

Zheng-Hao Liu

Exploring Quantum Contextuality with Photons

A doctoral Thesis accepted by
University of Science and Technology of China,
Hefei, China

June 13, 2023

Springer Nature

Parts of this thesis have been published in the following journal articles:

1. **Zheng-Hao Liu**, Hui-Xian Meng, Zhen-Peng Xu, Jie Zhou, Sheng Ye, Qiang Li, Kai Sun, Hong-Yi Su, Adán Cabello, Jing-Ling Chen, Jin-Shi Xu, Chuan-Feng Li, and Guang-Can Guo, “Experimental observation of quantum contextuality beyond Bell nonlocality”. *Physical Review A* **100**, 042118 (2019).
2. **Zheng-Hao Liu**, Wei-Wei Pan, Xiao-Ye Xu, Mu Yang, Jie Zhou, Ze-Yu Luo, Kai Sun, Jing-Ling Chen, Jin-Shi Xu, Chuan-Feng Li, and Guang-Can Guo, “Experimental exchange of grins between quantum Cheshire cats”. *Nature Communications* **11**, 3006 (2020).
3. **Zheng-Hao Liu**, Jie Zhou, Hui-Xian Meng, Mu Yang, Qiang Li, Yu Meng, Hong-Yi Su, Jing-Ling Chen, Kai Sun, Jin-Shi Xu, Chuan-Feng Li, and Guang-Can Guo, “Experimental test of the Greenberger–Horne–Zeilinger-type paradoxes in and beyond graph states”. *npj Quantum Information* **7**, 66 (2021).
4. **Zheng-Hao Liu**, Xiao-Bin Liang, Kai Sun, Qiang Li, Yu Meng, Mu Yang, Bo Li, Jing-Ling Chen, Jin-Shi Xu, Chuan-Feng Li, and Guang-Can Guo, “Photonic implementation of quantum information masking”. *Physical Review Letters* **126**, 170505 (2021).
5. **Zheng-Hao Liu**, Kai Sun, Jiannis K. Pachos, Mu Yang, Yu Meng, Yu-Wei Liao, Qiang Li, Jun-Feng Wang, Ze-Yu Luo, Yi-Fei He, Dong-Yu Huang, Guang-Rui Ding, Jin-Shi Xu, Yong-Jian Han, Chuan-Feng Li, and Guang-Can Guo, “Topological contextuality and anyonic statistics of photonic-encoded parafermions”. *PRX Quantum* **2**, 030323 (2021).
6. **Zheng-Hao Liu**, Hui-Xian Meng, Zhen-Peng Xu, Jie Zhou, Jing-Ling Chen, Jin-Shi Xu, Chuan-Feng Li, Guang-Can Guo, and Adán Cabello, “Experimental test of high-dimensional quantum contextuality based on contextuality concentration”. *Phys. Rev. Lett.* **130**, 240202 (2023).

*To my younger self, and to those
mourning the road not taken —
unperformed experiments have no results.*

Foreword

In the last summer, Zheng-Hao finished his Ph.D. research at the University of Science and Technology of China and left our group to pursue a postdoc overseas. Working with him has always been a pleasure for us, and after one year we are very glad to see his excellent doctoral dissertation would be published in the Springer Theses. The thesis investigated quantum contextuality from the experimental aspect. Contextuality is a peculiar property of quantum mechanics, it reveals the conflict between the quantum and the hidden-variable theories and constitutes a crucial resource for quantum computing. While contextuality has been observed in various quantum systems, many of its applications remain to be explored, and its interplay with quantum nonlocality is a heated research topic in recent years.

The thesis has focused on the above topics. Specifically, the first three sections of the thesis explain the concept of contextuality with clear math and vivid examples. We have found the metaphors there highly informative and original. The latter five sections present the author's experimental research about how a stronger, novel contextual correlation can be distilled from known nonlocal correlations, how contextuality gives rise to a quantum paradox analogous to the story of Cheshire cat in *Alice in Wonderland*, and how contextuality in the topological system enables universal quantum computation.

The thesis represents some best results from our lab, and we believe all the studies in this thesis are systematic, well-presented, interesting to the community of quantum information, and come with unique perspectives. We hope the readers will enjoy reading the thesis.

Hefei
May 2023

Prof. Jin-Shi Xu
Prof. Chuan-Feng Li

Preface

Quantum contextuality is one of the most intriguing and peculiar predictions of quantum mechanics. In layman's terms, it refers to the fact that the result of a single measurement of a physical quantity in quantum mechanics depends on the way the measurement is carried out. More precisely, the measurement outcome of an observable does not only reflect the pre-defined value of the observable itself; instead, the knowledge about the context—the set of compatible observables that are actually measured—is indispensable to determine the measurement result. Quantum contextuality is also a cornerstone in modern quantum information science. It is the origin of the famous quantum nonlocality and various nonclassical paradoxes. It is also a resource for many quantum information processing tasks and even universal quantum computing. Therefore, the study of quantum contextuality not only advances the comprehension of the foundations of quantum physics, but also facilitates the practical applications of quantum information technology.

In the last fifteen years, the study of quantum contextuality has developed from a purely theoretical level to a stage where direct experimental tests become amenable. However, the experimental research on contextuality at the current stage largely focuses on direct validations of some most famous predictions of contextuality, while other forms of contextuality and its practical applications in quantum information science are rarely involved. The researches in this thesis are committed to bridge this gap from two directions: (1) to construct and test stronger forms of contextuality and relieve the requirements of contextuality experiments on experimental platforms, and (2) to explore the connections between contextuality and the other concepts in quantum information science and directly demonstrate the application of contextuality in broader scenarios.

I will present the following aspects of research findings in this thesis:

The first topic is about the relationship between quantum contextuality and nonlocality. Since nonlocality is the manifestation of quantum contextuality in spacelike-separated systems, it is possible to enhance the nonlocal quantum correlation caused by nonlocality by lifting the constraint of the spacelike-separation from the measurement operator. I have experimentally realised an example of quantum contextuality beyond nonlocality. At the same time, lifting the additional constraints could reduce the dimensionality of the state space required for demonstrating the quantum correlation. I have further constructed and experimentally realised the embedding of the strongest known quantum correlations into low-dimensional systems.

The second topic is regarding the “all-versus-nothing” paradoxes from quantum contextuality. Every logical proof of quantum contextuality can be transformed into an “all-versus-nothing” type quantum-classical paradox. The graph states are a class of highly entangled quantum states and the basic building blocks for measurement-based quantum computing. I have constructed and observed the “all-versus-

nothing” paradox applicable to graph states, and show its applications in quantum state verification and the witness of quantum entanglement and quantum steering.

The third topic concerns the pre- and post-selection paradoxes from quantum contextuality. The investigation of quantum systems with pre- and post-selection provides additional information about its evolution process, which is reflected in the quantity called “weak values”. Quantum contextuality allows quantum processes to have strange behaviours; one example is the quantum Cheshire cat effect—the separation of the properties of a quantum system from the object itself. I have developed methods for weak value extraction without weak measurements to experimentally observe the exchange of grins between quantum Cheshire cats.

Finally, I will present the results about the topological protection and braiding dynamics of quantum contextuality in quasiparticle systems. The combination of quantum contextuality and the topological protection of quasiparticle systems is expected to pave the way for a universal fault-tolerant quantum computing architecture. I have studied the encoding of quantum information in parafermions—a type of quasiparticles—which topologically protects quantum information. I have designed and implemented a dedicated optical quantum simulator, in which the geometric phase from the braiding of parafermions, the topological protection of quantum contextuality resources, and the conservation properties of quantum contextuality in braiding operations are investigated to illustrate the potential of this system for universal fault-tolerant quantum computing.

The thesis is originally written in Chinese, and I have self-archived the Chinese version of the thesis on my website: https://manekimeow.github.io/document/zhliu_PhDthesis.pdf. It is my great pleasure that I can present a translated version of the thesis to a broader audience one year after my graduation. I wish all the readers a nice experience in exploring the world of quantum contextuality.

Copenhagen, Denmark
May 2023

Zheng-Hao Liu

Acknowledgements

This thesis encapsulates my study experience and research endeavors at the University of Science and Technology of China and, in particular, within the CAS Key Laboratory of Quantum Information. I joined the lab in 2016 and began to conduct doctoral research under the guidance of my supervisors, Prof. Jin-Shi Xu and Prof Chuan-Feng Li. I have benefited a lot from my seven-year experience here—both professionally and personally. I would like to take this opportunity to express my gratitude and acknowledge the invaluable contributions and support I received throughout my research journey, which culminated in this thesis. Although the process of scientific research is always accompanied by setbacks and doubts, I still deeply feel that my doctoral career is extremely lucky, and all the experiences in it are precious. I am immensely fortunate to have spent seven years in this thriving environment, where quantum information stands as a dynamic field of scientific advancement. Our laboratory, equipped with cutting-edge research facilities, has provided me with the opportunity to delve into experimental research aligned with my interests. Moreover, the friendly and vibrant atmosphere fostered by our dedicated colleagues has been a constant source of inspiration and encouragement. Their presence has been invaluable, guiding me through moments of uncertainty and extending a helping hand during challenging experiments. I am indebted to this cohesive, family-like lab, without which I would not have achieved the research outcomes and acquired the profound knowledge of quantum information science I possess today.

Therefore, first of all, I would like to thank the Academician Guang-Can Guo for establishing the Key Laboratory of Quantum Information, which serves as a preeminent scientific research platform. This visionary initiative has provided us with unparalleled opportunities for exploration and discovery, setting the stage for groundbreaking research in the field. I extend my heartfelt thanks to Prof. Chuan-Feng Li for granting me the privilege to engage in scientific research within the Laboratory. His dedication to both my scientific endeavors and my personal well-being has been truly remarkable. I am grateful for his continuous guidance, his presence as a mentor, and the invaluable assistance and suggestions he has provided during times of confusion.

I owe a tremendous debt of gratitude to Prof. Jin-Shi Xu, for his expertise, unwavering support, and insightful advices all the time. From the day I joined the research group, Jin-Shi has been my guide in scientific research. His guidance is indispensable for every tiny achievement of mine. As the leader of our team, Jin-Shi has not only guided me in scientific research but has also set a remarkable example of dedication and diligence. He has not only taught us how to carry out the research topic from determining the direction to starting the experiment and pushing it all the way to the end, but also endowed the whole team with an enterprising spirit and an active atmosphere. I am truly grateful for his continuous presence when we spent countless nights in the laboratory to overcome experimental challenges. I also clearly remember that Jin-Shi would gather the team together at every festival and never missed a single time. The quantum optics conference at Chuzhou in 2019 he organized is always a good memory for me: he led

us to work for several days to carry out various detailed tasks in an ordered manner, and the conference really fostered a thriving research atmosphere.

I extend my deepest thanks to Kai Sun and Ze-Di Cheng who gave me advice and help in the experiment. Their tireless guidance and patience, especially in the early stages of my research, have been crucial in shaping my experimental skills. Thanks to Prof. Xiao-Ye Xu who led me to complete the quantum Cheshire cat experiment. He imparting valuable insights into transforming preliminary ideas into significant research outcomes and improving the quality of papers through addressing reviewers' inquiries. I would also like to express my gratitude to Prof. Junfeng Wang, whose guidance has been instrumental in both my scientific research and career planning, providing me with invaluable perspectives beyond the realm of scientific exploration as a Ph.D. student.

I would like to thank Prof. Jing-Ling Chen at Nankai University for his support. Jing-Ling has not only all the skill as a theoretician, but the attainments for mentoring and teaching. From several dialogues with him, I got my initial understanding of quantum information theory, and finally fell in love with this interesting research direction. Jing-Ling and his team have collaborated with me on six projects so far, none of which would have been possible without their efforts. My gratitude also goes to Prof Yong-Jian Han for his guidance and inspiration. Yong-Jian is approachable and has unique insights into the field of quantum computing. The discussions with him have provided some theoretical skills, and my choice of research direction is also a result of his guidance. I would like to thank Zhen-Peng Xu at Siegen University, Prof. Jie Ma from the Department of Mathematics at University of Science and Technology of China, Prof. Bo Li and Dr. Xiao-Bin Liang at Shangrao Normal University, Prof. Weidong Tang at Shaanxi Normal University, Prof. Huangjun Zhu at Fudan University, Prof. Xianmin Jin at Shanghai Jiaotong University, Prof. Fu-Lin Zhang at Nankai University, Dr. Jintao Fan at Tianjin University, Prof. Dong Wang at Anhui University, and Dr. Kai Wen from Beijing Bose Quantum and his colleagues. The discussions with them not only helped my research but also broadened my horizons.

I also want to thank all the teachers in the Key Laboratory of Quantum Information. Profs. Jian-Shun Tang, Yu-Chun Wu, Guo-Yong Xiang, Bi-Heng Liu, and Drs. Xiang-Jun Ye, Chao Zhang, Zhibo Hou, and Xiao-Min Hu guided me in theory and experiments; Tao Wang, Li Wu, Kaimin Duan, Hongmei Li and Yuchan Wang and Yong-Mei Zhu contributed so much to the management of the laboratory. Sheng Wang working at the geospace building provided a lot of logistical help. I am grateful to the students and colleagues in the laboratory. Mu Yang is my good friend and is just like a younger cousin to me. Our discussions at work are of great benefit to each other's research. After get off work, we are free and easy together, pursuing the happiness of life. Wei Liu and Ze-Yan Hao are my good roommates. Sharing the joys and sorrows of life with them makes life no longer boring every night. Drs. Ya Xiao, Shang Yu, Qiang Li and Yu Guo are all role models for my scientific research. No matter how long I have graduated, I will always think of their hard work and thank them for their support on all my learnings and works. Fei-Fei Yan, Wei-Wei Pan, Jun Hu, En-Ze Li and Zhi-Peng Li shared with me the 329 office where I enjoyed all the laughter. I would thank my fellow Ph.D. students Zhi-Jin Ke, Chao Liu, Qin-Qin Wang, Shikai Liu, Peng Yin, Chen Yang, Yan Wang, Jian-Yin Huang, Ji-Yang Zhou, Yu-Wei Liao, Zhao-An Wang, Yi Zheng, Zhen-Xuan He and also my friends at Nankai University—Hui-Xian Meng, Jie Zhou, Wei-Min Shang, Xing-Yan Fan—and Shihao Ru, a friend from Xi'an Jiaotong University, we study together and make progress together.

I am grateful to my international collaborators Prof. Adán Cabello, Prof. Jiannis Pachos, and Prof. Rosario Lo Franco for their theoretical support of my work and tireless revision of the paper. I would also like to thank Profs. Jonas S. Neergaard-Nielsen and Ulrik L. Andersen of the Technical University of Denmark, and Prof. Giuseppe Carleo of the Federal Institute of Technology in Lausanne, Switzerland, for their discussions on continuous variable quantum optics and variable quantum algorithms, giving me the passionate for the future research directions.

Thanks to Hefei, the city that hosts the University of Science and Technology of China. Many years later, I will still think of Zhao's BBQ and Niu Niu BBQ, the delicious intestines and noodles at the east gate, and brother Chun's Jinglaihe hot pot. Thanks to Dr. Shutian Ma of the Traditional Medicine Department of the university hospital and his colleagues. Thanks to my friends Haoran Geng, Junheng Ma, Shang Xu, Xiongguan Zhang, Peng Xie and Hong Cheng in my hometown. Due to the impact of the epidemic, I spent a national quarantine for nearly half a year and cannot go to the university. Playing basketball and chatting with them every day not only strengthens my body, but also provided me with perspective beyond scientific research. Thanks to my undergraduate friends Yiming Xu, Rui Ding, Fei He, Zhipeng Yao, Weiguo Jing, Yu Cao, Mingxiao Feng, Shihao Jing, Jie Wang, Yijie Sun, and Hongbo Chen. Every time I hear news from old friends, it is special to meet them happy. Thanks to my good friends Yangu Guo, Hang Zhou and Huayu Zhang. Communicating with them can always inspire each other and see a different world.

Finally, although my time at the university is like being in a big family, the family outside of work is another strong support behind me. I would like to take this chance to express my deepest gratitude to my father and mother. Without their nurturing and cultivation for 27 years, I would not have had the opportunity to become a mature person. Their silent support is my strongest backing. My heartfelt thanks go to my grandparents: even if they don't know what scientific research is, they always support me silently. Thanks to my aunt, Jing, for providing me with a different perspective on life. Finally, I want to express my deepest appreciation to my wife, Yu. As both a fellow researcher and a life partner, she has been my companion in scientific endeavors and a soul mate in every sense. Every conversation with her serves as a driving force for personal growth, and every moment spent together is a source of immeasurable happiness. To all of these incredible individuals, I am forever grateful for the love, encouragement, and support they have provided me throughout my academic and personal endeavors. Without them, I would not have reached this point in my life, and for that, I am eternally thankful.

Contents

Part I Theoretical and experimental foundations

1	Prelude	3
2	Quantum contextuality in a nutshell	9
2.1	Mathematical foundations of quantum mechanics	10
2.1.1	Quantum states, evolution, and observables	11
2.1.2	Quantum measurements, Born's rule, and Lüders' rule	12
2.1.3	Gleason's theorem	13
2.2	Hidden variable theories and contextuality	13
2.2.1	The Kochen–Specker theorem	14
2.2.2	<i>Friend A</i> 's confusion: the Peres–Mermin square	15
2.2.3	Experimentally testable noncontextuality inequality	18
2.2.4	Graph-theoretic approach to contextuality	21
2.3	Contextuality in quantum information science	24
2.3.1	Relation with the reality of wavefunction	24
2.3.2	Contradiction with classical causal models	25
2.3.3	Role in pre-post-selection quantum paradoxes	27
2.3.4	Interplay with nonlocality	27
2.3.5	Application in quantum computation	28
3	Linear optics quantum information	31
3.1	Encoding quantum information on photons	32
3.1.1	Generation of photon pairs	32
3.1.2	Photonic degrees of freedom	33
3.2	Evolution and operation of photons	35
3.2.1	Operation of polarization qubits	35
3.2.2	Coupling of polarization with other degrees of freedom	38
3.2.3	Imaginary-time evolution and non-Hermitian Hamiltonians	39
3.3	Measurement of optical quantum states	42
3.3.1	Polarization mode measurement	42
3.3.2	Spatial mode measurement	43
3.3.3	Temporal mode discrimination	44
3.3.4	Orbital angular momentum mode detection	45

Part II Interplay between contextuality and nonlocality

4 Stronger contextuality beyond nonlocality	49
4.1 Nonlocality, contextuality, and the graph of exclusivity	50
4.1.1 Case study: quantum correlations in the pentagon	51
4.2 Contextuality from measurement-repreparation experiments	52
4.3 Contextuality beyond nonlocality	55
4.3.1 Sketch of the theory	55
4.3.2 Experimental implementation	56
4.4 Contextuality concentration: strongest correlation in lower dimensions	61
4.4.1 Mermin inequality: exponential quantum–classical gap	61
4.4.2 Reducing the dimensionality of measurements in the Mermin inequality	62
4.4.3 Experimental implementation	64
4.5 Summary	68
5 “All-versus-nothing” contextuality in graph states	69
5.1 Theoretical foundations: “all-versus-nothing” paradox and the graph states	70
5.2 A generic construction of “all-versus-nothing” contextuality in graph state	73
5.2.1 Application: quantum State verification	74
5.3 Preparation of four-qubit graph states	76
5.4 Observation and applications of the “all-versus-nothing” paradox	79
5.4.1 Fidelity estimation and witnessing entanglement	80
5.4.2 “All-versus-nothing” paradox for quantum steering	82
5.5 Summary	83

Part III Contextuality in quantum information science

6 Contextuality and pre-post-selection paradoxes: the exchanged grins between quantum Cheshire cats	87
6.1 Tracking the Evolution Process of Quantum Systems	88
6.1.1 Two-state vectors and weak values	89
6.1.2 An example: photon, where have you been?	89
6.2 Contextuality and the quantum Cheshire cat paradox	93
6.2.1 Where is the grin of the quantum Cheshire cat?	93
6.2.2 Finding the quantum Cheshire cat from the Peres–Mermin square	94
6.2.3 “Quantum” Cheshire cat?	97
6.3 Quantum Cheshire cats’ exchanged grins	97
6.3.1 Observable and pre- and post-selection states	97
6.3.2 Weak value extraction without weak measurement	100
6.3.3 Experimental implementation	102
6.3.4 Experimental results	103
6.4 Summary	104
7 Topologically-protected contextuality: simulation of photonic-encoded parafermions	107
7.1 From parafermions to optical quantum simulation	107
7.1.1 Mathematical description of parafermions	108
7.1.2 Braiding operations of parafermions	109
7.1.3 Quantum simulation in brief	111

Contents	xix
7.1.4 Mapping parafermions to optical systems	112
7.2 Optical simulation of parafermion braiding statistics	114
7.2.1 Designing a dedicated quantum simulator	115
7.2.2 Experimental results.....	115
7.3 From contextuality to universal topological quantum computing	116
7.3.1 Dynamics of contextuality under braiding operations	119
7.3.2 Topologically-protected braiding of parafermions	120
7.3.3 Noise-resilience of parafermion contextuality	123
7.4 Conclusion	126
8 Postlude	129
A Case Study of Optical Quantum Information Experiment: Implementing Quantum Information Masking	133
A.1 Theoretical foundation	133
A.2 Experimental design	134
A.3 Results	135
A.4 Applications	138
References	141

Acronyms

CSW	Cabello–Severini–Winter
GHZ	Greenberger–Horne–Zeilinger
KCBS	Klyachko–Can–Binicioğlu–Shumovsky
KLQI	CAS Key Laboratory of Quantum Information
LHV	Local hidden-variable
CHSH	Clauser–Horne–Shimony–Holt
NCHV	Noncontextual hidden-variable
OAM	Orbital angular momentum
USTC	University of Science and Technology of China

Part I

Theoretical and experimental foundations

Chapter 1

Prelude

Quantum mechanics has dramatically changed the development of human society in less than a hundred years of history. Since the second world war, modern integrated circuits began to enter people's lives with the help of the development of the principles of quantum mechanics, becoming one of the driving forces for the progress of human civilization, and bringing human society into the information age. At the end of the twentieth century, the combination of the principles of quantum mechanics and information theory gave birth to a new interdisciplinary subject—the quantum information science. As one of its most prominent applications, quantum computing based on manipulating quantum systems holds the promise of a further leap in human computing power [1]. The ongoing “second quantum revolution” has further improved people’s ability to operate a single quantum entity, so quantum technology can be widely used in many aspects of human society. For example, quantum computers have shown a decisive speedup in solving some highly contrived problems when compared with classical computers [2, 3].

Despite countless significant achievements since its inception in the 1920s, quantum mechanics has had a complex developmental history. Over time, our understanding of it has continuously evolved, leading to numerous disruptive new insights. For instance, von Neumann, who played a pioneering role in the development of quantum mechanics and computer science, is considered the founder of quantum information science. However, his views on non-commuting variables were criticized as “silly” by Mermin [4] in the 1990s. We now know that the criticized viewpoint of von Neumann actually represents a common problem within a class of phenomenological descriptions known as hidden variable theories, which seek to replace quantum mechanics. Hidden variable theories attempt to reflect the outcomes of measurements in quantum mechanics using pre-determined (although unknown) hidden variables. In this way, the process of quantum measurement simply reveals the physical quantities as functions of their corresponding hidden variables. Indeed, according to the Copenhagen interpretation, quantum mechanics is an intrinsically uncertain theory based on the wave function. Thus, there is sufficient motivation to reframe quantum mechanics as a deterministic theory by employing hidden variable theories, where all uncertainty arises from our ignorance of the hidden variables. However, this comes at the price of introducing additional assumptions, thereby weakening the credibility of the theory. Some hidden variable theories can be proven to be completely compatible with standard quantum mechanics, such as de Broglie and Bohm [5]’s pilot wave theory, which differ only in their interpretation of the same physical phenomena. However, this is not the case for other hidden variable theories, primarily because the (seemingly reasonable) assumptions employed in these theories are fundamentally incompatible with quantum mechanics itself. One representative example of such theories is the noncontextual hidden variable theory.

Quantum contextuality¹ refers to the dependence of measurement outcomes in quantum mechanics on the measurement context, specifically on how commuting observables are selected for simultaneous measurement. In contrast, noncontextual hidden variable theories posit models where such contextuality is absent. Both empirically and mathematically, noncontextual assumptions appear reasonable: empirically, the measurement result for an object's length should not depend on the choice of the ruler, as we have good reason to question the accuracy of the ruler if it does. Mathematically, the fundamental correspondence between wave functions and probabilities in quantum mechanics—the Born rule—is noncontextual. Readers familiar with relativity may notice a similarity to Einstein's conundrum [6]: when a measurement collapses the wave function of a particle in an entangled multi-particle system, the wave function of a distant particle appears to instantaneously change, seemingly contradicting relativity. However, when considering the average effects of measurements, the collapse of the wave function does not result in any effective information being transferred. An observer of the distant particle not only gains no information about the measurement outcome but cannot even ascertain whether a measurement has occurred. The distinction lies in whether single events or statistical results are considered. Statistically, all properties of quantum mechanics are local and noncontextual. However, quantum mechanics' description is not only applicable to ensembles but also valid for individual events. When focusing on the outcomes of single measurements, quantum mechanics exhibits its nonlocal and contextual characteristics [7]. The work of Fine [8] demonstrates that by considering the intercorrelations among probability distributions of multiple observables, it is possible to manifest this contextual feature in the probability distribution.

From the above discussion, it is apparent that the experimental studying of contextuality often requires more than just the knowledge of the overall probability distribution of a system under various measurement bases. Instead, it necessitates precise measurements of individual quantum systems to obtain a wealth of information about their evolution (which may include repeated measurements). Research on contextuality imposes highly demanding requirements on experimental techniques, as evident in the milestones of this field depicted in [Figure 1.1](#). It can be observed that theoretical investigations in the upper part attracted significant attention early on, while experimental research in the lower part only made groundbreaking progress in recent years. Particularly, at the turn of the century, according to discussions by Meyer [24], Kent [25], and Clifton and Kent [26], it seemed impossible to detect any distinction between quantum mechanics and noncontextual hidden variable theories within finite precision. Consequently, contextuality would be reduced to a philosophical question. Indeed, during that era, all experiments on contextuality, such as the works of Hasegawa *et al.* [27] and Bartosik *et al.* [28], relied on certain assumptions of quantum mechanics, leading to some degree of circular reasoning. A turning point occurred in 2008 when Cabello [13] proposed a method to experimentally detect contextuality without relying on any assumptions of quantum mechanics. Furthermore, they considered how to incorporate the effects of experimental errors [29]. Since then, experimental research on contextuality has made significant strides, enabling the testing of various noncontextual inequalities in different systems.

Due to its crucial role in various aspects of quantum information science, particularly in the context of future scalable universal quantum computation [16], contextuality has remained a focal point of research in the field. The Key Laboratory of Quantum Information (KLQI) at the University of Science and Technology of China (USTC) has long been devoted to studying contextuality, with a history spanning two decades in experimental detection [18]. Considering the hardware requirements for performing contextuality experiments, the commonly used platforms worldwide are primarily based on superconducting

¹ Later in this thesis, I will omit the attributive “quantum” in front of “contextuality” on a large number of occasions to enforce concise and smooth writing without causing ambiguity. However, I also point out that in the generic probability theory, there is a broad class of contextual probability models, so contextuality is by no means a unique property of quantum mechanics. In addition, the content in the footnotes contains my personal opinions and does not necessarily reflect the objective and correct science. If these contents do not conform to the reader's understanding, the reader's understanding should prevail.

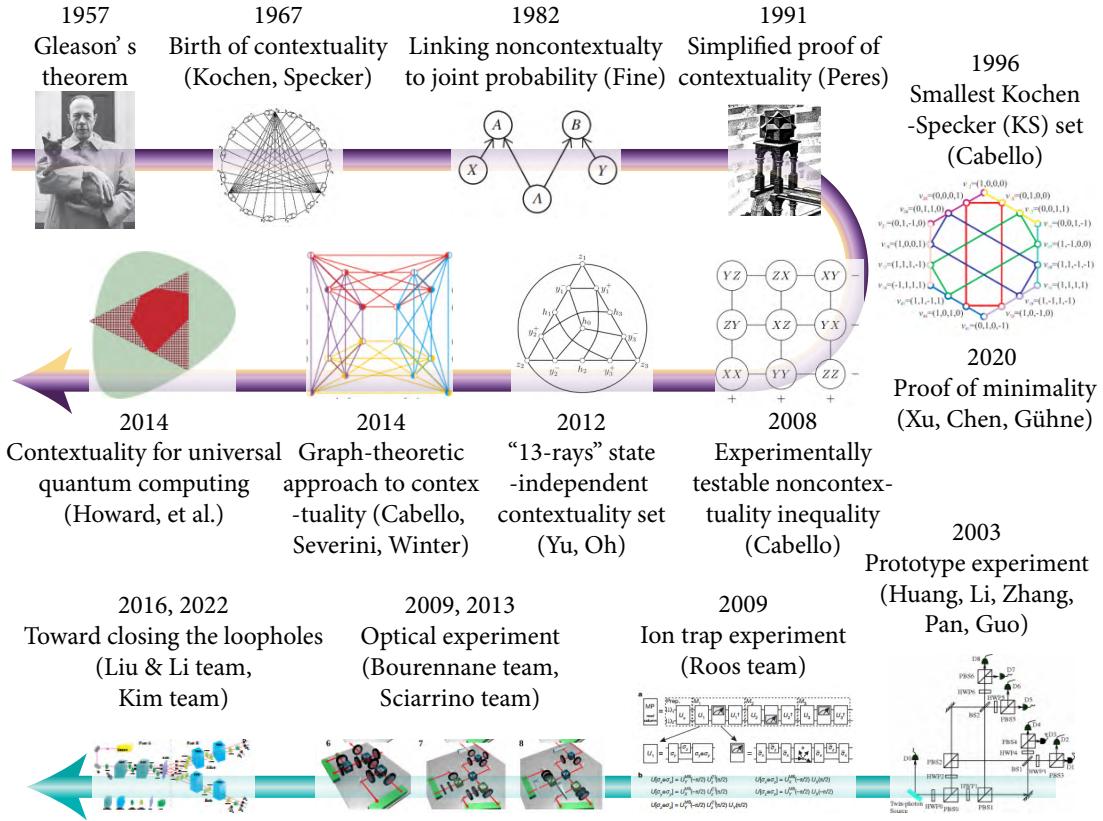


Fig. 1.1 History of the development of contextuality. The figure highlights the theoretical advancements by Fine [8], Gleason [9], Kochen and Specker [10], Peres [11], Cabello *et al.* [12], Cabello [13], Yu and Oh [14], Cabello *et al.* [15], Howard *et al.* [16], Xu *et al.* [17] and the experimental contributions by Huang *et al.* [18], Kirchmair *et al.* [19], Amselem *et al.* [20], D'Ambrosio *et al.* [21], Hu *et al.* [22], and Wang *et al.* [23].

circuits, ion traps, and optical systems. During my doctoral studies, my experimental work in the field of quantum contextuality was conducted using the internationally leading optical platform available in our laboratory, specifically focusing on linear optical systems [30]. In linear optical setups, the Hamiltonian describing the evolution process contains at most terms linear in the creation and annihilation operators of photons. Therefore, it neither generates additional photons nor induces interactions between photons. In summary, this system possesses the following advantages:

- Extremely long coherence time:** Compared to matter qubits, photons are minimally affected by decoherence due to electromagnetic field disturbances. The coherence time of photons is significantly longer relative to gate operation times.
- Rich intrinsic degrees of freedom:** Quantum information can be encoded in multiple degrees of freedom of photons. Particularly, photons have a plethora of degrees of freedom that can encode high-dimensional quantum information, which is crucial for contextuality.

3. **High precision in quantum state preparation, manipulation, and measurement:** This advantage facilitates precise testing of quantum mechanical predictions.

Therefore, linear optical systems provide a solid foundation for exploring new advancements in contextuality research. In fact, linear optical systems have been widely utilized as experimental platforms in fundamental quantum mechanics research and quantum communication applications. However, the aforementioned advantages come with a critical cost: since there are no interactions between photons, post-selection is necessary to achieve equivalent two-qubit gate operations, with a success rate that cannot exceed 50%. Consequently, if linear optics and an all-optical setup are employed without nonlinear interactions like the Kerr effect, quantum entanglement is inherently unscalable. Additionally, the lack of reliable two-qubit gates limits the ability to perform continuous measurements in optical systems to some extent.

The work presented in this study explores, to some extent, methods for addressing the aforementioned challenges. It allows for meaningful research on contextuality and its potential future applications using the available resources of current optical platforms, even without achieving true scalability of optical systems². This thesis will dedicate considerable space to focus on the following research areas:

1. Theoretical investigation: the verification of quantum contextuality is theoretically reduced to counting statistics in single-system prepare-and-measure experiments, thus avoiding the reliance on multi-photon interference and sequential measurements. In [chapter 4](#), I will present two achievements using this approach to study the relationship between contextuality and nonlocality, demonstrating how the strongest known quantum correlations can be embedded into lower-dimensional systems without weakening their strength.
2. Experimental exploration: leveraging the coupling between multiple degrees of freedom of photons to expand the accessible state space. In [chapter 5](#), I will describe the construction of graph states and the study of the “all-versus-nothing” contextuality paradox, enabling state verification and entanglement witnessing.
3. Development of measurement techniques that do not rely on additional pointers, further reducing the coupling requirements between quantum systems during the measurement process. In [chapter 6](#), I will introduce the extraction of weak values, a process quantity characterizing quantum system evolution, using this method. This will allow the observation of a novel phenomenon known as the quantum Cheshire cat.
4. Application of quantum simulation to expand the range of problems that can be studied. In [chapter 7](#), I will present experimental simulations on the optical platform to explore quasiparticle braiding statistics and topological protection properties, demonstrating the prospects of quasiparticle systems for future quantum computation.

Furthermore, to provide a foundation for the central results mentioned above, [chapter 2](#) will provide a brief overview of the theoretical basis of quantum contextuality, while [chapter 3](#) will outline the methods used in linear optical experiments. In [chapter 8](#), I will offer a further outlook on the potential future developments in these directions and provide a brief introduction to ongoing work.

Next, let us proceed to the introduction of the theoretical foundation of quantum contextuality. Compared to well-established conclusions in quantum mechanics such as quantum entanglement and nonlocality, the implication of contextuality is more elusive, and discussions about it often involve profound

² I believe that optical systems will be the most scalable quantum computing architecture in the future—beyond the scale of 10^6 qubits, matter-based systems will have to face the effects of decoherence due to coupling with the environment, whereas the decoupled advantage of optical systems will be decisive. However, it currently appears that achieving large-scale optical quantum computation may require further combinations with techniques such as the optomechanical interface [31], continuous variable cluster states [32, 33], and measurement-based quantum computing [34], rather than solely relying on linear optical systems.

mathematical concepts, which may discourage readers who are not familiar with the relevant background. Therefore, the theoretical part of this thesis will employ a small story to illustrate the peculiar effects caused by contextuality, supplemented by fundamental knowledge of quantum mechanics and simple calculations. In doing this, I imagine the concept of contextuality could become more illustrative to the readers. Although this approach may not be as rigorous as a framework based solely on axiomatic logic, I hope that through vivid and imaginative storytelling, readers can quickly develop an intuitive understanding of contextuality and enhance their comprehension.

Chapter 2

Quantum contextuality in a nutshell

Imagine there is a room with a solid-colored cat, either black or white, and we want to know the color of the cat. We can look into the room from the doorway, or the window, or place a camera in the corner and observe when the cat comes in front of it. Regardless of the method we use, we will always see either a black cat or a white cat. It is impossible to see the color of the cat change from different perspectives, or else we would suspect that there are at least two cats in the room. The color of the cat should be objective and independent of the observer's viewpoint.

If the cat in the room is a “Schrödinger’s cat” [35], the situation becomes more interesting. Before observation, the cat may be in a superposition of black and white states, and our observation causes it to collapse into either a black cat or a white cat¹. However, if the observation itself does not affect the behavior of the cat, there seems to be ample reason to believe that whether the Schrödinger’s cat becomes a black cat or a white cat should not depend on whether we observe it from the doorway, window, or camera. To test this hypothesis, we can create many identical Schrödinger’s cats and observe them randomly from each viewpoint. As expected, the probabilities of the Schrödinger’s cats becoming a black or white cat are completely the same for each viewpoint, still independent of the observer’s perspective.

One of my curious biologist friends (whom I will refer to as *Friend A*) is very interested in Schrödinger’s cat. He encountered a Schrödinger’s cat as shown in Figure 2.1 and would like to further describe the principle of the color change of Schrödinger’s cat in depth. He assumes the existence of an unobserved mechanism on the cat, which determines in advance what color the cat will eventually become in each experiment, and the human observation process only reveals the color that has already been predetermined. However, during his research, *Friend A* found that he couldn’t eliminate the dependence of the color-changing mechanism on the observation method: when observed from the doorway or the window, the way the color changes in the Schrödinger’s cat is different! After analyzing the experimental data, *Friend A* was puzzled and came to me for advice on the intricacies involved. Upon hearing *Friend A*’s description, I smirk, knowing that the mischief at play in the Schrödinger’s cat was nothing other than contextuality.

For a long time, contextuality has been a central issue in fundamental research in quantum mechanics. It is an essential requirement of any potential theoretical framework to account for the observed experimental results. However, its counterintuitive nature has sparked exploration into the physical principles of quantum theory. The fundamental role of contextuality in quantum theory gives prominent significance to research efforts related to it, both in terms of theoretical understanding and practical applications. To study contextuality, it is crucial to first understand its origins and implications. This chapter will begin by elucidating the natural consequence of contextuality as a mathematical structure in quantum mechanics from a foundational scientific perspective. It will then discuss the wide-ranging applications

¹ I will use a slightly modified reference to Schrödinger’s cat since we clearly prefer the cat to be alive.



Fig. 2.1 A Schrödinger's cat is a cat that exists in a superposition of black and white states, following the description of quantum mechanics. The cat, at present, has undergone wavefunction collapse due to measurement during filming and has transformed into a pure black cat.

of contextuality in fields such as quantum information processing and quantum computing from an interdisciplinary standpoint. As for the story of me and my *Friend A*, it will be continued in section 2.2.

2.1 Mathematical foundations of quantum mechanics

As possibly the most successful scientific theory in human history, quantum mechanics has an extraordinary mathematical form. The entire edifice of non-relativistic quantum mechanics is built upon

non-commutative matrix multiplication and the foundation of C^* -algebras. Understanding these theoretical foundations will help grasp the problems that the experiments in this paper are attempting to address. Therefore, we will now briefly review the basic knowledge of these theories.

2.1.1 Quantum states, evolution, and observables

A pure state of a d -dimensional quantum system $|\psi\rangle$ can be described using its corresponding ray as $|\psi\rangle = \sum_{k=1}^d c_k |\phi_k\rangle$, where $|\phi_k\rangle, k \in 1, \dots, d$, is a set of orthogonal basis vectors in the given Hilbert space \mathcal{H}^d , commonly referred to as the computational basis. In the computational basis, we often represent the ray as a vector:

$$|\psi\rangle = (c_1^* \quad c_2^* \quad \cdots \quad c_{d-1}^* \quad c_d^*)^\dagger.$$

Additionally, we use the Kronecker symbol $\delta_{jk} = \begin{cases} 1, & \text{if } j = k \\ 0, & \text{if } j \neq k \end{cases}$ to express the orthogonality and normalization conditions of the ray as $\langle \phi_j | \phi_k \rangle = \delta_{jk}$ and $\sum_{k=1}^d |c_k|^2 = 1$.

If a quantum system consists of multiple subsystems in pure states $|\psi_A\rangle_A, |\psi_B\rangle_B, \dots$, the state of the entire system can be described using the tensor product as $|\psi AB \dots\rangle = |\psi_A\rangle_A \otimes |\psi_B\rangle_B \otimes \dots$, often abbreviated as $|\psi_A\rangle |\psi_B\rangle$ or $|\psi_A, \psi_B\rangle$. However, composite systems in pure states may not always be decomposable into a direct product of subsystems. In such cases, we refer to the composite system as entangled. For a system of two qubits, there are four quantum states that exhibit the strongest entanglement between the two particles, known as Bell states. Mathematically, they can be expressed as:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad |\Phi^-\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \quad |\Psi^+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad |\Psi^-\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}.$$

A quantum system in a mixed state cannot be described as a ray, but can be represented by a density operator $\hat{\rho}$, which describes a probabilistic mixture of pure states:

$$\hat{\rho} = \sum_k w_k |\psi_k\rangle \langle \psi_k|, \quad \sum_k w_k = 1.$$

The density operator can be expanded in matrix form as:

$$\begin{aligned} \hat{\rho} &= \sum_{mn} |\phi_m\rangle \langle \phi_n|, \quad \text{with} \\ \rho_{mn} &= \langle \phi_m | \hat{\rho} | \phi_n \rangle = \sum_k w_k \langle \phi_m | \psi_k \rangle \langle \psi_k | \phi_n \rangle. \end{aligned}$$

The density operator is a Hermitian operator ($\rho^\dagger = \rho$) with a trace of $\text{Tr}\hat{\rho} = 1$. For a pure state $|\psi\rangle$, its density operator is $|\psi\rangle \langle \psi|$. In this case, the spectrum of the density operator matrix has only one non-zero element. Therefore, the proximity of a density operator to a pure state can be determined by the square of the density operator: the purity of a quantum state $\hat{\rho}$ is defined as $\text{Tr}(\hat{\rho}^2)$, and states with purity less than 1 are mixed states [36].

Quantum systems evolve under the constraints of the Schrödinger equation. The most general form of the Schrödinger equation is:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle,$$

where \hat{H} is the Hamiltonian that governs the dynamics of the quantum system. In quantum information science, it is common to set $\hbar = 1$, and hence we will omit this symbol in the following discussions. If \hat{H} is time-independent, the evolution of the state vector follows the equation $|\psi(t)\rangle = U(t) |\psi(0)\rangle$, where $U(t) = \exp(-i\hat{H}t)$.

Traditionally, all Hamiltonians are required to be Hermitian operators, which ensures that the trace of the evolved quantum state remains normalized. However, in recent years, there has been a growing interest in the study of non-Hermitian Hamiltonians. They describe the dynamics of open quantum systems and exhibit novel physical phenomena and topological properties [37]. One equivalent approach to describe non-Hermitian Hamiltonians is to preserve the Hermiticity of the Hamiltonian but perform the evolution in imaginary time [38]. The corresponding Schrödinger equation for this “imaginary-time evolution” is:

$$-\frac{\partial}{\partial \not{t}} |\psi(\not{t})\rangle = \hat{H} |\psi(\not{t})\rangle, \quad (2.1)$$

where the slashed \not{t} is used to emphasize the evolution in imaginary time. In this case, the system’s evolution operator due to the time-independent Hamiltonian is given by:

$$U(\not{t}) = \exp(-\hat{H}\not{t}). \quad (2.2)$$

The major difference between this evolution and unitary evolution is that the probability amplitudes are not normalized. This leads to an exponential decay of probability amplitudes corresponding to eigenstates with eigenvalues E_k of the Hamiltonian to $\exp(-E_k\not{t})$ at the initial time. However, this type of evolution is not only experimentally feasible in quantum optics experiments but also serves certain purposes that cannot be achieved solely through unitary evolution. In the work presented in Chapters 6 and 7 of this thesis, imaginary-time evolution is extensively employed to achieve specific experimental effects.

Observable quantities reflect the properties of quantum systems and establish correspondences with classical physics. Each observable can be represented as a Hermitian matrix, thus making its eigenvalues real. For a quantum system, the expectation value of an observable \hat{O} is given by:

$$\langle \hat{O} \rangle_{\psi} = \langle \psi | \hat{O} | \psi \rangle.$$

The Hamiltonian of a closed system is also an observable, corresponding to the system’s energy.

Each observable \hat{O} can be diagonalized in its own corresponding basis, where the diagonal elements represent the spectrum of the operator:

$$\hat{O} = \sum_{\lambda} \lambda \hat{\Pi}_{\lambda},$$

where $\hat{\Pi}_{\lambda} = \sum_k |\phi_{\lambda,k}\rangle \langle \phi_{\lambda,k}|$ is the operator that projects onto the eigenstate $|\phi_{\lambda,k}\rangle$ with eigenvalue λ . If the eigenvalue λ is non-degenerate, the summation has only one element, and in that case, $\hat{\Pi}_{\lambda} = |\phi_{\lambda}\rangle \langle \phi_{\lambda}|$.

2.1.2 Quantum measurements, Born’s rule, and Lüders’ rule

Born’s rule provides the probabilities of measurement outcomes for quantum states. For a quantum state $\hat{\rho}$, the measurement of an observable $\hat{O} = \sum_{\lambda} \lambda \hat{\Pi}_{\lambda}$ will yield an eigenvalue λ of the observable with a probability $\text{Pr}(\lambda)$. The probability is calculated as $\text{Pr}(\lambda) = \text{Tr}(\hat{\rho} \hat{\Pi}_{\lambda})$.

When two observables \hat{O}_1 and \hat{O}_2 commute ($[\hat{O}_1, \hat{O}_2] = \hat{O}_1\hat{O}_2 - \hat{O}_2\hat{O}_1 = 0$), both observables can be measured simultaneously without affecting each other's measurement probabilities given by Born's rule. However, measuring two non-commuting observables will disturb their values, and meaningful results cannot be obtained simultaneously. Orthogonal projection operators are examples of commuting observables.

Lüders' rule describes how quantum measurements induce changes in quantum states. Specifically, when the measurement of an observable \hat{O} yields a result λ , the measurement causes the quantum system to "collapse" to the corresponding eigenstate:

$$\hat{\rho} \rightarrow \hat{\rho}' = \frac{\hat{\rho}\hat{\Pi}_\lambda}{\text{Pr}(\lambda)}.$$

2.1.3 Gleason's theorem

The aforementioned Born's rule and Lüders' rule provide a complete description of measurement in quantum mechanics, but when they were proposed, they were given as postulates. Why is the measurement probability in quantum mechanics uniquely defined by the trace of matrix multiplication? The following Gleason's theorem precisely aims to demonstrate the uniqueness of this particular form.

Theorem 2.1 (Gleason, 1957) *Given a Hilbert space H^d with $d \geq 3$, the mapping*

$$\boldsymbol{\Pi} \rightarrow \mathbf{P},$$

from a set of projection operators $\boldsymbol{\Pi} = \{\hat{\Pi}_1, \dots, \hat{\Pi}_d\}$ with $\sum_k \hat{\Pi}_k = \mathbb{I}_d$ to a set of probability distributions $\mathbf{P} = \{p_1, \dots, p_d\}$ with $0 \leq p_k \leq 1$ and $\sum_k p_k = 1$ can be equivalently represented by a density operator $\hat{\rho}$ as

$$p_k = \text{Tr}(\hat{\rho}\hat{\Pi}_k).$$

The proof of Gleason's theorem is rather intricate and is beyond the scope of this paper. This theorem demonstrates that Born's rule is the only possible theory to explain quantum measurement outcomes. Considering the constraints of special relativity, the quantum state evolution induced by measurement must be determined by Lüders' rule. Therefore, Gleason's theorem provides a mathematical foundation for the measurement process in quantum mechanics.

Due to Gleason's theorem, even for pure states, any measurement of an observable that is not in an eigenstate of the observable can only yield results in a random manner. Thus, quantum mechanics, conforming to this mathematical language, must possess inherent uncertainty. In the next section, we will see that this requirement for randomness imposes strong constraints on the mathematical form satisfied by hidden variable models capable of reproducing all predictions of quantum mechanics.

2.2 Hidden variable theories and contextuality

A deterministic hidden variable theory refers to a scenario where the outcomes of a set of orthogonal projection measurements $\boldsymbol{\Pi}$ in a single experiment are predetermined by an unknown hidden variable μ . We focus only on deterministic hidden variable models. Non-deterministic hidden variable models can

be constructed using multi-level deterministic hidden variables, but such constructions are redundant. Ultimately, there must be a level of deterministic hidden variables that can fully describe physical reality, regardless of whether this reality can be observed experimentally in any way.

The hidden variable theories attribute the randomness of quantum mechanics to a mapping from observables to measurement outcomes:

$$\text{Pr}(\lambda) \rightarrow \int_{\mu} v(\lambda, \Pi, \mu) d\mu.$$

Here, the response function $v(\lambda, \Pi, \mu) \in 0, 1$ determines the measurement outcome that can be predetermined for the orthogonal projection measurement Π corresponding to a hidden variable value μ .

Noncontextual hidden variable theories are a paradigm within deterministic hidden variable theories. Specifically, a projection operator $\hat{\Pi}_0$ with an eigenvalue λ_0 can belong to several different contexts (different choices of orthogonal projection measurements), denoted as Π and Π' for convenience. In noncontextual hidden variable theories, the response function is independent of the choice of context. Mathematically, this can be expressed as:

$$v(\lambda_0, \Pi, \mu) \equiv v(\lambda_0, \Pi', \mu), \quad \forall \mu. \quad (2.3)$$

In other words, the predetermined measurement outcome is only affected by the measured projection operator and the hidden variable itself irrespective of the other projection operators that constitute a complete orthogonal measurement along with the measured projection operator.

2.2.1 The Kochen–Specker theorem

Due to Gleason's theorem which requires that the probabilities of all projection measurements in a $d \geq 3$ dimensional Hilbert space be mapped to the closed interval $[0, 1]$, there is an incompatibility between the continuous mapping of deterministic hidden variable theories and the discrete nature of the response functions, which can only take values in $0, 1$ for a single experiment. Kochen and Specker pioneered in providing a concrete example that demonstrated the irreconcilable contradiction arising from this mathematical discrepancy in noncontextual hidden variable theories that, crucially, do not employ the information of the complete sets of orthogonal projection operators. Formally, we have:

Theorem 2.2 (Kochen and Specker, 1967) *Quantum theory is contextual. A hidden variable description of quantum measurements must include information about the context, i.e., all the projection operators that constitute a complete orthogonal measurement.*

When Kochen and Specker initially proved the above theorem over 50 years ago, they used 117 state vectors corresponding to projection measurements to illustrate the contradiction between a noncontextual hidden variable model and the predictions of quantum mechanics. With the development of quantum information science, the proof process has gradually become simpler. In 2012, Yu and Oh constructed a proof using only 13 projection measurements to exclude noncontextual hidden variable theories for measurements on any quantum state [14]. Moreover, their record for the simplest proof of state-independent contextuality has been shown to be unbreakable [29]. If the aim is only to demonstrate the contradiction with noncontextual hidden variable models for a certain subset of quantum states, the requirement can be further reduced to using only 5 projection measurements [39].

2.2.2 *Friend A's confusion: the Peres–Mermin square*

The previous explanations, which are based entirely on the foundations of quantum mechanics and abstract mathematics, may still be insufficiently intuitive. Now, I shall return to the example of Schrödinger's cat mentioned at the beginning of this chapter to clarify why quantum mechanics is incompatible with noncontextual hidden variable theories. To study the pattern of color changes in "Schrodinger's cat" when observed, *Friend A* employs a controlled variable approach to establish a connection between fur color and other observable characteristics of the cat. *Friend A* discovers that this dependence can be completely represented by the colors of the cat's two eyes and the colors of its two front paws. Specifically, the colors of the cat's two eyes can be either (1) both yellow, (2) both green, or (3) one yellow and one green, while the colors of its paws can be either (1) both pink, (2) both black, or (3) one pink and one black.

