, and contextuality*, arXiv:quant-ph/0907.5449 (2009).

[RSA1978] R. Rivest, A. Shamir and L. Adleman, *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*, Communications of the ACM **21**, 2, 120-126 (1978).

[Rowe2001] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe and D. J. Wineland, *Experimental violation of a Bell's inequality with efficient detection*, Nature **409**, 791-794 (2001).

[Sangouard2011] N. Sangouard, J.-D. Bancal, N. Gisin, W. Rosenfeld, P. Sekatski, M. Weber, and H. Weinfurter, *Loophole-free Bell test with one atom and less than one photon on average*, Phys. Rev. A **84**, 052122 (2011).

[Schrödinger1936] E. Schrödinger, *Probability relations between spatially separated systems*, Proceedings of the Cambridge Philosophy Society, **32**, 446 (1936).

[SeDuMi] J. Sturm, *SeDuMi, a MATLAB toolbox for optimization over symmetric cones, http://sedumi.mcmaster.ca.*

[Shannon1949] C. Shannon, *Communication Theory of Secrecy Systems*, Bell System Technical Journal **28**, 4, 656-715 (1949).

[Shepherd2009] D. J. Shepherd and M. J. Bremner, *Instantaneous Quantum Computation*, Proc. R. Soc. A **465**, 1413-1439 (2009).

[Shih1988] Y. H. Shih and C. O. Alley, *New Type of Einstein-Podolsky-Rosen-Bohm Experiment Using Pairs of Light Quanta Produced by Optical Parametric Down Conversion*, Phys. Rev. Lett. **61**, 2921 (1988).

[Shor1997] P. W. Shor, *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer*, SIAM J. Comput. **26** (5), 1484-1509 (1997).

[Silman2011] J. Silman, A. Chailloux, N. Aharon, I. Kerenidis, S. Pironio, and S. Massar, *Fully Distrustful Quantum Cryptography*, Phys. Rev. Lett. **106**, 220501 (2011).

[Son2006] W. Son, J. Lee, and M. S. Kim, *Generic Bell Inequalities for multipartite Arbitrary Dimensional Systems*, Phys. Rev. Lett. **96**, 060406 (2006).

[Svetlichny1987] G. Svetlichny, *Distinguishing three-body from two-separability by a Bell-type inequality*, Phys. Rev. D **35**, 3066 (1987).

[Tittel1998] W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, *Violation of Bell Inequalities by Photons More Than 10 km Apart* , Phys. Rev. Lett. **81**, 3563 (1998).

[Toner2006] B. Toner, F. Verstraete, *Monogamy of Bell correlations and Tsirelson's bound*, arXiv:quant-ph/0611001 (2006).

[Turing1937] A. M. Turing, *On Computable Numbers, with an Application to the Entscheidungsproblem*, Proceedings of the London Mathematical Society, **2**, 42, 230-65, (1937).

[vanDam2000] W. van Dam, *Nonlocality and Communication Complexity*, PhD thesis, University of Oxford, Department of Physics (2000).

[VandenNest2006] M. Van den Nest, A. Miyake, W. Dür, H. J. Briegel, *Title: Universal resources for measurement-based quantum computation*, Phys. Rev. Lett. **97**, 150504 (2006).

[Vértesi2010] T. Vértesi, S. Pironio, and N. Brunner, *Closing the Detection Loophole in Bell Experiments Using Qudits*, Phys. Rev. Lett. **104**, 060401 (2010).

[Vidal2003] G. Vidal, *Efficient classical simulation of slightly entangled quantum computations*, Phys. Rev. Lett. **91**, 147902 (2003).

[Vidick2011] T. Vidick and S. Wehner, *More nonlocality with less entanglement*,Phys. Rev. A **83**, 052310 (2011).

[vonNeumann1944] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, (1944).

[vonNeumann1955] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*, Springer-Verlag (Berlin) (1955).

[Weihs1998] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter and A. Zeilinger, *Violation of Bell's inequality under strict Einstein locality conditions*, Phys. Rev. Lett. **81**, 5039 (1998).

[Werner1989] R. F. Werner, *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*, Phys. Rev. A **40**, 4277-4281 (1989).

[Werner2001] R. F. Werner and M. M. Wolf, *All multipartite Bell-correlation inequalities for two dichotomic observables per site*, Phys. Rev. A **64**, 032112 (2001).

[Wiesner1983] S. Wiesner, *Conjugate coding*, ACM SIGACT News **15**, 1 (1983).

[Wootters1982] W. K. Wootters and W. H. Zurek, *A Single Quantum Cannot be Cloned*, Nature **299**, 802-803 (1982).

[Yalmip] J. Lofberg, *Yalmip : A toolbox for modeling and optimization in MATLAB, http://users.isy.liu.se/johanl/yalmip/.*

[Zanardi1999] P. Zanardi and M. Rasetti, *Holonomic Quantum Computation*, Phys. Lett. A **264**, 94 (1999).

[Żukowski1999] M. Żukowski and D. Kaszlikowski, *Greenberger-Horne-Zeilinger paradoxes with symmetric multiport beam splitters*, Phys. Rev. A **59**, 3200 (1999).

[Żukowski2002] M. Żukowski and Č. Brukner, *Bell's Theorem for General N-Qubit States*, Phys. Rev. Lett. **88** 210401 (2002).