Friend A also observes that when observing the cat from the front through a door, they can observe the colors of the cat's two eyes and whether the colors of its paws are the same. When observing the cat from the side through a window, they can observe the colors of each eye and paw on each side. Furthermore, *Friend A* summarizes the following rules for the color changes in Schrödinger's cat:

-
1. When observed from the door, if Schrödinger's cat ultimately becomes a white cat, then either both eyes and both paws are either the same color or different colors. Conversely, if Schrödinger's cat ultimately becomes a black cat, then if the iris colors are different, the paw colors are the same; if the iris colors are the same, the paw colors are different.
 2. When observed from the window, if Schrödinger's cat ultimately becomes a white cat, then if one side has (1) yellow eyes and pink paws or (2) green eyes and black paws, the other side either has (a) green eyes and pink paws or (b) yellow eyes and black paws. If one side has (1) yellow eyes and black paws or (2) green eyes and pink paws, the other side either has (a) green eyes and black paws or (b) yellow eyes and pink paws. The situation is the opposite for Schrödinger's cat that ultimately becomes a black cat, compared to the description above for the white cat.
-

Friend A, who is good at summarizing the observed patterns, will use some variables to describe the color changes of Schrödinger's cat. Specifically, since the color of the cat itself, the color of its eyes, and the color of its paws only have two possibilities each, the Boolean variables are well suited to describe these relationships. *Friend A* defines the following variables:

$$\begin{cases} x_1 = x_1(\text{color of left eye}) & : x_1 = +1(-1) \Leftrightarrow \text{left eye is yellow (green)} \\ x_2 = x_2(\text{color of right eye}) & : x_2 = +1(-1) \Leftrightarrow \text{right eye is yellow (green)} \\ z_1 = z_1(\text{color of right paw}) & : z_1 = +1(-1) \Leftrightarrow \text{right paw is pink (black)} \\ z_2 = z_2(\text{color of left paw}) & : z_2 = +1(-1) \Leftrightarrow \text{left paw is pink (black)} \\ y = y(\text{cat's color}) & : y = +1(-1) \Leftrightarrow \text{cat is black (white)} \end{cases} \quad (2.4)$$

Friend A wants to describe the pattern of color changes of Schrödinger's cat using the relationships between these variables as $y = y(x_1, x_2, z_1, z_2)$. Specifically, when observed from the front door, we have:

$$y = -(x_1 x_2)(z_1 z_2). \quad (2.5)$$

When observed from the window, we have:

$$y = (x_1 z_2)(z_1 x_2). \quad (2.6)$$

It can be verified that the above equations (2.5) and (2.6) fully reproduce the effects of color changes on Schrödinger's cat when observed from the front door and the window, respectively. However, no matter how hard *Friend A* tries, he cannot derive a relationship of the form $y = y(x_1, x_2, z_1, z_2)$. He found that there is a contradiction when multiplying both sides of equations (2.5) and (2.6):

$$+1 = y^2 = -(x_1 x_2)(z_1 z_2)(x_1 z_2)(z_1 x_2) = -(x_1)^2 (x_2)^2 (z_1)^2 (z_2)^2 = -1!$$

In other words, without specifying whether the observation is made from the front door or the window, it is fundamentally impossible to define a function $y = y(x_1, x_2, z_1, z_2)$ that describes the relationship between the color of Schrödinger's cat and the colors of its eyes and paws.

Friend A summarized his observations in an intuitive chart, shown in Figure 2.2, to further describe his confusion. He discovered that finding a function that holds true for both observations from the front door and the window, describing the color changes of Schrödinger's cat, is equivalent to filling each cell in the table with a ± 1 value, such that the product of variables in each row and each column equals $+1$, except for the last row, which equals -1 . However, this problem has no solution. Furthermore, *Friend A* still believes that the observation method should have no effect on the behavior of Schrödinger's cat because he found that the probabilities of the cat's color, eye color, and paw color are exactly equal under both observation methods.

x_1	z_2	$x_1 z_2$
x_2	z_1	$x_2 z_1$
$x_1 x_2$	$z_1 z_2$	y

$\xrightarrow{\hspace{2cm}}$

Observable from the door

\downarrow

Observable from the window

Fig. 2.2 Illustration of the difficulties encountered when attempting to describe the color changes of Schrödinger's cat using a function. When observed from the front door, it is possible to measure variables within the same row, while when observed from the window, variables within the same column can be measured. The product of variables in each row and each column should equal $+1$, except for the last row, which equals -1 . It is impossible to fill the entire square with a set of ± 1 values that satisfy the two conditions mentioned above.

Now, I will begin to address *Friend A*'s confusion using a series of concepts from the perspective of quantum mechanics. Unlike cats we often encounter in daily life, Schrödinger's cat is a creature of the

quantum world, and its description can only be done using a state vector or wave function $|\psi\rangle$. The color characteristics of Schrödinger's cat are determined by the measurement results of certain observables in the state $|\psi\rangle$. Specifically, these observables are the Pauli operators frequently used in quantum mechanics:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

the set of observables was first discovered by Peres and Mermin [4] and can also be represented as a square grid within a 3×3 matrix, known as the Peres–Mermin square, as shown in [Figure 2.3](#). We use the arrow symbol \rightarrow to denote the measurement results of an observable. The relationship between these results and the colors of Schrödinger's cat is as follows:

$$\left\{ \begin{array}{l} \sigma_{x1}(\text{left eye color}) : \sigma_{x1} \rightarrow +1(-1) \Leftrightarrow \text{left eye is yellow (green)}; \\ \sigma_{x2}(\text{right eye color}) : \sigma_{x2} \rightarrow +1 \Leftrightarrow \text{right eye is yellow (green)}; \\ \sigma_{z1}(\text{right paw color}) : \sigma_{z1} \rightarrow +1(-1) \Leftrightarrow \text{right paw is pink (black)}; \\ \sigma_{z2}(\text{left paw color}) : \sigma_{z2} \rightarrow +1(-1) \Leftrightarrow \text{left paw is pink (black)}; \\ \sigma_{y1} \otimes \sigma_{y2} \equiv \mathcal{Y}(\text{cat color}) : \mathcal{Y} \rightarrow +1(-1) \Leftrightarrow \text{cat is black (white)}. \end{array} \right. \quad (2.7)$$

Based on the commutation relations of the Pauli operators, $[\sigma_{xi}, \sigma_{yj}] = \delta_{ij}\sigma_{zi}$, *Friend A*'s chosen function does not involve commuting all four independent variables. Therefore, seeking a function form of \mathcal{Y} solely depending on the four binary variables $x1, x2, z1, z2$ would be futile—it is only possible to achieve this through the two sets of correlation functions in the last row and last column of [Figure 2.3](#).

$\sigma_{x1} \otimes \mathbb{I}_2$	$\mathbb{I}_2 \otimes \sigma_{z2}$	$\sigma_{x1} \otimes \sigma_{z2}$
$\mathbb{I}_2 \otimes \sigma_{x2}$	$\sigma_{z1} \otimes \mathbb{I}_2$	$\sigma_{z1} \otimes \sigma_{x2}$
$\sigma_{x1} \otimes \sigma_{x2}$	$\sigma_{z1} \otimes \sigma_{z2}$	$\sigma_{y1} \otimes \sigma_{y2}$

Fig. 2.3 Peres–Mermin square is considered one of the most well-known and elegant examples of contextuality. It is obtained by replacing the binary variables in [Figure 2.3](#) with products of Pauli operators. Each cell in the Peres–Mermin square represents a 4-dimensional observable, and for any quantum state, the expected values of the product of observables in each row and column are equal to +1, except for the last row where the expected value is -1. This property holds regardless of the exact quantum state being considered.

It is easy to observe that regardless of whether we observe from the doorway or the window, the observables that can be measured simultaneously in each row or column commute with each other, which is in line with the requirements of quantum mechanics. The probabilities of any observation also do not depend on the measurement method. For example, according to the Born rule, the probability of measuring the left eye of the Schrödinger's cat to be green is always given by

$$\text{Pr}(\text{left eye is green}) = \text{Tr}(\Pi_{x1}^- |\psi\rangle\langle\psi|) = \text{Tr}\left(\frac{1 + \sigma_{x1}}{2} |\psi\rangle\langle\psi|\right),$$

which is independent of whether the measurement is made from the doorway or the window.

However, the Schrödinger's cat described in quantum mechanics no longer leads to paradoxes like $+1 = -1$. According to the multiplication relation of the Pauli operators, $\sigma_i\sigma_j = i\epsilon_{ijk}\sigma_k$, where ϵ_{ijk} is the Levi-Civita symbol, the product of the measurement results corresponding to the observables in each row and column of the square is not only well-defined but also consistent with the observed results of *Friend A* for the product of each row and column:

$$\left\{ \begin{array}{ll} \text{Row 1: } \text{Tr}[(\sigma_{x1} \otimes \mathbb{I}_2) \cdot (\mathbb{I}_2 \otimes \sigma_{z2}) \cdot (\sigma_{x1} \otimes \sigma_{z2}) |\psi\rangle\langle\psi|] & = +1, \\ \text{Row 2: } \text{Tr}[(\mathbb{I}_2 \otimes \sigma_{x2}) \cdot (\sigma_{z1} \otimes \mathbb{I}_2) \cdot (\sigma_{x2} \otimes \sigma_{z1}) |\psi\rangle\langle\psi|] & = +1, \\ \text{Row 3: } \text{Tr}[(\sigma_{x1} \otimes \sigma_{x2}) \cdot (\sigma_{z1} \otimes \sigma_{z2}) \cdot (\sigma_{y1} \otimes \sigma_{y2}) |\psi\rangle\langle\psi|] & = -1, \\ \text{Column 1: } \text{Tr}[(\sigma_{x1} \otimes \mathbb{I}_2) \cdot (\mathbb{I}_2 \otimes \sigma_{x2}) \cdot (\sigma_{x1} \otimes \sigma_{x2}) |\psi\rangle\langle\psi|] & = +1, \\ \text{Column 2: } \text{Tr}[(\mathbb{I}_2 \otimes \sigma_{z2}) \cdot (\sigma_{z1} \otimes \mathbb{I}_2) \cdot (\sigma_{z1} \otimes \sigma_{z2}) |\psi\rangle\langle\psi|] & = +1, \\ \text{Column 3: } \text{Tr}[(\sigma_{x1} \otimes \sigma_{z2}) \cdot (\sigma_{z1} \otimes \sigma_{x2}) \cdot (\sigma_{y1} \otimes \sigma_{y2}) |\psi\rangle\langle\psi|] & = +1. \end{array} \right. \quad (2.8)$$

Thus, everything is clear now regarding the color-changing behavior of Schrödinger's cat. The difficulties that *Friend A* encountered in establishing a descriptive theory are essentially the challenges faced when describing quantum systems using noncontextual hidden variable theories. In order to describe how the color of Schrödinger's cat relates to the colors of its eyes and paws, it is crucial to consider whether the observation is made from the doorway or the window. In order to provide a hidden variable description for a quantum system, the hidden variables must incorporate information about the context in which measurements are performed.

2.2.3 Experimentally testable noncontextuality inequality

The story of *Friend A* has already demonstrated the incompatibility between quantum mechanics and noncontextual hidden variable theories. However, how can we observe the contextual nature carried by Schrödinger's cat in a realistic test? In 2008, Cabello [13] found a way to demonstrate the contextuality in the Peres–Mermin square via experimentally registered statistics. Consider the product of observable pairs in each row and each column of Figure 2.3. For convenience, I shall denote the three observables in the first row as A, a, α , the second row as B, b, β , and the third row as C, c, γ . Furthermore, consider the following linear combination of observables:

$$\text{PM} := \langle ABC \rangle_\psi + \langle abc \rangle_\psi + \langle \alpha\beta\gamma \rangle_\psi + \langle Aa\alpha \rangle_\psi + \langle Bb\beta \rangle_\psi + (-\langle Cc\gamma \rangle_\psi). \quad (2.9)$$

According to the requirements of noncontextual hidden variable (NCHV) theories, before each experiment, all observables have predetermined results of ± 1 , independent of the choice of the measurement context. By considering the 2^9 possible combinations of values for these nine basic variables, it is found that it is impossible to choose the values of each variable in a way that all six monomials in Equation

(2.9) are positive. In other words, the following inequality:

$$\text{PM} \stackrel{\text{NCHV}}{\leqslant} 4, \quad (2.10)$$

holds for *any* noncontextual hidden variable theory. However, according to quantum theory (Q), it is calculated that $\forall \psi, \text{PM} \stackrel{Q}{\leq} 6$. Therefore, Equation (2.10) is a *state-independent* noncontextual hidden variable inequality, which is violated by any quantum state—even the maximally mixed state $\hat{\rho} = \mathbb{I}_4/4$. Cabello's work has sparked a significant amount of experimental research progress. Within the following year, this theory was tested in ion trap systems by Kirchmair *et al.* [19], and in optical systems by both Amselem *et al.* [20] and Liu *et al.* [40]. To date, the Peres–Mermin square constructed noncontextual hidden variable inequality remains one of the most commonly used tools to demonstrate the contextual nature of quantum theory in different physical systems.

Now, I shall briefly comment on how to experimentally test contextuality using Equation (2.10). We will see that although the form of the noncontextual inequality is simple, there are many possible pitfalls in experiments. I will largely paraphrase the excellent discussion by Budroni *et al.* [41]. Firstly, since each observable in Figure 2.3 can be written as a tensor product of two single-qubit operators, it seems that their measurements can be performed as local measurements on two separate qubits. Consequently, each observable's measurement would yield four outcomes instead of the two degenerate outcomes. However, doing so will put us in exactly the same situation as *Friend A*—simultaneous measurements of noncommuting variables—and thus we would observe the expected results according to the noncontextual hidden variable theory, rather than being able to detect contextuality experimentally.

Secondly, because the observables in each row and column of Figure 2.3 can be measured simultaneously, in principle, for each row and each column, it is always possible to find a set of measurements such that the measurement outcomes of the composite measurement determine the values of each individual observable. Taking the last row as an example, the common eigenstates of C, c, γ happen to be the four Bell states. This is because that:

$$\begin{aligned} (\sigma_x \otimes \sigma_x) |\Phi^+\rangle &= +|\Phi^+\rangle, & (\sigma_y \otimes \sigma_y) |\Phi^+\rangle &= -|\Phi^+\rangle, & (\sigma_z \otimes \sigma_z) |\Phi^+\rangle &= +|\Phi^+\rangle; \\ (\sigma_x \otimes \sigma_x) |\Phi^-\rangle &= -|\Phi^-\rangle, & (\sigma_y \otimes \sigma_y) |\Phi^-\rangle &= +|\Phi^-\rangle, & (\sigma_z \otimes \sigma_z) |\Phi^-\rangle &= +|\Phi^-\rangle; \\ (\sigma_x \otimes \sigma_x) |\Psi^+\rangle &= +|\Psi^+\rangle, & (\sigma_y \otimes \sigma_y) |\Psi^+\rangle &= +|\Psi^+\rangle, & (\sigma_z \otimes \sigma_z) |\Psi^+\rangle &= -|\Psi^+\rangle; \\ (\sigma_x \otimes \sigma_x) |\Psi^-\rangle &= -|\Psi^-\rangle, & (\sigma_y \otimes \sigma_y) |\Psi^-\rangle &= -|\Psi^-\rangle, & (\sigma_z \otimes \sigma_z) |\Psi^-\rangle &= -|\Psi^-\rangle. \end{aligned}$$

Therefore, if we use a Bell state measurement to distinguish between the four Bell states, we can simultaneously determine the values of the three observables, C, c , and γ . However, it can be observed that the measurement results of the Bell state will inevitably imply a product $\langle Cc\gamma \rangle \equiv -1$. Similar discussions for other rows/columns will reveal that it is impossible to assign consistent values to the nine observables within the framework of noncontextual hidden variable theory. The reason for this situation is that when we use Bell state measurements to calculate $\langle Cc\gamma \rangle$, it already assumes the use of Clifford algebra employed in quantum mechanics, thus making the argument circular. On the contrary, in order to accurately reflect the essence of contextual experiments, it is necessary to perform separate measurements on each observable without destroying or decohering the system. This places considerable technical demands on contextual experiments.

Next, I will introduce two more noncontextual hidden variable inequalities. In the same year as Cabello's work, Klyachko *et al.* [39] constructed a very simple model. Let A_0, \dots, A_4 be binary operators with eigenvalues ± 1 such that $[A_i, A_{i+1 \bmod 5}] = 0$. Then the inequality

$$\text{KCBS} := \langle A_0 A_1 \rangle_\psi + \langle A_1 A_2 \rangle_\psi + \langle A_2 A_3 \rangle_\psi + \langle A_3 A_4 \rangle_\psi + \langle A_4 A_0 \rangle_\psi \stackrel{\text{NCHV}}{\geqslant} -3.$$

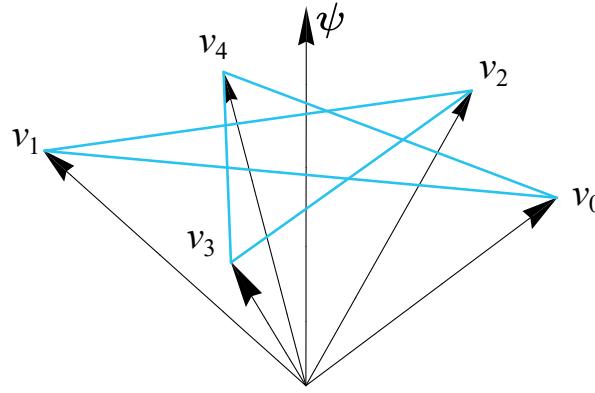


Fig. 2.4 The set of rays realizing the maximal quantum violation of the KCBS inequality. The coordinates are: $|\psi\rangle = \{0, 0, 1\}^\dagger$, $|v_k\rangle = \{\sin \theta \cos k\varphi, \sin \theta \sin k\varphi, \cos \theta\}^\dagger$, where $\theta = \arccos(1/\sqrt[4]{5})$, $\varphi = 2\pi/5$. Figure taken from Reference [41], copyright American Physical Society.

holds for any noncontextuality theory, and is known as the KCBS inequality. However, if we take the five three-dimensional rays $|v_k\rangle$, $k \in 1, \dots, 5$ from Figure 2.4 and the quantum state $|\psi\rangle$, and set $A_k = 2|v_k\rangle\langle v_k| - \mathbb{I}$, we have $\text{KCBS}(|\psi\rangle) = 5 - 4\sqrt{5} \approx -3.94$. Note that the violation of the KCBS inequality is *state-dependent*, unlike the previous case where it was independent of the initial state. It requires specific quantum states to observe contextuality. This property of the KCBS inequality makes it useful for testing whether a quantum state is close to a target quantum state [42].

In 2012, Yu and Oh constructed a proof that can exclude noncontextual hidden variable theories for any quantum state using only 13 projective measurements in a three-dimensional Hilbert space [14]. The configuration of these 13 vectors and their orthogonal relationships are shown in Figure 2.5. With the concept of response functions defined earlier, I will now briefly describe their clever proof. In a noncontextual hidden variable theory, the response function value $v(\hat{x}) = 1$ for a vector \hat{x} indicates that the measurement result of the corresponding projection measurement $\hat{\Pi}_x$ will be 1. Consider the following two cases:

1. If an experiment yields $v(\hat{h}_0) = 1$ and $v(\hat{h}_1) = 1$, according to the orthogonal relationship, we must have $v(\hat{y}_2^\pm) = 0$ and $v(\hat{y}_3^\pm) = 0$. However, due to the completeness of the three-dimensional vectors, we must have $v(\hat{z}_2) = 1$ and $v(\hat{z}_3) = 1$. This means that the hidden variable theory requires the projection operators for two orthogonal vectors to have simultaneous values of 1, which leads to a contradiction.
2. If an experiment yields $v(\hat{h}_1) = 1$ and $v(\hat{h}_2) = 1$, according to the orthogonal relationship, we must have $v(\hat{y}_1^\pm) = 0$ and $v(\hat{y}_2^\pm) = 0$. In this case, $v(\hat{z}_1) = 1$ and $v(\hat{z}_2) = 1$, resulting in a similar contradiction.

Taking into account the symmetry, the four response functions $v(\hat{h}_0)$, $v(\hat{h}_1)$, $v(\hat{h}_2)$, and $v(\hat{h}_3)$ can only have one value of 1. Therefore,

$$\text{YO} := \langle |\hat{h}_0\rangle\langle\hat{h}_0| + |\hat{h}_1\rangle\langle\hat{h}_1| + |\hat{h}_2\rangle\langle\hat{h}_2| + |\hat{h}_3\rangle\langle\hat{h}_3| \rangle_{\psi}^{\text{NCHV}} \leq 1.$$

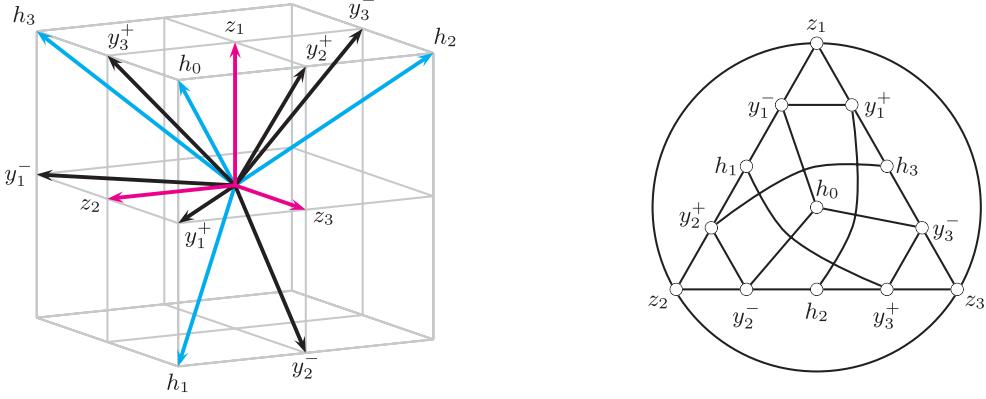


Fig. 2.5 The 13-rays state-independent proof of quantum contextuality. Left: visualization of the rays in the coordinate space: $\hat{y}_1^- = \{0, 1, -1\}^\dagger$, $\hat{y}_2^- = \{-1, 0, 1\}^\dagger$, $\hat{y}_3^- = \{1, -1, 0\}^\dagger$, $\hat{y}_1^+ = \{0, 1, 1\}^\dagger$, $\hat{y}_2^+ = \{1, 0, 1\}^\dagger$, $\hat{y}_3^+ = \{1, 1, 0\}^\dagger$, $\hat{h}_0 = \{1, 1, 1\}^\dagger$, $\hat{h}_1 = \{-1, 1, 1\}^\dagger$, $\hat{h}_2 = \{1, -1, 1\}^\dagger$, $\hat{h}_3 = \{1, 1, -1\}^\dagger$, $\hat{z}_1 = \{1, 0, 0\}^\dagger$, $\hat{z}_2 = \{0, 1, 0\}^\dagger$, $\hat{z}_3 = \{0, 0, 1\}^\dagger$. Right: relation of orthogonality between the rays. The two connected vectors are mutually orthogonal, so their response functions v cannot be simultaneously equal to 1. The three vectors forming a closed triangle span a complete basis, so their response functions v can have one and only one equal to 1. Figure taken from Reference [14], copyright American Physical Society.

However, since $|\hat{h}_0\rangle\langle\hat{h}_0| + |\hat{h}_1\rangle\langle\hat{h}_1| + |\hat{h}_2\rangle\langle\hat{h}_2| + |\hat{h}_3\rangle\langle\hat{h}_3| = 4\mathbb{I}_3/3$, for any quantum state $|\psi\rangle$, $\text{YO} \stackrel{Q}{=} 4/3$. Thus, the noncontextual hidden variable inequality is violated for any quantum state. Furthermore, Cabello *et al.* [29] proved that this construction is minimal: it is impossible to construct a state-independent proof of quantum contextuality using fewer than 13 vectors for projection measurements.

2.2.4 Graph-theoretic approach to contextuality

The last example in the previous section demonstrated how to construct noncontextual hidden variable inequalities using the orthogonal relationships between the vectors corresponding to the projection measurements. This method is highly representative and an important tool in the study of contextuality. Now I will present a result about how to bound the correlations allowed by noncontextual hidden variable theories and quantum theory using a given set of projection measurements.

Definition 2.1 Given a set of measurements $\tilde{\Pi}_k$, $k \in \{1, \dots, n\}$ in an abstract algebraic space, their *graph of exclusivity* $G = G(V, E)$ is an undirected graph defined as follows:

- The vertex set $V(G)$ of the graph is in one-to-one correspondence with the set of abstract measurements $\tilde{\Pi}_k$, and $|V(G)| = n$.
- The edges of the graph connect the vertices corresponding to measurements that are mutually exclusive: $(i, j) \in E(G)$, $\forall \tilde{\Pi}_i \tilde{\Pi}_j = 0$.

For quantum measurements, the measurement operators can be represented by the ray vectors in the Hilbert space: $\tilde{\Pi}_k \rightarrow \hat{\Pi}_k = |v_k\rangle\langle v_k|$. For noncontextual hidden variable theories, the explicit

form of the measurement operators cannot be written, but the response functions of the generalized measurement operators that are mutually exclusive satisfy a very simple relation: $v(\tilde{\Pi}_i, \mu) + v(\tilde{\Pi}_j, \mu) \leq 1$, $\forall \mu, (i, j) \in E(G)$: suppose not, there would be at least one point μ where both measurements respond with 1, contradicting their mutual exclusivity.

Definition 2.2 An *independent set* of an undirected graph $G = G(V, E)$ is a collection of vertices that are pairwise non-adjacent. The independence number of an undirected graph is defined as the cardinality of its largest independent set:

$$\alpha(G) := \max |S|, \quad S \subseteq V(G), \quad \forall i, j \in S, (i, j) \notin E(G). \quad (2.11)$$

Definition 2.3 (Lovász, 1979) An *orthogonal representation* [43] of an undirected graph $G = G(V, E)$ is a set of rays $|v_1\rangle, \dots, |v_n\rangle$ that satisfies the following condition: $\forall (i, j) \in E(G), \langle v_i | v_j \rangle = 0$. The *Lovász number* of an undirected graph is defined as the maximum value of the sum of the squared projections of a vector onto all possible rays in the orthogonal representation of the graph [44]:

$$\vartheta(G) := \max_{v_k, \psi} \sum_{k=1}^n |\langle v_k | \psi \rangle|^2. \quad (2.12)$$

With the definitions of the graph of exclusivity and graph-theoretic constants in hand, we can proceed to describe their relationship to noncontextual hidden variable theory and quantum correlations.

Theorem 2.3 (Cabello, Severini, and Winter, 2014) A set of generalized measurements $\tilde{\Pi}$ that satisfy the graph of exclusivity G has an upper bound on the sum of probabilities in noncontextual hidden variable theories given by the independence number of the graph of exclusivity, $\alpha(G)$. In quantum theory, the upper bound on the sum of probabilities is given by the Lovász number of the graph of exclusivity, $\vartheta(G)$. In particular, since the Lovász number is always greater than or equal to the independence number for any graph, we have:

$$\Pr(\tilde{\Pi}) \stackrel{\text{NCHV}}{\leqslant} \alpha(G) \stackrel{\text{Q}}{\leqslant} \vartheta(G). \quad (2.13)$$

This theorem, known as the CSW method for studying contextuality, is highly general and applicable.

In chapter 4 and chapter 7 of this thesis, the CSW method will be extensively used to construct noncontextual inequalities. However, before applying the method in any experiment, it's important to consider that the measurements implemented in experiments may not perfectly satisfy the relationships given by the graph of exclusivity. In a realistic experiment, there will be a certain probability of obtaining results 1 for two projection measurements that should ideally be orthogonal. How can we include this imperfection within the scope of Equation (2.13)? Cabello *et al.* [29] and Cabello [45] further proved that the following noncontextual hidden variable inequality holds for measurements with imperfect orthogonal relationships:

$$\sum_{k \in V(G)} \Pr(\tilde{\Pi}k) - \sum_{(i, j) \in E(G)} \Pr[v(\tilde{\Pi}_i) = 1, v(\tilde{\Pi}_j) = 1] \stackrel{\text{NCHV}}{\leqslant} \alpha(G). \quad (2.14)$$

By including measurements for the second term, the CSW method can be widely used for experimental investigations of contextuality.

The computation of graph theory constants required in the CSW method can be performed using software such as Mathematica or Python. Solving the independence number is a computationally complex problem, reaching NP-complete complexity, so it can be time-consuming for complex graph structures. On the other hand, the Lovász number can be efficiently calculated using convex optimization techniques.

Below we demonstrate the usage of the method through two specific examples. First, consider the graph of exclusivity, G_{YO} in Figure 2.5 again. The computation results indicate that

$$\alpha(G_{YO}) = \vartheta(G_{YO}) = 5.$$

This computation result can be understood as follows: $\alpha(G_{YO}) = 5$ implies that if we color the vertices of the graph of exclusivity, G_{YO} , ensuring that no two adjacent vertices are colored simultaneously, we can color at most 5 vertices; $\vartheta(G_{YO}) = 5$ indicates that if we find a set of vectors that correspond one-to-one with the vertices in G_{YO} , and the vectors corresponding to adjacent vertices are mutually orthogonal, then the upper bound for the sum of squared inner products between a vector and each vector in this set is 5.

Since $\alpha(G_{YO}) = \vartheta(G_{YO})$, directly using the CSW method cannot yield contextual correlations allowed by quantum theory in the exclusivity structure of G_{YO} . However, we can assign weights to the vertices of the undirected graph, which is equivalent to repeating different vertices multiple times. Xiao *et al.* [46] have proven that when the vertices h_k , for $k \in \{0, 1, 2, 3\}$, have a weight of 2, and the vertices y_k^\pm and z_k , for $k \in \{1, 2, 3\}$, have a weight of 3, the resulting modified graph of exclusivity, G'_{YO} achieves the maximum quantum–classical correlation ratio $\frac{\vartheta}{\alpha} = \frac{35}{33}$. This difference has been observed experimentally as state-independent contextuality.

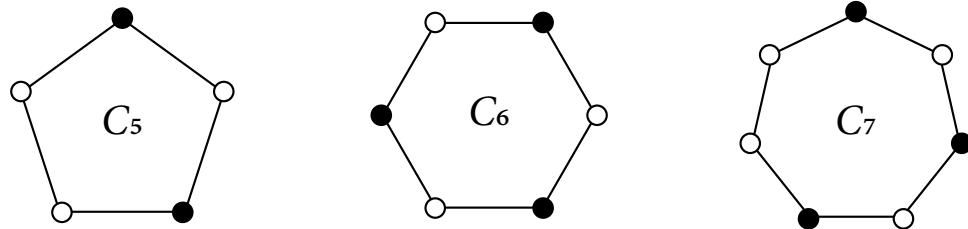


Fig. 2.6 Circular graphs with lengths of 5, 6, and 7. It can be seen from the coloring of vertices that it is impossible to color any more vertices while keeping all adjacent vertices not being colored at the same time. Therefore, $\alpha(C_5) = 2$, $\alpha(C_6) = \alpha(C_7) = 3$.

Next, we consider a more general case: the circular graph C_n with n vertices. In Figure 2.6, the specific forms of the circular graphs for $n = 5, 6, 7$ are shown. The independence number of a circular graph can be easily determined: by selecting one vertex every other vertex and considering the periodic boundary condition, we have $\alpha(C_n) = \lfloor n/2 \rfloor$. On the other hand, the Lovász number of a circular graph is a proven result:

$$\vartheta(C_n) = \begin{cases} \frac{n \cos(\pi/n)}{1 + \cos(\pi/n)}, & n \text{ is odd}, \\ n/2, & n \text{ is even}, \end{cases} \quad (2.15)$$

For the circular graph C_{2m+3} with an odd number of vertices $n = 2m + 3$, where $m \in \mathbb{Z}^+$, it always holds that $\alpha(C_n) < \vartheta(C_n)$. Therefore, the exclusivity structures constructed based on these circular graphs imply a meaningful class of noncontextual hidden variable inequalities.

2.3 Contextuality in quantum information science

Contextuality has always been a hot topic in quantum information science, and many practical applications in the field of quantum information rely on contextuality as a foundation. For example, Spekkens *et al.* identified the advantage of generalized contextuality in oblivious transfer [47] and quantum state discrimination [48]. Gühne *et al.* discovered that violations of noncontextual inequalities can be used to estimate the minimal dimension of a system [49]. Ma *et al.* extended contextuality to randomness generation, thus producing trusted random numbers guaranteed by quantum mechanics. They collaborated with Kim *et al.*'s experimental team working on ion trap systems [50] and with Pan *et al.*'s team on optical systems [51] for experimental implementations.

Due to length constraints and the focus of this thesis, I will only briefly introduce five prominent aspects of contextuality: the relationship between contextuality and the reality of wave functions in quantum mechanics, the contradiction between contextuality and classical causal models for describing quantum measurements, the relationship between contextuality and pre-post-selection quantum paradoxes, the relationship between contextuality and nonlocality, and the applications of contextuality in quantum computation. I have conducted experimental studies on the latter four aspects, and the discussions regarding the last three aspects form the main body of this paper.

2.3.1 Relation with the reality of wavefunction

Quantum mechanics and one of the core contradictions of various hidden variable theories is: what is the essence of the wave function? Is it a correspondence to physical reality or merely a description of people's knowledge of the underlying physical essence? If any hidden variables other than the wave function itself are used, it indicates to some extent that the wave function itself is not a direct correspondence to physical reality, but rather a description of physical reality. However, only when physical reality and the wave function are not in one-to-one correspondence, do hidden variable theories provide more information than quantum mechanics. Therefore, it is necessary to study this correspondence. We use the definition from Harrigan and Spekkens [52] to discuss this correspondence.

Definition 2.4 (Harrigan and Spekkens, 2010) Suppose a hidden variable μ corresponds one-to-one with physical reality. Take two different wave functions $\psi(\mu)$ and $\phi(\mu)$. If $\text{supp}(\psi) \cup \text{supp}(\phi)$ is always a zero-measure set, then the quantum theory is called ψ -ontic. Conversely, if there exists a non-zero measure segment μ such that $\psi(\mu)\phi(\mu) \neq 0$, then the quantum theory is called ψ -epistemic.

If the quantum theory is ψ -epistemic, then the wave function will no longer directly correspond to physical reality as in the Copenhagen interpretation. Leifer and Maroney [53] further defines that if the probabilities of quantum state projection measurements in the Born rule can be fully determined by the overlap of hidden variables, then the quantum theory is maximally ψ -epistemic. The concepts of ψ -ontic, ψ -epistemic, and their correspondence with hidden variable models are illustrated clearly in Figure 2.7.

Shortly after the theoretical proposal by Harrigan and Spekkens, Pusey, Barrett, and Rudolph constructed a concrete example [54] supporting that quantum mechanics is ψ -ontic, meaning that the wave function should correspond one-to-one with physical reality. However, the proof process requires strong assumptions. Subsequently, Leifer and Maroney [53] further analyzed their research and found a more convincing explanation:

Theorem 2.4 (Leifer and Maroney, 2013) *A theory of maximally ψ -epistemic must be noncontextual.*

Therefore, the contextuality of quantum mechanics supports a direct correspondence between the wave function and physical reality. The theory proposed by Leifer and Maroney was further developed by

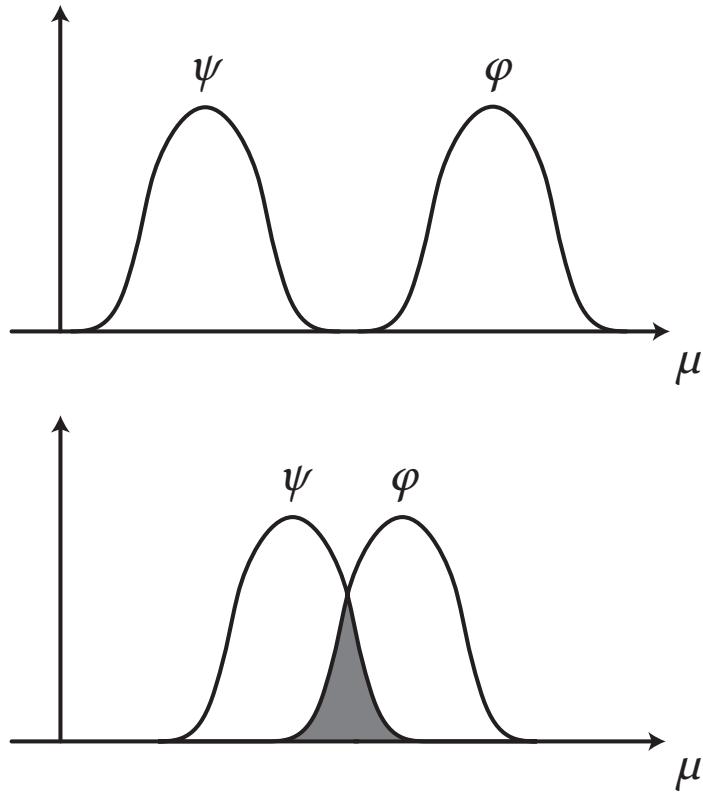


Fig. 2.7 Concepts of ψ -ontic and ψ -epistemic models. The upper and lower figures respectively show the correspondence between the quantum states ψ , φ and the physical reality μ in ψ -ontic and ψ -epistemic models.

Barrett *et al.* [55] and others, who constructed a noncontextual hidden variable inequality based on the distinguishability between quantum states and demonstrated experimental violations in optical systems [56].

2.3.2 Contradiction with classical causal models

In many cases, there is a correlation between random variables, such as the positive correlation often observed between the prices of gold and oil. However, correlation does not imply a direct causal mechanism between the price of gold and oil. In fact, due to factors such as international politics and military situations, investors may simultaneously desire to purchase both gold and oil as a hedge against energy crises and for asset preservation. Therefore, the correlation between the price of gold and oil is actually a result of the common cause of international politics and military situations.

Causal modeling is aimed at describing the conceptual models of causal mechanisms between variables. In Pearl's graphical causal models [57], a directed acyclic graph is used to represent the causal relationships between variables. Variables x are represented by nodes, and the causal relationships be-

tween variables are represented by directed arrows connecting the nodes. We refer to the pair of nodes connected by a directed arrow as a parent node and a child node. In the framework of causal models, the value of each variable depends only on the values of its parent nodes and a local random variable that is not associated with any other nodes. Therefore, the probability distribution of all random variables can be written in a Markovian form as:

$$\Pr(\mathbf{x}) = \prod_k \Pr [x_k | \text{parent}(x_k)] .$$

We can use causal models to describe contextual experiments. Considering the Peres–Mermin square, since we can simultaneously measure three variables in each row or column, there are a total of six possible measurement settings. We represent the selection of measurement settings by a random variable Ω , and the specific observables as X_1, X_2, X_3 with corresponding measurement results A_1, A_2, A_3 . Additionally, we denote the physical state used for the measurement as Λ . If we use a model to describe the experimental process, the choice of measurement settings determines the observables, and the observables, together with the physical state, determine the measurement results. Therefore, the entire experiment can be represented by a directed acyclic graph with the causal structure shown in [Figure 2.8](#).

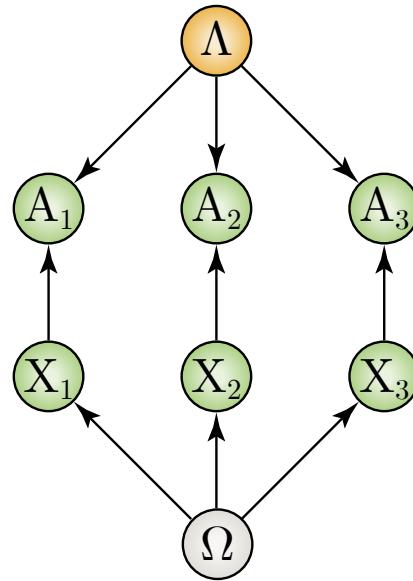


Fig. 2.8 Causal model for the Peres–Mermin square experiment.

By utilizing the Markov condition of causal structures, the probability distribution satisfied by the measurement results can be written in the following product form:

$$\Pr(A_1, A_2, A_3) = \iint \Pr [A_1 | X_1(\Omega), \Lambda] \Pr [A_2 | X_2(\Omega), \Lambda] \Pr [A_3 | X_3(\Omega), \Lambda] d\Omega d\Lambda.$$

However, Cavalcanti [58] found, based on Fine’s theorem [8], that any probability distribution factorizable as a product form can always be written as a joint distribution of all measurement results, where each observable value is a marginal probability of the joint distribution. The joint probability distribution of

all variables can always be described by a noncontextual hidden variable model. Further refinement of the proof reveals [59]:

Theorem 2.5 (Pearl and Cavalcanti, 2021) *Classical causal models cannot describe the contextuality of quantum mechanics. All classical causal models will yield probability distributions consistent with noncontextual hidden variable models for contextuality tests based on sequential measurements.*

The above statement briefly presents the equivalence relation while omitting some constraints related to the faithfulness of causal models. Considering the significance of contextuality from the perspective of causal models allows us to go beyond the framework of hidden variable theories, revealing contradictions between quantum mechanics and a broader class of classical mathematical models, such as Bayesian networks, in describing physical phenomena. The discussion of this contradiction has given rise to the flourishing interdisciplinary field of quantum causality theory [60]. Unfortunately, due to the limited scope of the topic, further exploration cannot be discussed here.

2.3.3 Role in pre-post-selection quantum paradoxes

In 1987, Vaidman *et al.* [61] published a brief yet sensational paper in *Physical Review Letters*. The paper presented a puzzling scenario: a mean king asks a scientist to simultaneously determine the values of σ_x , σ_y , σ_z for a spin-1/2 particle. This seems impossible because the three Pauli operators do not commute pairwise. However, Vaidman and colleagues considered a pre-post-selection scheme. Specifically, they entangled the system to be measured with an ancillary system and sent it to the king. After the king returned the system, they performed a Bell state measurement. By doing so, the scientist could indeed determine the values of the three Pauli operators for the system in the king's possession.

Since 1987, a large number of similar problems have been constructed, focusing on the paradoxical and counterintuitive phenomena in quantum evolution. These scenarios are referred to as pre-post-selected systems, where the initial and final states of the quantum system are well defined. The most representative examples include the violation of the pigeonhole principle, where three spin-1/2 particles can be prepared in states with opposite spins [62], and the phenomenon of quantum entities being disconnected from their properties, as discussed in chapter 6 with the “quantum Cheshire cat” [63].

Pre-post-selection quantum paradoxes are actually consequences of quantum contextuality. Regarding the “mean king” problem, Mermin [64] first pointed out that the scientist’s method in this problem reveals the coordination of an observable in two different contexts, thereby uncovering the results given by noncontextual hidden variable theories. Mermin showed that this method cannot be used to obtain the product of a whole row / column of operators in the Peres–Mermin square because complex multiplication is no longer sufficient to reproduce the operation rules of Clifford algebra. Leifer and Spekkens [65] further developed this result and demonstrated that all pre-post-selection quantum paradoxes can be equivalently attributed to proofs of contextuality, making these paradoxes actual consequences of contextuality. Regarding the “mean king” problem and the connection between the quantum pigeonhole paradox, the quantum Cheshire cat paradox, and contextuality, detailed discussions have been provided by Yu and Oh [66] as well as Waegell and Tollaksen [67].

2.3.4 Interplay with nonlocality

Bell [68] nonlocality is a more widely known peculiarity in quantum mechanics compared to contextuality. It suggests that a hidden variable description of quantum mechanics must be global, even for correlations

between quantum entities separated by spacelike intervals. It cannot be described by two localized hidden variables, but instead requires a global nonlocal hidden variable. In simple terms, quantum nonlocality can be stated as “the explanation of one measurement result cannot be solely explained by local hidden variables; it depends on the choice and outcome of the measurement direction of the other party.” In recent years, the Bell nonlocality of quantum mechanics has received decisive experimental support [69–72].

Quantum nonlocality is a consequence of contextuality. The contradictions reflected by nonlocality can always be attributed to the results of contextuality under spacelike separation. Here, we only consider the relationship between nonlocality and contextuality in a two-particle system of arbitrary dimensions, which is similar for multipartite cases. For a nonlocal system, special relativity guarantees that the measurement operators on one side have no influence on the other side. Therefore, the measurements on one side can always be written in a global form as the direct product of the measurement operators on the other side: $\hat{\Pi}_\alpha^1 = \hat{\Pi}_\alpha \otimes \mathbb{I}_{d2}$, $\hat{\Pi}_\beta^2 = \mathbb{I}_{d1} \otimes \hat{\Pi}_\beta$. Regardless of the specific choice of projection operators, it is always true that $[\hat{\Pi}_\alpha^1, \hat{\Pi}_\beta^2] = 0$ is commutative. In a noncontextual hidden variable theory, the distribution of $\hat{\Pi}_\alpha^1$ should be completely independent of the choice and measurement outcome of $\hat{\Pi}_\beta^2$, thus quantum nonlocality can be considered a special form of contextuality under spacelike separation.

The proof of Bell nonlocality is often based on experimental violations of a local hidden variable inequality, commonly known as Bell inequalities. Any Bell inequality can be interpreted as a noncontextual inequality, and there are known methods to transform noncontextual inequalities into Bell inequalities violated by quantum theory [73, 74]. On the other hand, implementing the exclusivity structure of measurements in Bell inequalities using single systems may lead to stronger quantum correlations [75], and it can even reduce the required system dimension for achieving correlations. In chapter 4, we will provide a detailed introduction to research in this area. In chapter 5, our focus will be on another approach pioneered by Greenberger, Horne, and Zeilinger (GHZ) [76] and Hardy [77], which investigates “inequality-free methods” for studying Bell nonlocality and contextuality.

2.3.5 Application in quantum computation

Quantum computing can be implemented in various ways, such as gate-based or measurement-based, and with discrete or continuous variables. An interesting phenomenon is that in multiple implementations of quantum computing, contextuality plays a crucial role, particularly in its direct connection to enhancing the universality of quantum computation. Here, we briefly introduce three representative theoretical results.

1. In continuous variable quantum computing, the universality of quantum computation can only be achieved with non-Gaussian states or operations, which implies that the Wigner quasi-probability of the system must be negative on at least one non-zero-measure set [78]. Spekkens [79] pointed out that this requirement is equivalent to generalized contextuality.
2. In measurement-based quantum computing, the effect of gate operations is achieved by performing multiple quantum measurements on a graph state [80], utilizing quantum teleportation. Raussendorf [81] proved that if measurement-based quantum computing can achieve a high success rate for computing Boolean functions, the computational process must be entangled and cannot be replicated by noncontextual hidden variable models. Furthermore, Bravyi et al. [82, 83] demonstrated that with the help of contextuality, one can achieve a speed advantage over classical computers for specific problems in shallow-depth quantum circuits that include noise.
3. Magic state distillation [84] allows the extraction of a “magic state” farthest from the eigenstates of Clifford operators using only Clifford operations from multiple copies of noisy resource states,

enabling universal quantum computation [85]. Howard *et al.* [16] proved that it is possible to construct a noncontextual inequality using projection operators onto eigenstates of all Clifford operators through the CSW method. Quantum states that violate this inequality can be used for performing magic state distillation. Therefore, the ability to perform universal quantum computation based on magic state distillation depends on the contextuality resource contained in the quantum states. In [chapter 7](#), we will use a 3-dimensional quantum system as an example to provide a detailed explanation of this method and the experimental research based on it.

From the above discussions, it is evident that contextuality is not only an important concept in the foundations of quantum mechanics but also plays a key role in applications such as quantum computing, enabled by quantum information technology. Thus, research in this direction holds profound significance in both fundamental studies and practical applications. With this, we conclude the introduction to the foundations of quantum contextuality. Next, we will turn to the experimental platform and discuss how to investigate contextuality using the resources and techniques of optical platforms.

Chapter 3

Linear optics quantum information

In the “second quantum revolution”, as the ability to manipulate individual quantum systems has made significant progress, quantum information technology has gradually found widespread applications in various fields. In this stage, especially in the last decade, the physical systems have played a decisive role in determining the specific development directions of quantum information technology. For example, two representative experiments that demonstrate the superiority of quantum computation over classical computation *for specific problems*—random circuit sampling and Gaussian boson sampling [86]—were tailored for superconducting quantum systems and optical quantum systems, respectively. Therefore, the properties and characteristics of the physical platforms are crucial for the problems that can be addressed.

Now I shall introduce the methods of using optical systems for research in the field of quantum information. As mentioned in [chapter 1](#), optical systems have a series of advantages such as high precision, high dimensionality, and multiple degrees of freedom, but they also suffer from the significant drawback of lacking effective interactions between photons. Therefore, in this chapter, we will focus on discussing how to leverage the advantages of optical systems to carry out meaningful research in the foundations of quantum mechanics, quantum simulation, and quantum computing. We will also construct equivalent two-qubit gate operations to overcome the constraints of lacking couplings. I will introduce the experimental system of optical platforms from the aspects of state preparation, evolution, and measurement, along with the necessary theoretical descriptions of the techniques used. As an application case, I will exemplify in [Appendix A](#) how to completely hide quantum information in the correlation between two photons using basic optical operations, so that a single photon no longer carries any information about the previously encoded subspace.

In this chapter, we will frequently use the mathematical language of second quantization to describe the experimental processes since it is the most natural way to express multi-photon quantum interference. However, when it is necessary to establish the correspondence between the experimental and theoretical parts, we will revert to the framework of first quantization introduced in [chapter 2](#). With this transition, the aim is to provide a comprehensive and clear understanding of optical experiments from both theoretical and experimental perspectives.

3.1 Encoding quantum information on photons

3.1.1 Generation of photon pairs

In linear optical systems, the most commonly used method to obtain a single photon or a pair of correlated photons is through the process of parametric down-conversion in nonlinear crystals. The schematic diagram of this process is shown in [Figure 3.1](#). A pump photon passes through a crystal with a non-zero nonlinear coefficient $\chi^{(2)}$, converting into a signal photon and an idler photon. The governing Hamiltonian for the parametric down-conversion process can be expressed as:

$$H = g(a_p a_s^\dagger a_i^\dagger + \text{h.c.})$$

Here, g is the coupling strength, a_m and a_m^\dagger , where $m \in \{p, s, i\}$, are the annihilation and creation operators for the photons in mode m , respectively. They decrease or increase the number of photons in the corresponding mode's field by 1, specifically: $a^\dagger |n\rangle = \sqrt{n+1} |n+1\rangle$ and $a |n\rangle = \sqrt{n} |n-1\rangle$, where $|n\rangle$ represents the Fock state with n photons. The modes are denoted by subscripts p , s , and i , corresponding to the pump, signal, and idler photons, respectively. h.c. denotes the Hermitian conjugate of the remaining terms in the parentheses. Generally, the wavelengths of the signal and idler photons only need to satisfy the energy conservation condition $1/\lambda_p = 1/\lambda_s + 1/\lambda_i$. However, in this paper, we only consider the degenerate output of the parametric down-conversion process, where $\lambda_s = \lambda_i = 2\lambda_p$. After obtaining the correlated photon pairs through parametric down-conversion, they can be further encoded for quantum information experiments¹.

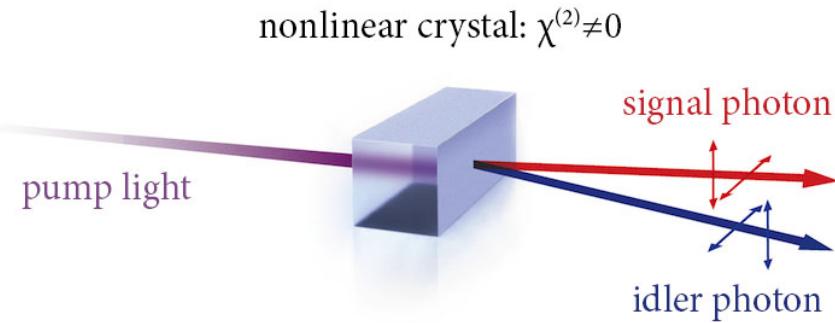


Fig. 3.1 Schematic diagram of the parametric down-conversion process.

Measuring photons refers to detecting the generated photons and studying their statistical properties. In quantum information experiments, we use single-photon avalanche detectors or superconducting nanowire single-photon detectors to convert individual photons into electric signals with very narrow time-domain distributions for detection. In this way, the total number of photons reaching the detector and the arrival times of each photon can be recorded over a period of time. In experiments using photon

¹ Parametric down-conversion is not within the realm of linear optics—as suggested by the name of the nonlinear crystal. However, almost the entire edifice of linear optical experiments is built upon parametric down-conversion/four-wave mixing light sources.

pairs, we perform coincidence measurements, where two detectors detect the two photons, respectively, and only consider the statistical correlations between photons with identical arrival times (or with a fixed delay caused by a fixed optical/circuit length difference). Since only photons with the same arrival time come from the same parametric down-conversion process, they can be described by the theoretically expected quantum states. Measurement results with only one photon arriving will be discarded. By using techniques such as quantum state tomography [87] and quantum process tomography [88], the quantum state of the arriving photons and the relevant information about their evolution can be analyzed.

Traditionally, if only one photon is used for all operations, the other photon is immediately detected by a single-photon counting detector, and the statistical properties of the evolving photon in the optical path are obtained using coincidence counting. This technique is called heralded single-photon. However, the photon evolution obtained by this method is, in most cases indistinguishable from using attenuated laser light for single-photon counting. Its advantage lies only in excluding the influence of dark counts when the detector has no photons. Our single-photon detectors produce around 80 dark counts per second, so I use attenuated laser light or photons generated by solid-state defects for single-photon experiments, which hardly affect the experimental precision.

3.1.2 Photonic degrees of freedom

Physics systems encode quantum information using different degrees of freedom. Degrees of freedom refer to the physical quantities in a physical system that can take on different values. Macroscopically, these can include position and momentum, while microscopically, they can include energy levels and angular momentum of electrons in atoms. Photons have a rich set of degrees of freedom. For example, intrinsic polarization determines the direction of the electric field carried by the photon, orbital angular momentum determines the phase shape of the wave function, and spatial and temporal modes determine the characteristics of photon evolution, among others. I point the readers to the Reference [89] which contains a schematic diagram of some of the commonly used photon degrees of freedom in optical experiments.

Next, we will introduce the methods of using each degree of freedom. Polarization is the most commonly used degree of freedom traditionally. It carries a natural qubit: the horizontal (H) and vertical (V) polarizations of photons can be one-to-one correspondence with the two computational basis states in the two-dimensional Hilbert space: $|0\rangle \leftrightarrow |H\rangle$ and $|1\rangle \leftrightarrow |V\rangle$. It is worth mentioning that for this type of qubit system, there is a particularly visual representation method, which is to use the isomorphism of the SU(2) and SO(3) groups to expand any qubit state $\hat{\rho}_2$ as:

$$\hat{\rho}_2 = \frac{1}{2} (\mathbb{I}_2 + x\sigma_x + y\sigma_y + z\sigma_z),$$

where the coordinates (x, y, z) represent a point on or within a unit sphere. This representation is traditionally called the Bloch sphere [36]. The points corresponding to pure states are on the surface of the sphere, where $|H\rangle$ and $|V\rangle$ correspond to the north and south poles, respectively, while the points corresponding to mixed states are inside the sphere. In the field of quantum information, this method is often used to describe qubit states.

Another reason polarization degree of freedom receives attention is that the polarization entanglement of two photons can be generated from parametric down-conversion processes. For example, by continuously pumping a beam-like type II nonlinear crystal [90] where two output photons are focused at one point, the polarization of the generated photons can be flipped by the first pump and then added with compensation to make the wavefunctions generated by the two pumps indistinguishable, resulting in a

bright entangled source [91]. The schematic diagram of this technique is shown in Figure 3.2, and the generated two-photon state is:

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|HV\rangle - e^{i\phi}|VH\rangle).$$

Additionally, polarization-dependent two-photon interference can be used to generate polarization entanglement from a nonlinear crystal, which will be detailed in the next section.

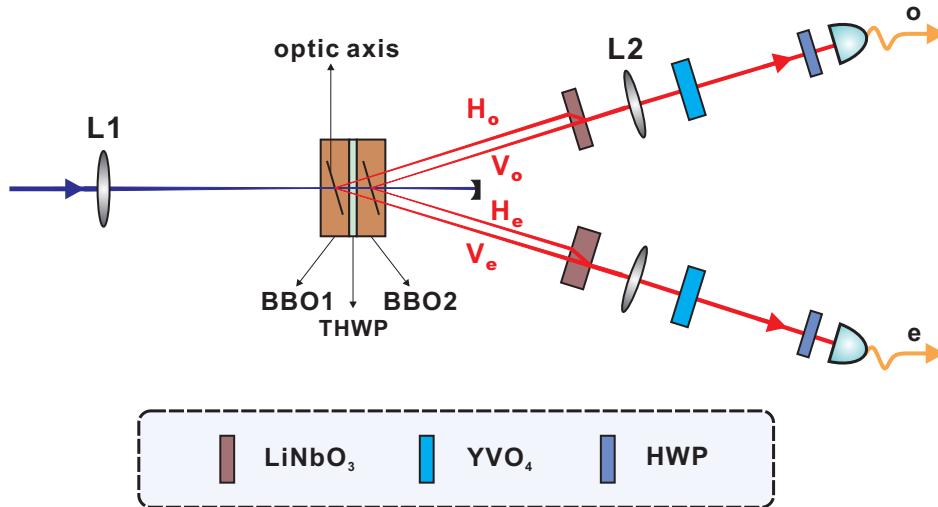


Fig. 3.2 Preparation of polarization qubit entanglement. BBO: Type II nonlinear crystal for parametric down-conversion, which generates a pair of photons with orthogonal polarizations; THWP: True-zero-order half-wave plate used to flip the polarization of the signal and idler photons; LiNbO₃: Lithium niobate crystal used for compensating the spatial separation of parametric photons generated by two pumps; YVO₄: Yttrium orthovanadate crystal used for compensating the temporal separation of parametric photons generated by two pumps. Figure taken from reference [91], copyright American Physical Society.

The path degree of freedom refers to encoding quantum information using the different propagation directions of photons. Each propagation direction \mathbf{k} of a photon can be used to carry a dimension of the Hilbert space, allowing in principle the encoding of quantum information in infinitely high dimensions. Moreover, it is easy to achieve the coupling between path and polarization qubits, enabling convenient conversion between the two. The main challenge of this degree of freedom lies in the phase noise between different paths caused by length fluctuations, which is particularly evident when the separation between two paths is large. Methods to address this issue include active phase locking and passive phase stabilization. Active phase locking refers to the addition of elements such as an optical delay line between paths and the introduction of feedback control using techniques like the Pound-Drever-Hall (PDH) technique [92] to maintain a constant phase difference between the paths. Passive phase stabilization involves constructing path quantum states with nearly identical optical components that are very close in position, such as the beam displacer array or the displaced Sagnac interferometer, effectively suppressing phase fluctuations between them. Recently, emerging technologies such as spatial light modulators [93] and metasurface arrays [94] have further increased the scalability of spatial mode quantum states encoded in the path degree of freedom under passive phase stabilization.

The time-bin quantum state refers to encoding the quantum state in the arrival time of photons at the detector. Using architectures with optical delay lines, in principle, arbitrary high-dimensional quantum states can also be prepared. This system is particularly suitable for studying quantum walks [95] and has the advantage of relatively low noise, making it promising for applications in quantum communication [96]. The main difficulty of this degree of freedom lies in designing methods to measure superposition states between noncomputational basis states. Finally, orbital angular momentum quantum states encode quantum information using the photon's orbital angular momentum, which intuitively refers to the shape (phase) of the wave function (wavefront) [97]. This also constitutes an infinite-dimensional Hilbert space and conveniently enables coupling with polarization qubits [98]. The main challenges of this degree of freedom are how to precisely and efficiently manipulate and detect photons with large orbital angular momentum and how to address the issue of greater photon loss for photons with larger orbital angular momentum during their evolution in optical systems.

3.2 Evolution and operation of photons

We found that different degrees of freedom of photons have distinct characteristics, such as high precision, good scalability, and convenient measurement. A current trend in linear optical experiments is to use multiple degrees of freedom to combine their advantages and expand the number of qubits in the system. Among them, due to the convenient use of wave plates for the operation and measurement of polarization qubits, and the possibility of interconversion with all other degrees of freedom, polarization qubits are often used as a medium to construct multi-degree-of-freedom hybrid systems in the form of “polarization + everything”.

3.2.1 Operation of polarization qubits

Generally, wave plates are used to implement arbitrary unitary operations on single polarization qubits. A wave plate is a birefringent crystal that introduces a certain phase delay for photons polarized along its optical axis. In optical experiments, the wave plates commonly used are half-wave plates and quarter-wave plates, which introduce phase delays of $\lambda/2$ and $\lambda/4$, respectively, for photons with a wavelength λ equal to their operating wavelength. When the optic axis is at an angle θ with the horizontal direction, they induce the following evolution on the polarization state of a photon $\alpha |H\rangle + \beta |V\rangle$:

$$U_{\text{HWP}}(\theta) = \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}, \quad U_{\text{QWP}}(\theta) = \begin{pmatrix} \cos^2 \theta + i \sin^2 \theta & (1-i) \sin \theta \cos \theta \\ (1-i) \sin \theta \cos \theta & \sin^2 \theta + i \cos^2 \theta \end{pmatrix}.$$

Englert *et al.* [99] proved that arbitrary evolutions U_{arb} on the polarization degree of freedom can be decomposed into a sequence of a quarter-wave plate sandwiched between two half-wave plates, which can be implemented by adjusting the angles of the wave plates (up to an overall phase difference):

$$U_{\text{arb}} = \exp(i\phi) U_{\text{QWP}}(\theta_3) U_{\text{HWP}}(\theta_2) U_{\text{QWP}}(\theta_1). \quad (3.1)$$

Since photons do not interact with each other, it is not possible for two photons to perform a controlled two-photon gate operation with 100% efficiency. However, there are ways to probabilistically generate entanglement between two photons. Next, I will introduce a very practical two-photon entangling gate:

the fusion gate. It projects a two-photon polarization state onto the $|\Phi^\pm\rangle$ states using a joint detection [100, 101].

3.2.1.1 Photon fusion gate

As shown in [Figure 3.3](#), a polarisation beam splitter (PBS) is used to achieve two-photon interference in the photon fusion gate. The PBS has two input ports and two output ports, and it changes the direction of photon propagation based on the photon's polarization state: horizontally polarized photons continue along their original path, while vertically polarized photons are reflected by 90 degrees to the other output port. We denote the creation operators for the photons propagating towards the upward observer Alice and the right observer Bob as a_κ^\dagger and b_κ^\dagger , respectively, where the subscript $\kappa \in \{H, V\}$ represents the polarization state of the photon. The evolution on the PBS can be summarized as the following transformations between the photon creation operators:

$$\begin{cases} a_H^\dagger \rightarrow a_H^\dagger, & a_V^\dagger \rightarrow i b_V^\dagger, \\ b_H^\dagger \rightarrow a_H^\dagger, & b_V^\dagger \rightarrow i a_V^\dagger. \end{cases} \quad (3.2)$$

Here, the creation operator for the reflected photon acquires an additional phase of $\pi/2$, which is a result of the Fresnel law. After the photon fusion gate, combined with coincidence detection, Alice and Bob select the case where each of them receives one photon. Consider the four possible combinations of the evolution of the creation operators when two photons enter the fusion gate through two different inputs:

$$a_H^\dagger b_H^\dagger \rightarrow a_H^\dagger b_H^\dagger, \quad a_H^\dagger b_V^\dagger \rightarrow i a_H^\dagger a_V^\dagger, \quad a_V^\dagger b_H^\dagger \rightarrow i b_H^\dagger b_V^\dagger, \quad a_V^\dagger b_V^\dagger \rightarrow -a_V^\dagger b_V^\dagger. \quad (3.3)$$

This indicates that only when the input photons are both in the $|H\rangle$ or $|V\rangle$ polarization state, the output photons generate a coincidence detection for Alice and Bob. Taking into account the additional π phase for the two vertically polarized photons, the evolution caused by the fusion gate on the two-photon state can be written as:

$$\Pi_{\text{fusion}} = |HH\rangle\langle HH| - |VV\rangle\langle VV|. \quad (3.4)$$

Therefore, the fusion gate implements a projection of the two photons onto the entangled basis. Note that the two input photons not only need to have the same wavelength but also need to arrive at the PBS at exactly the same time to ensure that the wave functions from the two photons at the output are indistinguishable, thus enhancing the visibility of quantum interference. Next, we use the two examples in [Figure 3.3](#) to illustrate the applications of the fusion gate:

- 1. Generation of photon pairs entanglement.**—Consider the conversion of two photons generated under degenerate parametric down-conversion. One of the photons is prepared in the maximally superposed polarization state $|D\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$, and the other photon is in an arbitrary polarization state $|\psi_2\rangle = \beta|H\rangle + \exp(i\theta)\gamma|V\rangle$, where $|\beta^2| + |\gamma^2| = 1$. After the action of the fusion gate, the two-photon state evolves as:

$$|D\rangle|\psi_2\rangle \xrightarrow{\text{fusion}} |\psi_{12}\rangle = \Pi_{\text{fusion}}|D\rangle|\psi_2\rangle = \beta|HH\rangle - \exp(i\theta)\gamma|VV\rangle. \quad (3.5)$$

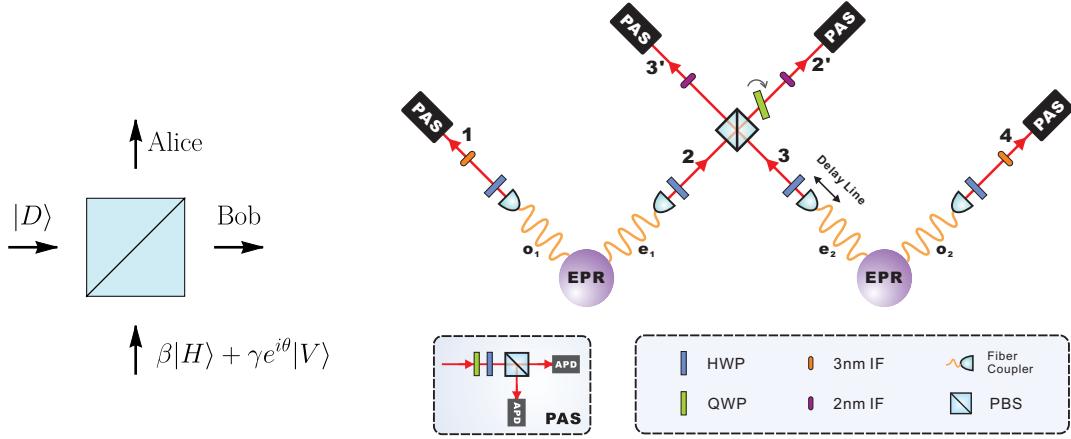


Fig. 3.3 Photon fusion gate. Left: working principle of the photon fusion gate. Right: preparation of a polarization-entangled GHZ state with the fusion gate, taken from Reference [91], copyright American Physical Society.

This results in an entangled state, especially when $|\beta| = |\gamma|$, $|\psi_{12}\rangle$ is a maximally entangled state. Note that the state here has been renormalized to compensate for the effect of the 50% efficiency of the fusion gate.

2. **Preparation of multi-photon entangled states.**—Consider the interaction between two independent pairs of entangled photons (e.g., two pairs of photons produced by the entangled photon source in Figure 3.2). Initially, both photon pairs are in the Bell state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$. The fusion gate acts on the 2nd and 3rd photons, so the overall evolution of the system is $\Pi_{1234} = \mathbb{I}_2 \otimes \Pi_{\text{fusion}} \otimes \mathbb{I}_2$. Therefore, after the fusion gate, the four-photon state evolves as:

$$|\Phi^+\rangle^{\otimes 2} \xrightarrow{\text{fusion}} \Pi_{1234} |\Phi^+\rangle^{\otimes 2} = \frac{1}{\sqrt{2}}(|HHHH\rangle - |VVVV\rangle).$$

The linear combination of these two completely opposite quantum states constitutes a multi-partite entangled state known as the GHZ state, named after Greenberger, Horne, and Zeilinger for their pioneering work in the field of multi-partite quantum entanglement.

The above techniques for multi-photon entangled states seem to be extendable indefinitely. However, in practice, due to the probabilistic nature of parametric down-conversion, the probability of simultaneous down-conversion in two nonlinear crystals is smaller, leading to very low photon counting rates in multi-photon experiments. Therefore, the development of techniques for the two-photon degrees of freedom becomes more prominent. Next, we turn to the introduction of techniques that couple photon polarization with other degrees of freedom.

3.2.2 Coupling of polarization with other degrees of freedom

3.2.2.1 Spatial degrees of freedom

The beam displacer architecture is a mainstream technique for encoding the path quantum state of light and coupling it with polarization qubits. The principle of the beam displacer is shown in Figure 3.4. It is a calcite crystal cut at a specific angle with so its optical axis is inclined at 42.7° to the cut surface, causing photons with polarization parallel to the optical axis to deviate from their propagation direction by 0.100 rad towards the optical axis; while photons with polarization perpendicular to the optical axis remain unaffected. Therefore, by using multiple beam displacers of the same length and controlling the polarization in between, the photon wave function can be distributed among multiple path states $|k\rangle$ as needed. For example, if the optical axis of the beam displacer is placed in the vertical direction, the following evolution occurs after the photon passes through the beam displacer:

$$|H\rangle_{\text{pol}} |k\rangle_{\text{path}} \rightarrow |H\rangle_{\text{pol}} |k\rangle_{\text{path}}, \quad |V\rangle_{\text{pol}} |k\rangle_{\text{path}} \rightarrow |V\rangle_{\text{pol}} |k+1\rangle_{\text{path}}.$$

Here, subscripts indicate the polarization state and the path state of the state vector. Furthermore, by establishing the correspondence between the computational basis and the physical states, such as $|2k\rangle \leftrightarrow |H\rangle_{\text{pol}} |k\rangle_{\text{path}}$, $|2k+1\rangle \leftrightarrow |V\rangle_{\text{pol}} |k\rangle_{\text{path}}$, the polarization degree of freedom and the path degree of freedom can be jointly used to encode a high-dimensional quantum state. Figure 3.4 shows an example of this encoding method. Finally, since the photon's path can take any point on the two-dimensional plane perpendicular to the photon's propagation direction, the encoding of the path degree of freedom can be very flexible.

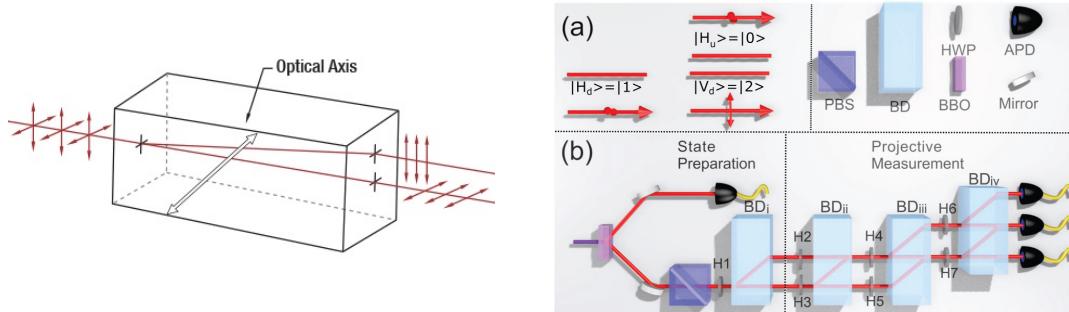


Fig. 3.4 Encoding spatial mode quantum states using beam displacers. *Left:* Principle of the beam displacer. *Right:* Encoding a polarization-path composite three-dimensional quantum state using beam displacers and achieving on-demand evolution. Figure taken from reference [102], copyright American Physical Society.

3.2.2.2 Temporal degrees of freedom

If the beam displacer is replaced with a birefringent crystal with its optical axis perpendicular to the propagation direction, the photon passing through it will exhibit temporal separation between the polarization components parallel and perpendicular to the optical axis. Therefore, the wave function of the temporal degree of freedom can be conveniently manipulated by controlling polarization. The operations here are very similar to the spatial mode case, so they will not be further elaborated. However, the detection

methods for temporal and spatial degrees of freedom are quite different, which will be clearly shown in the next section.

3.2.2.3 Orbital angular momentum

With the advancement of optical field control techniques, it is now possible to manipulate the wavefront of a single photon, allowing it to carry quantum information in the orbital angular momentum (OAM) mode. Several optical elements can be used to manipulate the OAM mode. The vortex phase plate was the earliest element used, where an ℓ -order vortex phase plate carries a phase distribution of $\exp(i\ell\theta)$ in polar coordinates, causing the OAM mode to increase or decrease by ℓ regardless of polarization. A q -plate, which is an ℓ -order waveplate with polarization-dependent OAM coupling, can be used to achieve polarization and OAM coupling evolution. For photons with input polarization states $|L\rangle_{\text{pol}} = (|H\rangle + i|V\rangle)/\sqrt{2}$ and $|R\rangle_{\text{pol}} = (|H\rangle - i|V\rangle)/\sqrt{2}$, the evolution induced by a q -plate is given by:

$$|L\rangle_{\text{pol}} |l\rangle_{\text{OAM}} \rightarrow |R\rangle_{\text{pol}} |l+2\ell\rangle_{\text{OAM}}, \quad |R\rangle_{\text{pol}} |l\rangle_{\text{OAM}} \rightarrow |L\rangle_{\text{pol}} |l-2\ell\rangle_{\text{OAM}}.$$

Finally, using a spatial light modulator, an arbitrary phase $\phi(r, \theta)$ can be directly imposed on a photon with polarization $|H\rangle$, and there are ways to make this phase evolve into any desired OAM wave function [103].

3.2.3 Imaginary-time evolution and non-Hermitian Hamiltonians

In the initial part of [chapter 2](#), the concept of imaginary-time evolution was introduced. Traditionally, linear optical experiments only consider unitary evolution and rarely employ imaginary-time evolution explicitly. Sometimes, this is attributed to the requirement of the quantum no-cloning theorem [104]: in linear optics, the probability amplitude of a quantum state cannot increase, as it would require cloning of an unknown photon state. However, this does not restrict the imaginary-time evolution for Hamiltonians whose spectra are entirely non-negative real numbers. My research group led by Prof. Chuan-Feng Li and Prof. Jin-Shi Xu at the KLQI, USTC has conducted in-depth experimental studies on imaginary-time evolution, particularly in the field of optical quantum simulation [105–107]. The following discussion aims to illustrate the wide application and immense potential of imaginary-time evolution in optical experiments.

Considering a diagonalized Hamiltonian $H = \sum_k \lambda_k |k\rangle \langle k|$, $\lambda \geq 0$, the evolution of the system can be calculated using Equation (2.2) as:

$$U(t) = \sum_k \exp(-\lambda_k t) |k\rangle \langle k|. \quad (3.6)$$

This corresponds to the attenuation of probability amplitudes of each eigenstate according to the magnitude of its eigenvalue. We discuss the applications of this attenuation operation in three cases for different values of t .

$t \rightarrow +\infty$: solving the ground state of the Hamiltonian

In this case, all eigenstates corresponding to non-zero eigenvalues of the Hamiltonian are discarded:

$$\lim_{t \rightarrow +\infty} U(t) = \sum_k \delta_{0,\lambda_k} |k\rangle \langle k|.$$

At this point, the effect of $U(t)$ on a wave function is to project it onto the ground state subspace of the Hamiltonian. This conclusion does not depend on whether the Hamiltonian was initially diagonalized or not. For Hamiltonians containing non-commuting terms, solving the ground state of a quantum system is a non-trivial problem. By employing imaginary-time evolution at large interaction times, projection onto the ground state subspace of the Hamiltonian can be achieved without explicitly diagonalizing the Hamiltonian. Therefore, continuous imaginary-time evolution can be used to implement algorithm cooling [105], quantum annealing, and other quantum algorithms. If the physical quantity of interest does not depend on the adiabaticity of the evolution process, even a few instances of imaginary-time evolution can be used to calculate the physical quantity, such as Berry phase [106, 107].

$t = O(1)$: special gate operations

In this case, the attenuation experienced by each eigenstate of the Hamiltonian can be controlled with specific amplitudes, providing powerful tools for implementing gate operations with special functionalities. Here, we provide a famous example. Consider $H = |H\rangle \langle H|$, $t = \ln(3)/2$. The action of this evolution is $|H\rangle \rightarrow |H\rangle/\sqrt{3}$, $|V\rangle \rightarrow |V\rangle$, which is equivalent to discarding horizontally polarized photons with a probability of 2/3 and always keeping vertically polarized photons. In experiments, this describes an optical beam splitter with transmittances $T_H = 1/3$ for horizontal polarization and $T_V = 1$ for vertical polarization. Next, we present a technique developed by Langford *et al.* [108], Kiesel *et al.* [109], and Okamoto *et al.* [110]: constructing a controlled-phase gate (CZ = controlled-Z gate) using this beam splitter, which is an entangling two-photon gate (note that CZ = $(\mathbb{I}_2 \otimes H) \cdot \text{CNOT} \cdot (\mathbb{I}_2 \otimes H)$), with the evolution matrix given by:

$$\text{CZ} = \hat{\Pi}_z^+ \otimes \mathbb{I}_2 + \hat{\Pi}_z^- \otimes \sigma_z = \text{Diag}(1, 1, 1, -1). \quad (3.7)$$

The experimental setup for implementing the controlled-phase gate is shown in [Figure 3.5](#). Similar to the previous analysis of the fusion gate, we directly calculate the evolution results of the photon creation operators for all possible input states:

$$\begin{aligned} \alpha_H^\dagger \beta_H^\dagger &\xrightarrow{\text{PDBS}_o} \alpha_H^\dagger \beta_H^\dagger & \xrightarrow{\text{PDBS}_{a/b}} \frac{1}{3} \alpha_H^\dagger \beta_H^\dagger, \\ \alpha_H^\dagger \beta_V^\dagger &\xrightarrow{\text{PDBS}_o} \alpha_H^\dagger \left(i\sqrt{\frac{2}{3}} \alpha_V^\dagger + \sqrt{\frac{1}{3}} \beta_V^\dagger \right) & \xrightarrow{\text{PDBS}_{a/b}} \frac{1}{3} \alpha_H^\dagger \beta_V^\dagger \left\{ +i\sqrt{\frac{2}{3}} \alpha_H^\dagger \alpha_V^\dagger, \right. \\ \alpha_V^\dagger \beta_H^\dagger &\xrightarrow{\text{PDBS}_o} \left(i\sqrt{\frac{2}{3}} \beta_V^\dagger + \sqrt{\frac{1}{3}} \alpha_V^\dagger \right) \beta_H^\dagger & \left. \xrightarrow{\text{PDBS}_{a/b}} \frac{1}{3} \alpha_V^\dagger \beta_H^\dagger \left\{ +i\sqrt{\frac{2}{3}} \beta_V^\dagger \beta_H^\dagger, \right. \right. \\ \alpha_V^\dagger \beta_V^\dagger &\xrightarrow{\text{PDBS}_o} \left(i\sqrt{\frac{2}{3}} \beta_V^\dagger + \sqrt{\frac{1}{3}} \alpha_V^\dagger \right) \left(i\sqrt{\frac{2}{3}} \alpha_V^\dagger + \sqrt{\frac{1}{3}} \beta_V^\dagger \right) & \left. \xrightarrow{\text{PDBS}_{a/b}} -\frac{1}{3} \alpha_V^\dagger \beta_V^\dagger. \right. \end{aligned} \quad (3.8)$$

Throughout the equations above, the terms that are discarded during measurement are highlighted in red using curly braces. By comparing with the definition of the CZ gate, it can be seen that the circuit achieves a controlled-phase gate between arbitrary polarization states with a success probability of 1/9. In principle, as long as there is a two-qubit controlled gate and single-qubit arbitrary operations, universal

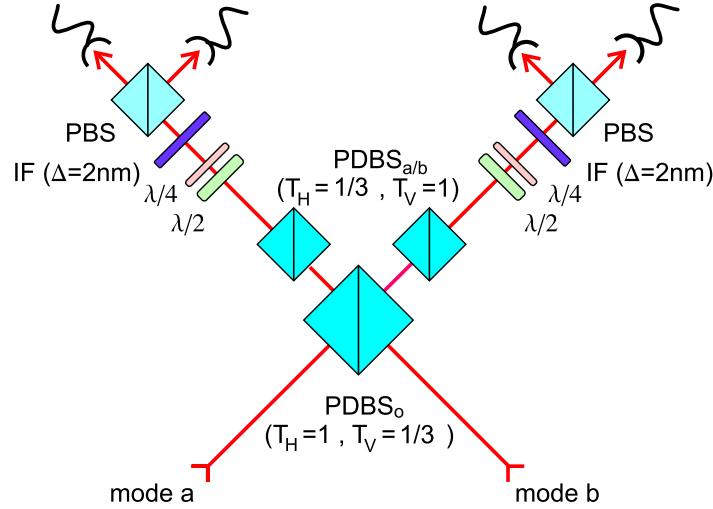


Fig. 3.5 Constructing the controlled-phase gate using imaginary-time evolution. PDBS represents the polarization-dependent beam splitter used to implement the imaginary-time evolution. T_H and T_V correspond to the transmittance of horizontal and vertical polarization photons on the polarization-dependent beam splitter. Figure taken from Reference [109], copyright American Physical Society.

quantum computation can be realized. Therefore, this device demonstrates the potential of imaginary-time evolution in linear optical quantum computation.

$\hbar \rightarrow 0^+$: weak value of quantum states

In this case, the exponential function in the evolution operator can be expanded using a Maclaurin series, resulting in the entire evolution being represented by the identity operator plus a perturbation term:

$$\lim_{\hbar \rightarrow 0^+} U(\hbar) = \mathbb{I} - \sum_k \lambda_k \hat{\Pi}_k \hbar = \mathbb{I} - H \hbar.$$

In this scenario, the action of $U(\hbar)$ leads to a small evolution of the quantum state towards the direction of decreasing the Hamiltonian. Therefore, imaginary-time evolution at small interaction times can be used as an alternative tool for unitary evolution to reconstruct the gradient of a quantum system's evolution.

Next, I calculate the probability of projecting the evolved quantum state onto another target state $|\phi\rangle$, according to the Born rule. It is given by:

$$\begin{aligned} \Pr[\phi|\psi(\hbar)] &= |\langle\phi|U(\hbar)|\psi\rangle|^2 = |\langle\phi|\mathbb{I} - H\hbar|\psi\rangle|^2 = \Pr[\phi|\psi(\hbar=0)] - 2\hbar \operatorname{Re} \langle\psi|\phi\rangle \langle\phi|H|\psi\rangle, \\ \frac{d}{d\hbar} \frac{\Pr[\phi|\psi(\hbar)]}{\Pr[\phi|\psi(0)]} &= 2\operatorname{Re} \langle H \rangle_w := 2\operatorname{Re} \frac{\langle\phi|H|\psi\rangle}{\langle\phi|\psi\rangle}. \end{aligned} \quad (3.9)$$

In Equation (3.9), we have used the weak value $\langle H \rangle_w := \frac{\langle \phi | H | \psi \rangle}{\langle \phi | \psi \rangle}$ defined by Aharonov, Albert, and Vaidman [111]. This quantity represents the strength of the perturbation due to the Hamiltonian operator when the quantum state evolves from ψ to ϕ . It will be the core mathematical tool for observing the counterintuitive effects of quantum measurement before and after the observation. The above calculation shows that the derivative of the projection measurement probability with respect to the evolution time under imaginary-time evolution is proportional to the real part of the Hamiltonian's weak value. I would like to point out that Dressel et al. [112] have already found that the derivative of the projection measurement probability with respect to the evolution time under unitary evolution is proportional to the imaginary part of the Hamiltonian's weak value. However, it was previously unknown how to use perturbation theory to measure the real part of the weak value. Therefore, my work, together with Prof. Xiao-Ye Xu, completes the derivation of this conclusion, enabling perturbation theory to be used fully for measuring the weak value of the Hamiltonian. In chapter 6, we use the perturbation theory introduced here to experimentally observe the “quantum Cheshire cat” paradox.

In addition, although optical systems cannot increase probability amplitudes, the normalization of the counting rates in the final basis, which is an orthonormal basis, does not affect the amplification effect of imaginary-time evolution on unattenuated eigenstates. Thus, this method can be used equivalently to investigate systems described by non-Hermitian Hamiltonians, which have received significant attention in recent years, such as the physics of \mathcal{PT} -symmetric evolutions [113–115].

3.3 Measurement of optical quantum states

Now I explain how to explore how to extract the quantum information encoded in photons and convert it into classical information. As mentioned in the previous section, due to the convenient coupling between polarization qubits and all other qubits, polarization qubits are often used as a medium for measuring quantum information in other degrees of freedom. Therefore, we will start by discussing measurements of photon polarization modes.

3.3.1 Polarization mode measurement

The measurement of polarization qubits is essentially the reverse process of evolution. Since a polarizing beam splitter can distinguish between horizontally and vertically polarized photons, the target states after evolution are typically $|H\rangle$ and $|V\rangle$. For a photon, if we want to perform a projection measurement on $|\psi\rangle_{\text{pol}} = \tilde{U} |H\rangle$, we can place a waveplate group in front of the polarizing beam splitter to implement \tilde{U}^{-1} . This way, the photon will ultimately be transformed into $|H\rangle$ and transmit through the polarizing beam splitter. By using fiber collection and single-photon detection, we can perform photon counting, and then normalize the counting rate or use quantum state tomography to obtain the measurement probability or reconstruct the initial photon wavefunction. Additionally, since there is no difference between detecting the photon to $|H\rangle$ and $\exp(i\phi) |H\rangle$ during the measurement, the measurement of polarization can be further simplified compared to arbitrary unitary evolution. It only requires a quarter-wave plate and a half-wave plate:

$$\tilde{U}_{\text{arb}}^{-1} = \exp(i\phi) U_{\text{HWP}}(\theta_2) U_{\text{QWP}}(\theta_1).$$

In [Table 3.1](#), we list the angles of the waveplate optical axes relative to the horizontal direction for projecting several common measurement bases onto $|H\rangle$.

Table 3.1 Waveplate angle settings for measuring common polarization states. The top six states from top to bottom are eigenstates of σ_z , σ_x , σ_y with eigenvalues ± 1 . The bottom two states are states on the equator and meridian of the Bloch sphere, where ϑ and φ represent the polar angle and azimuthal angle on the sphere, respectively.

Target state $ \psi\rangle$	QWP angle $\theta_1/^\circ$	HWP angle $\theta_2/^\circ$
$ H\rangle$	0	0
$ V\rangle$	0	45
$ D\rangle = (H\rangle + V\rangle)/\sqrt{2}$	45	22.5
$ A\rangle = (H\rangle - V\rangle)/\sqrt{2}$	45	-22.5
$ L\rangle = (H\rangle + i V\rangle)/\sqrt{2}$	0	-22.5
$ R\rangle = (H\rangle - i V\rangle)/\sqrt{2}$	0	22.5
$ \vartheta\rangle = \cos(\vartheta/2) H\rangle + \sin(\vartheta/2) V\rangle$	$\vartheta/2$	$\vartheta/4$
$ \varphi\rangle = H\rangle + \exp(i\varphi) V\rangle$	$22.5 - \varphi/4$	45

3.3.2 Spatial mode measurement

For quantum states encoded in photon spatial modes, if it is only necessary to distinguish the intensity of each spatial mode, one can collect the photons of each mode separately by moving single-mode fiber couplings or directly detect the corresponding photon numbers on each spatial mode using an intensified charge-coupled device (ICCD) camera. In this case, using the other photon in the photon pair for prediction can greatly reduce noise. My research group has used this technique to achieve Bohmian trajectory guidance, measurement-disturbance relationship in double-slit experiments, and weak value observations without post-selection, among other achievements. They have further developed this technique to achieve phase resolution between different spatial modes, enabling wavefront sensing based on photon momentum measurement.

If it is necessary to measure the superposition state of different spatial modes, it often requires using a beam displacer to recollect the separated spatial modes and gradually convert them to polarization measurements. Here, we discuss a highly representative scenario: the joint measurement of a path qubit and a polarization qubit, as illustrated in [Figure 3.6](#). In principle, the measurement of high-dimensional path states only requires stepwise conversions. Initially, the quantum information is encoded in the polarization qubit $|\cdot\rangle_{\text{pol}}$ and a 2-dimensional subspace $\{|k-1\rangle_{\text{path}}, |k\rangle_{\text{path}}\}$ of the path degree of freedom. The specific measurement procedure is as follows:

- First, the polarization qubit is measured, and then the part of $|k-1\rangle_{\text{path}}$ that should be retained after measurement is transferred to the state $|k\rangle_{\text{path}}$, while the part to be discarded remains in the state $|k-1\rangle_{\text{path}}$. The part of $|k\rangle_{\text{path}}$ that should be retained after measurement remains in the state $|k\rangle_{\text{path}}$, while the part to be discarded is transferred to the state $|k+1\rangle_{\text{path}}$. First, the polarization qubit is measured, and then the part of $|k-1\rangle_{\text{path}}$ that should be retained after measurement is transferred to the state $|k\rangle_{\text{path}}$, while the part to be discarded remains in the state $|k-1\rangle_{\text{path}}$. The part of $|k\rangle_{\text{path}}$ that should be retained after measurement remains in the state $|k\rangle_{\text{path}}$, while the part to be discarded is transferred to the state $|k+1\rangle_{\text{path}}$.
- Therefore, the polarization measurement on the $|k-1\rangle_{\text{path}}$ path should rotate the measured basis to $|V\rangle_{\text{pol}}$, which can be achieved by increasing the angle of each half-wave plate by 45° as shown in

Table 3.1. The polarization measurement on the $|k\rangle_{\text{path}}$ path should rotate the measured basis to $|H\rangle_{\text{pol}}$, which can be achieved directly using the operations given in [Table 3.1](#).

- After completing the polarization measurement, the previous path state $|k\rangle_{\text{path}}$ has been converted to the polarization state $|H\rangle_{\text{pol}}$, and $|k-1\rangle_{\text{path}}$ has been converted to the polarization state $|V\rangle_{\text{pol}}$, differing by at most a phase. Therefore, by performing another polarization measurement, we obtain the information of the previous path qubit.

3.3.3 Temporal mode discrimination

Due to the correspondence between time and spatial modes discussed in the previous section, in principle, the measurement of superposition states in all time modes can also be converted to polarization measurements. The method is very similar to spatial mode measurement, except that delay-type birefringent crystals are used. When we are interested in measuring the distribution of time modes in the computational basis, Hong, Ou, and Mandel [116] developed a highly representative technique known as the Hong–Ou–Mandel interference. The method involves interfering two indistinguishable photons on a polarization-independent 50:50 beam splitter. When the two photons coincide in time, they always end up on the same side, resulting in a noticeable drop in the coincidence count rate. This method can convert photon time differences on the order of ps (10^{-12} s) to path length differences on the order of $100 \mu\text{m}$, greatly facilitating time mode detection.

We will continue to analyze the Hong–Ou–Mandel interference using the mathematical language of second quantization. In the absence of polarization dependence, the transformation between the creation operators when two indistinguishable photons arrive at the beam splitter at the same time is given by:

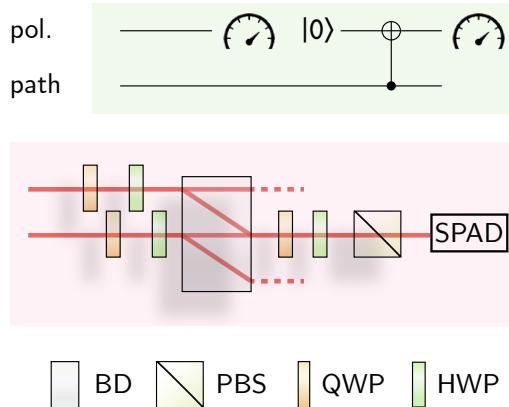


Fig. 3.6 Joint measurement of path and polarization qubits. Top: measurement setup. First, the polarization degree of freedom is measured, and then the information of the path degree of freedom is transferred to the reinitialized polarization degree of freedom using a CNOT operation. Finally, the polarization degree of freedom is measured again. Bottom: the experimental setup. BD: Beam displacer, PBS: Polarizing beam splitter, QWP: Quarter-wave plate, HWP: Half-wave plate, SPAD: Single-photon avalanche diode.

$$\begin{cases} a^\dagger \rightarrow (a^\dagger + ib^\dagger)/\sqrt{2}, \\ b^\dagger \rightarrow (ia^\dagger + b^\dagger)/\sqrt{2}. \end{cases} \quad (3.10)$$

After the beam splitter, we still consider the case where Alice and Bob each receive one photon. The evolution of the creation operators is:

$$a^\dagger b^\dagger \rightarrow \frac{1}{2}(a^\dagger + ib^\dagger)(ia^\dagger + b^\dagger) = \frac{i}{2}(a^\dagger a^\dagger + b^\dagger b^\dagger).$$

This explains why the two photons always end up on the same side due to the additional phase introduced by the reflection according to the Fresnel law, and the wave functions interfere destructively on the other side. Typically, in experiments, photons with a wavelength of approximately $\lambda = 800$ nm are used, and an interference filter is used to narrow the photon's spectrum to $\Delta\lambda = 3$ nm, resulting in a spatial broadening of the photon on the order of $\lambda^2/\Delta\lambda \approx 200$ μm . Hong–Ou–Mandel interference can be used to distinguish time modes when the temporal separation between the two photons exceeds this value.

3.3.4 Orbital angular momentum mode detection

As mentioned earlier, orbital angular momentum can, in principle, be detected by converting it to polarization modes. However, there is another technique that can greatly simplify the detection of this degree of freedom. Taking advantage of the fact that single-mode fibers can only receive photons with orbital angular momentum $|0\rangle_{\text{OAM}}$, any target qubit can be projected onto $|0\rangle_{\text{OAM}}$ in a single step using a spatial light modulator. When performing the projection operation, a phase pattern with a total phase sum of 2π corresponding to the preparation operation is used to implement the projection measurement of the quantum state associated with the preparation operation [117]. By performing phase corrections and compensating for higher-order orbital angular momentum collection efficiency, the measurement accuracy can be further improved.

To this point, the basic building blocks of linear optics experiments have been presented. In [Appendix A](#), an experimental case using linear optics systems will be provided, describing representative experimental details. In the other chapters of this thesis, the emphasis will be on the combination of experiment and theory. In fact, linear optics experiments can vary greatly, and each work has different purposes, so the specific processes can vary widely. It is impossible to fully apply the methods presented in this chapter to all situations. However, the arsenal shown in this chapter can be applied in a wide range of situations, so a thorough understanding of the basics of linear optics experiments is very helpful in accurately translating new theoretical results into experimental research plans. Starting from the next chapter, I will focus on how to use linear optics experiments to study the contextuality in quantum mechanics introduced in [chapter 2](#), hoping to fully demonstrate the flexibility and power of linear optics for fundamental research in quantum mechanics through four experiments.

Part II

Interplay between contextuality and nonlocality

Chapter 4

Stronger contextuality beyond nonlocality

We now proceed to the experimental study of contextuality in the linear optical platform, starting with two works aimed at investigating the correlation between contextuality and nonlocality. As mentioned in [chapter 2](#), nonlocality can be seen as a manifestation of contextuality under spacelike separation. However, spacelike separation restricts quantum measurements to take the form of operator products, thereby limiting the possible forms and strengths of quantum correlations.

Closely related to this, a motivation for studying the relationship between contextuality and nonlocality¹ is that, given the existing knowledge of humanity, determining whether a specific behavior—the probability distribution of measurements on a quantum system—is allowed by Bell nonlocality is *intractable*—even using the best-known method proposed by Navascués *et al.* [118], it is still impossible to distinguish some behaviors that are actually superquantum [119]. In contrast, as shown in [Theorem 2.3](#), determining whether a specific behavior is allowed by quantum contextuality is computationally tractable—using semidefinite programming and other methods, solving the Lovász number of a given graph can be done in polynomial time. Therefore, the intractability of the set of nonlocal correlations lies in the restrictions imposed by spacelike separation on measurement operators. To demonstrate this difference, a natural approach is to start with graphs of exclusivity and find experimental scenarios corresponding to the same graph of exclusivity that exhibit both Bell nonlocality and contextuality, thus characterizing the differences in behavior in these scenarios.

The two works described in this chapter both describe the differences in nonlocality and contextuality behaviors. These differences can be manifested in two ways:

1. For a given exclusive graph, the maximum violation of noncontextual hidden variable inequalities allowed by Bell nonlocality is smaller than the maximum violation allowed by contextuality. We refer to this as “contextuality beyond nonlocality.”
2. For a given exclusive graph, the system dimension required for achieving the maximum violation of noncontextual hidden variable inequalities allowed by Bell nonlocality is lower than the system dimension required for contextuality. We refer to this as “contextuality concentration.”

In Sections [4.3](#) and [4.4](#), we study examples of these two differences using a combination of theory and experiments. These two works demonstrate how techniques from graph theory, linear algebra, and linear optical interferometers can be tightly integrated in the study of fundamental aspects of quantum mechanics, providing new insights into the understanding of nonclassical behavior.

¹ This insightful idea is entirely credited to Adán Cabello; I am simply restating it here.

4.1 Nonlocality, contextuality, and the graph of exclusivity

In the theoretical part, the author introduces the concept of contextuality through a modified version of Schrödinger's cat story, where the observed color change of the cat depends on the method of observation used by friend A. However, in many other cases, there is no such “black or white” pattern of change, and it is necessary to transform the predictions of noncontextual hidden variable models into statistical regularities using inequality methods. The famous theorem by Fine [8] provides the way for this transformation: noncontextual hidden variable theories (which manifest as local hidden variable (LHV) theories under spacelike separation) allow *all* variables in the experiment to have a predetermined probability distribution, even if these variables do not commute! Therefore, by constructing probability combinations of these variables, one can construct an inequality in both the framework of noncontextual hidden variable models and local hidden variable models by searching for linear combinations of variables that lead to extreme values of this linear combination. For noncontextual hidden variable models, this is equivalent to the noncontextual hidden variable inequality mentioned earlier, while for local hidden variable models, such combinations yield a Bell inequality.

I shall first introduce one of the most famous local hidden variable inequalities: the Clauser–Horne–Shimony–Holt (CHSH) inequality [120]. It is expressed in terms of the following combination of expectation values:

$$\text{CHSH} := \langle a_0 b_0 \rangle_{\psi} + \langle a_0 b_1 \rangle_{\psi} + \langle a_1 b_0 \rangle_{\psi} - \langle a_1 b_1 \rangle_{\psi} \stackrel{\text{LHV}}{\leq} 2.$$

Here, $a_{\kappa}, b_{\kappa} = \pm 1$ are two physical quantities that observers at spacelike separation can measure. In the CHSH setup, $\kappa \in \{0, 1\}$ can take two values. Therefore, we say that the CHSH inequality is a Bell inequality with two observers, two measurement settings, and two possible outcomes, and in the research of quantum information, it is often denoted as a (2,2,2)-Bell inequality. In local hidden variable theory, its upper bound is 2, which can be obtained by considering the 16 possible combinations of a_{κ} and b_{κ} .

In quantum mechanics, the CHSH inequality can be violated. Take $|\psi\rangle = |\Phi^+\rangle$ and four measurement operators as the Pauli operators and their linear combinations:

$$a_0 = \sigma_z, \quad a_1 = \sigma_x, \quad b_0 = (\sigma_z + \sigma_x)/\sqrt{2}, \quad b_1 = (\sigma_x - \sigma_z)/\sqrt{2}. \quad (4.1)$$

In this case, we obtain $\text{CHSH} = 2\sqrt{2}$, which violates the predictions of local hidden variable theory. We define the quantum–classical ratio as the maximum value allowed by quantum mechanics divided by the right-hand side of the inequality. The quantum–classical ratio for the CHSH inequality is as high as $\sqrt{2}$, making it highly suitable for experimental observations. To this day, the degree of violation of this inequality in linear optical experiments is considered one of the “gold standards” for assessing the quality of entangled light sources.

Now I consider the CHSH inequality from the perspective of graphs of exclusivity. The measurements used in the graph of exclusivity are projection measurements, so the CHSH inequality needs to be transformed into a form based on combinations of elementary event probabilities. Note the correspondence between binary operators and projection operators: $\hat{\Pi}_{\kappa}^{\pm} = (\mathbb{I}_2 \pm \sigma_{\kappa})/2$, $\kappa \in \{x, y, z\}$, which gives:

$$\begin{aligned} \text{CHSH}^{\text{CSW}} := & \Pr(a_0 = +1, b_0 = +1) + \Pr(a_0 = -1, b_1 = -1) \\ & + \Pr(a_1 = -1, b_1 = +1) + \Pr(a_1 = +1, b_0 = +1) \\ & + \Pr(a_0 = -1, b_0 = -1) + \Pr(a_0 = +1, b_1 = +1) \\ & + \Pr(a_1 = +1, b_1 = -1) + \Pr(a_1 = -1, b_0 = -1) \stackrel{\text{LHV}}{\leq} 3. \end{aligned} \quad (4.2)$$

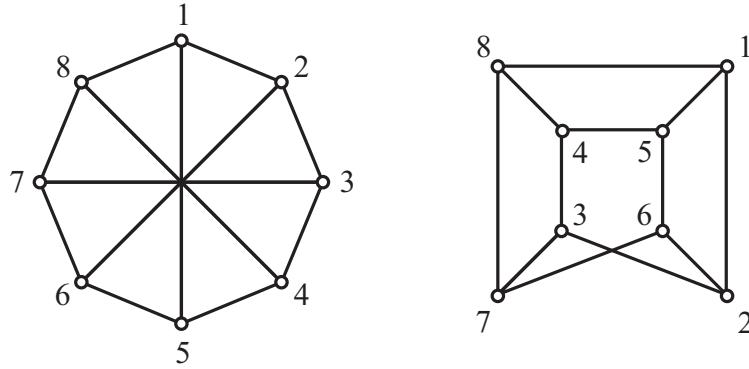


Fig. 4.1 Graph of exclusivity of a subset of events in the CHSH experiment. The left and right figures show two equivalent representations. The eight vertices correspond to the eight measurement events in Equation (4.2).

Here, we only need to consider the event combinations that contribute positively to the original CHSH inequality. Using the same quantum state and measurements as before, we find that the maximum value of CHSH^{CSW} allowed by nonlocal correlations is $2 + \sqrt{2}$. It is easy to see that the events that are mutually exclusive occur in the event combinations of adjacent terms and also in the four pairs of combinations that are 3 terms apart (the furthest apart). Therefore, a graph of exclusivity of these 8 events in the CHSH experiment can be depicted as shown in Figure 4.1: vertices 1 to 8 correspond to the 8 terms in CHSH^{CSW} , and this graph of exclusivity is the Wagner graph defined in graph theory. Interestingly, this graph is also equivalent to a Möbius ladder M_8 . From graph theory, it is known that

$$\vartheta(M_8) = 2 + \sqrt{2}.$$

Therefore, according to Theorem 2.3, if we only consider the restoration of the exclusivity relation in Figure 4.1 using 8 measurements without considering spacelike separation, the maximum value of CHSH^{CSW} that contextuality can provide is also $2 + \sqrt{2}$, the same as nonlocality. In the CHSH experiment, there is no contextuality beyond nonlocal correlations.

4.1.1 Case study: quantum correlations in the pentagon

Next, I consider a highly representative case of contextual correlations that transcend nonlocality. We consider the simplest graph that satisfies $\vartheta > \alpha$: the pentagon graph C_5 . It is depicted in Figure 4.2. The noncontextual hidden variable inequality corresponding to this graph is the probability form derived from the previously introduced KCBS inequality:

$$\text{KCBS}^{\text{QC}} := \sum_{k=0}^4 \Pr(A|k) = \sum_{k=0}^4 \Pr(A_k) \stackrel{\text{NCHV}}{\leqslant} 2. \quad (4.3)$$

Here, we use an alternative notation $\Pr(A|k)$, $A \in \{0, 1\}$, $k \in \{0, \dots, 4\}$ to represent the probability of obtaining result A when the measurement operator is chosen as k . Additionally, we have $\vartheta(C_5) = \sqrt{5}$.

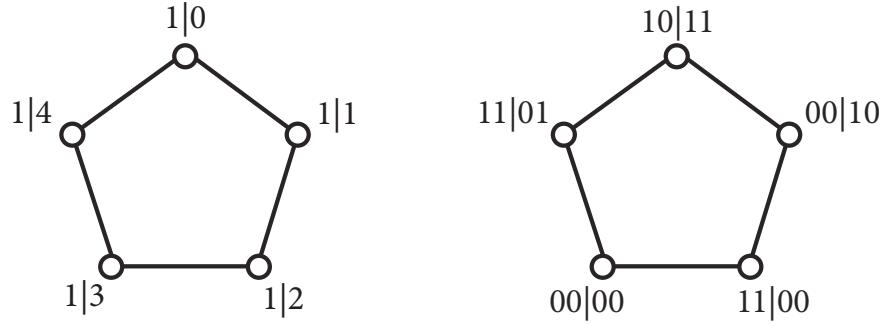


Fig. 4.2 The graph of exclusivity of a subset of events in the KCBS experiment. Left: events that need to be recorded when conducting the experiment using individual contextuality correlation. Right: the events that need to be recorded when conducting the experiment using two-body nonlocality correlations, taken from Reference [75], copyright American Physical Society.

Next, I shall discuss how to achieving the graph of exclusivity in Figure 4.2 from the perspective of nonlocality. In this case, each observer selects at least two projection measurement operators. We use $\text{Pr}(ab|xy)$ to denote the probability that the measurement operators chosen by the two observers, Alice and Bob, are x and y respectively, and the measurement results are a and b . In this scenario, Sadiq *et al.* [75] discovered a new Bell inequality:

$$\text{KCBS}^{\text{NL}} := \text{Pr}(00|00) + \text{Pr}(11|01) + \text{Pr}(10|11) + \text{Pr}(00|10) + \text{Pr}(11|00) \stackrel{\text{LHV}}{\leq} 2.$$

The maximum value allowed by quantum mechanics is $\text{KCBS}^{\text{NL}} \stackrel{\text{Q}}{\leq} (16 + \sqrt{13})/9$.² Therefore, there is a difference between the correlations provided by quantum contextuality (QC) and nonlocality (NL): $\Delta_{\text{NL}}^{\text{QC}}(\text{KCBS}) := \max \text{KCBS}^{\text{QC}} - \max \text{KCBS}^{\text{NL}} \approx 5.77 \times 10^{-2}$, indicating that the quantum correlations corresponding to the graph of exclusivity C_5 exhibit contextual aspects that transcend nonlocality. Additionally, since $\text{KCBS}^{\text{QC}} = \sqrt{5}$ can be achieved in a three-dimensional Hilbert space, while any Hilbert space dimension corresponding to nonlocality is at least 4, the quantum correlations in the graph of exclusivity C_5 also exhibit contextuality concentration. Overall, this scenario serves as an excellent starting point for studying the relationship between contextuality and nonlocality. We can also use similar methods to investigate other representative graphs of exclusivity, such as the famous nonlocality paradox by Hardy [77]. Mermin [121] has analyzed the violation of inequalities when this scenario is transformed into a contextuality experiment, where contextual aspects beyond nonlocality also appear with $\Delta_{\text{NL}}^{\text{QC}}(\text{Hardy}) \approx 2.90 \times 10^{-2}$. However, in this case, there is no contextuality concentration phenomenon.

4.2 Contextuality from measurement-repreparation experiments

The above examples were presented to illustrate the phenomena of contextual surpassing nonlocality and contextual concentration. It can be seen that research in this direction heavily relies on the language of

² Neither I nor Sadiq *et al.* [75] can provide analytical results for the measurement directions and quantum states. However, my numerical results agree with the analytical results of the maximum violation value of quantum mechanics up to 13 digits.

graphs of exclusivity to detect noncontextual hidden variable inequalities of the CSW type. Our starting point is a graph of exclusivity and its corresponding experimentally testable inequality, which is given by Equation (2.14). Using the notation of conditional probabilities introduced in the previous section, this equation can be written more concisely as:

$$\sum_{k \in V(G)} \Pr(1|k) - \sum_{(i,j) \in E(G)} \Pr(1, 1|i, j) \stackrel{\text{NCHV}}{\leq} \alpha(G). \quad (4.4)$$

However, there are still some difficulties in using this inequality in experiments. Specifically:

1. Equation (4.4) requires experimenters to have the ability to perform consecutive measurements. The quantum state obtained after the first measurement needs to be used for the second measurement in order to compensate for imperfect orthogonality between the two exclusive operators. Therefore, proper contextuality tests have heavily relied on sequential measurements since their development, which leads to a large number of required measurement devices and introduces additional couplings between degrees of freedom, thereby limiting experimental precision.
2. The assumption behind Equation (4.4) that allows compensating for imperfect orthogonality using joint probabilities of two consecutive measurements is that the choice of each measurement basis does not affect the outcome of the other measurement. Similar to the no-signaling condition in Bell nonlocality scenarios, this requirement is often referred to as the no-disturbance condition in sequential measurements. Specifically, the influence of the choice of measurement basis on the outcome of the other measurement can be expressed using the notion of absolute causality as the following signaling factor:

$$\left\{ \begin{array}{l} \varepsilon_{xy} := \left| \sum_a \Pr(a, b|x, y) - \Pr(b|y) \right|, \\ \varepsilon'_{xy} := \left| \sum_b \Pr(a, b|x, y) - \Pr(a|x) \right|. \end{array} \right. \quad (4.5)$$

In principle, contextuality tests require ensuring that for every pair of exclusive measurements (i, j) , $\varepsilon_{ij} = \varepsilon'_{ij} = 0$. However, due to the limitations of experimental precision, even in Bell nonlocality experiments conducted at spacelike separation, this value can significantly deviate from zero. Therefore, it is necessary to explicitly test Equation (4.5) in order to ensure testing Equation (4.4) under the no-disturbance condition.

Compared to early contextuality tests, the two-measurement scheme has reduced the requirements to some extent. Otherwise, the test for the no-signaling condition would be more stringent and challenging to analyze. However, even with the use of the two-measurement scheme, the fundamental reliance on sequential measurements cannot be fundamentally resolved. Particularly in high-dimensional systems, the resource consumption introduced by any sequential measurement is significant. Fortunately, Cabello [45] demonstrated that by accepting the measurement-repreparation experiment defined by Lüders rule, the sequential measurement experiments can be completely replaced. The principle of this method is shown in Figure 4.3. Except for the method of measuring the second term on the right side of Equation (4.4), which is different from standard sequential measurements, all other aspects follow the CSW framework.

Specifically, in this method, the first measurement result and probability are obtained using projection measurements, followed by the preparation of the quantum state based on Lüders rule. Then, the second projection measurement is performed to reproduce the results of the contextuality test in the CSW

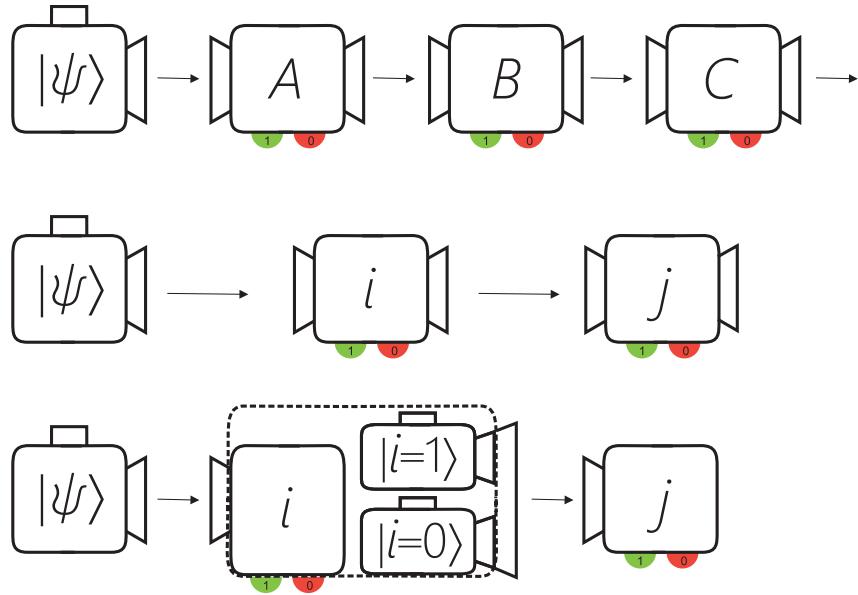


Fig. 4.3 Simplification of contextuality tests. Top: The original contextuality test requires a series of sequential measurements on the same quantum state. Middle: In the CSW framework, all contextuality tests can be represented in the form of at most one sequential measurement; the two operators in the sequential measurement must be orthogonal. Bottom: By additionally accepting the principle of state update using Lüders rule, the need for sequential measurements can be completely eliminated. The first measurement result and probability are obtained using projection measurements, followed by the state preparation based on Lüders rule, and then the second projection measurement is performed to reproduce the results of the contextuality test in the CSW framework.

framework. If we use the subscript Pr_ψ to denote the probability of measurement when the input state is $|\psi\rangle$, the inequality to be analyzed under this method can be expressed as:

$$\sum_{k \in V(G)} \text{Pr}_\psi(1|k) - \sum_{(i,j) \in E(G)} \text{Pr}_\psi(1|i)\text{Pr}_i(1|j) \stackrel{\text{NCHV+LR}}{\leq} \alpha(G). \quad (4.6)$$

Where the additional constraint condition LR = Lüders rule indicates that this testing method must be based on accepting Lüders rule. Therefore, this simplified method comes with a certain cost in terms of concept: it accepts part of the postulates of quantum mechanics as a prerequisite for testing quantum contextuality. However, the significance of this method lies in greatly expanding the range of noncontextual theories that can be tested. For example, when we test systems with dimensions increasing to the order of $d \sim 10$, there is currently no continuous measurement method that can achieve arbitrary measurements with high precision. On the other hand, the use of projection measurements based on optical field manipulation and path patterns is a mature technique. For example, Xiao *et al.* [122] used this method to study contextuality in the framework of Platonic solids. In section 4.4, we observed the phenomenon of contextual concentration using this method in a 7-dimensional quantum system encoded in paths.

4.3 Contextuality beyond nonlocality

Observing contextual phenomena that go beyond nonlocality is challenging experimentally. The reason is that in known cases, the difference in strength between contextuality and nonlocal behavior, denoted as $\Delta_{\text{NL}}^{\text{QC}}$, is extremely small, on the order of $O(10^{-2})$. This significantly limits the reliability of observing contextuality beyond nonlocality. Therefore, there are very few experimental works directly comparing the differences in these two types of correlation strength. Below, we present our research on this issue. This work is based on the study of generalized Bell inequalities corresponding to contextual graphs conducted by Prof. Jing-Ling Chen's research group.

4.3.1 Sketch of the theory

The earlier CHSH experiment considered a (2,2,2)-Bell inequality. In fact, Bell inequalities can be extended to more general scenarios. Here, we present a Bell inequality in the form of probabilities, involving two observers, three measurement settings, and two possible outcomes. It is a (2,3,2)-Bell inequality, where Alice and Bob can each perform three different measurements, and each measurement can yield two different outcomes. Therefore, this inequality is conventionally denoted as I_{3322} . In our study, we use a variant of this inequality in the CSW form:

$$\begin{aligned} I_{3322}^{\text{NL}} = & \Pr(0, 0|0, 1) + \Pr(0, 0|0, 2) + \Pr(0, 0|1, 0) + \Pr(0, 0|1, 2) \\ & + \Pr(0, 0|2, 0) + \Pr(0, 0|2, 1) + \Pr(0, 1|1, 1) + \Pr(1, 0|1, 1) \\ & + \Pr(1, 1|1, 1) + \Pr(0, 1|2, 2) + \Pr(1, 0|2, 2) + \Pr(1, 1|2, 2) \\ & + \Pr(1, -|0, -) + \Pr(1, -|1, -) + \Pr(-, 1|-, 0) + \Pr(-, 1|-, 1) \stackrel{\text{LHV}}{\leq} 6. \end{aligned} \quad (4.7)$$

Using the measurement events in the I_{3322} inequality, we can obtain the contextual graph G_{I3322} corresponding to the 16 measurements, as shown in [Figure 4.4](#). According to [Theorem 2.3](#), we can immediately derive a noncontextuality inequality from the contextual graph:

$$I_{3322}^{\text{QC}} = \sum_{k \in V(G_{I3322})} \Pr(1|k) - \sum_{(i,j) \in E(G_{I3322})} \Pr(1, 1|i, j) \stackrel{\text{NCHV}}{\leq} \alpha(G_{I3322}) = 6. \quad (4.8)$$

The upper bound of the noncontextuality inequality is also 6, which is not accidental but a consequence of Bell nonlocality being a special case of contextuality. For each term with coefficient 1 in the CSW form of a Bell inequality, the upper bound in the local model is the independence number of the contextual graph. The interesting feature of G_{I3322} is that it exhibits a strong instance of contextuality beyond nonlocality. Firstly, using semidefinite programming, we can trivially find $\vartheta(G_{I3322}) \approx 6.588$ and it requires a 5-dimensional quantum system. Moreover, using a 4-dimensional quantum system, we can achieve $I_{3322}^{\text{QC}} \stackrel{\text{Q}}{\approx} 6.571$. Secondly, the maximum violation of I_{3322}^{NL} is an unresolved question in the foundations of quantum mechanics. Pál and Vértesi [123] computed $I_{3322}^{\text{NL}} \stackrel{\text{Q}}{\approx} 6.251$ and it requires an *infinite*-dimensional system to achieve this maximum value. For qubit systems, the maximum achievable value is $I_{3322}^{\text{NL}} \stackrel{\text{Q}}{\approx} 25/4$, which is significantly different from the result in the contextuality scenario, giving $\Delta_{\text{NL}}^{\text{QC}}(I_{3322}) \approx 0.337$. This provides a significant example of contextuality beyond nonlocality.

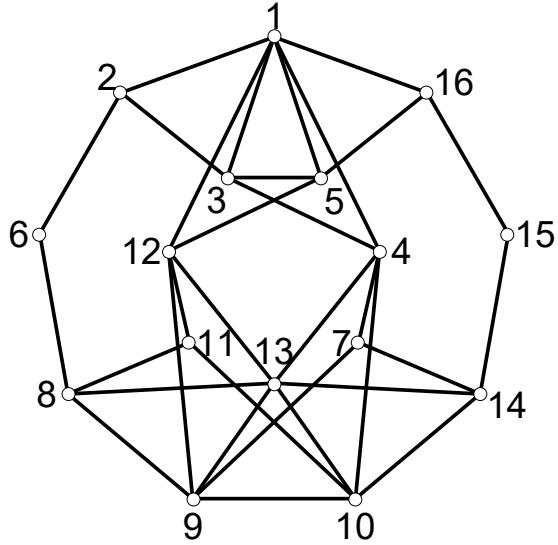


Fig. 4.4 The graph of exclusivity G_{I3322} corresponding to the observables in the Bell inequality I_{3322} . The events corresponding to the nonlocality experiments are defined as: **1** : $1, 1|2, 2$, **2** : $0, 0|2, 0$, **3** : $1, 0|2, 2$, **4** : $0, 0|2, 1$, **5** : $0, 1|2, 2$, **6** : $-, 1|-, 0$, **7** : $-, 1|-, 1$, **8** : $0, 0|1, 0$, **9** : $1, 0|1, 1$, **10** : $0, 1|1, 1$, **11** : $1, -, 1|-, 1$, **12** : $0, 0|1, 2$, **13** : $1, 1|1, 1$, **14** : $0, 0|0, 1$, **15** : $1, -, 0|-, 0$, **16** : $0, 0|0, 2$. For the contextuality experiment, we only required the relations of exclusivity of the events to conform to the graph without specifying the exact settings of each measurement event.

4.3.2 Experimental implementation

Due to G_{I3322} exhibiting a contextuality–nonlocality behavior gap that is nearly an order of magnitude larger than any other known contextual graph to date, this scenario is suitable for experimentally demonstrating contextuality beyond nonlocality. Based on inequalities (4.7) and (4.8), we conducted experiments using a two-qubit system with polarization-entangled encoding and a four-dimensional single-particle system with orbital angular momentum encoding to test the maximum quantum correlation achievable by nonlocality and contextuality, respectively. In this section, we describe these two experiments.

We first describe the contextuality experiment. In general, it is necessary to implement state preparation and measurements in a four-dimensional photon system in a set of bases. The experimental setup is shown in Figure 4.5. Photons enter the experimental setup through a single-mode fiber. By adding a holographic phase pattern on the spatial light modulator, any desired wavefront phase can be imprinted onto the photons. We select four computational basis states corresponding to the eigenstates of orbital angular momentum: $|i_0\rangle \leftrightarrow |l=3\rangle$, $|i_1\rangle \leftrightarrow |l=1\rangle$, $|i_2\rangle \leftrightarrow |l=-1\rangle$, $|i_3\rangle \leftrightarrow |l=-3\rangle$. This can be achieved by modulating the photon's wavefunction to the corresponding order Laguerre–Gaussian mode LG_p^l , where $p=0$. The spatial light modulator for state preparation is placed at the input plane of a $4f$ -system, and the second spatial light modulator for measurement is placed at the output plane. An iris diaphragm is inserted at the focal plane to filter out unwanted zeroth-order and high-order diffraction mode terms, retaining only the first-order diffraction mode corresponding to the target state [103]. The function of the second spatial light modulator is to convert the measured orbital angular momentum state back to a Gaussian mode $|l=0\rangle$, which is then selected by a single-mode fiber. As previously mentioned in chapter 3, photons carrying non-zero orbital angular momentum cannot be focused to a point, so they are filtered out by the single-mode fiber. Additionally, two lenses are inserted between the second spatial light modulator and the single-mode fiber to adjust the beam size and optimize fiber coupling efficiency.

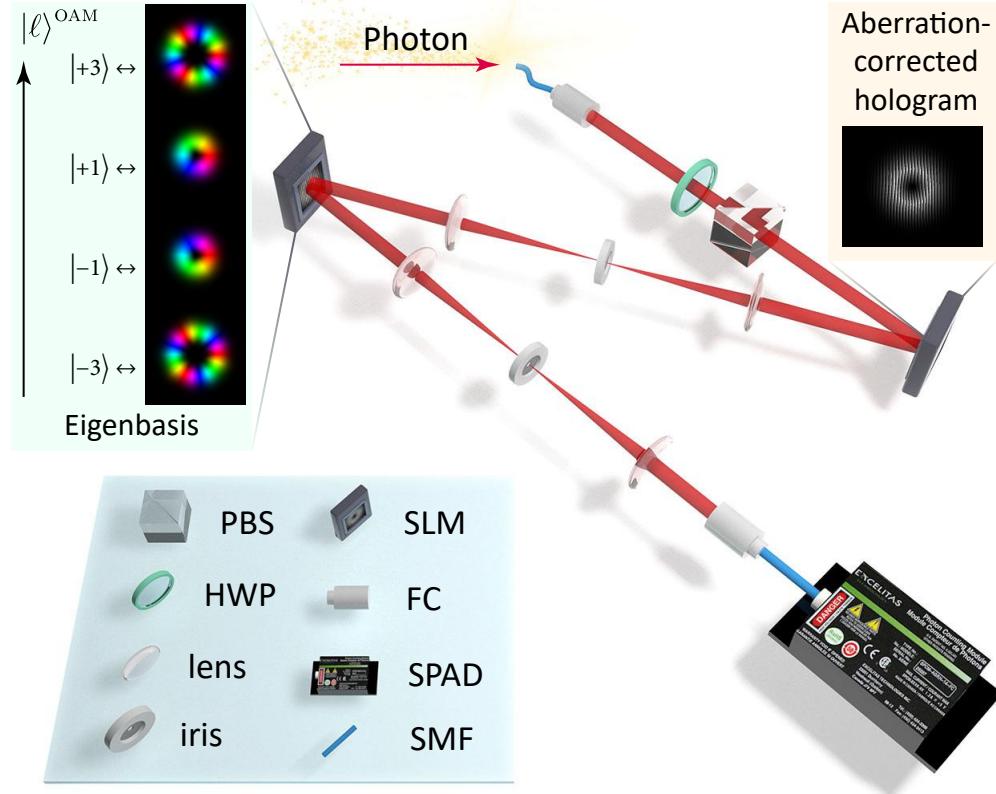


Fig. 4.5 Experimental setup for contextuality beyond nonlocality. The four-dimensional quantum state is encoded in the orbital angular momentum mode of photons. The orbital angular momentum mode preparation is achieved using a spatial light modulator, and precise measurements are performed using a single-mode fiber. Finally, a single-photon avalanche detector is used for single-photon counting to calculate the required measurement probabilities in Equation (4.8). PBS: polarizing beam splitter, HWP: half-wave plate, SLM: spatial light modulator, FC: fiber coupler, SPAD: single-photon avalanche detector, SMF: Single-mode fiber.

Finally, the photons are coupled to a single-photon avalanche detector. In each state preparation and measurement setting, the count rate is proportional to the measurement probability: $\text{Pr}_\psi(1|\phi) = |\langle \phi | \psi \rangle|^2$.

Using the apparatus shown in Figure 4.5, we measured each term in Equation (4.8) according to the setup described in Figure 4.5. For the first term, we measured the probabilities of the initial state in the basis of certain projection operators. For the second term, we additionally measured the probabilities of the $+1$ eigenstates of each projection operator in all the mutually exclusive basis states. By multiplying these probabilities with the probabilities obtained from the measurement of the initial state on the original projection operators, we can reconstruct the results for each term in Equation (4.8). The experimental results are shown in Figure 4.10, and they match the theoretical expectations very well, indicating good orthogonality between the mutually exclusive operators. Substituting the measurement results into

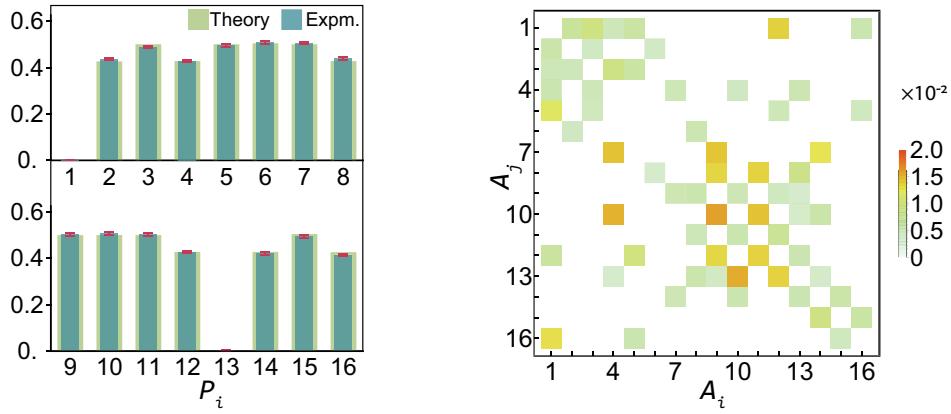


Fig. 4.6 Experimental data for contextuality beyond nonlocality: contextuality part. Left: measurement probabilities corresponding to the 16 projection operators that satisfy the mutual exclusivity relation shown in Figure 4.4. Right: measurement probabilities between the eigenstates of the projection operators $\hat{\Pi}_i$ and $\hat{\Pi}_j$ that should, in principle, be mutually exclusive. The corresponding probabilities are measured only when $(i, j) \in E(G_{13322})$.

Equation (4.8), we obtain:

$$I_{3322}^{\text{QC}} = 6.488 \pm 0.025. \quad (4.9)$$

Here, the error bar corresponds to the 1σ standard deviation estimated from Poisson counting statistics. This result violates the noncontextual hidden variable inequality (4.8) with a confidence of over 19 standard deviations, thus observing the contextuality of the quantum system. It should be noted that due to the efficiency of the single-photon detectors used in the experiment ($\eta \approx 60\%$) and the diffraction efficiency of photons on the spatial light modulator ($\eta \approx 30\%$), only a small fraction of photons resulted in detector responses. Therefore, the observation of contextuality relies on the assumption of fair sampling, i.e., assuming that the photons that arrive faithfully reflect the behavior of all the photons that should be detected.

Next, we quantify the disturbance between consecutive measurements in the contextuality experiment. In this experiment, we separate the two cases of measurement results 0 and 1 in Equation (4.5) into four signal transmission factors: $\varepsilon(_, 0|_, v_j)$, $\varepsilon(0, |v_i, _)$, $\varepsilon(_, 1|_, v_j)$, and $\varepsilon(1, |v_i, _)$. They are defined as follows:

$$\begin{cases} \varepsilon(_, 0|_, v_j) &= |\Pr_\psi(_0|_j) - \Pr_\psi(00|ij) - \Pr_\psi(10|ij)|, \\ \varepsilon(_, 1|_, v_j) &= |\Pr_\psi(_1|_j) - \Pr_\psi(01|ij) - \Pr_\psi(11|ij)|, \\ \varepsilon(0, |v_i, _) &= |\Pr_\psi(i_|0_0) - \Pr_\psi(00|ij) - \Pr_\psi(01|ij)|, \\ \varepsilon(1, |v_i, _) &= |\Pr_\psi(1_|i_0) - \Pr_\psi(10|ij) - \Pr_\psi(11|ij)|. \end{cases} \quad (4.10)$$

Here, the underscores represent marginal probabilities. The quantities on the right-hand side of these equations can all be expressed using the experimentally observable probabilities:

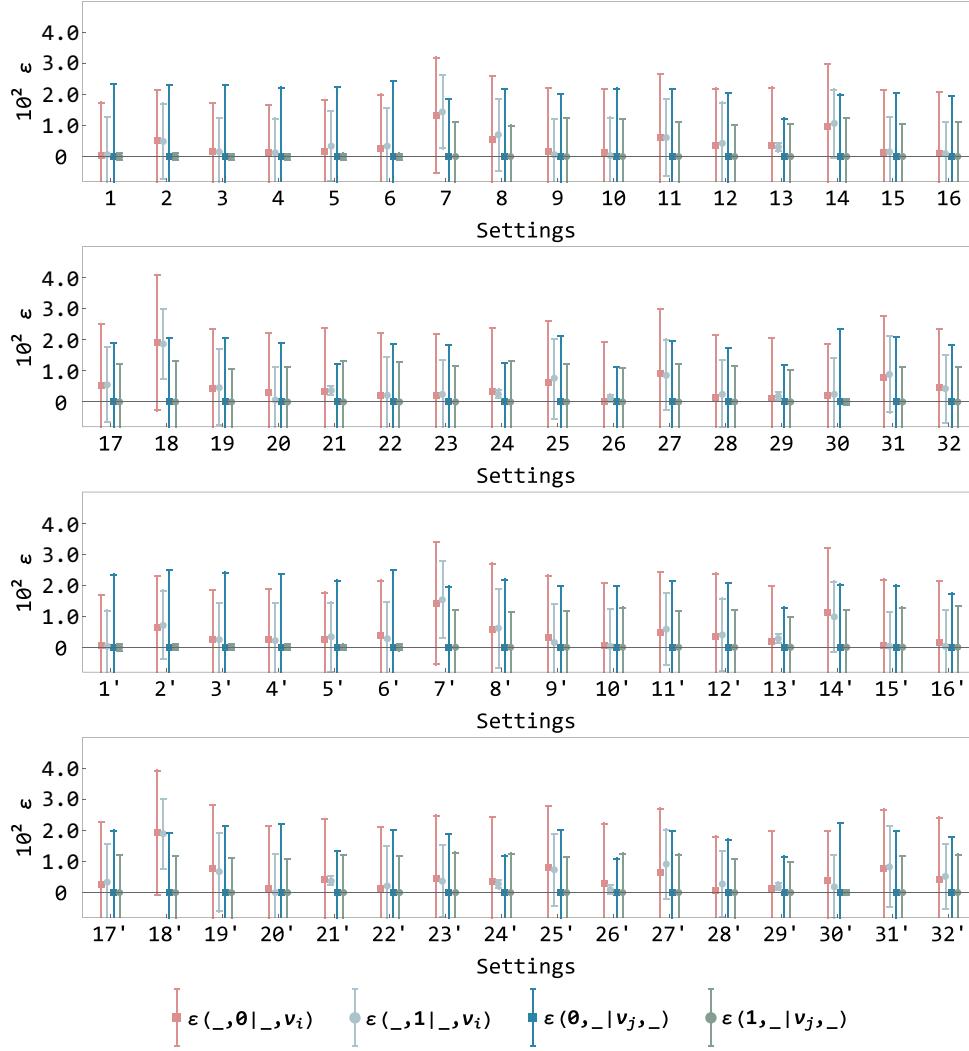


Fig. 4.7 Experimental data for contextuality beyond nonlocality: no-disturbance condition. The vertical axis represents the signal transmission factors calculated according to Equation (4.5), analyzed separately for measurement results of 0 and 1. Note that the two types of signal transmission factors that propagate towards the past automatically become 0. The horizontal axis corresponds to the 32 edges in Figure 4.4, arranged in increasing order of vertex numbers. The parameters without ' and with ' correspond to the forward ($i < j$) and backward ($i > j$) signal transmission factors, respectively.

$$\begin{cases} \Pr_{\psi}(0|i) &= 1 - \Pr_{\psi}(1|i), \\ \Pr_{\psi}(11|ij) &= \Pr_{\psi}(1|i)\Pr_i(1|j), \\ \Pr_{\psi}(01|ij) &= \Pr_{\psi}(0|i)\Pr_{i^\perp}(1|j), \\ \Pr_{\psi}(10|ij) &= \Pr_{\psi}(1|i) - \Pr_{\psi}(11|ij), \\ \Pr_{\psi}(00|ij) &= \Pr_{\psi}(0|i) - \Pr_{\psi}(01|ij). \end{cases} \quad (4.11)$$

It should be noted that in the third line of the above equations, $\Pr_{i^\perp}(1|j)$ appears, which represents the projection probability of the remaining states on j after a measurement on i yields 0. This is the point where the author mentioned earlier that Lüders' rule is used to reprepare the quantum state. The preparation is also done using the spatial light modulator: based on the measurement probabilities obtained in the first step, a hologram is displayed on the screen of the spatial light modulator to prepare:

$$|i^\perp\rangle = \frac{|\psi\rangle - \langle i|\psi\rangle|i\rangle}{||\psi\rangle - \langle i|\psi\rangle|i\rangle|} = \frac{|\psi\rangle - \langle i|\psi\rangle|i\rangle}{\sqrt{1 - |\langle i|\psi\rangle|^2}}. \quad (4.12)$$

In this way, all the required probabilities can be obtained using the preparation-measurement experiment. The calculated result shows that $|\varepsilon| = (0.22 \pm 1.44) \times 10^{-2}$. Due to the imperfections in the experiment, for a few data points, even with the addition of the 1σ error bars, ε does not satisfy $\varepsilon = 0$. However, if we slightly relax the statistical requirements and only consider the average of the data, it fits well with the no-disturbance condition, without significant effects of measurement disturbance. Therefore, it can be considered that the conditions required by Equation (4.5) are satisfied, and the violation of noncontextual hidden variable theory is valid. Furthermore, the impact caused by the imperfect no-disturbance condition can be attributed to the observed inequality results through causal modeling theory [124]. This issue goes beyond the scope of this work and will be an interesting research direction in future loophole-free contextuality experiments.

Finally, we observe the experimental results of Equation (4.7) to directly demonstrate the phenomenon of contextuality beyond nonlocality by comparing the results from the two experiments. The setup used at this point is the same as in Figure 4.5, where a half-wave plate is added to one path to generate a maximally entangled two-photon state, $|\Phi^+\rangle$. We analyze the quality of this entangled source using quantum state tomography, and the results are shown in Figure 4.8. The fidelity between the entangled source and the target state reaches 97.5%. The fidelity between two quantum states is defined as [36]:

$$F(\hat{\rho}, \hat{\rho}') := \text{Tr} \left[\left| \sqrt{\hat{\rho}} \sqrt{\hat{\rho}'} \right|^2 \right].$$

By measuring the probabilities corresponding to the two-body measurement operators in Figure 4.4, we calculate:

$$l_{3322}^{\text{NC}} = 6.165 \pm 0.012. \quad (4.13)$$

This result is slightly lower than the upper limit that l_{3322}^{NC} can reach under quantum mechanics, due to the limited fidelity between the input entangled state and the target state, as well as imperfections in the measurement devices. However, it still sufficiently demonstrates strong quantum nonlocality, violating local hidden variable inequalities by more than 13 standard deviations. Comparing the results from Equations (4.9) and (4.13), we obtain:

$$\Delta_{\text{NL}}^{\text{QC}}(l_{3322}) \approx 0.323 \pm 0.037. \quad (4.14)$$

Therefore, with a confidence level over 8.7 standard deviations, we have observed contextuality beyond nonlocality based on the graph of exclusivity G_{13322} . The research results indicate a deep connection between graph theory and quantum correlations, and further deepen the understanding of different types of quantum correlations.

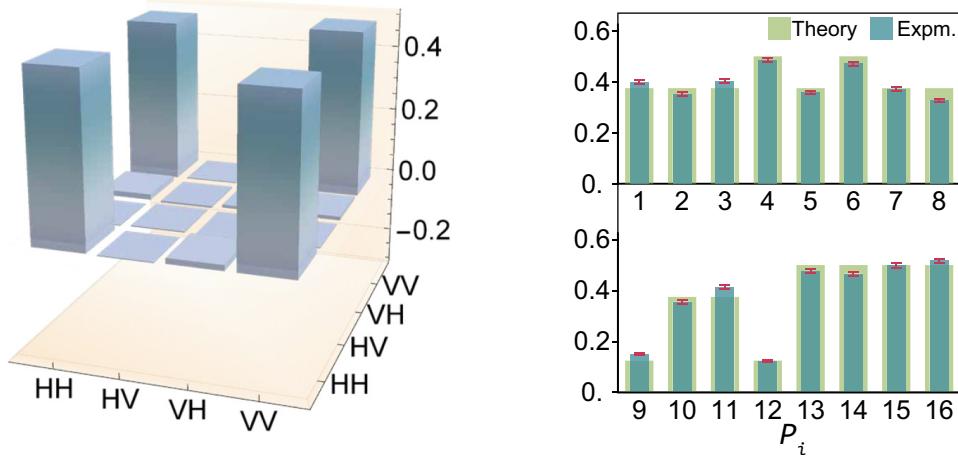


Fig. 4.8 Experimental data for contextuality beyond nonlocality: nonlocality part. Left: real part of the quantum state tomography result for the entangled source. Right: probabilities corresponding to the two-body measurement operators defined in Figure 4.4 and comparison with the theoretical results.

4.4 Contextuality concentration: strongest correlation in lower dimensions

Now we turn to the theoretical and experimental study of the phenomenon of contextual concentration. Compared to contextuality beyond nonlocality, contextual concentration is a more recent discovery and has been rarely observed beyond the KCBS experiment. Since 2021, advancements in graph theory and quantum correlations have led to increasing attention in this research direction. Among them, Ray *et al.* [125] discovered a maximum violation of a Bell inequality (commonly known as the Mermin [126] inequality) applicable to the three-qubit GHZ state. This violation can be equivalently achieved in the context of graphs of exclusivity using a $2^3 - 1 = 7$ -dimensional indivisible system. In this section, we demonstrate that the examples discovered by Ray et al. exist universally in Mermin inequalities for any number of particles. The n -particle Mermin inequalities can always be transformed into noncontextual hidden variable inequalities using the CSW framework, and the maximum violation can be achieved in a $2^n - 1$ -dimensional system. We utilize an optical system with path encoding and the optical field control methods developed in the previous section to construct an interferometer and directly observe contextual concentration for $d = 7$.

4.4.1 Mermin inequality: exponential quantum–classical gap

Previously, we introduced the implementation of the four-photon GHZ state using fusion gates. Now, let us formally define the GHZ state in a more general sense. The n -particle GHZ state has the form:

$$|\text{GHZ}_n\rangle = (\underbrace{|00\dots 0\rangle}_n + i \underbrace{|11\dots 1\rangle}_n)/\sqrt{2},$$

which is a superposition of two quantum states that are completely opposite. When n is large, it describes the coherent superposition of two macroscopic object states, similar to Schrödinger's cat. The GHZ state exhibits strong nonclassical features, such as entanglement among all n qubits. These nonclassical features are well demonstrated in the Mermin inequality. The Mermin inequality is an $(n, 2, 2)$ -Bell inequality, and its form for odd n is³:

$$\mathcal{M}_n = \frac{1}{2i} \sum_{\nu \in \{\pm 1\}} \nu \bigotimes_{j=1}^n \left(\sigma_x^{(j)} + i\nu \sigma_y^{(j)} \right) \stackrel{\text{LHV}}{\leq} 2^{\frac{n-1}{2}}. \quad (4.15)$$

In quantum mechanics, the upper limit allowed is $\mathcal{M}_n \stackrel{\text{Q}}{\leq} 2^{n-1}$, which can be achieved using the GHZ state. It can be seen that in this case, the ratio of quantum correlations to the maximum strength allowed by local hidden variables *increases exponentially* with the number of particles!

Next, we use the CSW framework to transform the Mermin inequality into a noncontextual hidden variable inequality. Similar to the operations in the CHSH experiment, we keep only the positive contributions from the projection measurements in Equation (4.15), represent the negative contributions using the positive ones, and add a constant corresponding to the sum of the total probabilities. The resulting probability sum is denoted as $\mu_n = \mathcal{M}_n/2 + 2^{n-2}$, and its corresponding graph of exclusivity is denoted as G_{M_n} . As an example, in Figure 4.9, I provide the form of the graph of exclusivity G_{M_3} corresponding to $n = 3$. According to the calculations by Cabello *et al.* [15]:

$$\alpha(G_{M_n}) = 2^{n-2} + 2^{\frac{n-3}{2}}, \quad \vartheta(G_{M_n}) = 2^{n-1}. \quad (4.16)$$

By substituting into Theorem 2.3, we have:

$$\mu_n \stackrel{\text{NCHV}}{\leq} \alpha(G_{M_n}) = 2^{\frac{n-3}{2}} + 2^{n-2} \stackrel{\text{Q}}{\leq} \vartheta(G_{M_n}) = 2^{n-1}. \quad (4.17)$$

In μ_n , the total number of contexts used is equal to the Lovász number. Because each complete context contains 2^n projection measurements, half of which appear in μ_n , we have $|V(G_{M_n})| = 2^{2n-2}$.

Studying the graph of exclusivity G_{M_n} corresponding to the Mermin inequality has the first significance that the ratio of allowed correlation strength between quantum theory and noncontextual hidden variable theory, $\vartheta/\alpha = 2 - 2 / \left(1 + 2^{\frac{n-1}{2}} \right)$, increases with n . Inequalities based on graphs of exclusivity inherit the property of high violation in Mermin inequalities and exhibit excellent robustness against potential experimental errors in quantum violation. Therefore, they are well-suited for observing contextuality in high-dimensional systems. Another example of the increase in contextual correlation strength with dimensionality is the work by Vidick and Wehner [128]. However, overall, there are very few known cases that satisfy this property.

4.4.2 Reducing the dimensionality of measurements in the Mermin inequality

The second significance of studying the graph of exclusivity G_{M_n} corresponding to the Mermin inequality is that it involves contextuality concentration. The quantum maximum of μ_n only requires utilizing a $2^n - 1$ dimensional Hilbert space. I provide a constructive proof: Let $\boldsymbol{\Pi} = \{\hat{\Pi}_1, \hat{\Pi}_2, \dots, \hat{\Pi}_{2^{2n-2}}\}^T$ be the concatenation of all projection operators in Equation (4.17). Based on knowledge from linear algebra:

³ We do not consider the case of even n where the Mermin inequality is not the strongest [127].

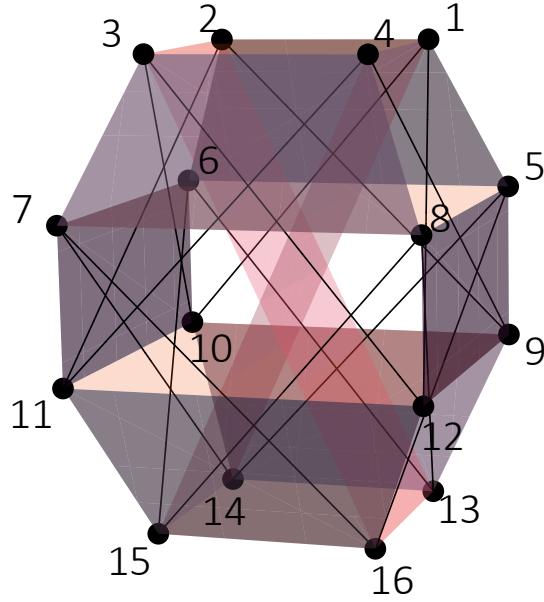


Fig. 4.9 Graph of exclusivity of the projection measurements in the Mermin inequality. Following the CSW method, only events with positive contributions to the inequality are shown in the graph. The points connected by line segments represent mutually exclusive events, and the four vertices of the shaded plane correspond to events that are all mutually exclusive. The case of $n = 3$ is shown in the figure.

$$\text{Rank}(\mathcal{A}) = 2^n - \text{Dim}(\ker \mathbf{\Pi}), \quad (4.18)$$

where $\ker \mathbf{\Pi}$ is the kernel of the matrix $\mathbf{\Pi}$, i.e., the solution space of $\mathbf{\Pi}\mathbf{x} = \underbrace{\mathbf{0}\mathbf{0}\dots\mathbf{0}}_n$ for 2^n dimensional vector \mathbf{x} . If we go back to considering Equation (4.15), we will find that this solution space actually corresponds to the requirement that each term in the sum yields -1 . Now, we already know that the GHZ state that gives the maximum violation makes each term in the sum yield $+1$. The following proposition demonstrates that there is a unique state that reverses the sign of each term in Equation (4.15), and it can be explicitly constructed using Clifford algebra.

Proposition 4.1 $\text{Dim}(\ker \mathbf{\Pi}) = 1$. In fact, by taking:

$$|\text{GHZ}'_n\rangle = (\sigma_z \otimes \mathbb{I}_2^{\otimes n-1}) |\text{GHZ}_n\rangle = (|\underbrace{00\dots 0}_n\rangle - i|\underbrace{11\dots 1}_n\rangle)/\sqrt{2},$$

we obtain

$$\langle \mathcal{M}_n \rangle_{\text{GHZ}'_n} = -2^{n-1},$$

which means that $|\text{GHZ}'_n\rangle$ is the -1 eigenstate of each binary operator in Equation (4.15).

Proof First, the product of Pauli operators satisfies the following relation:

$$\sigma_i \sigma_j \sigma_i = (-1)^{1-\delta_{ij}} \sigma_j, \quad i, j \in \{x, y, z\}. \quad (4.19)$$

This can be directly verified using the definition of Pauli operators. Now, we observe that in Equation (4.15), the Pauli operator acting on the first qubit is either σ_x or σ_y . Therefore, using the above relation, we can obtain: $\sigma_z \sigma_x \sigma_z = -\sigma_x$, $\sigma_z \sigma_y \sigma_z = -\sigma_y$. Hence,

$$\langle \mathcal{M}_n \rangle_{\text{GHZ}'_n} = \left\langle \text{GHZ}'_n \left| \frac{1}{2i} \sum_{\nu \in \{\pm 1\}} \nu \bigotimes_{j=1}^n (\sigma_x^{(j)} + i\nu \sigma_y^{(j)}) \right| \text{GHZ}'_n \right\rangle \quad (4.20)$$

$$= \left\langle \text{GHZ}_n \left| \frac{1}{2i} \sigma_z^{(1)} \sum_{\nu \in \{\pm 1\}} \nu \bigotimes_{j=1}^n (\sigma_x^{(j)} + i\nu \sigma_y^{(j)}) \sigma_z^{(1)} \right| \text{GHZ}_n \right\rangle \quad (4.21)$$

$$= \left\langle \text{GHZ}_n \left| -\frac{1}{2i} \sum_{\nu \in \{\pm 1\}} \nu \bigotimes_{j=1}^n (\sigma_x^{(j)} + i\nu \sigma_y^{(j)}) \right| \text{GHZ}_n \right\rangle \quad (4.22)$$

$$= -\langle \mathcal{M}_n \rangle_{\text{GHZ}_n} = -2^{n-1}. \quad (4.23)$$

Finally, we prove that the state satisfying $\langle \mathcal{M}_n \rangle = -2^{n-1}$ is unique. Consider the following n operators extracted from Equation (4.15):

$$-\sigma_y \otimes (\sigma_x)^{\otimes n-2} \otimes \sigma_y, \quad -\sigma_x^{\otimes k} \otimes \sigma_y^{\otimes 2} \otimes \sigma_x^{\otimes n-k-2}, \quad k \in \{0, \dots, n-2\}.$$

Each operator is linearly independent and has eigenvalues ± 1 . For each eigenvalue, there are $n/2$ eigenvectors. Therefore, the dimension of the subspace spanned by the eigenvectors with eigenvalue $+1$ is 2^{n-1} . Furthermore, since the eigenstates of the three Pauli operators form a mutually unbiased basis, the dimension of the subspace spanned by the eigenvectors with the common eigenvalue $+1$ for two operators is 2^{n-2} , and so on. Thus, the dimension of the subspace spanned by the eigenvectors with eigenvalue $+1$ for all n operators is $2^{n-n} = 1$. It is the explicitly constructed state $|\text{GHZ}'_n\rangle$. Therefore, the state satisfying $\langle \mathcal{M}_n \rangle = -2^{n-1}$ is unique. \square

Thus, we have proved that the projection measurements in Π can be spanned in a $2^n - 1$ -dimensional space, but it may require an indivisible system. In terms of the degree of violation of local hidden variable theories, the Mermin inequality represents the strongest known form of quantum nonlocality [129]. Our work indicates that it can be attributed to contextuality concentration already present in lower-dimensional Hilbert spaces.

4.4.3 Experimental implementation

The path degree of freedom of photons provides a powerful way to encode high-dimensional quantum information. With the help of spatial light modulators [93, 94, 130, 131], precise manipulation of high-dimensional quantum states has been achieved. Here, considering the required precision, we choose to implement the preparation-measurement experiment with the spatial mode encoded in the path degree of freedom, which has a high demand for the dimension of the state space.

The experimental setup is shown in Figure 4.10, which is almost identical to Figure 4.5, except for the quantum information encoding scheme. Now, a spatial light modulator displays holograms with seven circular sectors, with each sector corresponding to a computational basis state. Therefore, overall, the preparation-measurement of a seven-dimensional single-photon system is achieved. Within each sector, a diffractive grating is added to control the amplitude of the photon wavefunction in the respective sector by adjusting the maximum phase variation of the grating. To examine the precision of quantum state

preparation and measurement, we scan through all the holograms used in the experiment, measure the intensity distribution before the second spatial light modulator using a charge-coupled device camera, and compare them with the theoretical predictions. The comparison of the experimentally prepared path quantum states with the theoretical ones is done using the Pearson correlation function [132]. The computed results show an average Pearson correlation function value of 95.5%, indicating that the method can achieve precise initial state preparation and measurement.

We first measure the violation of the noncontextual hidden variable inequality in Equation (4.17). In the measurement of the first term, the first spatial light modulator prepares the same initial state, and the second spatial light modulator scans through all measurement bases to obtain all the required projection measurement probabilities. In the measurement of the second term, both spatial light modulators traverse all combinations of mutually exclusive measurements, record the projection measurement probabilities under these bases, and then calculate the final result. The experimental data is shown in Figure 4.11. The computed result shows that

$$\mu_3 = 3.821 \pm 0.012. \quad (4.24)$$

It violates the noncontextual hidden variable inequality (4.17) by 68.7 standard deviations.

There is another very intuitive way to present the quantum contextuality in G_{M_3} . Disregard the second term of Equation (4.17) and express the first term in a “constraint–conclusion” form yields:

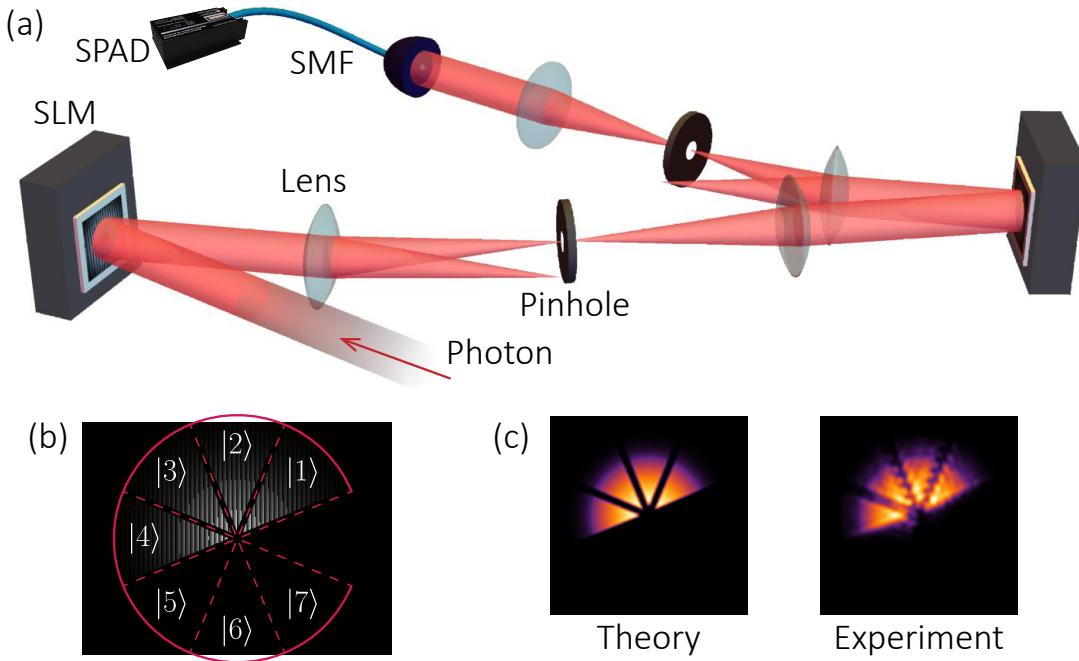


Fig. 4.10 Experimental setup for observing contextuality concentration. The upper panel shows the overall optical path diagram, which is the same as the one used for transcending nonlocality. The lower-left panel illustrates the path encoding scheme and the holograms used. The lower-right panel compares the theoretical and experimental intensity distributions obtained from hologram calculations. In all the figures below, the prepared quantum states are $(|1\rangle + |2\rangle + |3\rangle + |4\rangle)/2$.

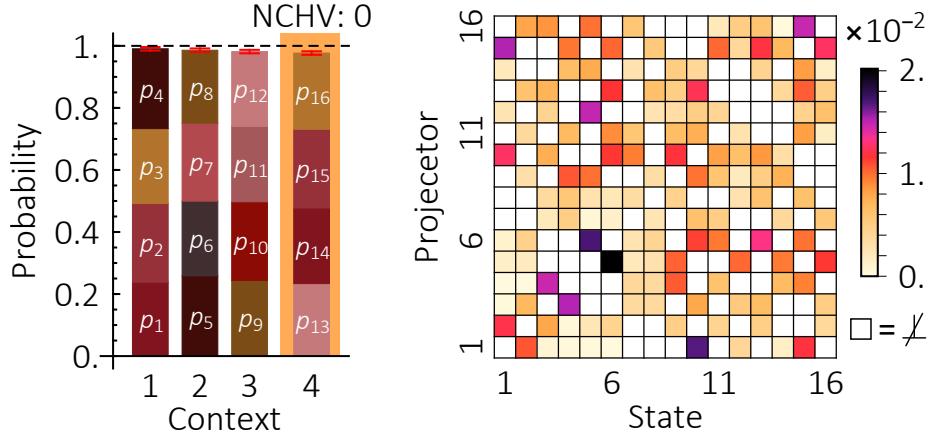


Fig. 4.11 Experimental results of contextuality concentration. Left: stacked bar graphs of measurement probabilities for each event. Each bar corresponds to four events in a layer in Figure 4.9. According to the predictions of the noncontextual hidden variable model, when the first three bars are nearly full, the last bar should be empty, thereby intuitively demonstrating contextuality without the need for inequalities. Right: orthogonality between mutually exclusive measurements, determined by preparing the eigenstates of each projection operator and measuring the corresponding projection probabilities of the mutually exclusive projection operators. Only the grids corresponding to mutually exclusive measurements are computed and colored.

$$\begin{aligned}
 \sum_{i=1}^4 \Pr(1|i) &= 1, \quad \sum_{i=5}^8 \Pr(1|i) = 1, \quad \sum_{i=9}^{12} \Pr(1|i) = 1, \\
 \hline
 1 \stackrel{Q}{=} \Pr(\text{success}) &= \sum_{i=13}^{16} \Pr(1|i) \stackrel{\text{NCHV}}{=} 0.
 \end{aligned} \tag{4.25}$$

It can be seen that the contradiction "1=0" arises between the quantum mechanical prediction of the sum of probabilities and the noncontextual hidden variable theory in the last equation. This is precisely the method demonstrated by Hardy [77] that almost applies to all entangled states, and it has been extended to the study of contextuality by Cabello *et al.* [133], now commonly known as the "inequality-free method". In the experiment, we often refer to the probability of this event, prohibited by noncontextual hidden variables, as the "success probability". Using the form of Equation (4.25), we express the measurement probabilities as stacked bar graphs, where the first three bars are almost completely filled, thus satisfying the constraints of Equation (4.25). The last completely filled bar provides strong evidence in favor of quantum mechanics, contradicting the predictions of nonlocal hidden variable theory for the measurement probabilities.

Finally, we present the results of the no-disturbance condition testing. The experimental procedure involves preparing the conditioned state, obtained from the measurement of the first projection operator, as the initial state again, followed by a measurement using the second projection operator for each edge connecting two measurement projections in the graph of exclusivity G_{M_3} . In this experiment, we directly use the original form of Equation (4.5) to characterize the no-disturbance condition. For the conditional probability when the result of the first measurement is 0, i.e., $P(0, 1|i, j)$, the required conditioned state

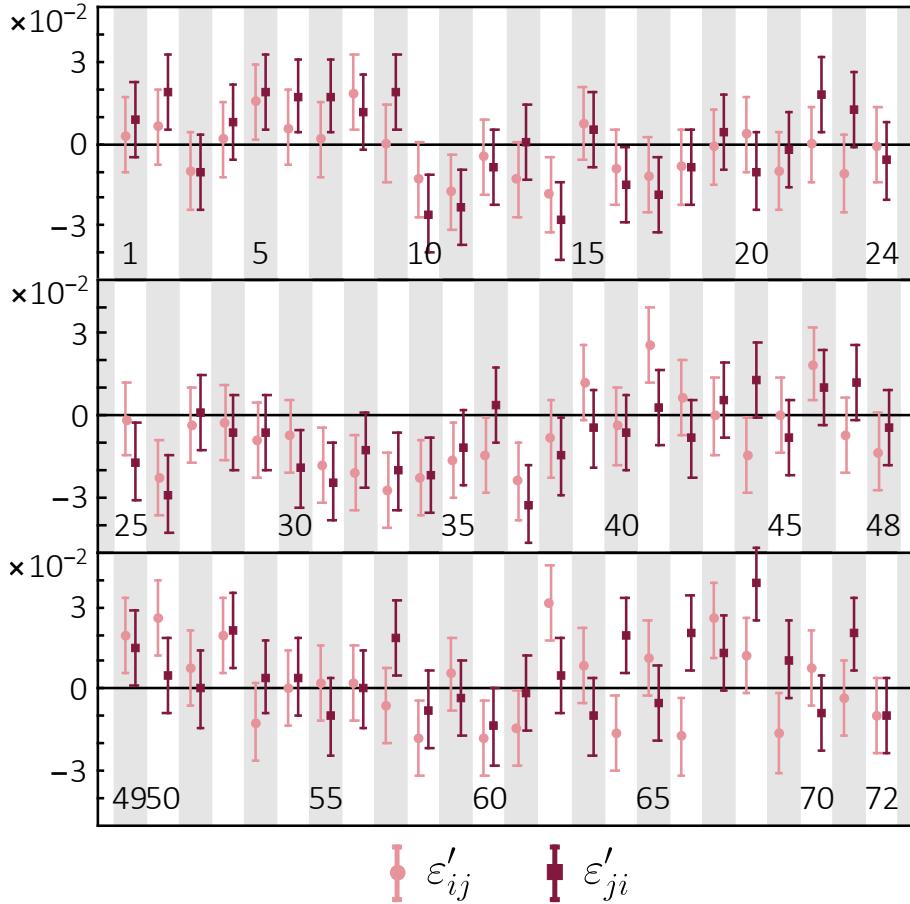


Fig. 4.12 Results of no-disturbance condition testing in the contextuality concentration experiment. The labeled numbers represent the order of edges $(i, j) \in G_{M_3}$ in the graph, sorted according to i and then j . The data points represent the signal transfer factors, and the error bars represent the 1σ standard deviation obtained through re-sampling based on Poisson counting statistics.

preparation method is also calculated based on the Lüders rule and implemented by directly repreparing the state using the spatial light modulator. The other necessary quantities have already been measured in the previous experiments. Since signals cannot propagate from future to past, the values of ε_{ij} and ε_{ji} , representing the influence of later measurements on earlier measurements, are always zero. The measured results of ε'_{ij} and ε'_{ji} for the 72 edges in G_{M_3} are shown in Figure 4.12. Most of the signal transfer factors deviate from 0 within only one standard deviation, and the quantitative result gives $|\overline{\varepsilon'}| = (1.17 \pm 1.39)\%$. Overall, the results are in good agreement with the no-disturbance condition.

4.5 Summary

We have investigated the relationship between contextuality and nonlocality and experimentally demonstrated contextuality concentration that goes beyond nonlocality. The results of the contextuality concentration experiment exhibit the strongest known form of nonlocality in arbitrary multipartite systems, which can always be transformed into contextuality in lower-dimensional systems. Our optical experiment, by constructing a 7-dimensional quantum system encoded in spatial modes, observes strong quantum contextuality resulting from the simplest case in the constructed scenario, thus enriching theoretical findings. Moreover, the studied contextuality based on the CSW framework and stabilizer is an important resource for implementing quantum computation in various quantum algorithms [16, 81, 84]. Therefore, our work has great potential to further advance high-dimensional quantum information processing and the development of novel quantum algorithms.

Chapter 5

“All-versus-nothing” contextuality in graph states

In this chapter, we focus on graph states—a class of highly entangled multipartite quantum states that serve as the foundation for measurement-based quantum computation [33, 34, 134, 135]. Studying the contextuality of graph states not only enriches our understanding of their manipulation and measurement techniques but also provides insights into the characteristics of graph-state-based quantum computation. It is important to note that although graph states are realized using multipartite quantum bits (qubits), their contextuality can manifest as nonlocality. However, the methods used to implement graph states often cannot guarantee that the individual qubits are in a spacelike-separated configuration, and in some cases, these qubits may not even exist simultaneously [136]! Therefore, in this chapter, we attribute these paradoxes to contextuality and no longer mention nonlocality, which has been discussed in the previous chapter. Graph states are often defined using undirected connected graphs, so please differentiate between the representation of graph states and the mutually exclusive graphs introduced earlier.

Various inequalities with desirable properties have been constructed for graph states, such as the Mermin inequalities by Gühne *et al.* [137], Tóth *et al.* [138], and Cabello *et al.* [139], and the graph-state self-testing inequalities based on the Ardehali inequality by Baccari *et al.* [140]. In this work, we take a different approach and study the contextuality of graph states using the “inequality-free method”¹. An example of this method was already presented in the previous chapter’s conclusion, demonstrated in Equation (4.25). In the inequality-free method, events that are prohibited by noncontextual hidden variable theories may occur with a certain probability of success in experiments described by quantum mechanics. If this success probability reaches 100%, then, under certain assumptions, quantum mechanics and noncontextual hidden variable models *must* contradict each other for the measurement results of individual events. We refer to this situation as the “all-versus-nothing” paradox induced by contextuality. In summary, the “all-versus-nothing” paradox can be expressed by the equation:

$$\pm 1_{\text{Quantum}} = \mp 1_{\text{Classical}},$$

where the equal sign on both sides represents the predictions of the two theories for some observable product. The classical theory refers to noncontextual hidden variable theories. Readers familiar with chapter 2 will immediately recognize that the peculiar properties of Peres–Mermin squares can be transformed into an “all-versus-nothing” paradox.

Although the study of the “all-versus-nothing” paradox has been conducted in various systems and has yielded certain findings, a unified method for constructing the paradox had not been discovered before our work. Our work demonstrates that graph states possess abundant resources for contextuality. We

¹ Although error analysis still relies on Equation (4.17), this method provides a completely new approach in theoretical analysis.

provide a unified construction method that allows each part of a graph state to be used for constructing “all-versus-nothing” paradoxes, and these paradoxes can be employed to verify the form of graph states and witness multipartite entanglement in many cases. In conjunction with theoretical results, we perform linear optics experiments to test the “all-versus-nothing” paradox using a two-photon four-qubit graph state and demonstrate the proposed applications. Our work deepens our understanding of the foundations of quantum mechanics and quantum paradoxes and has the potential to find applications in a wide range of quantum information tasks.

5.1 Theoretical foundations: “all-versus-nothing” paradox and the graph states

To introduce the “all-versus-nothing” paradox, I shall once again consider the three-qubit GHZ state $|\text{GHZ}_3\rangle = (|000\rangle + |111\rangle)/\sqrt{2}$ as our starting point and write down the outcomes of all four measurements in the Mermin inequality:

$$\left\{ \begin{array}{ll} [E_1 = \sigma_x^1 \sigma_x^2 \sigma_x^3] |\text{GHZ}_3\rangle & = + |\text{GHZ}_3\rangle, \\ [E_2 = \sigma_x^1 \sigma_y^2 \sigma_y^3] |\text{GHZ}_3\rangle & = - |\text{GHZ}_3\rangle, \\ [E_3 = \sigma_y^1 \sigma_x^2 \sigma_y^3] |\text{GHZ}_3\rangle & = - |\text{GHZ}_3\rangle, \\ [E_4 = \sigma_y^1 \sigma_y^2 \sigma_x^3] |\text{GHZ}_3\rangle & = - |\text{GHZ}_3\rangle, \end{array} \right. \quad (5.1)$$

Therefore, $\langle E_1 \rangle_{\text{GHZ}_3} \langle E_2 \rangle_{\text{GHZ}_3} \langle E_3 \rangle_{\text{GHZ}_3} \langle E_4 \rangle_{\text{GHZ}_3} = -1$. However, if we multiply the results of these four measurements together, we find that these outcomes cannot be explained by any noncontextual hidden variable model. These theories allow us to assign binary results $v(\cdot) = \pm 1$, predetermined by hidden variables, to each of the basic Pauli operators:

$$\sigma_\nu^i \rightarrow v_\nu^i, \quad \nu \in \{x, y, z\}, \quad i \in \{1, 2, 3\}.$$

However, this assignment leads to:

$$\begin{aligned} v(E_1) &= v_x^1 v_x^2 v_x^3, \\ v(E_2) &= v_x^1 v_y^2 v_y^3, \\ v(E_3) &= v_y^1 v_x^2 v_y^3, \\ \times & \qquad \qquad \qquad v(E_4) = v_y^1 v_y^2 v_x^3 \\ \hline v(E_1)v(E_2)v(E_3)v(E_4) &= \bigotimes_{i=1}^3 \prod_{\nu \in \{x, y\}} v_\nu^i = +1. \end{aligned}$$

Consequently, regardless of how we adjust the assignments of the observed values for each Pauli operator, the product of the four observables E will always have a value of 1, contrary to the predictions of quantum mechanics. This is the famous GHZ paradox, which is an example of the “all-versus-nothing” paradox and can also be seen as a state-dependent version of the Peres–Mermin square. By adding some single-qubit measurement operators, we can obtain a state-independent proof of quantum contextuality, known as the “magic pentagram”, as shown in [Figure 5.1](#). I provide only a brief introduction to this proof in the figure, and readers familiar with [chapter 2](#) will quickly recognize its subtleties.

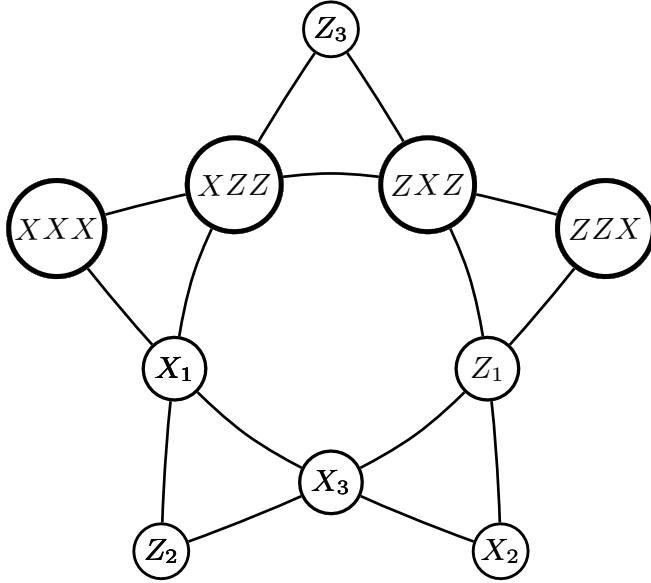


Fig. 5.1 “Magic Pentagram” - A state-independent proof of quantum contextuality. The product of operators connecting each side of the pentagram is $\mathbb{I}^{\otimes 3}$, but according to the analysis of the GHZ paradox above, the product of operators along the horizontal side is -1 in a noncontextual hidden variable model, while the other four sides are $+1$. Therefore, the product of operators along all sides of the “magic pentagram” is -1 in a noncontextual hidden variable theory. However, each operator appears on two different sides, so no matter how the hidden variable values of the operators are adjusted, it is impossible to consistently assign noncontextual hidden variable values to the entire “Magic Pentagram” - the overall product will always be $+1$. Figure taken from the reference [66].

Since its discovery in 1989 [76], the GHZ paradox has sparked great interest in studying the peculiar properties of such multipartite quantum systems. Cabello [141] and Scarani *et al.* [142] further extended this paradox to the linear cluster state of four qubits:

$$|LC_4\rangle = (|0000\rangle + |0011\rangle + |1100\rangle - |1111\rangle)/2.$$

In this case, the four operators that form the “all-versus-nothing” paradox take the following form:

$$\left\{ \begin{array}{ll} [E_1 = \sigma_x^1 \sigma_x^2 \sigma_z^3 \mathbb{I}_2^4] & |LC_4\rangle = + |LC_4\rangle, \\ [E_2 = \sigma_x^1 \sigma_y^2 \sigma_y^3 \sigma_x^4] & |LC_4\rangle = + |LC_4\rangle, \\ [E_3 = \sigma_y^1 \sigma_y^2 \sigma_z^3 \mathbb{I}_2^4] & |LC_4\rangle = - |LC_4\rangle, \\ [E_4 = \sigma_y^1 \sigma_x^2 \sigma_y^3 \sigma_x^4] & |LC_4\rangle = + |LC_4\rangle. \end{array} \right. \quad (5.2)$$

Similar to the previous analysis, it is found that it is impossible to use noncontextual hidden variables to determine the predetermined values of all the operators. Therefore, the “all-versus-nothing” paradox also appears in linear cluster states.

Which quantum states exhibit “all-versus-nothing” type of contextuality? Chen [143] showed that at least 3 qubits are needed to achieve this task, and the three-qubit GHZ state requires starting from

a product state and using at least two two-qubit entangling gates to construct it. This motivates us to search for a representation of quantum states using two-qubit entangling gates. The graph state [80, 144] to be introduced next is precisely such a representation, and the name “linear cluster state” for the aforementioned quantum state comes from the graphical representation using graph states. A graph state corresponds to a quantum state $|G\rangle$ and an undirected graph $G(V, E)$, where V and E are the sets of vertices and edges of the graph, respectively. Each vertex represents a qubit in the maximally superposed state in the computational basis, $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, and each edge corresponds to a CZ operation between the two corresponding qubits. Formally, the correspondence between a graph state and an undirected graph is given by:

$$|G\rangle = \prod_{(i,j) \in E} \text{CZ}^{\{i,j\}} |+\rangle^{\otimes V}.$$

In addition to this definition based on CZ operations along the edges of the graph, there is an equivalent definition based on the vertices of the graph: a graph state is a common $+1$ eigenstate of all its stabilizers. The stabilizers of a graph state are a set of mutually commuting operators defined according to the connectivity of each vertex:

$$S_i = \sigma_x^{(i)} \prod_{\{j|(i,j) \in E\}} \sigma_z^{(j)}.$$

Next, we explain that the quantum states appearing in the two examples above are locally equivalent to graph states. First, consider the complete graph of three qubits, K_3 , which is represented in Figure 5.2. Using the two computational methods given above, the graph state can be explicitly calculated as: $|K_3\rangle = \{1, 1, 1, -1, 1, -1, -1, -1\}^\dagger / 2\sqrt{2}$. However, by using only local unitary operations, we can transform $|K_3\rangle$ into the GHZ state:

$$(\mathbb{I}_2 - i\sigma_x) \otimes (\sigma_y + \sigma_z) \otimes (\mathbb{I}_2 - i\sigma_x) |K_3\rangle = |\text{GHZ}_3\rangle.$$

In fact, all multipartite GHZ states are locally equivalent to graph states represented by fully connected graphs². Hence, the GHZ state exhibits a high degree of contextuality.



Fig. 5.2 Graph state representations of the three-qubit GHZ state and the four-qubit linear cluster state. In order to convert them to the forms in the main text, we only need to apply local operations $(\mathbb{I}_2 - i\sigma_x) \otimes (\sigma_y + \sigma_z) \otimes (\mathbb{I}_2 - i\sigma_x)$ and $\sigma_z \otimes (\mathbb{I}_2 + i\sigma_y) \otimes \mathbb{I}_2 \otimes (\sigma_x + \sigma_z)$ to the two graph states, respectively.

² The standard representation is a star graph, but by applying a local complementing operation, which will be mentioned later, the two representations can be converted between each other. The star graph emphasizes the characteristic of the GHZ state where a σ_z measurement leads to disentanglement, while the fully connected graph better reflects the symmetry of the GHZ state.

Next, consider the path graph of four qubits, P_4 , which is represented in [Figure 5.2](#). The graph state can be explicitly calculated as: $|P_4\rangle = (|0+ +0\rangle + |0+ -1\rangle + |1- -0\rangle + |1- +1\rangle)/2$, where $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. Similar to the complete graph example above, we have:

$$\sigma_z \otimes (\mathbb{I}_2 + i\sigma_y) \otimes \mathbb{I}_2 \otimes (\sigma_x + \sigma_z) |P_4\rangle = |\text{LC}_4\rangle.$$

The name “linear cluster state” comes from the fact that the representation of the path graph is a linear path connecting a cluster of vertices.

Before concluding this section, it is important to note that the “all-versus-nothing” paradox in [Equation \(5.2\)](#) is not *irreducible* [[145](#)]. This is because the measurement of the last qubit has only one measurement setting, which means that at least one noncontextual hidden variable theory can explain the measurement result of this qubit. To understand this issue more intuitively, I apply the $\sqrt{\sigma_y}$ operation to the measurement operator of the third qubit, which results in the exchange of σ_x and $-\sigma_z$. The modified “all-versus-nothing” paradox becomes:

$$\left\{ \begin{array}{ll} [E_1 = \sigma_x^1 \sigma_x^2 \sigma_x^3 \mathbb{I}_2^4] & |\text{LC}_4\rangle = -\sqrt{\sigma_y} |\text{LC}_4\rangle, \\ [E_2 = \sigma_x^1 \sigma_y^2 \sigma_y^3 \sigma_x^4] & |\text{LC}_4\rangle = +\sqrt{\sigma_y} |\text{LC}_4\rangle, \\ [E_3 = \sigma_y^1 \sigma_y^2 \sigma_x^3 \mathbb{I}_2^4] & |\text{LC}_4\rangle = +\sqrt{\sigma_y} |\text{LC}_4\rangle, \\ [E_4 = \sigma_y^1 \sigma_x^2 \sigma_y^3 \sigma_x^4] & |\text{LC}_4\rangle = +\sqrt{\sigma_y} |\text{LC}_4\rangle. \end{array} \right.$$

However, upon comparing with [Equation \(5.1\)](#), it is evident that when the states on both sides of the equation are taken as $\sigma_z^{(1)} |\text{GHZ}_3\rangle \otimes |+\rangle$, all four relations are satisfied. The σ_z in front of the GHZ state is used to give all four equations a coefficient of -1 - this is the technique used in the previous chapter to prove [Proposition 4.1](#). Therefore, the “all-versus-nothing” paradox can be observed with a partially separable state or even a mixture of the two mentioned states. This naturally raises the question: under what circumstances does the “all-versus-nothing” paradox hold only for a specific quantum state? Furthermore, we know that the GHZ paradox is irreducible, while the “all-versus-nothing” paradox of linear cluster states can be reduced to the GHZ paradox. So, if we consider only the “all-versus-nothing” paradox composed of Pauli operators (discussions on the high-dimensional case have been extensively covered in the work by Tang *et al.* [[146](#)]), is there any other irreducible “all-versus-nothing” paradox apart from the GHZ paradox? The next section provides affirmative answers to these two questions.

5.2 A generic construction of “all-versus-nothing” contextuality in graph state

Now we present the construction of an “all-versus-nothing” paradox that is different from the GHZ paradox. The construction is based on the following idea: any operator composed of stabilizer products has $|G\rangle$ as its eigenstate with eigenvalue $+1$. Therefore, we can consider combinations of these operators such that their assignments in a noncontextual hidden variable theory always yield -1 , thus resulting in an “all-versus-nothing” paradox. Now the question is, how can we make the assignment of the noncontextual hidden variable theory differ from the expectation values of the observables in quantum mechanics by a negative sign? I shall start with the form of the GHZ paradox:

Lemma 5.1 *In the graph-state representation of undirected graphs, there exists a subgraph represented by a path graph P_3 of length 3, and when the products of all stabilizers in the graph are written in terms of Pauli operators, there will be a negative sign in front.*

Proof Denote this path graph as 2—1—3. First, for the vertices connected to vertices outside of this subgraph, the Pauli operators contributed by the stabilizers are always σ_z , which commute with each other and do not introduce negative signs. So we only need to focus on these three vertices, and they can only have two connection patterns: linear connection (2—3 disconnected) or complete connection (2—3 connected). For vertex 1, the operator on it is always $\sigma_z \sigma_x \sigma_z = -\sigma_x$, which introduces a negative sign. For the other two vertices, when they are not connected, the operators are $\sigma_x \sigma_z$ and $\sigma_z \sigma_x$, respectively, and the product of their coefficients is always 1; when they are connected, the operators are $\sigma_x \sigma_z \sigma_z$ and $\sigma_z \sigma_z \sigma_x$, respectively, where the two σ_z cancel each other out, and the product of their coefficients is still 1. Finally, when the product of the three stabilizers is expressed as Pauli operators, there is only one negative sign in front of vertex 1. \square

With this lemma, we can consider a star graph in the graph state, choose an odd number of path graphs passing through the central vertex, and combine them with the products of other operators, ensuring that each Pauli operator appears an even number of times. This guarantees that the assignment in a noncontextual hidden variable theory always yields -1 . The following [Theorem 5.1](#) presents our construction method.

Theorem 5.1 *Given a graph state $|G\rangle$ with $|V(G)| = m$ and its adjacency matrix in the undirected graph representation as C , where the first vertex is connected to vertices 2 to n : $C_{1k} = 1$, $k \in \{2, \dots, n\}$. Let S_i denote the stabilizer corresponding to vertex i . Then, for the measurement of the observable $\{E\}$ listed in [Table 5.1](#), an “all-versus-nothing” paradox arises.*

The proof of [Theorem 5.1](#) is rather involved, and interested readers can refer to my published paper for details. The proof follows a similar approach to the construction process and only requires demonstrating that each stabilizer appears an even number of times and that the product form corresponding to the subgraph P_3 appears an odd number of times.

5.2.1 Application: quantum State verification

There is a strong corollary to [Theorem 5.1](#):

Corollary 5.1 *When $n = m$ is odd, the quantum state that leads to the “all-versus-nothing” paradox described in [Theorem 5.1](#) is unique.*

Proof In [Theorem 5.1](#), the $n + 1$ operators, any n of which can be used to regenerate the stabilizers $\{S_1, \dots, S_n\}$, are unique due to the construction method of the graph state. \square

$\{E\} :=$	$\{(-1)^{a_i} S_1^{a_i} S_i S_{i+1} \mid i \in \{2, \dots, n-1\}\}$ $\cup \{S_1\} \cup \{S_1^{a_n} S_j \mid j \in \{2, n\}\}.$
$\{E\} :=$	$\{(-1)^{a_i} S_1^{a_i} S_i S_{i+1} \mid i \in \{2, \dots, n\}\} \cup \{S_1\}.$
Definitions	$a_i = 1 + C_{2n} + \sum_{k=2, k \neq i}^{n-1} C_{k(k+1)},$ $S_1^{a_i} = \begin{cases} S_1, & \text{Mod}[a_i, 2] = 1, \\ \mathbb{I}_2, & \text{Mod}[a_i, 2] = 0, \end{cases}, \quad S_{n+1} = S_2.$

Table 5.1 Construction of “all-versus-nothing” paradoxes in graph states. C is the adjacency matrix of the undirected graph representation of the graph state, and S_i is the stabilizer corresponding to the qubits at the vertex i .

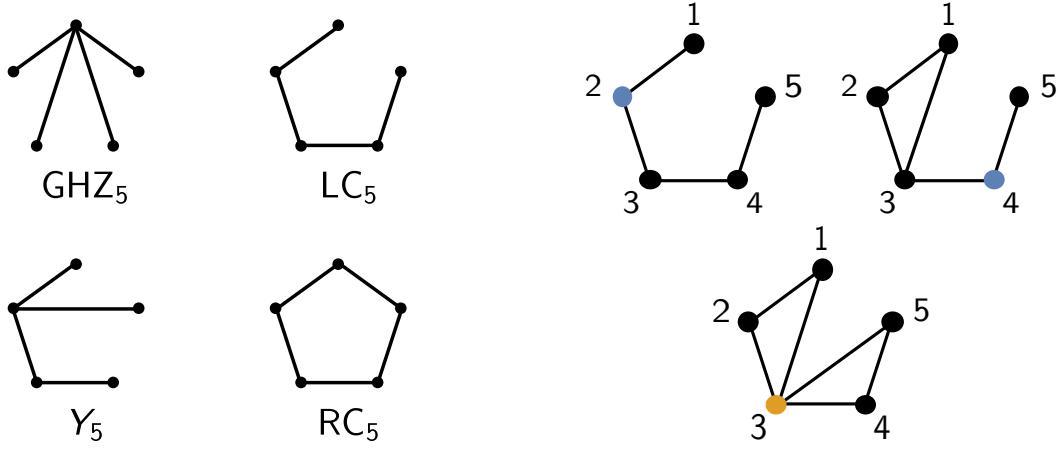


Fig. 5.3 Graph states of 5 qubits, local equivalence, and quantum state verification. Upper left panel: all the inequivalent graph states of 5 qubits under local complementation operations. Upper right panel: a process of local complementation operations, where two local complementations are performed at the blue nodes with the center node, resulting in a graph state that is locally equivalent to $|LC_5\rangle$. It is suitable for constructing the “all-versus-nothing” paradox using [Theorem 5.1](#) and can be verified as a quantum state using [Corollary 5.1](#). Lower panel: the number (in orange-red) and proportion (in purple) of graph states with a central node obtained by local complementation operations for qubit numbers $n \leq 9$ among all graph states.

As an application, we first present a construction method for the “all-versus-nothing” paradox in all graph states of 5 qubits. This construction aims to demonstrate the enormous potential of local complementation operations in conjunction with [Theorem 5.1](#). We first introduce the local complementation [144] operation on graph states. Applying a local complementation operation to a vertex a in the graph state $|G\rangle$ results in a new graph state $|\tau_a(G)\rangle = U_a^\tau(G)|G\rangle$, where the mathematical form is:

$$U_a^\tau(G) = e^{-i\frac{\pi}{4}\sigma_x^{(a)}} e^{i\frac{\pi}{4}\sigma_z^{N(a)}},$$

Here, $N(a) = \{b \in V(G) \mid (a, b) \in E(G)\}$ refers to the set of vertices adjacent to the vertex undergoing the local complementation operation. The effect of the local complementation operation is to reverse the connectivity of the vertices in the subgraph $N(a)$ —thus, it can be proven that star graphs and complete graphs describe the same graph state. Graph states that can be transformed into each other using local complementation operations are locally equivalent, resulting in a significant reduction in the number of independent graph states. The significance of this result in our work is that as long as a graph state can be transformed using local complementation operations to include a central vertex in its representation, we can construct an “all-versus-nothing” paradox that includes all vertices using [Theorem 5.1](#), thereby obtaining a large number of new “all-versus-nothing” paradoxes that are distinct from the GHZ paradox.

Most graph states do not directly include a central vertex. However, through local complementation operations, a high proportion of graph states can be transformed into equivalent forms that include a central vertex, as shown in [Figure 5.3](#). All possible equivalent forms of a graph state can be efficiently obtained using computer programming [[147](#)]. For $m \leq 9$, it can be observed that out of 591 possible graph states, 519 of them can be transformed into a form with a central vertex using local complementation. In the case of $m = 9$, there are 392 different graph states that can be used to construct unique “all-versus-nothing” paradoxes, which can then be used to verify these graph states.

In [Figure 5.3](#), all locally inequivalent graph states of 5 qubits are shown, and there are only four possible forms. They can all be transformed into graphs with a central vertex using local complementation operations, and thus, they can be verified as quantum states based on the “all-versus-nothing” paradox using [Corollary 5.1](#). Here, I provide an example that has not been reported in other literature. By applying local complementation operations to the linear cluster state of 5 qubits to obtain a central vertex, constructing the “all-versus-nothing” paradox, and then using the inverse operation of local complementation to transform both the graph state and the operators in the “all-versus-nothing” paradox back, we obtain:

$$\left\{ \begin{array}{ll} [E_1 = \sigma_z^1 \sigma_y^2 \sigma_x^3 \sigma_y^4 \sigma_z^5] & |LC_5\rangle = -|LC_5\rangle, \\ [E_2 = \mathbb{I}_2^1 \sigma_z^2 \sigma_y^3 \sigma_y^4 \sigma_z^5] & |LC_5\rangle = +|LC_5\rangle, \\ [E_3 = \sigma_z^1 \sigma_y^2 \sigma_y^3 \sigma_z^4 \mathbb{I}_2^5] & |LC_5\rangle = +|LC_5\rangle, \\ [E_4 = \sigma_x^1 \sigma_z^2 \mathbb{I}_2^3 \mathbb{I}_2^4 \mathbb{I}_2^5] & |LC_5\rangle = +|LC_5\rangle, \\ [E_5 = \mathbb{I}_2^1 \mathbb{I}_2^2 \mathbb{I}_2^3 \sigma_z^4 \sigma_x^5] & |LC_5\rangle = +|LC_5\rangle, \\ [E_6 = \sigma_x^1 \mathbb{I}_2^2 \sigma_x^3 \mathbb{I}_2^4 \sigma_x^5] & |LC_5\rangle = +|LC_5\rangle. \end{array} \right. \quad (5.3)$$

This type of “all-versus-nothing” paradox that occurs only in one quantum state of 5 qubits has been used to construct quantum error-correcting codes [[148](#)]. Therefore, our quantum state verification tool provides a new method for constructing quantum error-correcting codes.

5.3 Preparation of four-qubit graph states

Since observing the “all-versus-nothing” paradox requires at least 3 qubits, using a single-polarization-encoded entangled photon source is not sufficient to accomplish the task. Using multiple pairs of photon sources significantly reduces the counting rate. Fortunately, Reference [[149](#)] designed and implemented a polarization-path hybrid encoded two-photon four-qubit entangled photon source, which greatly improves the counting rate of the required quantum states. In this work, we draw inspiration from this mixed encoding technique to prepare all the three-qubit and four-qubit graph states using a unified experimental

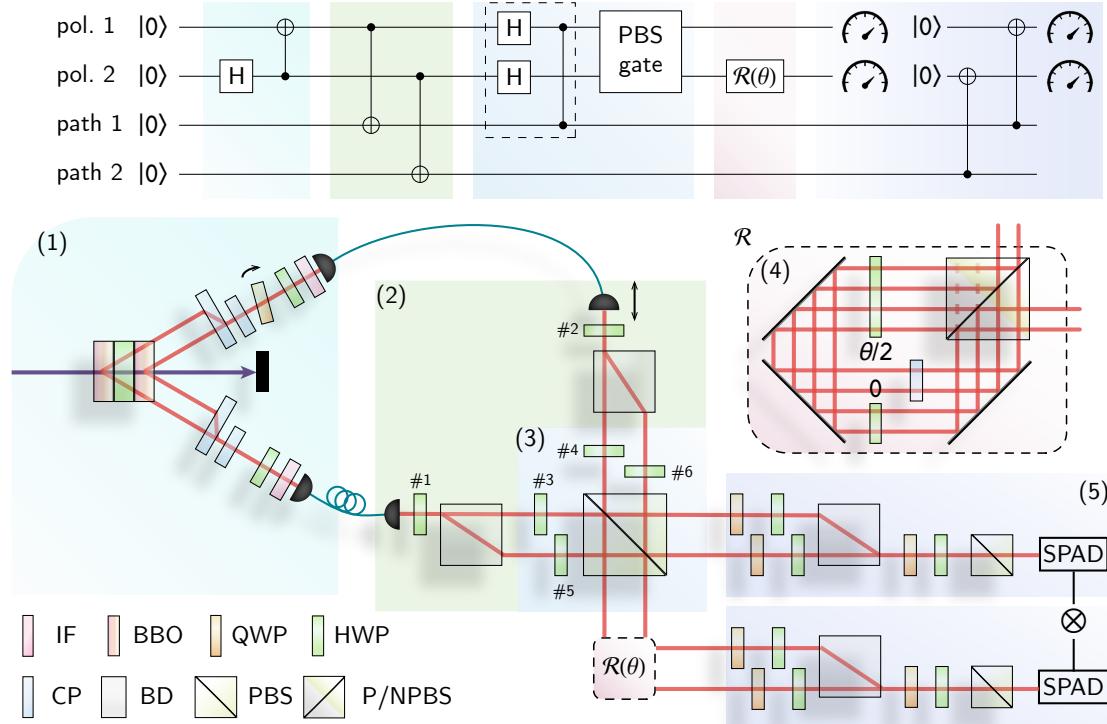


Fig. 5.4 Experimental setup of the quantum circuit for observing the “all-versus-nothing” paradox of graph states. The Top: quantum circuit for generating and measuring the four-qubit linear cluster state $|LC_4\rangle$ and the GHZ state $|GHZ\rangle$ using polarization and path degrees of freedom. The dashed box operations are used solely for generating the linear cluster state. Bottom: optical implementation of the quantum circuit. Each color block from left to right represents: (1) photon source, a pulsed laser pumping a sandwich structure nonlinear crystal to generate two entangled photons in polarization and spatial modes after spatial and temporal compensation; (2) state preparation, where the two photons are further encoded in polarization and path modes; (3) fusion gate; (4) a non-unitary evolution used to prepare more diverse quantum states, not necessary in this case; (5) joint measurement of polarization and path modes, as discussed in chapter 3. IF: interference filter, HWP: half-wave plate, QWP: quarter-wave plate, CP: spatial and temporal compensation crystal, BD: beam displacer, PBS: polarization beam splitter, P/NPBS: custom-made beam splitter, half PBS and half polarization-independent 50:50 beam splitter.

setup. These include the three-qubit and four-qubit GHZ states and the four-qubit linear cluster state, denoted as $|GHZ_3\rangle$, $|GHZ_4\rangle$, and $|LC_4\rangle$, respectively.

I shall first introduce the quantum circuits for state preparation and measurement. We start with a pair of qubits in the maximally entangled state $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ and use two polarization-to-path CNOT gates to obtain the four-qubit GHZ state $|GHZ_4\rangle$. To achieve a three-qubit GHZ state, we only need to perform a projection measurement on any one degree of freedom onto $|+\rangle$, and the remaining three qubits will then collapse into a GHZ state. To generate the linear cluster state $|LC_4\rangle$, we need to implement the CZ operation between two polarization qubits. However, we decompose it into a CZ operation between a polarization-path qubit and a fusion gate between two polarization qubits, which further increases the success rate by more than three times. The effect of the fusion gate is equivalent to projecting the polarization qubits onto $|00\rangle\langle 00| - |11\rangle\langle 11|$, and with some single-qubit operations, the entire system can be prepared into a four-qubit cluster state $|LC_4\rangle$.

Next, we will use linear optics to implement this circuit. The preparation of entangled photon pairs and the joint measurement of polarization and path have been discussed in [chapter 3](#), so we will focus on the most exciting part of the experiment: preparing the four-qubit cluster state using a fusion gate. First, two beam displacers shift the wave function of vertically polarized photons to the lower path, while horizontally polarized photons remain in the upper path. This operation effectively applies a CNOT gate between the two path qubits and the polarization qubit, resulting in the entire two-photon state being a GHZ state. At the center of the setup, the two paths of photons interfere at two different locations of a polarization beam splitter. Here, we again analyze the evolution of the optical quantum state using the language of second quantization. Denote the creation operators for the photons entering the polarization beam splitter from the left and top sides as $\alpha_{\mu\nu}^\dagger$ and $\beta_{\mu\nu}^\dagger$, respectively, where $\mu \in \{H, V\}$ distinguishes the polarization modes of the photons and $\nu \in \{u, d\}$ distinguishes the path modes of the photons. Similarly, denote the creation operators for the photons leaving the polarization beam splitter from the right and bottom sides as $\gamma_{\mu\nu}^\dagger$ and $\delta_{\mu\nu}^\dagger$, respectively. In this case, the following mode transformations occur at the polarization beam splitter:

$$\alpha_{H\nu}^\dagger \rightarrow \gamma_{H\nu}^\dagger, \quad \alpha_{V\nu}^\dagger \rightarrow i\delta_{V\nu}^\dagger, \quad \beta_{H\nu}^\dagger \rightarrow \delta_{H\nu}^\dagger, \quad \beta_{V\nu}^\dagger \rightarrow i\gamma_{V\nu}^\dagger.$$

Furthermore, denote the angles of the half-wave plates #3 ~ #6 in [Figure 5.4](#) as $\theta_3/2 \sim \theta_6/2$. We can explicitly write the wave function of the light field reaching the polarization beam splitter as:

$$\frac{1}{\sqrt{2}} \left[(\cos \theta_3 \alpha_{Hu}^\dagger + \sin \theta_3 \alpha_{Vu}^\dagger) (\cos \theta_4 \beta_{Hu}^\dagger + \sin \theta_4 \beta_{Vu}^\dagger) \right. \\ \left. + (-\sin \theta_5 \alpha_{Hd}^\dagger + \cos \theta_5 \alpha_{Vd}^\dagger) (-\sin \theta_6 \beta_{Hd}^\dagger + \cos \theta_6 \beta_{Vd}^\dagger) \right] |0\rangle.$$

By substituting the mode transformation relationships, we can obtain the wave function of the light field leaving the polarization beam splitter:

$$\frac{1}{\sqrt{2}} \left[(\cos \theta_3 \gamma_{Hu}^\dagger + i \sin \theta_3 \delta_{Vu}^\dagger) (\cos \theta_4 \delta_{Hu}^\dagger + i \sin \theta_4 \gamma_{Vu}^\dagger) \right. \\ \left. + (-\sin \theta_5 \gamma_{Hd}^\dagger + i \cos \theta_5 \delta_{Vd}^\dagger) (-\sin \theta_6 \delta_{Hd}^\dagger + \cos \theta_6 i \gamma_{Vd}^\dagger) \right] |0\rangle.$$

However, since only the output modes that simultaneously include both γ and δ are recorded during coincidence detection, we end up with the following wave function:

$$\frac{1}{\sqrt{2}} \left[(\cos \theta_3 \cos \theta_4 \gamma_{Hu}^\dagger \delta_{Hu}^\dagger - \sin \theta_3 \sin \theta_4 \gamma_{Vu}^\dagger \delta_{Vu}^\dagger) \right. \\ \left. + (\sin \theta_5 \sin \theta_6 \gamma_{Hd}^\dagger \delta_{Hd}^\dagger - \cos \theta_5 \cos \theta_6 \gamma_{Vd}^\dagger \delta_{Vd}^\dagger) \right] |0\rangle \\ = \cos \theta_3 \cos \theta_4 |HHuu\rangle - \sin \theta_3 \sin \theta_4 |VVuu\rangle \\ + \sin \theta_5 \sin \theta_6 |HHdd\rangle - \cos \theta_5 \cos \theta_6 |VVdd\rangle.$$

Thus, the problem of constructing the quantum state is reduced to solving for the angles of the wave plates. This setup can generate highly flexible quantum states. For example, setting all wave plate angles to 0 yields a GHZ state acted upon by σ_z . If we set $\{\theta_3, \theta_4, \theta_5, \theta_6\} = \{45^\circ, -45^\circ, 45^\circ, 45^\circ\}$, we obtain the linear cluster state precisely.

5.4 Observation and applications of the “all-versus-nothing” paradox

Before observing the specific “all-versus-nothing” paradox, we first characterize the operation of the fusion gate involving two paths. Its principle is very similar to the Hong–Ou–Mandel interference, but because the polarization beam splitter is polarization-dependent, a Hadamard operation is applied to the polarization using a half-wave plate to erase the polarization information carried by the photons. For the two paths, we perform the Hong–Ou–Mandel interference experiment separately to verify the temporal and spatial overlap of the two photons arriving at the polarization beam splitter. The measurement results are shown in Figure 5.5, with an average visibility of the two-path interference reaching 97.1%, indicating that the two photons arriving at the polarization beam splitter are highly indistinguishable and exhibit excellent simultaneity.

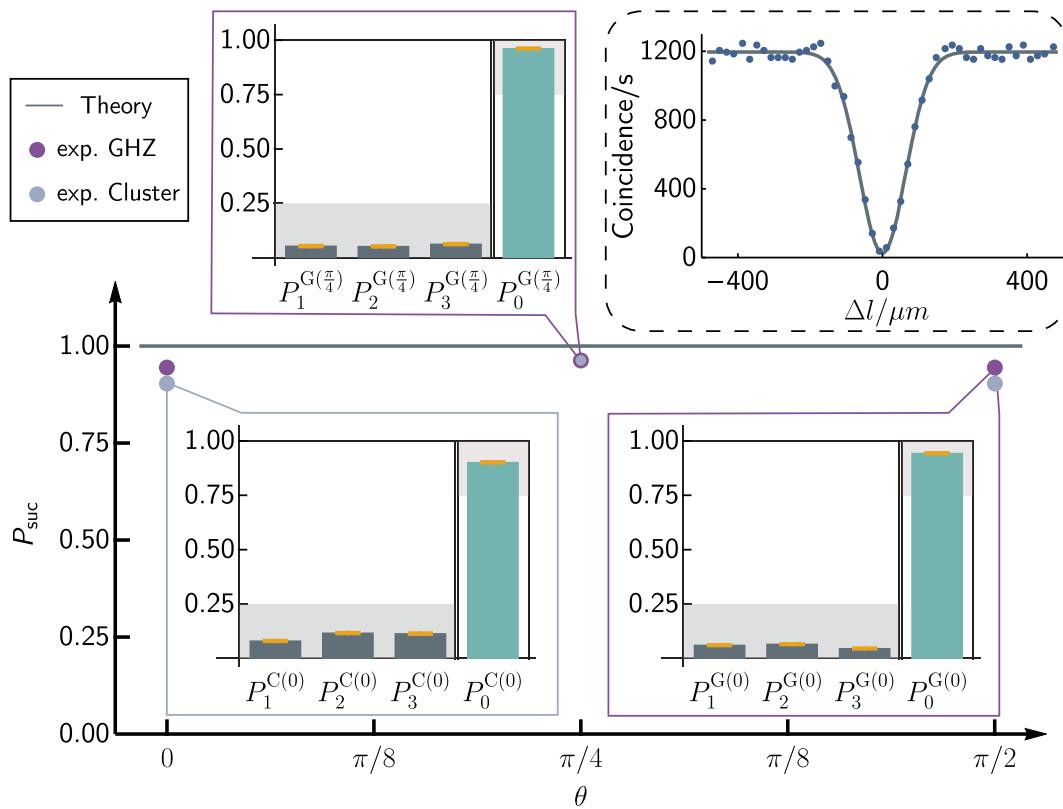


Fig. 5.5 Experimental results of the graph state “all-versus-nothing” paradox. Data points: Success probability of the “all-versus-nothing” paradox. Error bars correspond to 1σ standard deviation calculated using Poisson counting statistics. Gray points correspond to the linear cluster state, and purple points correspond to the GHZ state. Subfigure: Probability of each observed event. The shaded region represents the region that can effectively rule out noncontextual hidden variable theories, given a 75% visibility threshold. Dashed box: Hong–Ou–Mandel interference data for the two photons (visibility of 0.971) and its Gaussian fit curve.

Next, we observe the “all-versus-nothing” paradox for the three types of graph states. Although there may be some errors in the experiment and not all probabilities can be perfectly 0 or 100%, a small amount of imperfection does not affect the persuasiveness of the paradox. By translating the “all-versus-nothing” paradox back to noncontextual hidden variable inequalities, Ryff [150] pointed out that for this “all-versus-nothing” paradox involving four contradictory conditions, an interference visibility of only 75% is sufficient to indicate the existence of contextuality. Our experimental results, as shown in [Figure 5.5](#), not only match the counting rates expected by quantum mechanics but also significantly exceed the critical values of interference visibility: the probabilities of events that would not occur in noncontextual hidden variable models reach 96.2%, 94.4%, and 90.1% for $|\text{GHZ}_3\rangle$, $|\text{GHZ}_4\rangle$, and $|\text{LC}_4\rangle$, respectively, surpassing the critical visibility thresholds by 93.9, 72.7, and 56.0 standard deviations. Since $|\text{GHZ}_3\rangle$ satisfies the conditions of [Corollary 5.1](#), the experimental results can be used to verify that the prepared quantum states should be $|\text{GHZ}_3\rangle$ within a certain accuracy range. A natural question arises: can we infer the fidelity between the prepared quantum states and the target states based on the experimental results? We further investigate this question using the obtained experimental data.

5.4.1 Fidelity estimation and witnessing entanglement

Multiplying all the operators and their corresponding eigenvalues in [Equation \(5.1\)](#) and summing them up, we obtain a noncontextual hidden variable inequality:

$$\langle \mathcal{E} \rangle_{\psi} := \sum_{k=1}^4 \langle f_k E_k \rangle_{\psi} \stackrel{\text{NCHV}}{\leq} 2. \quad (5.4)$$

Here, $f_k := \langle \text{GHZ}_3 | E_k | \text{GHZ}_3 \rangle$ is the eigenvalue corresponding to the operator. We note that due to the quantum state verification in [Corollary 5.1](#), the maximum eigenvalue of $\sum_{k=1}^4 f_k E_k$ must be non-degenerate, and since each stabilizer appears an even number of times in the elements of \mathcal{E} , the second largest eigenvalue of $\sum_{k=1}^4 f_k E_k$ is 0. This implies that any positive value of $\langle \mathcal{E} \rangle_{\psi}$ represents an overlap between the used quantum state $|\psi\rangle$ and the target state $|\text{GHZ}_3\rangle$, and we can easily calculate a lower bound on fidelity based on $\langle \mathcal{E} \rangle_{\psi}$:

$$F(\psi, \text{GHZ}_3) \geq \frac{1}{4} \langle \mathcal{E} \rangle_{\psi}. \quad (5.5)$$

Using experimental data, we obtain $\langle \mathcal{E} \rangle_{\psi^*} = 3.792$ for the experimentally prepared quantum state $|\psi^*\rangle$. By applying [Equation \(5.5\)](#), we find:

$$F(\psi^*, \text{GHZ}_3) \geq 89.6\%. \quad (5.6)$$

Therefore, we can determine the fidelity between the prepared quantum state and the target state without relying on quantum state tomography, significantly reducing resource consumption.

Furthermore, we extend this conclusion to the scenario of quantum state verification with an arbitrary number of parties. Since the spectral characteristics of \mathcal{E} in this case are similar to the case of GHZ states mentioned above, we have the following corollary:

Corollary 5.2 *When $n + 1$ operators that can be verified by the “all-versus-nothing” paradox are multiplied by their eigenvalues under graph states and summed, a noncontextual hidden variable inequality is obtained:*

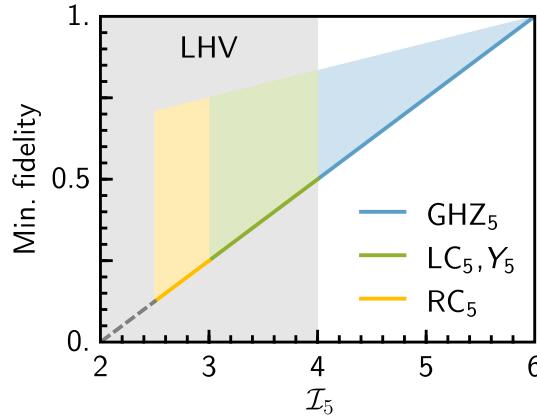


Fig. 5.6 Violation of noncontextual hidden variables and the fidelity between quantum states and target states. The figure shows the case of all graph states for a 5-qubit target state. The colored region indicates the presence of at least two qubits in an entangled state, violating the noncontextual hidden variable inequality, and thus observing genuine multipartite entanglement.

$$I_n = \sum_{k=1}^{n+1} \langle f_k E_k \rangle_{\psi} \stackrel{\text{NCHV}}{\leq} n - 1. \quad (5.7)$$

The value of the inequality obtained and the fidelity between the quantum state and the target state have the following relationship:

$$\frac{1}{4}(I_n - n + 3) \leq F(\psi, G_n) \leq \frac{I_n + n + 1}{2n + 2}. \quad (5.8)$$

Furthermore, the violation of the noncontextual hidden variable inequality (5.7): $I_n \geq n - 1$ leads to:

1. The fidelity between the quantum state and the target state reaches at least 50%.
2. The quantum state cannot be written as a tensor product of any two subsystems.

Proof Equation (5.8) is a result of the spectral characteristics of \mathcal{E} . Note that the second largest and smallest eigenvalues of \mathcal{E} are $n - 3$ and $-n - 1$, respectively. By combining them with the largest eigenvalue and calculating the fidelity between the obtained state and the target state, as well as the violation of inequality (5.7), the desired result is obtained.

For the proof of the last two propositions:

1. It is a direct inference from the fidelity estimation above. In Figure 5.6, the results applied to the four types of 5-qubit graph states are given. Note that these graph states have different Schmidt numbers. Therefore, when the value obtained according to the corresponding inequality is greater than the left boundary of the colored region, it indicates the presence of entanglement in the quantum system—no product state can achieve such a high fidelity with the target state.
2. It is because if a state that can be decomposed into a tensor product form of subsystems is obtained by linearly combining the target state with another eigenstate, the proportion of the target state must not exceed 50%.

Therefore, the above proof process demonstrates the application potential of the “all-versus-nothing” paradox we constructed in fidelity estimation and witnessing entanglement. \square

5.4.2 “All-versus-nothing” paradox for quantum steering

In the previous sections, we have conducted a series of experiments on the “all-versus-nothing” paradox and quantum state verification for the case of odd n . In contrast, the case of even n has not been thoroughly investigated. When the “all-versus-nothing” paradox cannot be used for quantum state verification, does it have any other implications? Chen *et al.* [151] first pointed out that the “all-versus-nothing” paradox is also suitable for studying quantum steering [152], an asymmetric nonclassical feature. This work has been experimentally verified in Reference [153]. In this work, we find that this method can also be extended to multi-qubit graph state systems to observe quantum steering phenomena between an even number of qubits in the form of the “all-versus-nothing” paradox. Since quantum steering has looser requirements on hidden variable models compared to contextuality and nonlocality, our work enables the observation of a wider range of quantum nonclassical features using the “all-versus-nothing” paradox.

We present the construction of a simple case as an example. Consider the four-qubit linear cluster state $|LC_4\rangle$, where the first two particles are held by one observer, Alice, and the last two particles are held by another observer, Bob. Alice can steer Bob’s state, which means that when Alice performs measurements on her particles, Bob’s state cannot be explained as being selected from a local hidden state ensemble [152]. Similar to the formulation in Fine’s theorem, this phenomenon can be described

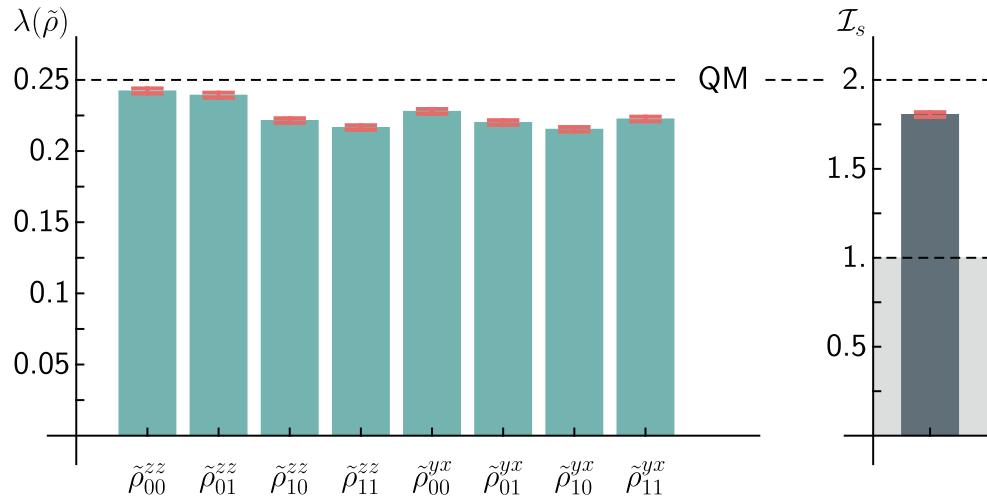


Fig. 5.7 Experimental results of the “all-versus-nothing” paradox with graph states. Left: Main eigenvalues of Bob’s conditional states calculated based on Alice’s measurement choices and results. Right: Contradiction between the calculated steering parameter I_s and the prediction of the local hidden state model $I_s = 1$. The error bars correspond to the 1σ standard deviation calculated through counting statistics.

as the joint probabilities obtained by Alice and Bob's observations cannot be described by the following local hidden variable-hidden state model:

$$\Pr(a, b|x, y) = \int_{\mu} \Pr(a|x, \mu) \text{Tr}(\rho(\mu) \hat{\Pi}_{b,y}) d\mu.$$

In the case of the four-qubit linear cluster state, this local hidden state model would imply that for any two sets of measurements performed by Alice on her two qubits, the sum of the largest eigenvalues of Bob's hidden state would always be 1. However, in practice, Alice can choose measurements that make Bob's state close to a pure state, in which case the sum of the largest eigenvalues of the hidden state would be 2, leading to a contradiction of “2 = 1”.

To achieve this steering paradox experimentally, we set Alice's two measurement bases as $\sigma_z \otimes \sigma_z$ and $\sigma_y \otimes \sigma_x$. Corresponding to these two sets of measurement bases, Bob measures the expectation values of two operators $\sigma_x \sigma_x + \sigma_y \sigma_y$ and $\sigma_y \otimes \sigma_x + \sigma_x \otimes \sigma_y$. Since the spectra of these two operators are {2, 0, 0, -2}, the eigenvalues with larger magnitudes can be used to determine how close Bob's state is to a pure state—this is exactly the technique we used in the fidelity estimation section above. In the experiment, we measured the eigenvalues of eight conditional states as shown in [Figure 5.7](#), and their sum reached 1.805 ± 0.014 , exceeding the upper limit of 59 standard deviations of the local hidden state model, thus achieving the witnessing of quantum steering using the “all-versus-nothing” paradox.

5.5 Summary

By constructing the “all-versus-nothing” paradox, we experimentally observed the presence of nonlocality in graph states. Our theoretical methods include quantum state verification, fidelity estimation, and witnessing of entanglement and quantum steering, which could further develop into methods for constructing novel quantum error-correcting codes. We designed a flexible optical setup in experiments to achieve a large class of four-qubit states, including the preparation and arbitrary measurements of all four-qubit graph states, thus facilitating the observation of the contextuality in the form of the “all-versus-nothing” paradox with high precision. We hope that this research can further expand the “all-versus-nothing” paradox as a novel means of observing contextuality in the foundations of quantum mechanics and quantum computing.

Part III

Contextuality in quantum information science

Chapter 6

Contextuality and pre-post-selection paradoxes: the exchanged grins between quantum Cheshire cats

The research introduced in the previous two chapters derived some characteristics of contextuality from very rigorous mathematical tools and then observed them experimentally. This chapter introduces a work that is completely different from the previous two in terms of ideas. Similar to [chapter 2](#), a cat will be the protagonist of this research. First, I introduce the character of the Cheshire cat from “Alice’s Adventures in Wonderland” [154]. It has the ability to change its appearance and can disappear, leaving only a mischievous grin floating in the air:

‘All right’, said the Cat; and this time it vanished quite slowly, beginning with the end of the tail, and ending with the grin, which remained some time after the rest of it had gone.

‘Well! I’ve often seen a cat without a grin’, thought Alice, ‘but a grin without a cat! It’s the most curious thing I ever saw in my life!’

—*Alice’s Adventures in Wonderland*, by Lewis Carroll



Fig. 6.1 The story of the Cheshire cat. The Cheshire cat is a mysterious creature that can appear and disappear at will. When the Cheshire cat disappears, only a lonely grin remains floating in the air. Figure taken from the movie *Alice in Wonderland*.

Alice, seeing such a scene, cannot help but be surprised. In real life, even if a cat can grin, that grin is still an attribute of the cat. There should not be a mischievous grin that does not belong to the cat itself. Just like in optics and quantum mechanics, polarization is an attribute of photons, and it is very strange to have a photon without polarization or a polarization that is not attached to any photon. However, Aharonov *et al.* [63] designed a thought experiment using quantum ensembles with pre- and post-selection, which led to the polarization of photons appearing in a place where photons do not exist at all. This phenomenon, where matter and its properties are separated in pre- and post-selection systems, is called the quantum Cheshire cat phenomenon.

Once the concept of the quantum Cheshire cat was proposed, it received widespread attention due to its counterintuitive nature and peculiar predictions. It was first observed experimentally by Denkmayr *et al.* [155] using a neutron interferometer. However, the focus of the research on this phenomenon lies in the fact that similar phenomena can be observed using classical light [156] or single-photon interference [157], and the observed phenomena can be fully explained by classical interference theory [158]. Therefore, it is worth questioning why this phenomenon is called the quantum Cheshire cat phenomenon rather than the “classical Cheshire cat” phenomenon. Yu and Oh, who discovered the simplest-state-independent quantum contextuality proof [66], pointed out that the quantum Cheshire cat phenomenon is a manifestation of quantum contextuality in pre- and post-selection systems. Therefore, if contextuality can be observed using classical light or post-selected single photons, it can certainly be observed using classical light or single photons in the case of the quantum Cheshire cat phenomenon. From another perspective, if the entity that makes up the Cheshire cat cannot be described by classical wave theory, then a true quantum Cheshire cat can be created. This is precisely the theoretical work of Das and Pati [159] and the starting point of our experimental work. However, another major contribution of our experimental work is the development of a method for extracting the real part of weak values based on imaginary-time evolution, which greatly reduces the quantum resource consumption of weak measurements.

Our experimental work on the exchange of mischievous grins between two quantum Cheshire cats has been summarized in a research highlight story written by Prof. Xiao-Ye Xu and others (in Chinese). The content is rich and detailed, so I recommend that readers who are only interested in the exchange of mischievous grins between Cheshire cats directly read the research highlight. In this paper, the content of this work will not be presented in the order of experimental design. Instead, I will use contextuality as the core supporting the theoretical work and imaginary-time evolution as the focus of the experimental introduction. I hope to tell a different story of the quantum Cheshire cat from these two new perspectives.

6.1 Tracking the Evolution Process of Quantum Systems

In the world of the Cheshire cat, people are interested in the path taken by this magical creature. For example, if we know that the Cheshire cat traveled from Stockholm to Copenhagen today, we would be interested in whether the Cat traveled directly from the Baltic Sea to its destination, or if it took a detour through Linkoping and Malmo, or perhaps it actually went all the way across the Atlantic to Ann Arbor. In the world of quantum Cheshire cats, a similar situation arises. People are interested in understanding what happens to a quantum system during its evolution when the initial and final states are known. If one waits at a certain point, will they catch a glimpse of the Cheshire cat? Such systems, with known initial and final states, are referred to as pre- and post-selected systems. Studying them requires the use of mathematical language described by two-state vectors, and the trace of the Cheshire cat is determined by a quantity known as weak value.

6.1.1 Two-state vectors and weak values

Assume that a quantum system is prepared in the initial state $|\psi\rangle$ and measured to be in the final state $|\phi\rangle$. We denote the description of this process using two-state vectors as $\langle\phi|\cdot|\psi\rangle$. During the process, the state of the quantum system depends simultaneously on its past state ψ and its future state ϕ . At first glance, this seems to violate special relativity, as the future measurement determines the current state of the system. However, it is important to note that the establishment of two-state vectors requires knowledge of the system's final state. Therefore, all the results are obtained after the evolution has already finished, selecting a successful subset of ensembles that have evolved to the final state through post-selection, while discarding another subset of ensembles that have failed to evolve, thus altering the events that may have occurred during the evolution process. Thus, two-state vectors provide a retrocausal inference of the already occurred evolution, offering an illustrative description of the past evolution.

Weak values are a mathematical language used to describe the already occurred evolution. Loosely speaking, the weak value of an observable can be understood as the average value of that observable during the evolution process described by two-state vectors. Formally, the weak value of an observable \hat{O} under the two-state vector description $\langle\phi|\cdot|\psi\rangle$ [111] is defined as:

$$\langle\hat{O}\rangle_w = \frac{\langle\phi|\hat{O}|\psi\rangle}{\langle\phi|\psi\rangle}.$$

From an operational point of view, the weak value of an operator can be seen as a measure of the strength of the coupling between the operator and the evolving system [112]. Therefore, if a quantum system is coupled to a pointer state using an operator, the rate of change of the pointer state during the coupling is proportional to the weak value. However, weak values possess peculiar properties. For example, the weak value of a Hermitian operator can be imaginary, and its range of values extends far beyond the operator's own spectrum. In the field of optics, this effect was first demonstrated by Pryde *et al.* [160] and others, who measured weak values of a photon's σ_z approaching 20 under nearly orthogonal pre- and post-selection conditions. How can weak values, which are not even observables, correspond to physical reality? The next section aims to explain the connection between them.

6.1.2 An example: photon, where have you been?

Since weak values can characterize the coupling between observables and biorthogonal vectors during the evolution process, when a projection operator is used as the observable, its weak value at a certain point in the evolution represents whether the biorthogonal vector has reached that point. Here, “reaching a point” means that the specific position is simultaneously supported by the initial and final wave functions. To illustrate this concept, consider the famous experiment by Danan *et al.* [161] that tracks a pre- and post-selected photon's trajectory in a Mach–Zehnder interferometer.

We start by considering a standard Mach–Zehnder interferometer as shown in Figure 6.2. A mirror system is added to each arm of the interferometer. By introducing a periodic oscillating signal with a certain frequency to the mirrors, the collection efficiency can be periodically changed by tilting, resulting in power variations recorded by the detector due to this modulation. Conversely, if the detector records a power variation at a frequency that matches the oscillation frequency of a particular mirror, it indicates that the mirror at that position affects the photon—meaning that the photon has reached that location. On the other hand, if the power spectrum at a corresponding mirror frequency is 0, it means that the photon has not reached the labeled position of that mirror. This is consistent with the observations in Figure 6.2:

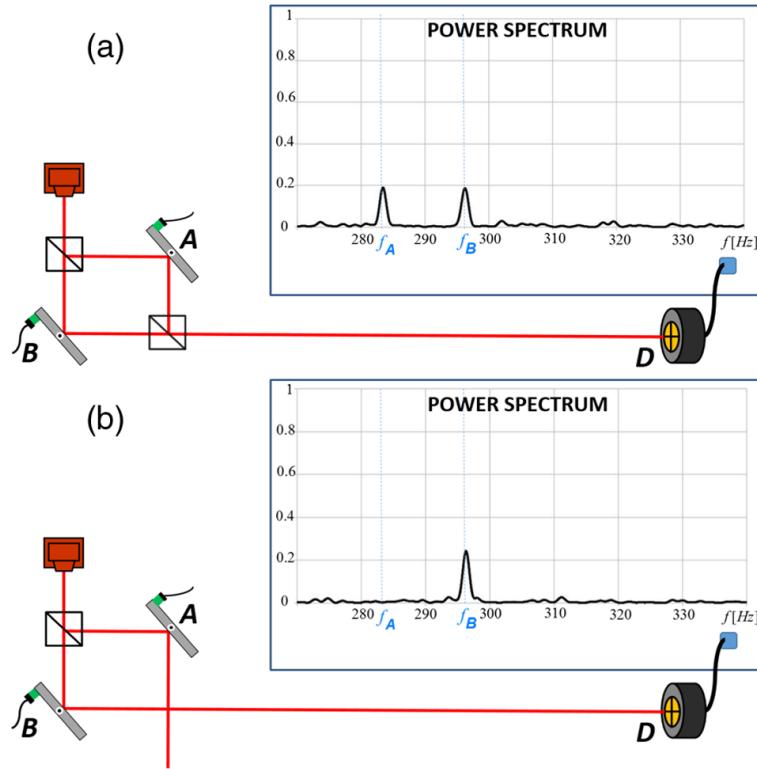


Fig. 6.2 Tracking the trajectory of a photon, part one. The mirror system can mark the positions where the photon has arrived. The power meter at the output reveals that the photon simultaneously passes through both arms of the Mach–Zehnder interferometer. If the beamsplitter that closes the interferometer is removed, only the power modulation corresponding to the mirror facing the detector is recorded. Figure taken from reference [161], copyright American Physical Society.

when the interferometer is adjusted to maximize the output power, the power spectrum observed by the detector shows non-zero values at the frequencies corresponding to both mirrors. When the final 50:50 beamsplitter of the interferometer is removed, the power spectrum only contains the frequency component corresponding to the mirror facing the input detector.

Now we use weak values to explain the observed phenomena. We use simple labeled states and biorthogonal vectors. For example, $|A\rangle$ represents the photon wave function at location A. Corresponding to the upper panel of the figure, the initial and final states of the biorthogonal vectors are both $|\psi\rangle = |\phi\rangle = (|A\rangle + |B\rangle)/\sqrt{2}$. The photons at locations A and B correspond to the projection operators $\Pi_A = |A\rangle \langle A|$ and $\Pi_B = |B\rangle \langle B|$. According to the definition of weak values, we have:

$$\langle \hat{\Pi}_A \rangle = \frac{1}{2}, \quad \langle \hat{\Pi}_B \rangle = \frac{1}{2}.$$

This means that each photon always passes through both mirrors. For the lower panel of the figure, the initial and final states of the biorthogonal vectors are $|\psi\rangle = (|A\rangle + |B\rangle)/\sqrt{2}$ and $|\phi\rangle = |B\rangle$, respectively. In this case, we have:

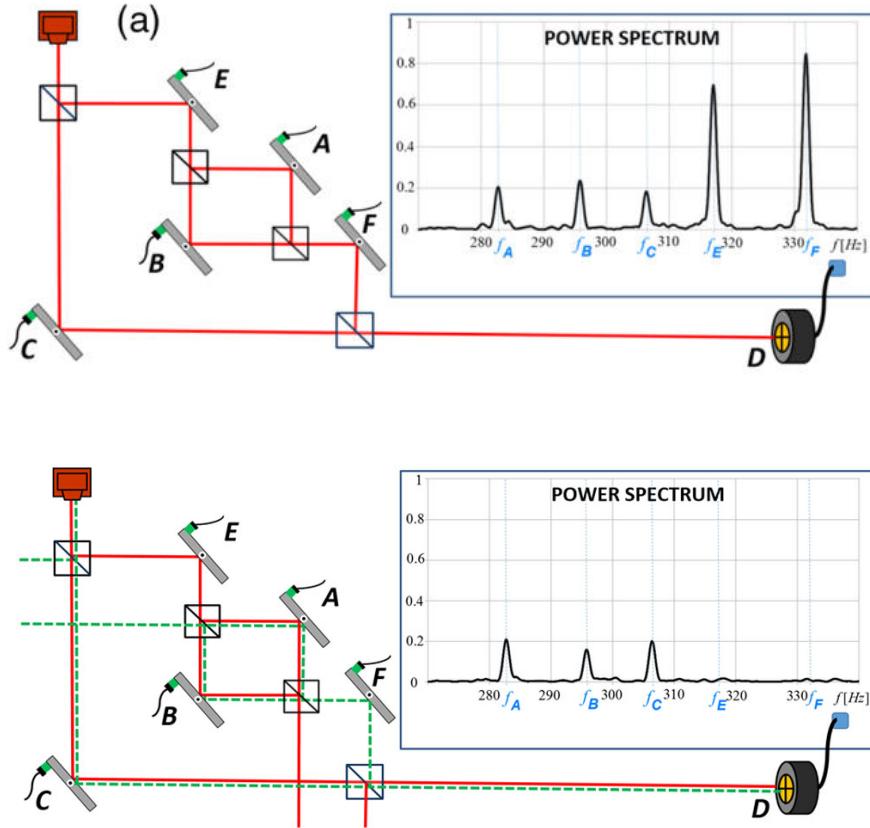


Fig. 6.3 Tracking the trajectory of a photon, part two. Using a cascaded Mach–Zehnder interferometer to demonstrate the peculiar trajectory of photons in the interferometer. In the lower panel, the photon appears to have not passed through locations E and F but reaches locations A and B , indicating a discontinuous photon trajectory. Figure adapted from reference [161], copyright American Physical Society.

$$\langle \hat{\Pi}_A \rangle = 0, \quad \langle \hat{\Pi}_B \rangle = 1.$$

This means that the photon can no longer reach the receiving end via mirror A , and all the photons that arrive come from B . Thus, we have demonstrated in a very straightforward way the relationship between weak values and photon trajectories: if the weak value is 1, it means the photon has completely reached the final state via a certain location; if the weak value is 0, it means the photons that reach the final state do not pass through that location; if the weak value is between 0 and 1, it means that the photons reach the final state not only through that location but also through other paths.

Next, we proceed to consider the fascinating example shown in [Figure 6.3](#), where a Mach–Zehnder interferometer is nested within another Mach–Zehnder interferometer. In the upper panel, the inserted interferometer does not have any effect, and both interferometers result in the photons completely reaching the detector. As expected, we see that the modulation of all mirrors is again recorded in the final power spectrum.

In the lower panel, the phase settings of the interferometer are different. Compared to the upper panel, only the phase between the paths A and B of the inner interferometer is adjusted, causing the photons to be completely blocked and no longer reach the detector. The other phase settings remain unchanged. A peculiar phenomenon occurs: although the photons reaching the detector now exclusively come from C , the recorded spectrum still contains contributions from A and B . What's even more interesting is that if the photons travel from the initial state to the final state via A and B , they must pass through two mirrors E and F on the necessary path. However, the power modulation of these two mirrors is not recorded in the final result. In other words, the photons seemingly teleport from locations E and F and arrive at the inner interference loop composed of A and B . Thus, a highly counterintuitive phenomenon is constructed.

To solve the puzzle, we need to examine the context and understand the paradox of the pre-post selection quantum system. Using weak values provides a possible explanation. Once again, I write down the initial and final states of the photons for the lower panel case:

$$\begin{cases} |\psi\rangle = \frac{1}{\sqrt{3}}(|C\rangle + \sqrt{2}|E\rangle) = \frac{1}{\sqrt{3}}(|C\rangle + |A\rangle + |B\rangle), \\ |\phi\rangle = \frac{1}{\sqrt{3}}(|C\rangle + \sqrt{2}|F\rangle) = \frac{1}{\sqrt{3}}(|C\rangle + |A\rangle - |B\rangle). \end{cases}$$

By substituting these values into the definition of weak values, we calculate:

$$\langle \hat{\Pi}_A \rangle = 1, \quad \langle \hat{\Pi}_B \rangle = -1, \quad \langle \hat{\Pi}_C \rangle = 1, \quad \langle \hat{\Pi}_E \rangle = 0, \quad \langle \hat{\Pi}_F \rangle = 0.$$

This explains the strange phenomenon observed in the power spectrum. The photons seem to “circle” inside the internal interferometer. The photons that travel from A contribute to the final state, but they are canceled out by the photons that “return” from B to the initial state. Thus, it is reasonable that the photons do not pass through E and F . It is important to note that the spectrum of projection operators is binary, with values $\{0, 1\}$. The value $\langle \hat{\Pi}_B \rangle = -1$ is a singular weak value that falls outside the spectrum of the projection operator. This leads to the fascinating behavior in the pre-post selection system [162].

Furthermore, it has been proposed that we can imagine a photon, as shown by the green dashed line in [Figure 6.3](#), going from the detector back to the laser. The overlapping region of the forward and backward paths represents the actual region that the photon passes through. This visualization is consistent with the predictions provided by weak values.

Of course, the explanations given above are based on a two-state vector description, which may be more familiar to readers familiar with field theory. Readers focusing on quantum information may consider this description as an application of Occam’s razor—“Entities should not be multiplied without necessity”. How can we interpret it according to the standard description of quantum mechanics? According to the standard interpretation of quantum mechanics, there is no photon returning to the initial state from B . The observed phenomena are all results of single-photon interference. Considering only the case of the lower panel in [Figure 6.3](#), the reason why the photons reaching the detector via E and F are blocked is that the interference phase between A and B causes the photons to deviate from the path. Adjusting the angles of the mirrors at E and F cannot change the destructive interference inside the interferometer. Therefore, the recorded power fluctuations are caused by the modulation of A and B , which disrupt the conditions for interference cancellation. Thus, we have recovered the conclusion from the two-state vector description using the standard description of quantum mechanics.

From the above discussion, we can see that the two-state vector description provides an intuitive and convenient understanding of the pre-post selection quantum system, but it is not a necessary interpretation. The paradox based on the two-state vector description reflects the contradiction between quantum systems and classical cognition or physical intuition. The question that follows is whether it can be elevated

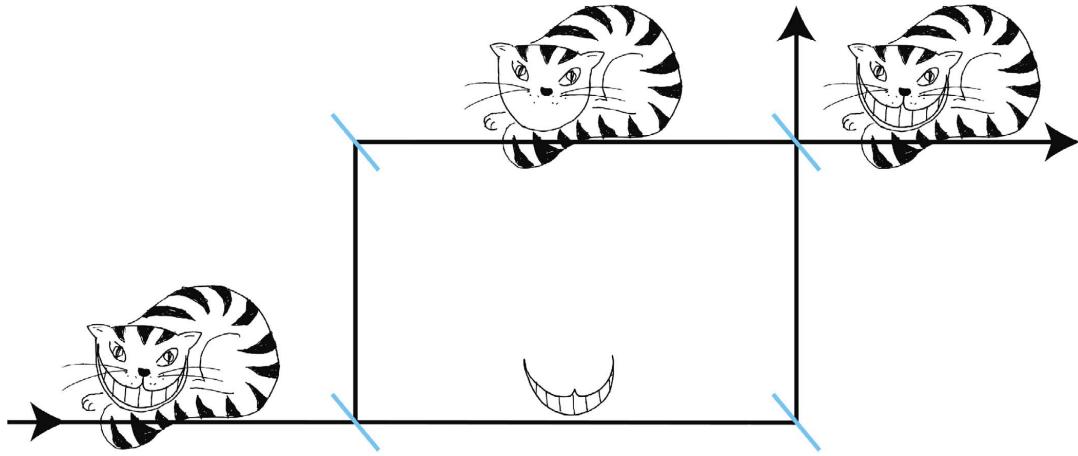


Fig. 6.4 Conceptual illustration of the quantum Cheshire cat. In a Mach–Zehnder interferometer, the Cheshire cat and its grin progress along two different paths, resulting in the grin appearing in a place where the Cheshire cat itself is entirely absent. Figure taken from [155] under a Creative Commons Attribution 4.0 International License.

to a quantitative mathematical theory similar to the contradiction between quantum mechanics and noncontextual hidden variable models. Next, we will take the example of the quantum Cheshire cat to introduce how the pre-post selection paradox is actually a contextuality paradox.

6.2 Contextuality and the quantum Cheshire cat paradox

Now that we have introduced the two-state vector description and weak values, we can apply them to investigate the quantum Cheshire cat paradox mentioned at the beginning of this chapter. An illustrative depiction is shown in Figure 6.4, where the Cheshire cat and its grin progress along two different paths, resulting in the grin appearing in a place where the Cheshire cat itself is entirely absent. I shall first explain how this paradox is realized in the thought experiment proposed by Aharonov *et al.* [63]. In this case, the Cheshire cat corresponds to a photon, while the grin corresponds to the photon’s polarization. We will see that by appropriately choosing the two-state vectors for pre-post selection, we can separate the photon and its polarization, just like the Cheshire cat and its grin are separated in two different spatial regions.

6.2.1 Where is the grin of the quantum Cheshire cat?

We will use the polarization-path composite system introduced in Chapter 3 to describe the experimental setup for achieving the quantum Cheshire cat effect. We denote the state of a photon as $|\cdot\rangle_{\text{pol}}|\cdot\rangle_{\text{path}}$,

where the first qubit represents polarization and the second qubit represents the path, which can take values $|L\rangle$ or $|R\rangle$, indicating which side of the Mach–Zehnder interferometer the photon goes through.

To achieve the quantum Cheshire cat effect, we can use the linear optical setup shown in [Figure 6.5](#). For the pre-selection (state preparation) part, we input a horizontally polarized photon, and the principles of the other operations are self-evident. The prepared quantum state is:

$$|\psi\rangle = |H\rangle_{\text{pol}} (|L\rangle + |R\rangle)_{\text{path}}/\sqrt{2}.$$

The back selection part is slightly more complex to implement. The principle of back selection is to select the photons that trigger detector D_1 . For these photons, before passing through the polarization beam splitter, they should be in the polarization state $|H\rangle_{\text{pol}}$. Thus, before BS_2 , the state of these photons should be $|H\rangle_{\text{pol}} (|L\rangle + |R\rangle)_{\text{path}}/\sqrt{2}$. However, there is a half-wave plate on the right side, which performs the σ_x operation on the polarization. Therefore, the photons reaching the detector should be in the following state at the beginning of the back selection:

$$|\phi\rangle = \frac{1}{\sqrt{2}} (|H\rangle_{\text{pol}} |L\rangle_{\text{path}} + |V\rangle_{\text{pol}} |R\rangle_{\text{path}}).$$

Now, the question is: During this two-state vector evolution, which paths do the photon and its polarization traverse in the Mach–Zehnder interferometer?

Since we are studying a pre-post selection system, it is convenient to use weak values for the description. First, I define operators that can trace the quantum Cheshire cat (photon) and its grin (polarization). For the photon itself, we still use the projection operator $\Pi_\nu = |\nu\rangle\langle\nu|$, $\nu \in \{L, R\}$. For the polarization, since determining the polarization also involves a path, we use the tensor product of the polarization Pauli operator and the projection operator to represent the polarization's location. Specifically, the observable chosen in this case is $\sigma_x \otimes \Pi_\nu$, $\nu \in \{L, R\}$. With these two operators, we can calculate the weak values by substituting the two-state vectors and determining the photon's position.

I first consider the position of the photon itself. The calculation results are:

$$\langle \Pi_L \rangle_w = 1, \quad \langle \Pi_R \rangle_w = 0.$$

In layman's terms, the photon that reaches the final state must have taken the left path in the interferometer.

Next, I consider the position of the photon's polarization. The calculation results are:

$$\langle \sigma_x \otimes \Pi_L \rangle_w = 0, \quad \langle \sigma_x \otimes \Pi_R \rangle_w = 1.$$

Again, this can be simply explained as follows: the polarization σ_x of the photon reaching the final state must come from the right path of the interferometer. However, an anomalous phenomenon occurs: this means that the photon and the polarization have separated in the interferometer, and on the right side of the interferometer, there is a spin that is not attached to the photon, which is a grin without the Cheshire cat!

6.2.2 Finding the quantum Cheshire cat from the Peres–Mermin square

In the previous section, we have presented Aharonov's thought experiment, providing a complete description of the quantum Cheshire cat concept. However, it only remained at the level of two-state vector description and counterintuitive observations. Even without relying on concepts such as weak values, it is still possible to perfectly explain the observed phenomenon using classical electromagnetic wave

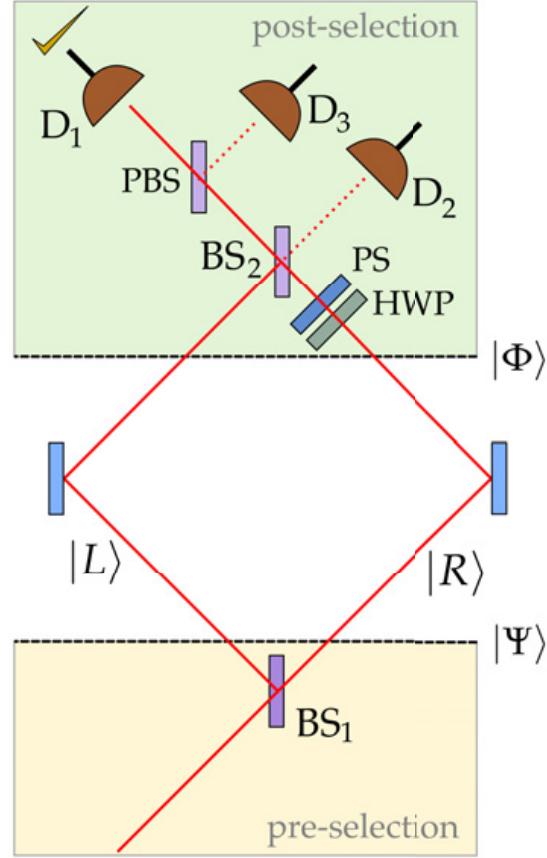


Fig. 6.5 Thought experiment setup for the quantum Cheshire cat. In a Mach–Zehnder interferometer, by selecting appropriate two-state vectors for pre-post selection, the photon and spin progress along two different paths, resulting in a spin without the presence of the photon and a photon appearing out of nowhere. PS: Phase shifter, HWP: Half-wave plate, BS: 50:50 beam splitter, PBS: Polarizing beam splitter. Figure taken from [63] under a Creative Commons Attribution 3.0 International License.

interference. So, what kind of new knowledge does the phenomenon of the quantum Cheshire cat bring? In this section, we will demonstrate how to connect the counterintuitive paradox of the quantum Cheshire cat with the contradiction between contextuality and noncontextual hidden variable theories.

Let us reconsider the variables that appear in the paradox of the quantum Cheshire cat and eliminate the disguise of the Cheshire cat from their relationships, reducing it to a contextuality paradox. Specifically, we need to consider which quantities have definite values for the initial state and final state, respectively. Now, without considering a specific physical system, we directly change the subscripts for the corresponding polarization and path to 1 and 2. Since the variables defined for weak values are either the projection operators for the path or the Pauli operators for the polarization, for convenience, I define their mappings to the computational basis as $|H\rangle_{\text{pol}} \leftrightarrow |0\rangle_{1,2} \leftrightarrow |L\rangle_{\text{path}}$, $|V\rangle_{\text{pol}} \leftrightarrow |1\rangle_{1,2} \leftrightarrow |R\rangle_{\text{path}}$. This way, we can conveniently perform operations using Pauli operators. For the initial state, we can write down the eigenvalues of all the Pauli operators in their respective eigenstates:

$$|\psi\rangle = |0\rangle(|0\rangle + |1\rangle)/\sqrt{2} : \sigma_z^{(1)} |\psi\rangle = +|\psi\rangle, \sigma_x^{(2)} |\psi\rangle = +|\psi\rangle, \sigma_z^{(1)} \sigma_x^{(2)} |\psi\rangle = +|\psi\rangle. \quad (6.1)$$

As the final state is a Bell state, it is not an eigenstate of any of these operators. This is quite reasonable since if the initial state and final state were eigenstates of the same operator with different eigenvalues, they would be orthogonal, and thus the projection probability would be 0. However, we can rewrite the formulas in terms of projection operators:

$$\begin{aligned} \langle \phi | \Pi_x^{-(1)} | \psi \rangle &= 0, \\ \langle \phi | \Pi_z^{-(2)} | \psi \rangle &= 0, \\ \langle \phi | \Pi_x^{+(1)} \Pi_z^{+(2)} + \Pi_x^{-(1)} \Pi_z^{-(2)} | \psi \rangle &= 0. \end{aligned} \quad (6.2)$$

where $\Pi_\kappa^\pm = (\mathbb{I}_2 \pm \sigma_\kappa)/2$. These operators correspond to the first row of the Peres–Mermin square (cf. Figure 2.3). This indicates that the following three statements hold for the photons that reach the final state:

- The polarization of the photons must be $\sigma_x = +1$.
- The photons must come from the left path $|L\rangle$.
- The polarization and the path of the photons are “anti-correlated”: if the photons come from the left path $|L\rangle$, then the polarization must be $\sigma_x = -1$; if the polarization is $\sigma_x = +1$, then the photons must come from the right path $|R\rangle$.

However, we can immediately see that when the polarization is attached to the photons, the three statements are contradictory: it is impossible for a polarization and path state to simultaneously satisfy all three requirements. The quantum Cheshire cat is a compromise between these three statements, considering that the polarization can exist independently of the photons in a counterintuitive way, so that all the predictions in Equation (6.2) can be compatible.

From the perspective of contextuality, Equation (6.2) clearly reflects the contradiction between quantum mechanics and noncontextual hidden variable theories. If we were to describe the system using a noncontextual hidden variable theory, then for the binary measurements Π_κ^\pm , the result must be one of the two possibilities, meaning that the response functions satisfy $v(\Pi_\kappa^+) + v(\Pi_\kappa^-) = 1$. According to the description of noncontextual hidden variable theories,

$$\begin{aligned} \langle \phi | \Pi_x^{-(1)} | \psi \rangle = v(\Pi_x^{-(1)}) \langle \phi | \psi \rangle &= 0 \Rightarrow v(\Pi_x^{-(1)}) \equiv 0, \\ \langle \phi | \Pi_z^{-(2)} | \psi \rangle = v(\Pi_z^{-(2)}) \langle \phi | \psi \rangle &= 0 \Rightarrow v(\Pi_z^{-(2)}) \equiv 0, \end{aligned} \quad (6.3)$$

the response functions of the other two operators that are dual to the observable in the weak values, $v(\Pi_x^{+(1)})$ and $v(\Pi_z^{+(2)})$, must always be 1. However, this implies that

$$\langle \phi | \Pi_x^{+(1)} \Pi_z^{+(2)} | \psi \rangle = \langle \phi | \psi \rangle = \frac{1}{\sqrt{2}} \neq 0,$$

which contradicts the last conclusion in Equation (2). Thus, we have transformed the quantum Cheshire cat paradox into a state-dependent contextuality paradox within the Peres–Mermin square.

6.2.3 “Quantum” Cheshire cat?

As mentioned earlier, all the observable predictions in the quantum Cheshire cat experiment can be reproduced using classical electromagnetic wave theory. Atherton *et al.* [156] provide a remarkable example. They conducted the experiment using lasers instead of individual quantum objects and observed the phenomenon of laser “separation” from its polarization. Their measurement principle involved using neutral density filters and wave plates to change the intensity and polarization of the beam. If changes made in a particular path had no effect on the final detected power, it indicated that the laser or its polarization did not pass through that position. The experimental results showed that all the lasers reaching the detector came from the left path, while the polarizations reaching the detector came exclusively from the right path.

The study of the “classical” Cheshire cat experiment actually has profound implications for both weak values and quantum contextual features. It should be noted that weak values are merely a language, a descriptive method for phenomena involving pre- and post-selection. They are not physical quantities, nor are they intrinsic characteristics of quantum mechanics. For classical systems involving pre- and post-selection, weak values can also be defined similarly, and peculiar effects can be observed [163]. Moreover, even the measurement methods used in this experiment can be described by the system of imaginary-time evolution introduced in Chapter 3. The case of the “classical” Cheshire cat demonstrates that if two systems are governed by the same mathematical laws, they should exhibit the same behavior. On the other hand, to construct a genuine “quantum” Cheshire cat experiment, it is necessary to incorporate at least one experimental step that has no classical correspondence. One approach is to use double-photon interference for weak measurements in the measurement process, a technique realized by Kim *et al.* [164]. Another approach is to use two entangled Cheshire cats in the preparation process to create a quantum system without a classical description, which is the theoretical approach pursued in our work.

6.3 Quantum Cheshire cats' exchanged grins

There are several clear motivations for constructing a genuine quantum Cheshire cat experiment without classical correspondence. Theoretically, the separation of multiple quantum objects from their properties could lead to even more peculiar phenomena. For example, if it were possible to exchange the properties of two particles out of thin air, it would have a stronger effect than entanglement swapping [165]. It is also counterintuitive for quantum objects to acquire properties that do not exist within themselves from other objects. From an experimental perspective, observing the phenomenon of multiple quantum objects separating from and exchanging with their properties presents technical challenges that drive the development of new technologies. From an application standpoint, the ability to separate the properties of multiple objects can contribute to achieving noise-immune communication. For instance, if the wave-particle duality of photons is separated [166], it is possible to achieve robustness against phase noise and intensity noise separately. Here, I present the experimental design and results of implementing the quantum Cheshire cat's grinning swap based on the theoretical work of Das and Pati [159].

6.3.1 Observable and pre- and post-selection states

I will start by introducing the experimental design based on a conceptual cartoon. As shown in Figure 6.8, two quantum Cheshire cats can choose to enter a one-way channel from the upper path ($u = \text{up}$) or the

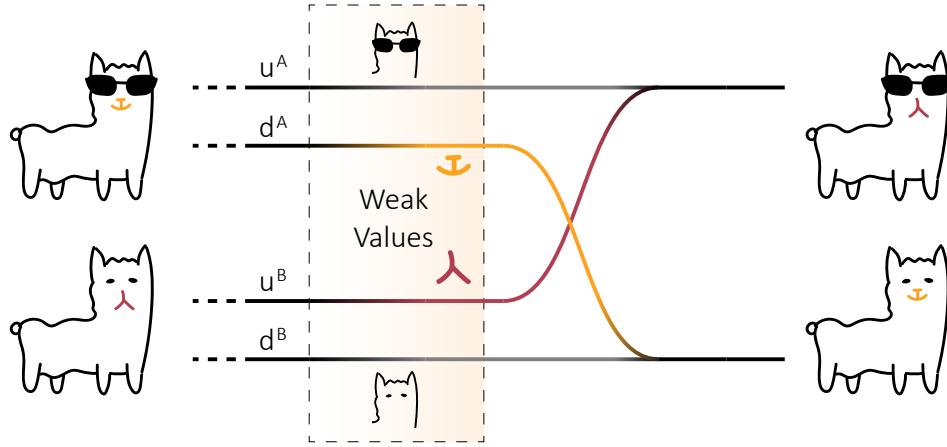


Fig. 6.6 Cartoon illustration of the quantum Cheshire cat grinning swap. The two Cheshire cats enter a two-input-one-output one-way channel, but their grinning is separated and directed to the other cat, resulting in an exchange of grinning between the two Cheshire cats. The separation of grinning from the body can be witnessed using weak values.

lower path (d = down). As Cheshire cats, they prefer to separate their grinning from themselves and each choose a separate channel to enter the device. However, the consequence is that when they leave the device, their own grinning is transformed into the other Cheshire cat's grinning. This is because after leaving the region that represents the separation of their grinning using weak values, the path containing the grinning of each Cheshire cat is redirected to the other Cheshire cat.

Now I will show how we can achieve the above fairy tale in a physical system by choosing suitable observables and pre- and post-selection states. We still use a composite system of photon polarization and path, denoted as $|\cdot\rangle_{\text{pol}}|\cdot\rangle_{\text{path}}$, to represent the states of the two Cheshire cats. The observable for photon position is given by $\langle \mathbb{I}_2 \otimes \Pi_\mu^\nu \rangle$, where $\mu \in \{u, d\}$ and $\nu \in \{A, B\}$, representing that photon ν is in path μ . The observable for photon polarization is $\langle \sigma_z \otimes \Pi_\mu^\nu \rangle$, where $\mu \in \{u, d\}$ and $\nu \in \{A, B\}$, representing that the polarization σ_z of photon ν is in path μ . The correspondence between physical and computational basis states is given by $|H\rangle_{\text{pol}} \leftrightarrow |0\rangle_{1,2,3,4} \leftrightarrow |u\rangle_{\text{path}}$ and $|V\rangle_{\text{pol}} \leftrightarrow |1\rangle_{1,2,3,4} \leftrightarrow |d\rangle_{\text{path}}$. Our device needs to achieve two functions: first, to separate the grinning of the two Cheshire cats from themselves, and second, to implement the exchange process such that each Cheshire cat merges with the previously unassociated grinning. For the first purpose, we choose the pre- and post-selection states to be of the form $|\cdot\rangle_{\text{pol}1}|\cdot\rangle_{\text{pol}2}|\cdot\rangle_{\text{path}1}|\cdot\rangle_{\text{path}2}$:

$$\begin{aligned} |\xi\rangle &= \frac{1}{\sqrt{2}} [-|\Phi^-\rangle \otimes |u^A d^B\rangle + |\Phi^+\rangle \otimes |d^A u^B\rangle], \\ |\zeta\rangle &= |D\rangle^{\otimes 2} \otimes |\Psi^-\rangle. \end{aligned} \quad (6.4)$$

Here, the superposition state of polarization $|D\rangle = (|H\rangle + |V\rangle)/\sqrt{2}$. If we switch from physical basis states to computational basis states, we find that the pre-selection states are the 4-qubit linear cluster states mentioned in Chapter 6. By substituting the definition of weak values, we obtain:

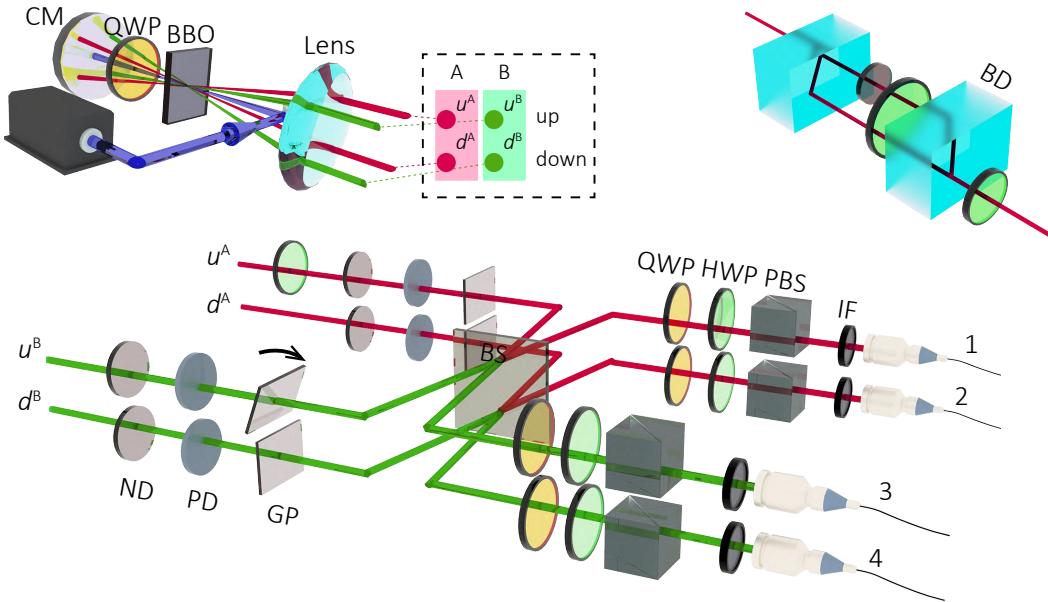


Fig. 6.7 Experimental setup for the quantum Cheshire cat grinning swap. *Top left:* Entangled photon source used in the experiment, producing photon pairs with maximally entangled polarization and path. *Top right:* Principle of the tunable polarization filter, consisting of two beam displacers, two wave plates, and a tunable neutral density filter. It attenuates the amplitude of vertically polarized photons. *Bottom:* Weak value extraction and post-selection device. The desired weak values are obtained using a tunable neutral density filter and a polarization filter, followed by a joint polarization-path measurement using a beam splitter and a coincidence detection between single-photon counters {1, 4}. ND: tunable neutral density filter, PD: tunable polarization filter, BD: beam displacer, QWP: quarter-wave plate, HWP: half-wave plate, PBS: polarization beam splitter, IF: interference filter.

$$\begin{aligned} \langle \Pi_\mu^\nu \rangle_w &= \delta_{\mu d} \delta_{\nu A} + \delta_{\mu u} \delta_{\mu B}, \\ \langle \sigma_z^\nu \otimes \Pi_\mu^\nu \rangle_w &= \delta_{\mu u} \delta_{\nu A} + \delta_{\mu d} \delta_{\nu B}. \end{aligned} \quad (6.5)$$

In this way, similar to the case in Equation (6.2), we achieve an inverse correlation between the weak values of the two position operators and the two polarization operators.

To achieve the second objective, some ingenuity is required in the experimental setup. Our experimental setup is shown in Figure 6.7, focusing on the bottom panel. After the weak value analysis, we perform two subsequent post-selections for polarization and path according to Equation (6.4). However, the post-selection for path is achieved through two-photon interference in our setup: each photon's path interferes with the spin path of the other photon on a 50:50 beam splitter, followed by a coincidence detection between single-photon counters {1, 4}. The effect of this operation is to collect the singlet state of the path qubits, $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|u^A d^B\rangle - |d^A u^B\rangle)$, and discard the three triplet states: for $|\Phi^\pm\rangle$, Hong–Ou–Mandel interference occurs at the beam splitter, causing both photons to be detected in the same detector; for $|\Psi^+\rangle$, the additional π phase leads to both photons being detected on the same side of the two-layer detectors. The specific principles can still be derived from the evolution of operators generated by second quantization. According to the requirements of Equation (6.5), due to the quantum entanglement, the two

Cheshire cats are necessarily located on the upper and lower layers, and the two grinning states are also distributed on the upper and lower layers, opposite to the Cheshire cats. Finally, during the detection, each layer receives a complete photon, which is a superposition of the Cheshire cat from that layer and the grinning originally belonging to the Cheshire cat on the other layer. In this way, we use the coincidence detection technique to achieve both post-selection and the exchange process simultaneously.

To briefly explain the quantum photon source used in the experiment, as shown in the top left panel of [Figure 6.7](#), it is different from all the photon sources we have used before. It is obtained by pumping a type-I nonlinear crystal with continuous-wave ultraviolet laser [167]. This photon source produces photon pairs with the same polarization and maximal entanglement in path. The principle of this photon source is as follows: for each pump, two photons with the same polarization are generated. Between two pumpings, a quarter-wave plate at the parametric wavelength rotates the polarization of the photons generated during the first pumping by 90 degrees. Therefore, the final polarization state of the photons is $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$. However, its most notable feature is that it produces a ring-shaped output region for parametric photons. Thus, the pairs of parametric photons can be distributed at any pair of symmetric positions on the ring. By selecting two pairs distributed at the four vertices of a rectangle, the photons will always appear in pairs at the symmetric vertices. Therefore, the final path state of the photons is $|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|u^A d^B\rangle + |d^A u^B\rangle)$. In summary, this photon source produces a hyper-entangled state where both polarization and path are maximally entangled. By introducing a half-wave plate placed at 0 degrees in one of the output paths, a CZ operation is effectively performed, creating a 4-qubit linear cluster state.

Why do we need to use this pump-return photon source instead of the light path already used in [chapter 5](#) to generate linear cluster states? The answer is that in that light path, the preparation of the linear cluster state already involves two-photon interference. Using it again would cause the coincidence detection relied upon by the first gate operation to fail. On the other hand, the device here only relies on the spatial indistinguishability of the photons to prepare entanglement [168]. The generated photon pairs can be subjected to Bell state measurements in the final state, which is necessary for our post-selection. Therefore, using this pump-return photon source that relies on the photon's indistinguishability in space is the inevitable choice for this experiment.

6.3.2 Weak value extraction without weak measurement

With the pre- and post-selected biorthogonal states and the experimental setup, the quantum Cheshire cat is already undergoing grinning swap in the optical path. However, how can we observe this effect? If we follow the methods of Pryde *et al.* [160] or Kim *et al.* [164] and use CNOT/CZ gates for weak measurement, at least one additional photon is required. But in the process of parametric down-conversion, only one photon pair can be obtained at a time. If we use pulsed pumping instead, additional compensations at the photon source are needed to ensure the temporal overlap of the wave function of the photons generated in two parametric processes. Moreover, the CZ gate itself is highly dependent on the parameters of imaginary-time evolution, which imposes high requirements on experimental resources and precision. Therefore, we consider an alternative method to implement weak value extraction without relying on controlled gates.

In Chapter 3, we have derived the relationship between the change in post-selection probability caused by imaginary-time evolution and the Hamiltonian as well as the evolution time when $\beta t \rightarrow 0^+$. For the sake of completeness, we rewrite Equation 3.9 again:

$$\frac{d}{dt} \frac{\Pr[\phi|\psi(t)]}{\Pr[\phi|\psi(0)]} = 2\text{Re} \langle H \rangle_w := 2\text{Re} \frac{\langle \phi|H|\psi \rangle}{\langle \phi|\psi \rangle}. \quad (6.6)$$

In plain language, the derivative of the normalized post-selection probability with respect to the evolution time due to imaginary-time evolution is twice the real part of the weak value of the imaginary-time evolution Hamiltonian. Since only the real part of the weak values is used when observing the quantum Cheshire cat, the imaginary-time evolution method is particularly suitable for our work. To eliminate the imaginary part of weak values, it is sufficient to verify that the pre- and post-selected biorthogonal states do not contain any imaginary part.

Now, we derive the required forms of imaginary-time evolution for path and polarization observables, respectively. For the path operator, we have:

$$U_{\mu,\text{path}}^v(\pm t_n) = \exp(-\Pi_\mu^v \pm t_n) = \mathbb{I}_2 - \Pi_\mu^v [1 - \exp(-\pm t_n)],$$

which means attenuating the amplitude of photon v on path μ to $\exp(-\pm t_n)$. In the experiment, this can be realized by using a tunable neutral density filter on the photon's path, which transmits with a transmission coefficient of $\gamma_n = \exp(-2\pm t_n)$. To solve for the weak value of polarization, the situation is slightly different. If we directly use the polarization operator for imaginary-time evolution, we would have:

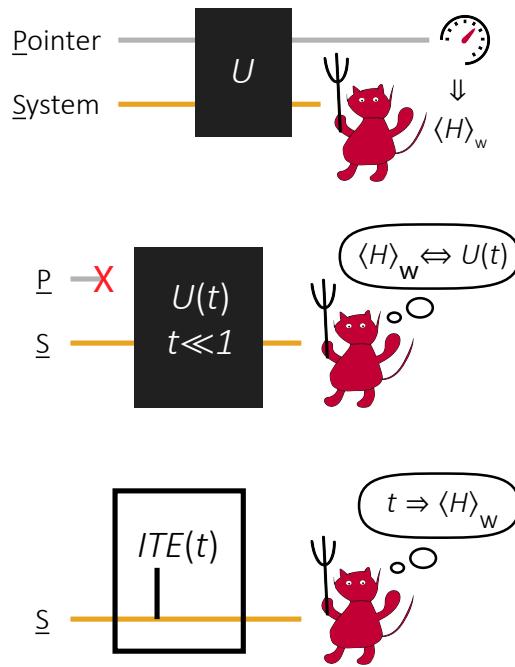


Fig. 6.8 Weak value extraction without reliance on weak measurement. *In standard weak measurement*, weak values are obtained by measuring the pointer state after interacting with the system. Here, we demonstrate that *without measuring the pointer state* but by allowing post-selection on the measurement probability, one can equivalently obtain the weak value corresponding to the Hamiltonian causing the probability change. In our work, by employing imaginary-time evolutions with different evolution times, the post-selection probability and evolution time are linearly fitted to extract the real part of the weak value.

$$\tilde{U}_{\mu,\text{pol}}^\nu(\not{t}_p) = \exp(-\sigma_z^\nu \otimes \Pi_\mu^\nu \not{t}_n) = (\mathbb{I}_2 - \Pi_\mu^\nu) + \Pi_\mu^\nu [\cosh(\not{t}_p) - \sigma_z^\nu \sinh(\not{t}_p)],$$

Both terms in the square brackets require amplification of probability amplitudes, which is not achievable in linear optics. However, we take a different approach and use the technique of interconversion between projection operators and Pauli operators. We select $(\mathbb{I}_2 - \sigma_z)^\nu \otimes \Pi_\mu^\nu / 2$ to generate this imaginary-time evolution. In this case, since the spectrum of the operator is all non-negative, the imaginary-time evolution can be fully realized using attenuation:

$$\begin{aligned} U_{\mu,\text{pol}}^\nu(\not{t}_p) &= \exp(-\Pi_\mu^\nu \not{t}_n) = \mathbb{I}_2 - \Pi_\mu^\nu (1 - \sigma_z)^\nu [1 - \exp(-\not{t}_p)] / 2 \\ &= \mathbb{I}_2 - |V\rangle \langle V|_{\text{pol}} \Pi_\mu^\nu [1 - \exp(-\not{t}_p)], \end{aligned}$$

which means attenuating the amplitude of the vertically polarized component of photon ν on path μ to $\exp(-\not{t}_n)$. In the experiment, this can be realized by adding a tunable polarizing filter on the photon's path, which allows full transmission of horizontally polarized photons and has a transmission coefficient of $\gamma_p = \exp(-2\not{t}_p)$ for vertically polarized photons. The tunable polarizing filter is actually composed of two beam displacers and a tunable neutral density filter, and it attenuates the vertical polarization component of the photon based on a very intuitive principle.

With the defined forms of imaginary-time evolution and their experimental implementations, we can solve weak values using perturbation theory. Substituting the relationships between the observables, evolution strengths, and changes in counting rates into Equation (6.6):

$$\langle \Pi_\mu^\nu \rangle_w = -\frac{1}{2} \frac{dN_{\mu,n}^\nu}{d\not{t}_n}, \quad (6.7)$$

$$\langle \sigma_z^\nu \otimes \Pi_\mu^\nu \rangle_w = -\frac{1}{2} \frac{dN_{\mu,n}^\nu}{d\not{t}_n} + \frac{dN_{\mu,p}^\nu}{d\not{t}_p}. \quad (6.8)$$

So far, we have transformed the measurement of weak values into a problem of fitting the response of counting rates to attenuation operations in the optical path. With all the tools in place to observe the quantum Cheshire cat's grinning swap, I shall move on to the formal introduction of the experimental process.

6.3.3 Experimental implementation

Before formally observing the grinning swap of the quantum Cheshire cat, we first examine the pre-selection of the initial state and the post-selection of the final state in the experiment to demonstrate that the biorthogonal state description is consistent with theoretical expectations. For pre-selection, we temporarily remove the 50:50 beam splitter, which allows us to separately analyze the polarization states corresponding to the $|u^A d^B\rangle$ and $|d^A u^B\rangle$ path combinations using quantum state tomography. According to the requirements of Equation (6.4), they should correspond to $|\Phi^+\rangle$ and $|\Phi^-\rangle$, respectively. The experimental results are shown in Figure 6.9, and the fidelity of the measured quantum states on both paths with the target states is as high as 99.0

To check the execution of the post-selection, the main task is to calibrate the phase of the two-photon interference. We introduce a tunable polarizing filter with $\not{t}_p = O(1)$ on one of the photon's paths in the interferometer and adjust the angle of a glass plate in the interferometer to optimize the projection accuracy of the path state towards $|\Psi^-\rangle$ during post-selection. Theoretical calculations show that when the tunable polarizing filter is placed on the path that allows spin σ_z to pass, the coincidence count rate

will increase, while placing it on the path that does not allow spin to pass will decrease the coincidence count rate. Therefore, we place the tunable polarizing filter on the path that allows spin σ_z to pass and scan the angle of the glass plate to maximize the coincidence count rate, thus achieving post-selection that matches the theoretical expectations. We record the coincidence count rate for the entire angular range and fit it with the theoretical predictions. The results are shown in [Figure 6.10](#), and the variations in the coincidence count rate after placing the tunable polarizing filters on both paths are in good agreement with the theoretical expectations, demonstrating the reliability of our post-selection operation in the experiment.

6.3.4 Experimental results

After all the preparations are completed, the experimental observation of the grinning swap of the quantum Cheshire cat can proceed smoothly. By adjusting the position and transmittance of the tunable neutral density filters and tunable polarizing filters added to the optical path and using Equation (6.8) for linear fitting, we can calculate the weak values of all the observables of interest from the changes in the coincidence count rate. This demonstrates the state where both quantum Cheshire cats are separated from their respective grins. Our experimental results, as shown in [Figure 6.11](#), yield the following weak values:

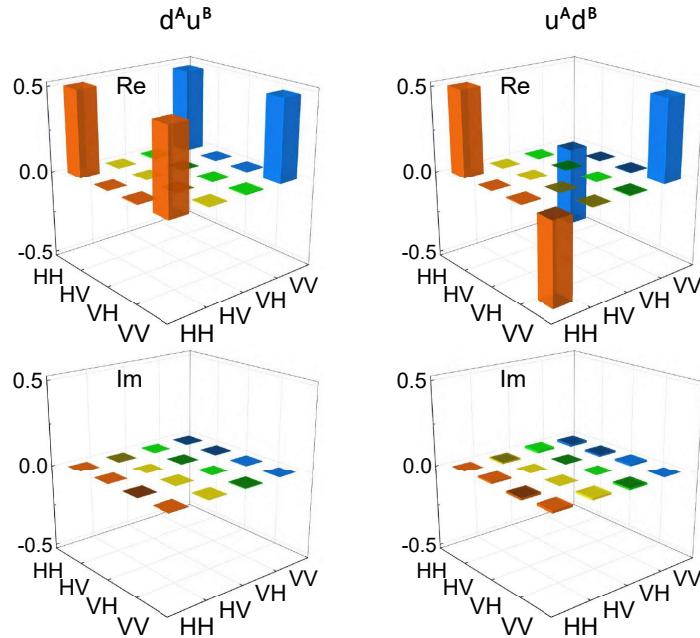


Fig. 6.9 Preparation of the initial state for the grinning swap of the quantum Cheshire cat. The left and right figures correspond to the quantum state tomography results of the polarization states on the $|u^A d^B\rangle$ and $|d^A u^B\rangle$ paths, respectively.

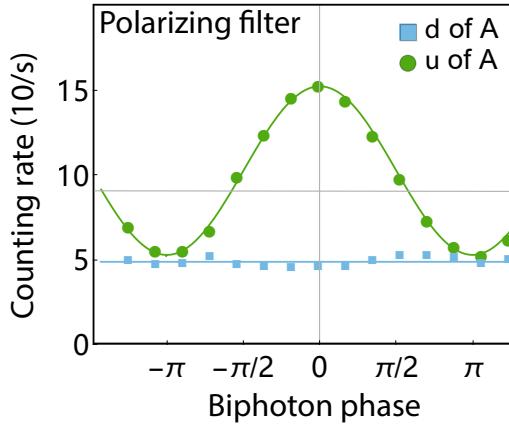


Fig. 6.10 Measurement calibration for the grinning swap of the quantum Cheshire cat. By introducing a tunable polarizing filter in the optical path, the phase of the two-photon interference can be accurately measured, and by tilting a glass plate pre-placed in the optical path, the phase can be adjusted to the position required by the post-selected state.

$$\begin{aligned} \langle \Pi_u^A \rangle_w &= -0.01(3), \quad \langle \Pi_d^A \rangle_w = 1.04(4), \quad \langle \Pi_u^B \rangle_w = 1.11(4), \quad \langle \Pi_d^B \rangle_w = 0.06(4), \\ \langle \sigma_u^A \rangle_w &= 1.01(3), \quad \langle \sigma_d^A \rangle_w = -0.04(4), \quad \langle \sigma_u^B \rangle_w = 0.10(2), \quad \langle \sigma_d^B \rangle_w = 0.04(3). \end{aligned}$$

The uncertainty in parentheses represents the uncertainty of the last digit after the decimal point, which is the 1σ standard deviation of the weak values obtained by re-sampling all the experimental data 100 times using Poisson counting statistics and then fitting. These results are in good agreement with the predictions of Equation (6.5).

In Figure 6.11, I provide a visual interpretation of the experimental results. It can be seen that according to the calculated weak values, each finally detected photon is a combination of the self of the previously prepared photon and the polarization of the other photon—the purple photon carries the orange photon’s polarization, and the orange photon carries the purple’s polarization. At this point, we have completed the observation of the intriguing phenomenon where two quantum Cheshire cats each obtain the grin of the other. Using the means of higher-order quantum interference, we have found a Cheshire cat paradox that has no classical correspondence and only occurs in the quantum world. The phenomenon of attribute separation demonstrated in the experiment is also immune to certain forms of noise specific to the environment, making the separation of the Cheshire cat and its attributes highly robust.

6.4 Summary

After the completion of the experimental observation of the grinning swap of the quantum Cheshire cat, it has received extensive attention from the academic community and the media. Many readers are curious about the background of the story and questions such as “Why can two photons exchange their polarization properties across space?” Given the opportunity to write this article, I emphasize that the grinning swap of the quantum Cheshire cat is an contextuality paradox and a challenge to human intuition and noncontextual hidden variable theories that conform to human intuition. The term “paradox” implies

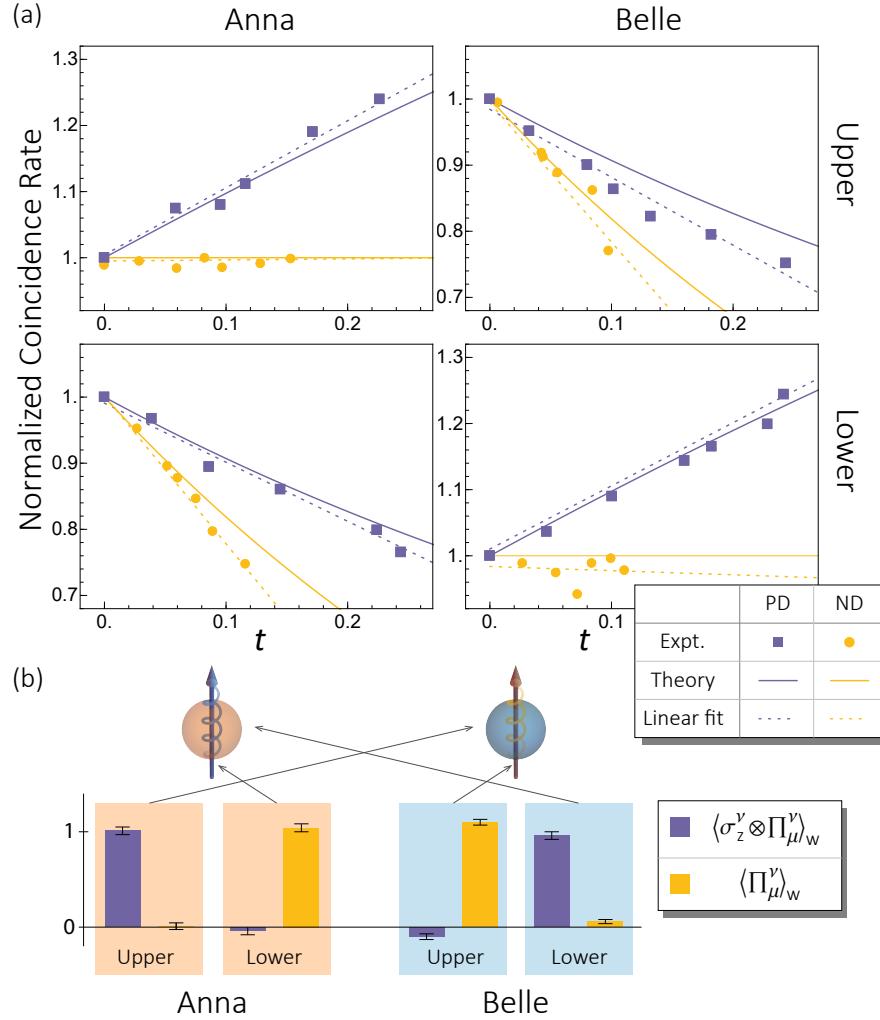


Fig. 6.11 Experimental results of the grinning swap of the quantum Cheshire cat. Top: normalized coincidence count rate as a function of the type, transmittance, and placement position of the tunable filters, where ND and PD represent the results with neutral density filters and polarizing filters, respectively. Bottom: weak values obtained after linear fitting and the visual results of the spin exchange between the two photons observed based on these weak values.

that it is not a new discovery of the unknown physical world but the discovery of a contradiction between two existing laws, forcing people to make choices and at least discard one of them. The exchange we observed in the experiment is the exchange of quantum information carried on the polarization degree of freedom, and we cannot yet exchange material properties such as mass or charge itself—nor do we know how to achieve the exchange or even superposition of different properties for non-identical particles! However, although good scientific work requires a good background story to package it, the inability to reproduce every detail of the background story does not prevent a study from being meaningful. The weak measurement technique we developed based on imaginary time evolution in our research will have broad

applications in a wide range of scenarios, greatly saving the resources required for quantum measurement tasks. Our discussion of the “classical” Cheshire cat illustrates the idea that *if two systems are governed by the same mathematical laws, they must exhibit the same behavior.* This idea has given rise to the emerging research direction of quantum simulation [169], which has made great strides in recent years. In the following [chapter 7](#), I will continue to demonstrate our use of quantum simulation tools to study topological systems in linear optical systems.

Chapter 7

Topologically-protected contextuality: simulation of photonic-encoded parafermions

In the previous chapters, we discussed various types of quantum bits (qubits) or higher-dimensional quantum bits. In this chapter, we turn our attention to a new type of system: anyon systems. In topological systems, there exist excitations called non-abelian anyons¹. Similar to how vibrational modes in solids are called phonons, we often consider anyons as quasi-particles. The interest in studying this system stems from the ability to encode quantum information in nonlocal quantum correlations that are topologically protected, which, in principle, can be immune to *any* local noise [170–172]. Furthermore, the non-abelian statistical properties of such systems [173] make the wavefunction of the system evolve unitarily in the encoding space when two particles are exchanged, rather than remaining invariant like in a bosonic system or acquiring a phase shift of π as in a fermionic system. This allows for the construction of quantum gate operations using braiding—the exchange of positions between quasi-particles. These properties make topological quantum computation an attractive approach for realizing fault-tolerant quantum computation.

The edge zero-energy modes of Majorana fermions, also known as Majorana bound states (but here we will refer to them simply as Majorana zero modes), are the most well-known examples of non-abelian anyons. Although there is experimental evidence supporting the existence of Majorana zero modes², the non-abelian statistical properties of these systems are still challenging to observe. One major obstacle in achieving braiding of Majorana fermions is the contamination from quasi-particles leaking into the system from the environment, a mechanism known as quasi-particle poisoning [174]. This mechanism greatly reduces the coherence time of quantum states encoded in Majorana zero modes, posing a significant challenge for implementing topological quantum computation based on Majorana zero modes.

7.1 From parafermions to optical quantum simulation

To overcome the aforementioned limitations of Majorana fermions, some scientists have proposed using non-abelian anyons with even more exotic properties to encode quantum information. One representative example of non-abelian anyons is parafermions. They possess \mathbb{Z}_n symmetry (n being a positive integer) and are a natural extension of Majorana fermions with \mathbb{Z}_2 symmetry. Fendley [175] first pointed out that parafermions provide symmetrically protected n -fold degenerate ground states that can be used to encode n -dimensional quantum bits. Interestingly, due to the existence of composite topological charges consisting of pairs of parafermions in fractional quantum Hall systems [176, 177], the quasi-particle poisoning

¹ Having one's name an uncapitalized adjective is the highest honor a scientist can get

² Let us assume in this article that topological quantum computation is ultimately achievable. Or at least the breakthrough from 0 to 1 qubits is achievable.

can be effectively suppressed in parafermion-based quantum computation [178]. Similar to Majorana zero modes, braiding operations of parafermions can only generate evolutions corresponding to Clifford operators, and thus, they cannot independently realize universal quantum computation using braiding operations. However, various methods have been proposed to use parafermions for universal quantum computation. One representative approach, as mentioned earlier, is the use of magic-state distillation [84] to asymptotically construct non-Clifford gates such as the $\pi/8$ -phase gate. To determine whether magic-state distillation is feasible, contextuality can be used as a criterion to test the resourcefulness of universal quantum computation [16].

In principle, the study of parafermions can be realized by manipulating fractional quantum Hall states. However, this approach poses significant technical challenges, and it was not until 2020 that evidence of anyons was observed [179, 180]. On the other hand, quantum simulation [181, 182] has made a series of achievements in showing the properties of anyon systems, such as demonstrating the braiding effects of anyons [106, 183] and the optical simulation of Majorana zero mode quantum gates by my research group [107]. In this chapter, we will introduce our use of optical quantum simulation to investigate two key elements of universal quantum computation based on \mathbb{Z}_3 parafermions:

1. Braiding operations for generating Clifford gates, and
2. Contextuality for implementing non-Clifford gates as topologically protected resources.

Due to the long coherence time and precise gate operations of our optical quantum simulator, it is well-suited for observing the Berry phase that determines the statistical properties of anyons. We encode parafermion states using a beam displacer network. We will demonstrate the robustness of their operations against local noise that arises in gate operations due to the topological protection of parafermions. Since single photons have limited scalability, the encoding of parafermions in the beam displacer network cannot be extended using a direct product form, and thus, the size of the achievable Hilbert space increases only polynomially with the number of spatial mode resources used. However, it has recently become possible to manipulate and measure multiple spatial modes simultaneously using spatial light modulators [131] and wavefront sensing [184], which greatly expands the prospects for large-scale optical quantum simulation. We hope that this photon simulation will soon gain the power of executing simple topological quantum algorithms.

7.1.1 Mathematical description of parafermions

I start from the math of Majorana fermions, which appear as zero modes at the ends of one-dimensional fermion chains with suitable coupling coefficients [171]. Using the Jordan–Wigner transformation, a fermion chain can be mapped to a chain of spin-1/2 particles. The mathematical form of the Jordan–Wigner transformation is as follows:

$$\begin{aligned}\gamma_{ka} &= \sigma_{z,k} \prod_{j < k} \sigma_{x,j}, \\ \gamma_{kb} &= \sigma_{z,k} \prod_{j \leq k} \sigma_{x,j}.\end{aligned}$$

Here, γ_{km} represents the creation and annihilation operators of the Majorana mode $m = a, b$ on the k -th particle, which also act as the annihilation operators. The Jordan–Wigner transformation is a nonlocal transformation, where the state of a fermion mode is mapped to the state of all spin-1/2 particles on its one side. Due to this property, the local properties of the original fermion system and the resulting

spin-1/2 system after mapping are fundamentally different: the localized noise in the fermion system is no longer confined to a single position in the spin system, resulting in the loss of topological protection in the spin system that exists in the fermion system. However, the Jordan–Wigner transformation preserves the spectrum of the fermion operators before and after the transformation, thereby keeping the exchange operations and the geometric phases detected invariant.

Similar as in the study of Majorana systems, we investigate the mathematical correspondence of \mathbb{Z}_n parafermion systems and spin systems with higher spin numbers, but in the opposite direction—by considering how to obtain parafermions from spin chains. The mathematical language used here is called the Fradkin–Kadanoff transformation [185]. We first define the translation operator τ and the clock operator σ , which are generalizations of the Pauli operators σ_x and σ_z in higher dimensions:

$$\tau = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & \omega & 0 & \dots & 0 \\ 0 & 0 & \omega^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \omega^{n-1} \end{pmatrix},$$

$$\text{where } \omega = e^{2\pi i/n}. \quad (7.1)$$

The commutation relation between these two operators is $\sigma\tau = \omega\tau\sigma$. Similarly to the Jordan–Wigner transformation, we define the annihilation operators α_{ka} and α_{kb} for the two modes a and b of the k -th parafermion as:

$$\alpha_{ka} = \sigma_k \prod_{j < k} \tau_j, \quad (7.2)$$

$$\alpha_{kb} = \sigma_k \prod_{j \leq k} \tau_j. \quad (7.3)$$

Note that for parafermions with $n \neq 2$, the creation and annihilation operators are no longer degenerate. Conversely, the inverse transformation from parafermion operators to spin operators is given by:

$$\begin{aligned} \tau_k &= \alpha_{ka}^\dagger \alpha_{kb}, \\ \sigma_k &= \alpha_{ka} \prod_{j < k} \alpha_{jb}^\dagger \alpha_{ja}. \end{aligned}$$

7.1.2 Braiding operations of parafermions

In this work, we consider a one-dimensional chain composed of 3 pairs of \mathbb{Z}_3 parafermions, where each pair of parafermions contains two \mathbb{Z}_3 parafermion modes (α_{ja} and α_{jb}). They satisfy the commutation relations $\alpha_{ja}\alpha_{ka} = \omega\alpha_{ka}\alpha_{ja}$, $\alpha_{jb}\alpha_{kb} = \omega\alpha_{kb}\alpha_{jb}$, and $\alpha_{ja}\alpha_{kb} = \omega\alpha_{kb}\alpha_{ja}$, $\forall j < k$, where $\omega = e^{2\pi i/3}$. We will consider the braiding operations between these parafermions to obtain the Berry phase during the evolution, thereby manifesting the non-abelian statistics of parafermions. Specifically, we first consider the ground state of the following Hamiltonian:

$$\mathcal{H}_0^{\text{Pf}} = -e^{i\pi/6}(\alpha_{1b}\alpha_{2a}^\dagger + \alpha_{2b}\alpha_{3a}^\dagger) + \text{h.c..} \quad (7.4)$$

This Hamiltonian leads to the appearance of two \mathbb{Z}_3 parafermion edge zero modes (α_{1a} and α_{3b}) at the ends of the chain, as shown in Figure 7.1. We call them “zero modes” because these two modes do not appear in the Hamiltonian, and therefore, their occupations can freely change in the ground state.

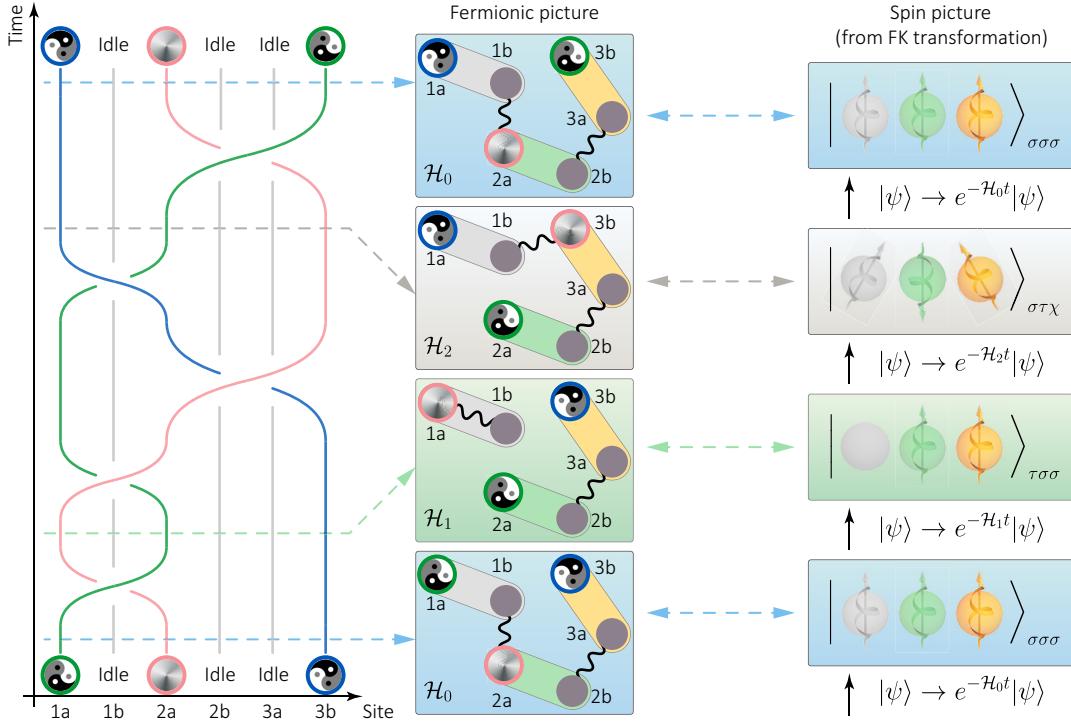


Fig. 7.1 Braiding of parafermion edge zero modes. Left panel: Braiding of two zero modes (Taiji symbols) located at the ends of the parafermion chain using auxiliary parafermions. The braiding operation generates a Clifford gate on the three-dimensional quantum system encoded by the parafermion zero modes. Middle panel: Physical implementation of the braiding process on an interacting parafermion chain. The system configuration evolves as the adiabatic evolution between the Hamiltonians $\mathcal{H}_0^{\text{Pf}}$, $\mathcal{H}_1^{\text{Pf}}$, $\mathcal{H}_2^{\text{Pf}}$, and $\mathcal{H}_0^{\text{Pf}}$. This evolution leads to the exchange of the two parafermion zero modes. Right panel: Mapping the parafermion chain to the spin-1 quantum Potts model with \mathbb{Z}_3 symmetry using the Fradkin–Kadanoff transformation. The corresponding evolution can be effectively simulated by imaginary-time evolution with a series of evolution times $t \rightarrow +\infty$, projecting the system onto the ground state of the new Hamiltonian. In the entire figure, the horizontal dashed lines connect the configurations of the three different panels corresponding to different stages.

The Hamiltonian $\mathcal{H}_0^{\text{Pf}}$ has a three-fold degenerate ground state, where a nonlocal encoding of a three-dimensional quantum system can be realized. We define the parity operator as

$$Q = \prod_k \alpha_{ka}^\dagger \alpha_{kb},$$

and choose the basis $|\psi_l^{\text{Pf}}\rangle$ ($l = 0, 1, 2$) as the computational basis; they are eigenstates of the parity operator with eigenvalues ω^l . The non-abelian statistics of parafermions lead to different phase factors

between the eigenstates $|\psi_l^{\text{Pf}}\rangle$ when the two parafermion edge zero modes α_{1a} and α_{3b} are exchanged [186]. To braid these two parafermion edge zero modes, we can adiabatically, cyclically evolve between three Hamiltonians: $\mathcal{H}_0^{\text{Pf}} \rightarrow \mathcal{H}_1^{\text{Pf}} \rightarrow \mathcal{H}_2^{\text{Pf}} \rightarrow \mathcal{H}_0^{\text{Pf}}$. Here,

$$\begin{aligned}\mathcal{H}_1^{\text{Pf}} &= -e^{i\pi/6} \alpha_{2b} \alpha_{3a}^\dagger - \alpha_{1a} \alpha_{1b}^\dagger + \text{h.c.}, \\ \mathcal{H}_2^{\text{Pf}} &= -e^{i\pi/6} (\alpha_{2b} \alpha_{3a}^\dagger + \alpha_{3b} \alpha_{1b}^\dagger) + \text{h.c.}.\end{aligned}\quad (7.5)$$

Note that $\alpha_{1a} \alpha_{1b}^\dagger$ term represents a ferromagnetic interaction, which can be realized in a physical system using an external field, while the other terms represent paramagnetic interactions between modes. Therefore, in our designed evolution, all the Hamiltonians have a clear physical picture and can be naturally implemented in quantum Hall systems.

Now I analyze the effects of these adiabatic evolutions with reference to Figure 7.1:

1. First, by turning off the interaction between mode 1 and mode 2, the parafermion edge zero on mode 1a and the auxiliary parafermion at mode 2a will be exchanged.
2. Second, the interaction is switched on between the initially isolated mode 1 and mode 3. At this point, the parafermion edge zero mode originally located at mode 3b moves to mode 1a, which is now located at the end of the interacting parafermion chain.
3. Finally, by resetting the couplings of the system to their original configurations, one parafermion edge zero mode is transferred from mode 2a to mode 3b.

The above evolution results in the exchange of the two parafermion edge zero modes α_{1a} and α_{3b} . In addition, the evolution process repeatedly moves the auxiliary parafermion, but it eventually returns to its original position. Since the braiding operator and the parity operator defined above can be simultaneously diagonalized [186], the braiding induces a relative phase between $|\psi_0^{\text{Pf}}\rangle$, $|\psi_1^{\text{Pf}}\rangle$, and $|\psi_2^{\text{Pf}}\rangle$, thereby manifesting the non-abelian statistical properties of parafermions. Specifically, the evolution caused by the braiding operation in Figure 7.1 can be written as:

$$|\psi^{\text{Pf}}\rangle \rightarrow \mathcal{B} |\psi^{\text{Pf}}\rangle : \mathcal{B} = \text{diag}(1, 1, \omega).$$

7.1.3 Quantum simulation in brief

In the previous section, we have introduced the mathematical foundation of parafermions, braiding operations, and their non-abelian statistical properties. However, it is apparent that these studies are highly theoretical and seem to differ significantly from the optical system experiments discussed earlier. Now we aim to bridge this gap. The complex mathematical nature of parafermions and the experimental challenges in realizing them in quantum Hall effect platforms precisely motivate us to investigate them using optical platforms [169]. Feynman first pointed out that the dynamics of quantum systems can be effectively *simulated*: by using a quantum computer or other artificial quantum systems, it is possible to accurately reproduce the evolution of quantum systems in nature. To achieve quantum simulation, the most important aspect is to recreate the dynamics of the target system, which means establishing a mapping between the Hamiltonian of the system under study and the Hamiltonian of the quantum simulator:

$$H_{\text{sys}} \leftrightarrow H_{\text{sim}} \approx f H_{\text{sys}} f^{-1}, \quad (7.6)$$

where f is a unitary mapping. This establishes a correspondence between the evolutions of the two systems. As shown in Figure 7.2, we want to study the evolution of a quantum system but cannot directly

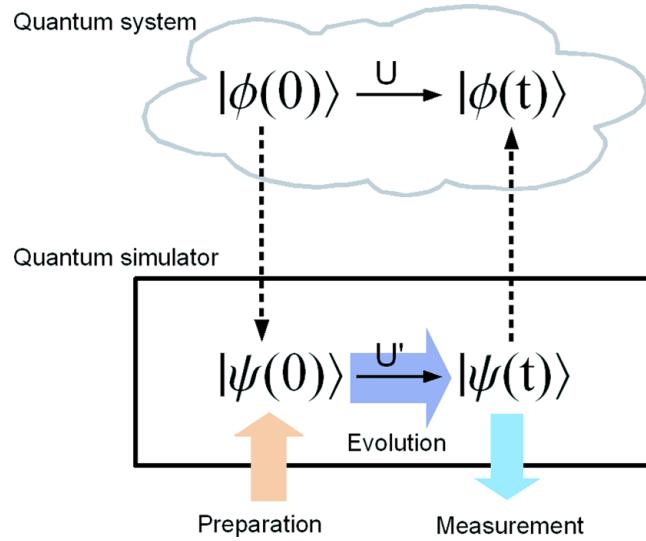


Fig. 7.2 Illustration of using quantum simulation to study quantum systems. Figure taken from reference [182], copyright American Physical Society.

implement the operation $U = \exp(-iH_{\text{sys}}t)$. The solution is to implement

$$U' = \exp(-iH_{\text{sim}}t) = \exp\left(-ifH_{\text{sys}}f^{-1}t\right) = f \exp(-iH_{\text{sys}}t) f^{-1} = fUf^{-1}.$$

In this way, if the initial state of the quantum simulator is set as $|\phi(0)\rangle$, to obtain the solution of the final state $|\phi(t)\rangle$, we can prepare the initial state $|\psi(0)\rangle = f|\phi(0)\rangle$ in the quantum simulator, evolve it using the operation U' to obtain the final state $|\psi(t)\rangle$, and then infer the final state of the original quantum system as $|\phi(t)\rangle = f^{-1}|\psi(t)\rangle$. Generally, convenient-to-manipulate quantum systems, such as linear optical systems, are chosen as quantum simulators. In quantum simulators, high-precision preparation, evolution, and measurement of quantum states can be achieved.

7.1.4 Mapping parafermions to optical systems

By utilizing the concept of quantum simulation, we can transform the problem of parafermion braiding statistics into a problem of linear optical evolution. However, first, we need to establish a correspondence between the two Hamiltonians. Since we are familiar with the Fradkin–Kadanoff transformation, it is natural to transform the parafermion system into a spin-1 system. Applying Equation (7.3) to the Hamiltonians in the three stages, we obtain their representations in the spin system:

$$\begin{aligned}\mathcal{H}_0^S &= -e^{i\pi/6}(\sigma_1\sigma_2^\dagger + \sigma_2\sigma_3^\dagger) + \text{h.c.}, \\ \mathcal{H}_1^S &= -e^{i\pi/6}\sigma_2\sigma_3^\dagger - \tau_1 + \text{h.c.}, \\ \mathcal{H}_2^S &= -e^{i\pi/6}(\sigma_2\sigma_3^\dagger + \sigma_1\tau_2^\dagger\tau_3^\dagger\sigma_3^\dagger) + \text{h.c.},\end{aligned}\quad (7.7)$$

Based on the knowledge of quantum simulation we discussed earlier, we know that by performing the following adiabatic evolution: $\mathcal{H}_0^S \rightarrow \mathcal{H}_1^S \rightarrow \mathcal{H}_2^S \rightarrow \mathcal{H}_0^S$, we can reproduce the Berry phase obtained from the braiding operation of parafermion zero modes.

Now, I introduce a definition: let $|k\rangle_\sigma$, $|k\rangle_\tau$, and $|k\rangle_\chi$ correspond to eigenstates of the operators σ , τ , and $\sigma\tau$ with eigenvalues ω^k . The following will show that this definition greatly simplifies the encoding of parafermions using optical systems. We represent $|\psi_0^{\text{Pf}}\rangle$, $|\psi_1^{\text{Pf}}\rangle$, and $|\psi_2^{\text{Pf}}\rangle$ in terms of the eigenstates of the clock operator:

$$\begin{aligned}|\psi_0^{\text{Pf}}\rangle \leftrightarrow |\psi_0^S\rangle &= \frac{1}{\sqrt{3}}(|000\rangle_\sigma + |111\rangle_\sigma + |222\rangle_\sigma), \\ |\psi_1^{\text{Pf}}\rangle \leftrightarrow |\psi_1^S\rangle &= \frac{1}{\sqrt{3}}(|000\rangle_\sigma + \omega|111\rangle_\sigma + \omega^2|222\rangle_\sigma), \\ |\psi_2^{\text{Pf}}\rangle \leftrightarrow |\psi_2^S\rangle &= \frac{1}{\sqrt{3}}(|000\rangle_\sigma + \omega^2|111\rangle_\sigma + \omega|222\rangle_\sigma).\end{aligned}\quad (7.8)$$

The first task of our quantum simulation is to obtain the Berry phase evolution that causes the braiding statistics of parafermion zero modes. Interestingly, solving the Berry phase does not necessarily require adiabatic evolution; it can be achieved solely using projection measurements. Let Π_k be the operator that projects onto the ground state subspace of \mathcal{H}_k^S . According to [187, 188], we have:

$$\phi_{B,l} = -\arg \left\langle \psi_l^S \middle| \Pi_1 \Pi_2 \middle| \psi_l^S \right\rangle, \quad (7.9)$$

where $|\psi_l^S\rangle$ corresponds to the three ground states of \mathcal{H}_0 given in Equation (7.8). When $\phi_{B,l}$ depends on l , it exhibits non-abelian statistical properties. We use a series of imaginary time evolutions to obtain the Berry phase in Equation (7.9):

$$e^{-\mathcal{H}_0^S t} \cdot e^{-\mathcal{H}_1^S t} \cdot e^{-\mathcal{H}_2^S t} \cdot e^{-\mathcal{H}_0^S t}, \quad t \rightarrow +\infty. \quad (7.10)$$

The significance of using imaginary time evolution is that it avoids solving for the ground state of each Hamiltonian. Normally, diagonalizing the Hamiltonian requires matrix diagonalization, which is difficult for high-dimensional systems. However, if the two terms commute, the imaginary time evolution can be expanded into a product of imaginary time evolution operators corresponding to each Hamiltonian, using the Baker–Campbell–Hausdorff formula. In this experiment, we find that since \mathcal{H}_1^S and \mathcal{H}_2^S each have one term identical to \mathcal{H}_0^S , the imaginary time evolution of these two terms is not needed at all. The imaginary time evolutions corresponding to the other terms are equivalent to projecting the system onto the following three subspaces successively:

$$|k\rangle_\sigma |k\rangle_\sigma |k\rangle_\sigma, \quad |0\rangle_\tau |k\rangle_\sigma |k\rangle_\sigma, \quad |k\rangle_\sigma |l\rangle_\tau |l+k \bmod 3\rangle_\chi, \quad k, l \in \{0, 1, 2\}.$$

Before the actual experiment, we can further simplify the form of imaginary time evolution through appropriate encoding. This is because when $|k\rangle_\sigma |l\rangle_\tau |l+k \bmod 3\rangle_\chi$ is projected onto the other two subspaces, the coefficients do not depend on l . In other words, the evolution results for different l values are identical. Therefore, when using quantum simulation, we can encode only one evolution path for $l = 0$

and ignore the evolution paths for $l = 1$ and $l = 2$. There is no difference between the two methods when we normalize the states again. As a result, the simulation of the geometric phase during parafermion braiding reduces to solving the coefficients during the cyclic transformation between orthogonal bases in the three subspaces. By encoding these three orthogonal bases using an optical system and measuring the coefficients when they are projected, we can calculate the Berry phase according to Equation (7.9) and observe the non-abelian statistics of the parafermion zero modes. Now, I turn to explore how to experimentally measure these coefficients in a linear optical system.

7.2 Optical simulation of parafermion braiding statistics

Since a spin- $(d - 1)/2$ system can be directly mapped to a d -dimensional quantum bit (qudit) and encoded into the degrees of freedom of photons, such as the path, linear optical systems can be used for quantum simulation of parafermion systems with the aid of the Fradkin–Kadanoff transformation. In this experiment, we encode the ground state of a spin-1 chain into different spatial modes of a single photon,

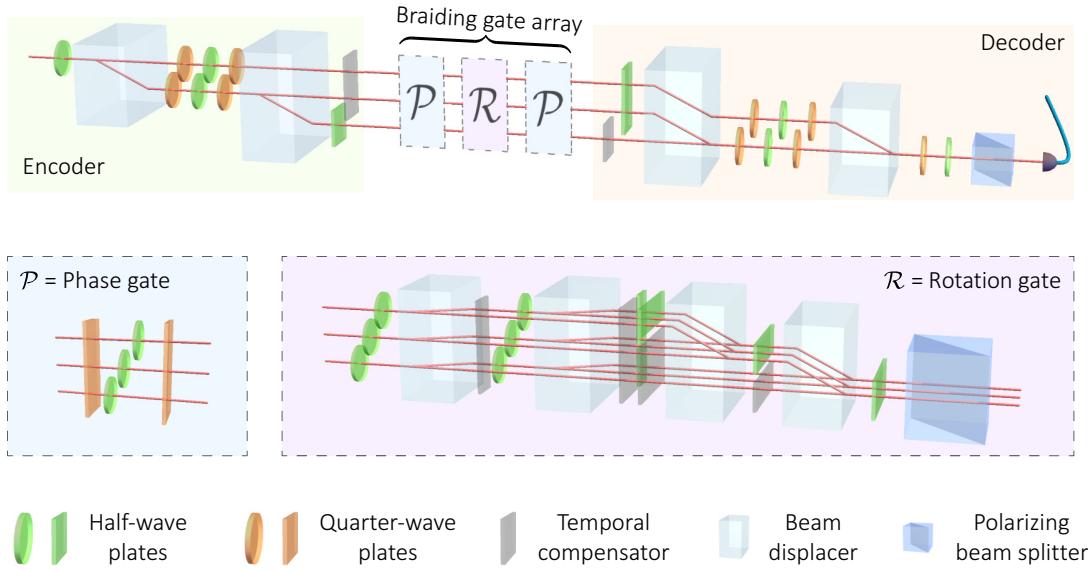


Fig. 7.3 Optical simulator for parafermions. The setup encodes the three-dimensional quantum state $|\psi^{\text{Pf}}\rangle$ carried by the parafermion zero modes onto the path modes of photons, followed by evolution and measurement operations. In the encoding part, a beam displacer array is used for preparing the quantum state. Half-wave plates and quarter-wave plates control the complex amplitudes of each mode. In the evolution part, the photon modes undergo the desired braiding evolution gates. These gates include two phase gates \mathcal{P} that apply phase evolutions between the three modes, and a rotation gate \mathcal{R} that implements arbitrary SU(3) rotations. In the decoding part, another beam displacer architecture is used to reconstruct the quantum state encoded by the parafermion zero modes via quantum state tomography. In the subfigures with dashed boxes, we provide detailed implementations of the gates \mathcal{P} and \mathcal{R} . Glass plates with adjustable tilt angles are inserted for time compensation to align the optical path lengths of different modes and optimize the interferometric visibility of the setup.

forming a three-dimensional quantum system. Since we only need to study the wave function evolution in the degenerate ground state space, we can repeatedly use the imaginary time evolution in Equation (7.9) in the device to evolve the wave function to a new ground state and dissipate the excited states. These continuously implemented photon imaginary time evolutions are the core processes of dissipative quantum simulation.

7.2.1 Designing a dedicated quantum simulator

In Figure 7.3, we present a parafermion optical simulator for the preparation, imaginary time evolution, and measurement of arbitrary three-dimensional quantum states. We use beam displacers to prepare the three-dimensional quantum states encoded in path modes. Half-wave plates and quarter-wave plates are inserted before and between the beam displacers to selectively control the complex amplitudes of the three spatial modes. Since we only consider the wave functions in the ground state subspace and employ an optimized encoding scheme, the number of modes required to encode the parafermion chain is reduced from the theoretical $3^4 = 81$ to only $3^2 = 9$ in the experiment. This encoding scheme significantly expands the applicability of our optical simulator and facilitates the further simulation of parafermion zero mode braiding.

The basic modules for measuring the Berry phase are unitary evolution gates and imaginary time evolution gates. When the projection operator obtained from Equation (7.9) shares the same eigenstates as the current Hamiltonian, the non-dissipative phase gate \mathcal{P} realizes the unitary evolution by generating relative phases between the three optical modes. Each \mathcal{P} gate consists of two quarter-wave plates with three adjustable half-wave plates sandwiched between them, where the optical axis of the quarter-wave plates is always fixed at 45° , and the angle of the half-wave plates determines the phase applied. When the projection operator does not coincide with the eigenstates of the current Hamiltonian, the dissipative imaginary time evolution must be implemented, which corresponds to the rotation gate \mathcal{R} . It consists of two parts: first, a basis transformation is performed by using beam displacers to spatially separate the three optical modes that were originally arranged vertically to encode the quantum state, then the complex amplitudes of the resulting nine modes are adjusted individually, and finally, these modes are combined vertically to obtain the horizontally arranged quantum state. During this merging process, the coherent superposition of wave functions corresponding to different initial modes and representing new ground states is achieved, enabling arbitrary SU(3) rotation operations, while the photons corresponding to excited states are dissipated due to the imaginary time evolution.

Finally, in the decoding stage, another pair of beam displacers recombine the three modes. Quantum state tomography is performed using half-wave plates, quarter-wave plates, and polarizing beam splitters. Since the measurement relies on interference between different optical modes, the optical path lengths of each evolution path must be aligned to maintain good temporal coherence and maximize the interferometric visibility of the setup. We also introduce additional temporal compensators in the setup to compensate for the path length differences introduced by imperfect alignment of waveplates and beam displacers and to compensate for the relative phase between different paths by tilting them.

7.2.2 Experimental results

First, we characterize the overall effect of the braiding operation. We use nine three-dimensional quantum states, $|\psi_j\rangle$, $j = 1, \dots, 9$, encoded in the triple degenerate ground states of \mathcal{H}_0^S . We perform braiding

evolution on these quantum states and then use quantum state tomography to obtain the final state results. The tomography results are shown in [Figure 7.4](#). By converting the braiding of parafermion zero modes into three consecutive basis rotations and imaginary time evolution, we experimentally measure the Berry phase values between $|\psi_2^S\rangle$ and $|\psi_0^S\rangle$ to be:

$$\delta\phi_{B,2} = 2.061 \pm 0.128,$$

while the phase between $|\psi_1^S\rangle$ and $|\psi_0^S\rangle$ is negligible with

$$\delta\phi_{B,1} = 0.066 \pm 0.104.$$

The experimental results, with an average fidelity of 94.9%, clearly demonstrate the non-abelian braiding statistics of parafermion zero modes, as expected from the \mathbb{Z}_3 parafermion braiding matrix $\mathcal{B} = \text{diag}(1, 1, \omega)$. To estimate the experimental errors, we assume that the photon counting statistics in the experiment follow a Poisson distribution and perform 100 resamplings of the coincident events to compute the 1σ standard deviation of the measured physical quantities.

Furthermore, we use quantum process tomography to fully characterize the dynamics of the braiding process. We calculate the process matrix χ_{jk} , which is expanded on the complete basis vectors composed of the Gell-Mann matrices $\{\lambda_1, \dots, \lambda_8\}$ and the identity \mathbb{I}_3 . The process matrix is shown in [Figure 7.4](#), which also reveals the relative phase of $2\pi/3$ between $|\psi_2^S\rangle$ and the other two states. The definitions of the eight Gell-Mann matrices are given. By comparing the braiding results with the theoretical expectations, we obtain a fidelity of 93.4% for the braiding operation in the optical quantum simulator. Therefore, the optical quantum simulation based on a series of imaginary time evolutions faithfully reproduces the non-abelian statistical properties of parafermion braiding. In [Figure 7.5](#), we provide more detailed experimental results.

7.3 From contextuality to universal topological quantum computing

In the previous section, we introduced the method of simulating the braiding of parafermion zero modes using linear optics and mathematical mappings. In this section, we return to the study of contextuality and demonstrate its connection to topological quantum computing. By employing different braiding operations, it is in principle possible to obtain all Clifford gates. However, at least one non-Clifford gate operation is required to achieve universal quantum computing [85], which cannot be obtained solely through braiding. By using “magic states” as auxiliary quantum systems, one can utilize controlled Clifford gate operations along with measurements on the auxiliary states to implement equivalent non-Clifford gate operations, as illustrated in [Figure 7.6](#). The essence is that if precise Clifford gate operations can be realized, it is possible to construct a quantum circuit that extracts a higher-fidelity magic state from a series of identical but lower-fidelity magic states, thereby iteratively improving the precision of the implementation of non-Clifford gates. This method is now known as magic state distillation.

In the work by Bravyi and Kitaev [84], the noise threshold for magic state distillation in qubit systems was calculated—the quantum states that are further than the noise threshold from ideal magic states cannot be used for magic state distillation. Since then, different magic state distillation protocols have been developed for various systems. Which systems are capable of magic state distillation? Howard *et al.* [16] derived a noncontextual hidden variable inequality using the CSW method, where measurements are constructed using projection operators onto eigenstates of certain Clifford group elements. Quantum states that violate this noncontextual hidden variable inequality can also be used for magic state distillation. We refer to these quantum states as exhibiting “Clifford contextuality”.

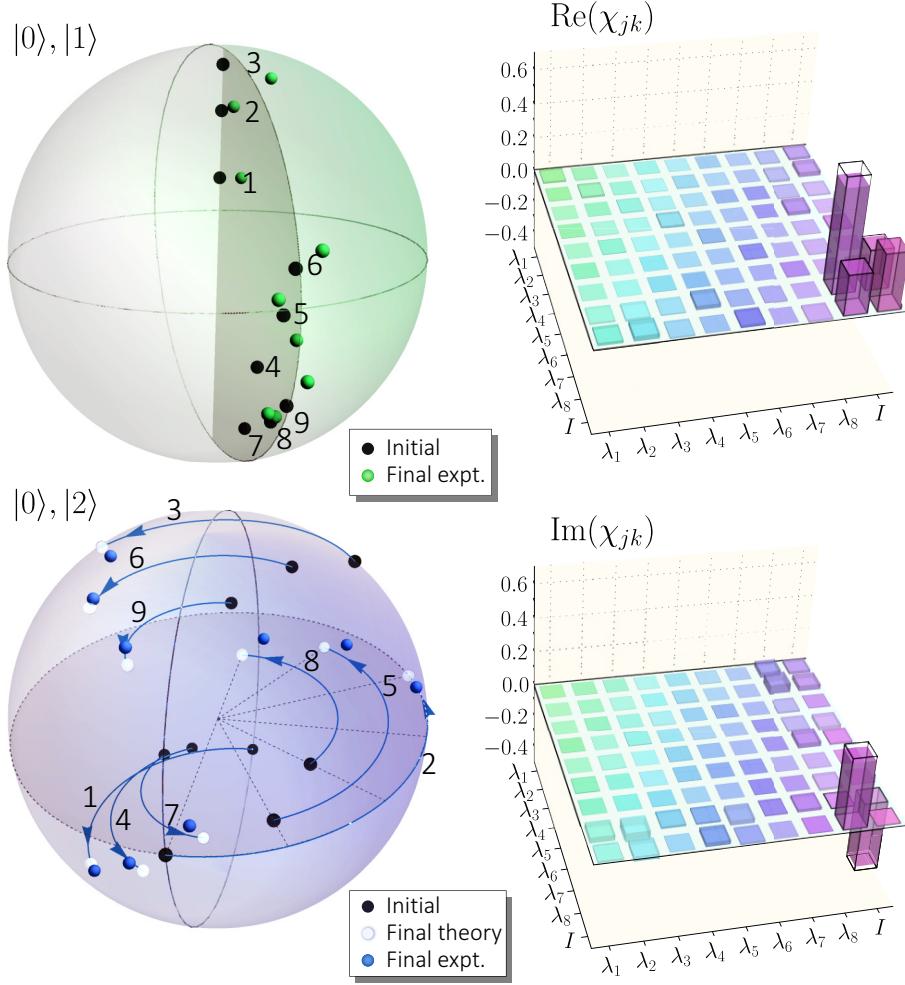


Fig. 7.4 Experimental results of parafermion braiding. Nine quantum states $|\psi_j\rangle$, $j = 1, \dots, 9$, are input into the optical quantum simulator, and quantum state tomography is performed to determine the evolution results. The upper and lower Bloch spheres represent the projections of the evolution results in the subspaces $\{|\psi_0^S\rangle, |\psi_1^S\rangle\} \otimes \{\langle\psi_0^S|, \langle\psi_1^S|\}$ and $\{|\psi_0^S\rangle, |\psi_2^S\rangle\} \otimes \{\langle\psi_0^S|, \langle\psi_2^S|\}$, respectively. It can be seen that there is no significant evolution phase in the upper Bloch sphere, while each point in the lower sphere traces an arc close to $2\pi/3$ central angle, consistent with the additional $2\pi/3$ Berry phase obtained between $|\psi_2^{\text{Pr}}\rangle$ and $|\psi_0^{\text{Pr}}\rangle$ or $|\psi_1^{\text{Pr}}\rangle$. Right panel: The evolution process matrix χ_{jk} obtained from quantum process tomography. It is represented by the coefficients of the process matrix expansion on the Gell-Mann matrices. The solid lines and color filling correspond to theoretical and experimental values, respectively.

Now, for the case of three-dimensional quantum systems, we use the Weyl–Heisenberg representation to explicitly give the criterion inequality for Clifford contextuality. First, we define the Weyl–Heisenberg displacement operators as:

$$D_{x,z} = \omega^{2^{-1}xz} \tau^x \sigma^z, \quad \{x, z\} \in \{0, 1, 2\}. \quad (7.11)$$

The eigenstates of these operators are all $\{1, \omega, \omega^2\}$. Now, we select 4 sets of displacement operators, denoted as $\mathbf{D} = \{D_{0,1}, D_{1,0}, D_{1,1}, D_{1,2}\}$. Their eigenstates can form a mutually unbiased basis in a three-dimensional Hilbert space. We define three auxiliary vectors: $\mathbf{a} = \{1, 0, 1, 2\}$, $\mathbf{b} = \{-0, 1, 1, 1\}$, $\mathbf{r} = x\mathbf{a} + z\mathbf{b}$, where $x, z \in \{0, 1, 2\}$. Furthermore, let $\hat{\Pi}_j^{rj}$ be the projection operator corresponding to the eigenstate with eigenvalue ω^{rj} of the j th element in \mathbf{D} . Then, the Clifford contextuality can be expressed as the violation of the following noncontextual hidden variable inequality:

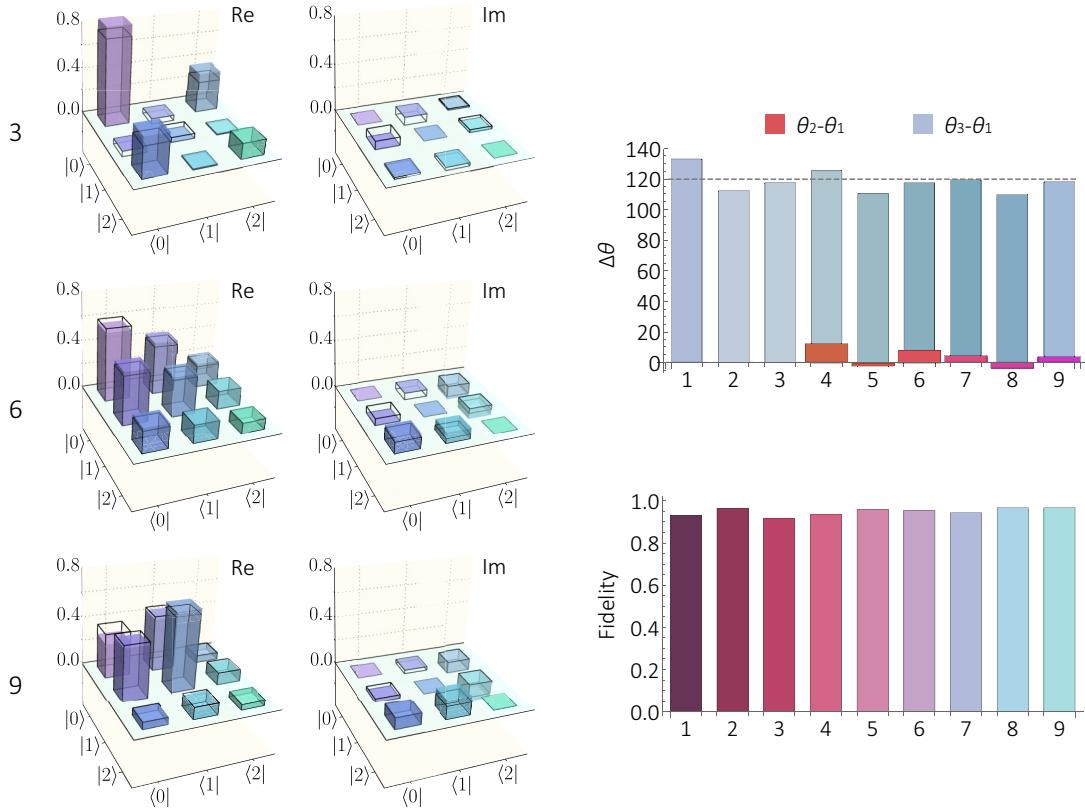


Fig. 7.5 Supplementary experimental results of parafermion braiding. Left panel: Density matrices (color filling) of three example quantum states after the braiding operation compared with the theoretical expectations (solid lines). Right panel: Additional Berry phase measurements obtained between $|\psi_1^S\rangle$ and $|\psi_2^S\rangle$ relative to $|\psi_0^S\rangle$, as well as the fidelity of all nine quantum states relative to the theoretical predictions.

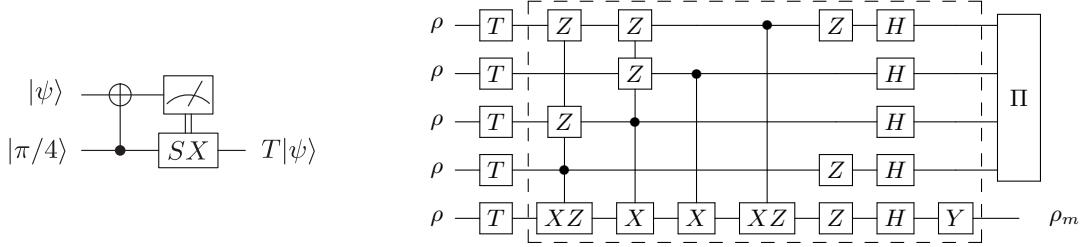


Fig. 7.6 Universal quantum computing with magic state assistance and magic state distillation. Left: schematic showing how a non-Clifford gate operation can be performed using a magic state and Clifford gate operation. Right: a quantum circuit for $5 \rightarrow 1$ magic state distillation.

$$\mathcal{M} := \max_{\mathbf{r}} \text{Tr} [A^{\mathbf{r}} \rho] \stackrel{\text{NCHV}}{\leqslant} 0, \quad (7.12)$$

$$\text{with } A^{\mathbf{r}} = \mathbb{I}_3 - \sum_{j=1}^4 \hat{\Pi}_j^{r_j}.$$

This inequality is maximally violated by ideal magic states, and the violation value happens to be the golden ratio: $\mathcal{M} \stackrel{Q}{\leqslant} (\sqrt{5} - 1)/2$. Therefore, the magnitude of the violation of inequality (7.12) can reflect the capability of a quantum state to be used as a magic state for universal quantum computing. When the violation is large, the quantum state itself can be used directly or with a few rounds of magic state distillation to implement non-Clifford gate operations with fidelity surpassing the fault-tolerant threshold [189]; when the violation is small, a significantly increased number of rounds of magic state distillation is required, and repeated magic state distillation is needed to surpass the fault-tolerant threshold, resulting in a substantial consumption of resources. When $\mathcal{M} \leqslant 0$, the quantum state cannot be approximated as a resource state through magic state distillation, and its behavior can be simulated by classical computation, thus failing to demonstrate the acceleration characteristics of quantum computation.

7.3.1 Dynamics of contextuality under braiding operations

Since all the projection operators defined in Equation (7.12) are eigenstates of Clifford operators, these operators evolve under the actions of Clifford group elements, resulting in closed evolution without generating non-Clifford operators. For all possible operator symmetries defined in Equation (7.12), its value remains invariant under Clifford operations. In other words, Clifford operations do not affect the ability of a quantum state to undergo magic state distillation.

We experimentally simulate the evolution of contextuality resources under braiding operations for a three-dimensional quantum system encoded with parafermion zero modes. For the initial and final states, we directly perform projection measurements to obtain the values of the nine Clifford contextuality witness operators in Equation (7.12). Theoretical analysis indicates that the values of the operators should satisfy the following relations before and after the evolution:

$$\begin{cases} \langle A_{0z} \rangle \rightarrow \langle A_{0(z+1) \bmod 3} \rangle, \\ \langle A_{1z} \rangle \rightarrow \langle A_{1z} \rangle, \\ \langle A_{2z} \rangle \rightarrow \langle A_{2(z-1) \bmod 3} \rangle. \end{cases}$$

Therefore, the value of \mathcal{M} determined by their minimum values remains unchanged. The experimental results are shown in [Figure 7.7](#). First, we examine the variation of the nine contextuality witness operators in a sample state $|\psi_5\rangle$, which satisfies the aforementioned cyclic symmetry. Next, we measure the Clifford contextuality witnesses \mathcal{M} before and after the evolution for the nine sample states used in the experiment. The results demonstrate that the contextuality witnesses for all sample states remain almost unchanged before and after the evolution, further highlighting the accurate reconstruction of the braiding process by the photonic quantum simulator.

The first upshot of studying the interaction between braiding operations and contextuality is that fault-tolerant universal quantum computation can be achieved in a quantum system encoded with parafermion zero modes using braiding operations and magic state distillation. As the contextuality resources are unaffected by braiding operations, magic state distillation can be performed at any time during quantum computation. By distilling a magic state through magic state distillation and then applying braiding, non-Clifford gate operations can be equivalently performed at any time during quantum computation, thus achieving universality in quantum computation.

7.3.2 Topologically-protected braiding of parafermions

The second upshot of studying the interaction between braiding operations and contextuality is the immunity of braiding operations in topological systems to local noise, which may enable *ideal* Clifford gate operations. This corresponds well to the theoretical requirements of magic state distillation (noiseless Clifford gate operations and noisy magic states [[84](#)]). Moreover, braiding operations do not introduce

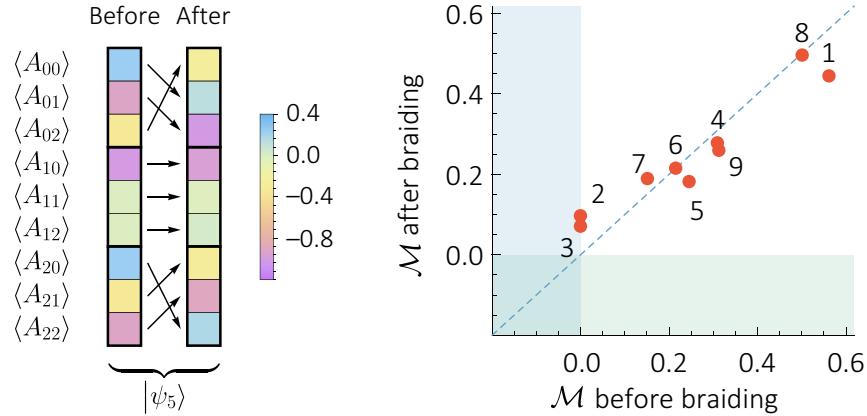


Fig. 7.7 Interplay between braiding operations and the Clifford contextuality. Left: correspondence between the nine contextuality witnesses in Equation (7.12), and the evolution of them before and after the braiding operation. Right: witness of Clifford contextuality, \mathcal{M} , before and after braiding, measured for the nine parafermionic states in [Figure 7.4](#).

additional errors to contextuality, reducing the resource consumption of magic state distillation during practical quantum computation.

To demonstrate this, we use a noise model to investigate the interaction between braiding operations and contextuality. We consider the response of three-dimensional quantum systems encoded with topological and spin codes to local hopping noise and phase noise. We will see that due to the effect of topological protection, local noise leads to the wavefunction of the topological system undergoing transitions, leaving the ground-state subspace that encodes quantum information. However, by performing projection based on imaginary-time evolution again in the ground-state subspace, simple error correction can be achieved. This error correction mechanism is similar to the qubit error correction of Gottesman–Kitaev–Preskill code in the harmonic oscillator system [190]. Below, we provide a specific description of the model we use.

First, we define the Fock fermion operators corresponding to parafermions. A Fock fermion corresponds to two parafermion modes located at the same position, with subscripts a and b , similar to how an ordinary fermion corresponds to two parafermions:

$$C_k = \frac{2}{3}\alpha_{ka} - \frac{1}{3}\sum_{m=1}^2 \omega^{m(m+1)/2}\alpha_{ka}^{m+1}\alpha_{kb}^{\dagger m}. \quad (7.13)$$

With the Fock fermion operators, we can express the noise that occurs in parafermion systems mathematically. Specifically, a hopping noise causes a Fock fermion to move to an adjacent position, and it is represented as:

$$X^{\text{Pf}} = C_k^\dagger C_{k+1}.$$

Here, we only consider the hopping noise between fermion positions 1 and 2, but the case for other positions is very similar. Using the Fradkin–Kadanoff transformation, it can be converted to the spin representation:

$$X^S = \frac{1}{9}\sigma_1^\dagger(2 - \tau_1 - \tau_1^\dagger)\tau_1\sigma_2(2 - \tau_2 - \tau_2^\dagger). \quad (7.14)$$

For a density matrix of a parafermion zero mode encoding, the effect of hopping noise with probability p is given by:

$$\rho \rightarrow X^S(p, \rho) = (1 - p)\rho + pX^S\rho X^{S\dagger}. \quad (7.15)$$

We now calculate the effects of these two noises on the computational basis. The effects on any arbitrary state can be obtained by using the evolution results of the computational basis and the superposition principle, similar to the method we used to calculate the output effects of fusion gates and other devices in Chapter 3. Note that the term projecting back to the ground-state subspace only makes sense after the evolution is completed. The results are as follows:

$$\begin{cases} \langle \psi_i^S | X^S(p, |\psi_j^S\rangle\langle\psi_j^S|) | \psi_i^S \rangle = (1 - p)\delta_{ij}, \\ \langle \psi_i^S | Z^S(q, |\psi_j^S\rangle\langle\psi_j^S|) | \psi_i^S \rangle = (1 - q)\delta_{ij}. \end{cases}$$

In other words, regardless of the strength of the local noise, all wavefunctions that are not dissipated by the noise and remain in the ground-state subspace remain completely unchanged compared to the wavefunctions before the noise occurred. We intuitively observe that topological protection makes the quantum information encoded in parafermion edge zero modes completely immune to *any* local noise!

For comparison, we consider the effects of hopping and phase noise on spin-encoded quantum systems without topological protection. In this case, corresponding to the physical meaning of the previous noises, we use clock and shift operators to construct two types of noise. Their superoperator forms are defined

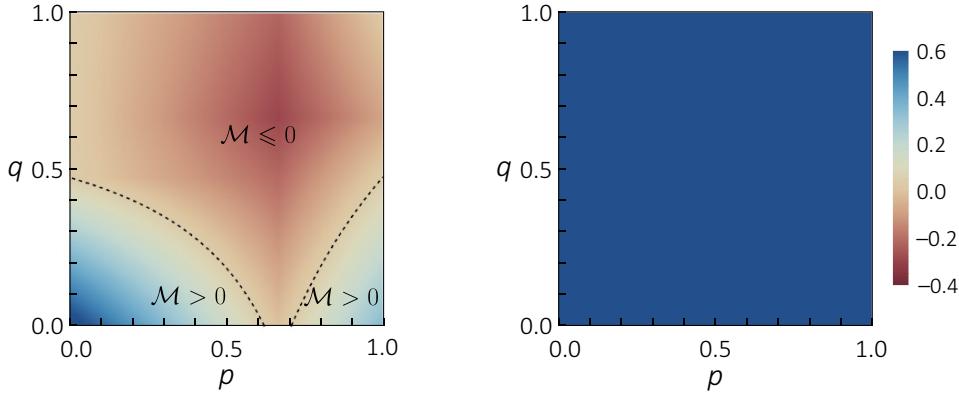


Fig. 7.8 Response of the Clifford braiding operator \mathcal{M} to noise. The figures compare the behavior of the value of the braiding operator \mathcal{M} under noise in non-topologically protected systems and topologically protected systems encoded with parafermion edge zero modes, where the dashed line represents the feasibility boundary of magic state distillation with $\mathcal{M} = 0$.

as:

$$T(p, \rho) = (1-p)\rho + \frac{p}{2}(\tau\rho\tau^\dagger + \tau^\dagger\rho\tau), \quad (7.16)$$

and

$$\Sigma(q, \rho) = (1-q)\rho + \frac{q}{2}(\sigma\rho\sigma^\dagger + \sigma^\dagger\rho\sigma). \quad (7.17)$$

In this case, due to the absence of the topological protection mechanism of imaginary-time evolution outside the ground-state subspace, both superoperators will cause changes in the form of states.

We now focus on the effects of noise on the sub-process of magic state distillation, where the most valuable aspect is the change in the strength of Clifford braiding. For spin-encoded non-topologically protected systems, we choose an initial state that is very close to the ideal magic state: $|\psi\rangle = \{1/2, 0, \sqrt{3}/2\}^\dagger$. We then apply noise constructed by the clock and shift operators, $T \circ \Sigma$, with probabilities p and q , respectively (note that these two superoperators commute). We calculate the change in the Clifford braiding operator \mathcal{M} , which characterizes the resources for magic state distillation. The results are shown in Figure 7.8, where both types of noise lead to a rapid decrease in the Clifford braiding and, when the probabilities exceed certain values, the quantum state completely loses its ability for magic state distillation. In contrast, when the same initial state is encoded with parafermion edge zero modes and noise is applied with probabilities p and q using hopping and phase operators, $X \circ Z$ (which also commute), the Clifford braiding contained in the quantum state remains completely unchanged.

Next, we observe this topological protection phenomenon experimentally. As a proof-of-concept experiment, we focus on hopping noise and noise constructed by shift operators as examples. To implement these two noises, we set both phase gates \mathcal{P} in the experimental setup to perform no operations and use rotation gates \mathcal{R} to apply the two noise superoperators X and T separately. Behind the evolution of the corresponding topological system, we also employ imaginary-time evolution to dissipate the excited state part. Finally, we use projection measurements to determine the values of the Clifford braiding operator.

The experimental results for the two systems are shown in Figure 7.9. For the system encoded with parafermion edge zero modes, the indicator of Clifford braiding obtained from the optical quantum simulation is at least $\mathcal{M} \geq 0.580 \pm 0.013$. In contrast, for the non-topologically protected system encoded

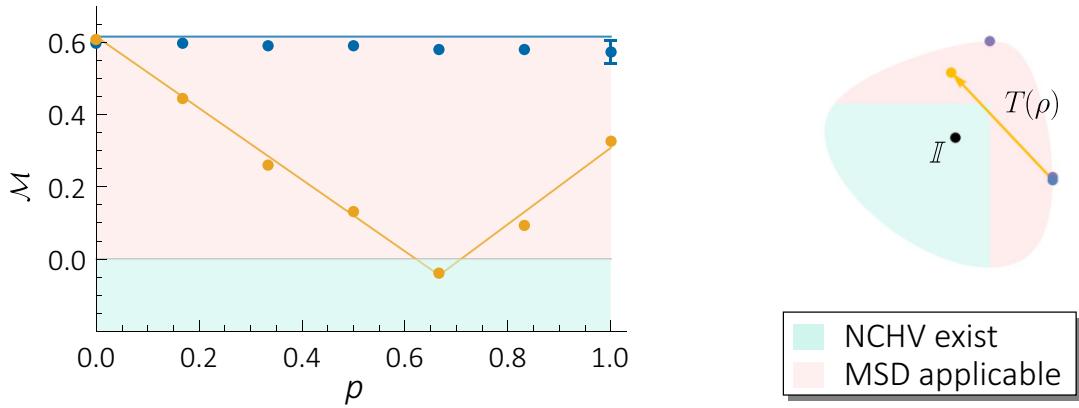


Fig. 7.9 Experimental results of topologically protected Clifford braiding. Left: measured values of Clifford braiding, with blue and yellow indicating the responses of the topological system encoded with parafermion edge zero modes and the non-topological spin-encoded system to noise, respectively. Right: trajectory of the quantum state in the state space when it is affected by noise. Note that the position of the state remains unchanged for the topological system, always staying slightly to the right. The far right and top represent the two magic states. The green region in both figures indicates that the corresponding quantum correlations can be explained by noncontextual hidden variable theories, and magic state distillation is not available. The red region has the opposite meaning.

with spins, this indicator of braiding decreases rapidly with noise and becomes less than 0 around $p = 2/3$, indicating that the quantum state cannot be used for magic state distillation. However, as the noise further increases, the braiding starts to increase again but cannot recover to its initial value. Here, we use a visualization of qutrit space as in Figure 7.9 to explain this phenomenon. By taking the two components of the braiding operator as coordinates in the state space, we construct a slice and find that as the strength of the shift operator increases, the quantum state initially departs from a magic state and approaches the maximum mixed state, which is a garbage state for magic state distillation. However, it then approaches another magic state, resulting in an increase in the Clifford braiding operator. Overall, the results indicate that the non-topological encoding of the quantum system is significantly affected by local noise, while the topological encoding of the system exhibits strong robustness against local noise. Supplementary results of this experiment are provided in Figure 7.10, further illustrating the topological protection characteristics of the parafermion edge zero mode encoding.

7.3.3 Noise-resilience of parafermion contextuality

Now, we turn to the use of contextuality as a tool to investigate the influence of local noise on the quantum state itself under two encoding schemes. Bharti *et al.* [42] found that contextuality can be used as a self-testing tool for quantum states. Here, I will provide a simplified explanation of widely used concepts such as device-independent and self-testing: self-testing refers to the fact that the maximum violation of an abstract inequality can only be achieved by a specific set of states and measurements, and any other state or measurement that leads to the maximum violation can be obtained by local operations on this specific combination. In other words, even without knowing any of the states or measurements,

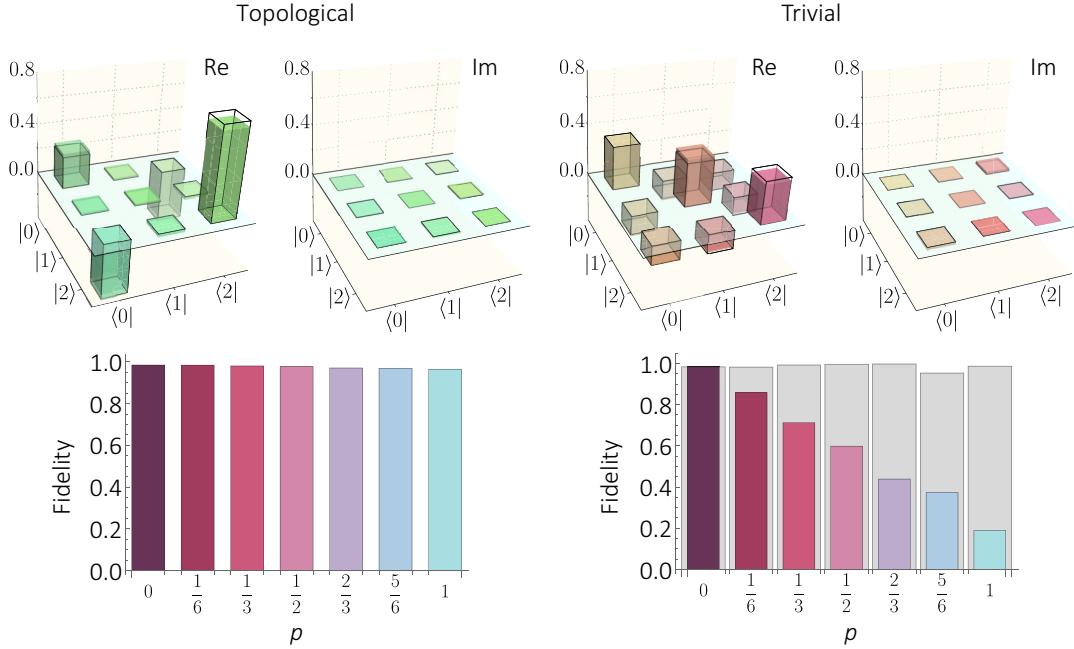


Fig. 7.10 Supplementary experimental results of topologically protected Clifford braiding. The left and right figures show the experimental results for the topological quantum system encoded with parafermion edge zero modes and the regular quantum system encoded with spins, respectively. Upper figure: After the evolution of the quantum state $|\psi\rangle$, quantum state tomography is performed to obtain the density matrix at noise probability $p = 2/3$. The solid lines and colored fillings correspond to the theoretical and experimental values, respectively. Lower figure: Fidelity between the experimental results and theoretical predictions for each data point. In the chart on the right, the gray and colored bars represent the fidelity of the final state with respect to the theoretical states, considering and not considering flip error effects, respectively.

the maximum violation immediately determines the relationship between the two; as long as one of them is known, the maximum violation immediately determines the other. For example, the previous CHSH experiment is a self-test of the maximally entangled state. Similarly, the KCBS experiment can serve as a criterion for whether a quantum state passes through the center of the pentagon measurement shown in Figure 2.4—the higher the violation of the KCBS inequality, the closer the unknown quantum state is to the state at the center of the pentagon.

In this work, we use the probability form in Equation (4.3) for the self-testing of quantum states. To ensure the completeness of each section, here is the specific form of the noncontextual hidden variable inequality once again:

$$\mathcal{K} := \Pr_{\phi}(1|k) \stackrel{\text{NCHV}}{\leq} 2. \quad (7.18)$$

We choose the eigenstates corresponding to the five projection operators of the noncontextual hidden variable inequality as:

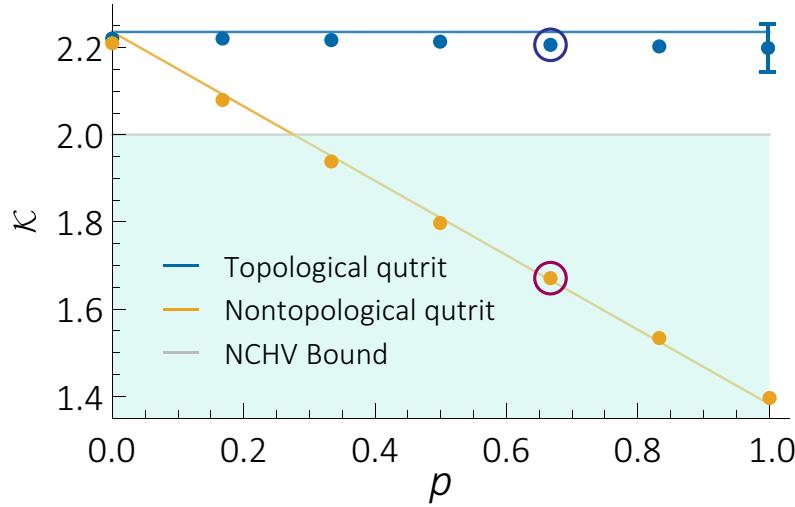


Fig. 7.11 Experimental results of topologically protected KCBS contextuality. The data points represent the values of the measured KCBS inequality (see Equation (7.18)) in response to noise. The blue and yellow colors represent the response of the topological system encoded with parafermion edge zero modes and the non-topological system encoded with spins, respectively. The green region represents the area where the noncontextual hidden variable model is not violated, and quantum states do not exhibit contextuality.

$$|k\rangle = \sqrt{1 - \frac{1}{\sqrt{5}}} \left\{ \cos\left(\frac{2k\pi}{5}\right), \sin\left(\frac{2k\pi}{5}\right), \frac{\sqrt{1+\sqrt{5}}}{2} \right\}^\dagger, \quad k = 1, 2, \dots, 5.$$

At this point, the quantum state that maximally violates the inequality is $|\phi\rangle = \{0, 0, 1\}^\dagger$. We use this state as the input to the optical quantum simulator and apply a imaginary-time evolution rotation gate \mathcal{R} to introduce noise, following the same method as the study of Clifford contextuality. Due to the self-testing nature of the KCBS inequality, when its value exceeds the theoretical bound of the noncontextual hidden variable model, the magnitude of the violation essentially provides the distance between the quantum state and the target state.

The experimental results are shown in Figure 7.11. For the system encoded with parafermion edge zero modes, the indicator of KCBS contextuality obtained by the optical quantum simulator is at least $\kappa \geq 2.199$. According to the calculations of Bharti *et al.* [42], this significant violation indicates that the quantum state encoded with parafermion edge zero modes, after experiencing noise, deviates from the target state by a trace distance of $T(\phi, \phi') \leq O(\sqrt{2.236 - 2.199}) = O(\sqrt{0.037})$ order. In contrast, in the non-topological system, noise causes the quantum state to remain far from the target state. Therefore, the value of the KCBS inequality monotonically decreases and does not re-increase after a certain critical noise intensity, as observed in the study of Clifford contextuality. The experimental results once again demonstrate, from the perspective of self-testing of quantum states, that non-topological encoded quantum systems are significantly affected by local noise, while topological encoded systems exhibit strong robustness against local noise.

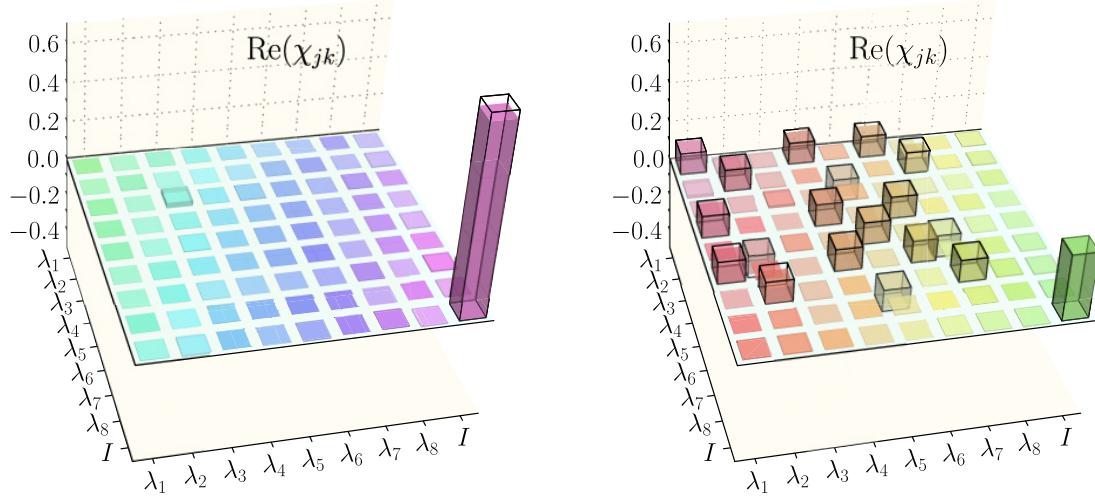


Fig. 7.12 Quantum process tomography results under noise in different systems. The left and right figures represent the evolution results of the topological system encoded with parafermion edge zero modes and the non-topological system encoded with ordinary quantum states under $p = 2/3$ noise, respectively. They are shown on the Gell-Mann matrix, displaying only the real part. The solid lines and color filling correspond to theoretical and experimental values, respectively.

Finally, we complement the self-testing results with the results of quantum process tomography, to clearly demonstrate the differences in the impact of local noise on the two different physical systems, namely, the topological encoding with parafermion edge zero modes and the non-topological encoding in the optical simulator. The two process matrices correspond to the evolution results of the topological system encoded with parafermion edge zero modes and the non-topological system encoded with ordinary quantum states under $p = 2/3$ noise. We observe that for the topological quantum system encoded with parafermion edge zero modes, the evolution process matrix is nearly an identity matrix, with a fidelity of 96.0%. This indicates that after being subjected to local noise, the wavefunction in the ground state subspace of the system undergoes almost no evolution. In contrast, the process matrix of the non-topological protected quantum system has deviated significantly from the identity matrix, indicating that local noise significantly disrupts the quantum information in the system.

7.4 Conclusion

In conclusion, the experimental research on parafermion optical quantum simulation and topologically protected quantum contextuality comes to an end. Now, I briefly discuss the significance of this experiment. Theoretically, topological systems are powerful candidates for achieving fault-tolerant quantum computation in the future. Whether based on Majorana fermions or parafermion contextuality, quantum computing systems face an obstacle in that the operations that can be achieved through contextuality belong to the Clifford group and do not constitute a universal set of quantum gate operations. To address this problem, we use magic state distillation as an approach to achieve quantum computational universality.

sality. In this architecture, we simulate two key elements of parafermion-based quantum computing: the braiding operations that generate topological Clifford gates and the topologically protected contextuality as a resource for magic state distillation. Although contextuality itself is an important resource for various quantum information tasks, the relationship between contextuality, topological systems, and noise immunity has not been well understood before. Our work opens up a new research direction in which contextuality is encoded in real topological systems and protected from environmental noise.

From a practical perspective, studying topological systems using quantum simulation not only provides new insights into topological quantum computation but also provides a testing platform for studying the fault tolerance of encoding systems. There has been some progress in this area of research. For example, Huang *et al.* [191] have used a superconducting quantum simulator to achieve the teleportation of a topologically protected qubit encoded with actively error-corrected Majorana zero modes. For our system, although we only use a single photon to encode the entire system, resulting in our simulator not being exponentially scalable, the research methodology in this experiment can be directly extended to several parafermion chains, thus enabling the testing of contextuality in composite systems consisting of two (higher-dimensional) qubits and directly benchmarking the theoretical work of Howard *et al.* [16]. Therefore, our approach is expected to be used in the future to simulate larger-scale parafermion chains and to study small-scale topological quantum algorithms.

Chapter 8

Postlude

I have introduced a series of works conducted in the research of contextuality. The experimental parts of these works have been carried out on linear optical platforms. They not only directly verify the latest theoretical advances (the first work in [chapter 4](#)), but also reflect a series of progress made by me and collaborators in theoretical research (the second work in [chapter 4](#) and [chapter 5](#)), innovative experimental methods ([chapter 6](#)), and exploration of novel physical system properties ([chapter 7](#)). These extensive and rich research contents can be seen as a microcosm of the central role of contextuality in fundamental and interdisciplinary subjects such as quantum mechanics, quantum information science, and quantum computation.

The significance of each chapter's content lies not only in the works themselves. On the contrary, each completed work brings about broader directions for further research, some of which have already achieved preliminary results. I will now look ahead to the exciting progress that can be expected in the future for the research directions corresponding to these works. Concretely:

1. In the relationship between contextuality and Bell nonlocality, we do not yet know what role contextuality beyond nonlocality and contextuality concentration can play in quantum computation and other aspects. The current stage of quantum computing prototypes almost exclusively operates on qubit systems and describes their mechanisms using the circuit model. However, high-dimensional quantum systems have demonstrated advantages in certain quantum information tasks, such as generating randomness beyond quantum bits [192]. So, is it possible to use high-dimensional quantum systems as subroutines for certain quantum algorithms (e.g., magic state distillation)? We know that Clifford contextuality is a resource for quantum computation, but does contextuality beyond Clifford class also exhibit concentration phenomena, allowing high-dimensional systems to replace multiple qubit systems and reduce resource consumption in magic state distillation?
2. How can we find more examples of contextuality concentration? The known examples are very rare, and in fact, we do not know if this is a universally existing mechanism or merely a special phenomenon caused by Clifford algebra. I can prove that for every “all-versus-nothing” paradoxical operator used for graph state verification in [chapter 5](#), a Bell inequality containing contextuality concentration phenomena can be constructed. However, these inequalities are not the strongest graph state inequalities [139], and even if all the strongest inequalities are constructed using Clifford algebra, it is not guaranteed that concentration can be achieved. Some terms need to be discarded, sacrificing the degree of violation of local hidden variable inequalities, to obtain distillable inequalities. To discover more examples of contextuality concentration and further explore their applications, it is necessary to systematically search for Bell inequalities that may exhibit similar phenomena, which requires the development of mathematical tools such as rank-constrained semi-definite programming. These tools can be applied in applications such as witnessing high-dimensional entanglement [125].

3. In the study of “all-versus-nothing” paradoxes, can we find further applications of such paradoxes in the fields of quantum information and quantum computation? Especially with the development of quantum communication technology, “all-versus-nothing” paradoxes should have certain applications in that direction. For example, Brassard *et al.* [193] discovered that Peres–Mermin squares can be used for quantum remote sensing to reduce communication complexity. Can similar techniques be applied in practical communication scenarios? Another question that tends towards fundamental scientific research: Can we find stronger, more counterintuitive, and noise-resistant “all-versus-nothing” paradoxes? For example, the GHZ paradox has a contradiction in the product of four terms, while the Peres–Mermin square requires six terms for a contradiction to occur. Is it possible to further reduce the number of terms required, such as using three terms to yield a probability distribution with a contradiction? This question already has an answer: I can prove that “all-versus-nothing” paradoxes in the form of syllogisms do exist, but currently, the simplest syllogism “all-versus-nothing” paradox requires a 37-dimensional system for verification, thereby posing a challenge for the further development of interferometer experimental techniques.
4. Is it possible for the quantum Cheshire cat effect to provide a new understanding of the physical essence of weak values in the context of pre- and post-selection experiments? Weak values are a very natural language for describing such systems and have evolved from playing the role of toy model parameters to being used in practical tasks such as wavefront sensing [184] and precision sensing [194]. However, due to its peculiar properties, this concept has been heavily criticized in the literature. In experiments involving tracking photons and tracking spin positions, the predictions provided by weak value theory can be fully reproduced by standard quantum mechanics even if nonzero weak values are not considered as indicators of which paths the particles or properties have taken [195], and there are thought experiments that provide examples that weak value theory cannot perfectly explain [196]. Therefore, to what extent can first-order weak values correspond to physical essence is a question worth further research. This research may contribute to a more intuitive and unified interpretation of quantum mechanics.
5. Is it possible to drive quantum computation in topological systems through contextuality other than magic state distillation? Obtaining quasi-particles in topological systems itself is a challenging task, requiring a breakthrough in the number of qubits from zero to one. Operations with high resource consumption, such as measurement-based quantum computation and magic state distillation, may not be feasible in the near future. However, it is already known that non-Clifford gate operations can be achieved using parafermions and interactions between parafermions [178, 197]. In these schemes, how can contextuality be utilized?
6. How can contextuality be used in near-term quantum computation? Even though we know that universal quantum computation is possible using magic state distillation, the number of qubits required for such computation is beyond the capabilities of current medium-sized noisy intermediate-scale quantum (NISQ) systems [198]. Contextuality can be considered as the most computationally difficult property to reproduce on classical computers [83]. Is it possible to leverage techniques such as variational quantum algorithms [199] combined with classical machine learning to further exploit the computational advantages of contextuality? There has been some progress in this direction [200], and studying contextuality using linear optical systems holds great potential.
7. Finally, how can we study quantum contextuality beyond the simplest cases, such as Peres–Mermin squares and KCBS experiments, outside of linear optical systems? This question addresses the essence of contextuality in quantum mechanics. It is well-known that the evolution of photon wavefunctions in single-photon interference experiments can be perfectly reproduced by the Maxwell equations, indicating that contextuality observed in linear optical systems does not necessarily require an explanation in terms of quantum mechanics. Quantum mechanics provides a possible language for interpretation. Therefore, developing experimental platforms for studying contextuality beyond linear

optics will further advance our understanding of the essence of quantum mechanics. Currently, linear optical platforms are almost the only choice for implementing precisely controllable high-dimensional systems. If alternative experimental platforms with the ability to encode high-dimensional quantum information can be developed, not only can these platforms be used to study contextuality, avoiding alternative descriptions other than quantum mechanics, but they can also directly transfer many conceptually validated results from linear optical platforms to new platforms. This can be applied to tasks such as universal, fault-tolerant, and scalable quantum computation, making contextuality a “magic” that goes beyond magic state distillation and truly advances human computational power.

I am well aware that the above perspectives are not comprehensive, and there may be omissions or biases due to my limited knowledge. However, I hope that this paper can inspire readers to think and explore in the relevant directions, ultimately promoting the development of contextuality, quantum anything¹, and even interdisciplinary sciences.

¹ Here I borrow Terry Rudolph’s wonderful highlight of his own work.

Appendix A

Case Study of Optical Quantum Information Experiment: Implementing Quantum Information Masking

In this appendix, we demonstrate how the optical experimental techniques described in chapter 3 can be applied to a specific scenario, achieving a theoretical scheme: using photons to implement quantum information masking. Since the concept of quantum information masking was proposed by Modi *et al.* [201], this field has experienced rapid development in recent years, making it impossible to cover all the research progress in this section. We kindly recommend interested readers to refer to our review article on this topic authored by our research group. Conversely, the content of this section aims to illustrate how to implement the study of parametric optical schemes on an optical platform, with the hope of providing a comprehensive view of the entire process from conceptualization and analysis to experimental design, data acquisition, and the search for applications.

A.1 Theoretical foundation

Quantum information masking¹ refers to the entangling operation performed on a quantum state ρ_s^A in a subspace of the single-party state space, such that the quantum information is diffused into the correlation between two qubits. When observing either of the two-qubit states ρ_s^{AB} , no information about the original subspace can be obtained. The operation to determine the state of an individual system in the composite system is obtained by taking the partial trace of the overall density matrix: in the above example, the states of the subsystems A and B of the two-qubit state are given by:

$$\tilde{\rho}_s^A = \text{Tr}^B \rho_s^{AB} := \sum_k {}_B \langle k | \rho_s^{AB} | k \rangle_B, \quad \tilde{\rho}_s^B = \text{Tr}^A \rho_s^{AB} := \sum_k {}_A \langle k | \rho_s^{AB} | k \rangle_A,$$

where the summation $|k\rangle$ traverses any complete orthonormal basis corresponding to the index of the system. The realization of masking operation is the implementation of a linear mapping \mathcal{U} , which transforms ρ_s^A into $\rho_s^{AB} = \mathcal{U}\rho_s^A \otimes |0\rangle\langle 0| \mathcal{U}^\dagger$, $s \in \{0, 1, 2, \dots\}$. After the application of this mapping, we have:

$$\tilde{\rho}_s^A = \tilde{\rho}_0^A, \quad \tilde{\rho}_s^B = \tilde{\rho}_0^B, \quad \forall s.$$

In the seminal work, Modi *et al.* [201] have already pointed out the relationship between quantum information masking, quantum secret sharing, and bit commitment. Therefore, it can be expected that the realization of quantum information masking will demonstrate certain value in quantum information

¹ Here we consider only two-party quantum information masking that utilizes a single ancillary system. In the following, I will omit the preceding adjectives.

research. However, Modi et al. also pointed out that it is impossible to mask the entire state space. Based on the results of theoretical research, the questions that experimental research in this direction should focus on, and also the difficulties that should be expected to be addressed, are:

- How to implement masking of quantum states in a subspace of the state space?
- The properties of the masked subspace, such as its geometric distribution in the state space.
- Applications of the information obtained from masking in a finite state space.

Shortly after the theoretical proposal, Liang *et al.* [202] pointed out that for the case of qubits, the set of masked states can be determined: the density matrices of the states that can be masked form a disk on the Bloch sphere. The most general masked set for qubits can be represented as:

$$\mathcal{D}_\alpha^\theta(\rho_0) = \{\rho : x \sin \alpha \cos \theta + y \sin \alpha \sin \theta + z \cos \alpha = c\},$$

where $c = x_0 \sin \alpha \cos \theta + y_0 \sin \alpha \sin \theta + z_0 \cos \alpha$, $\alpha \in [0, \pi]$ and $\theta \in [0, 2\pi]$. The corresponding form of the masking operation for this masked set can be written as:

$$\mathcal{U}_\alpha^\theta = \begin{pmatrix} \cos(\alpha/2) & 0 & e^{-i\theta} \sin(\alpha/2) & 0 \\ 0 & \cos(\alpha/2) & 0 & e^{-i\theta} \sin(\alpha/2) \\ 0 & \sin(\alpha/2) & 0 & -e^{-i\theta} \cos(\alpha/2) \\ \sin(\alpha/2) & 0 & -e^{-i\theta} \cos(\alpha/2) & 0 \end{pmatrix}. \quad (\text{A.1})$$

A.2 Experimental design

Now that we have mathematical guidance for theoretically implementing two-party quantum information masking, we can consider how to construct it on a linear optical platform. The paradigmatic thought process is as follows:

1. First, consider in which degree of freedom it can be implemented. The answer should be polarization since currently only the polarization degree of freedom allows the implementation of entangling gate operations for two photons with coincidence counting.
2. Second, note that by combining wave plates, any unitary operation on a single polarization qubit can be realized, and unitary operations correspond to rotations on the Bloch sphere, which preserve the disk shape. Therefore, it is sufficient to implement any masking operation $\mathcal{U}_{\alpha^*}^{\theta^*}$ and combine it with rotations of the polarization qubit to achieve arbitrary masking operations.
3. Now consider how to select special values for α^* and θ^* . We find that when $\alpha = 0$ and $\theta = \pi$, we have:

$$\mathcal{U}_0^\pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \text{CNOT}.$$

This means that by combining CZ and Hadamard gates, masking operations can be achieved, and we already know how to implement the most challenging CZ gate.

4. Finally, since the controlled quantum state $|0\rangle\langle 0|$ is fixed, we can consider how to further optimize the experimental design to reduce the dependence of the CZ gate on the accuracy of time evolution. By comparing the mapping of fusion gates and CNOT gates for the four possible input operators, we find that they only differ by a sign when the control photon is vertically polarized. After trying the replacement, we find that the functionality achieved by the fusion gate combined with a Hadamard

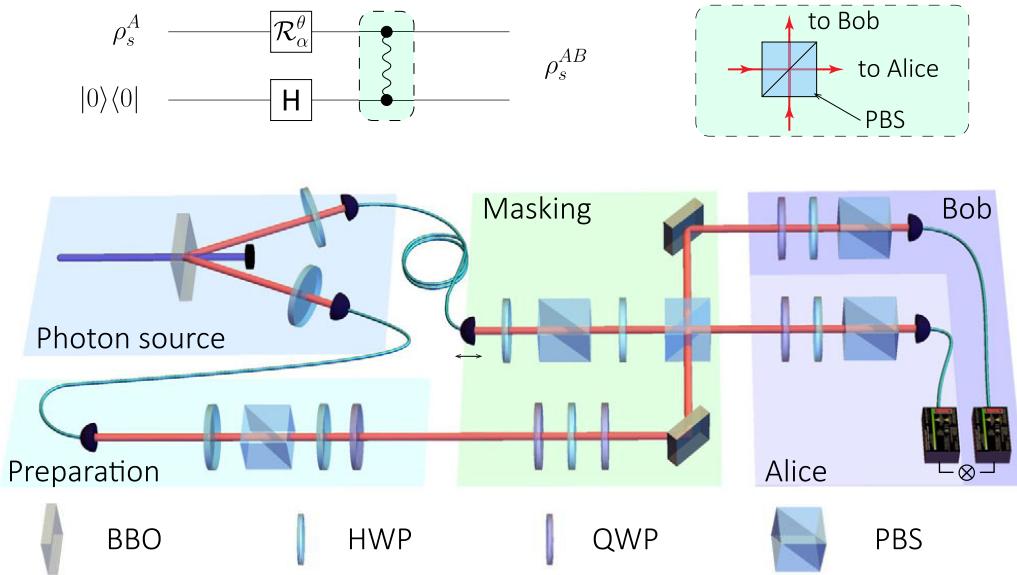


Fig. A.1 Implementation of quantum information masking in linear optics. *Top:* Circuit diagram for implementing quantum information masking, with the dashed line indicating the fusion gate and its optical implementation. *Bottom:* Overall experimental setup, including state preparation, arbitrary unitary rotations of polarization qubits, fusion, and coincidence detection.

gate $H = (\sigma_x + \sigma_z)/\sqrt{2}$ applied to the ancillary qubit is equal to \mathcal{U}_0^0 . Therefore, we use the fusion gate with the fewest components as the core of the quantum information masking optical path.

Finally, the optical path we designed for implementing polarization-encoded qubit quantum information masking in linear optics is shown in Figure A.1. By pumping a type-II beam-like nonlinear crystal, a pair of photons in a polarization product state is generated. One of the photons is used as an ancillary photon. By using a half-wave plate followed by a Hadamard operation, its polarization is prepared in the maximum superposition state $|D\rangle$ of the computational basis. The other photon is projected onto $|H\rangle$ using a polarizing beam splitter, and then prepared to any desired polarization state using a half-wave plate and a quarter-wave plate. Next, we enter the masking operation part. First, a set of quantum states to be masked is selected using a wave plate group, and this set is transformed to the equator plane of the Bloch sphere with the same $\langle\sigma_z\rangle$ value. Then, the fusion gate is used to interfere the two photons, and this achieves the masking of quantum information in the part where one photon is output on each side. To recover this part of information, we use two polarization analyzers and coincidence detection to perform joint measurements on the two-photon polarization qubits.

A.3 Results

After setting up the experimental apparatus, what data should we measure? In order to clearly demonstrate the working principle of the device and show the working effect, we need to describe the efficacy of the masking mechanism from both qualitative and quantitative aspects. "Qualitative" refers to key predictions

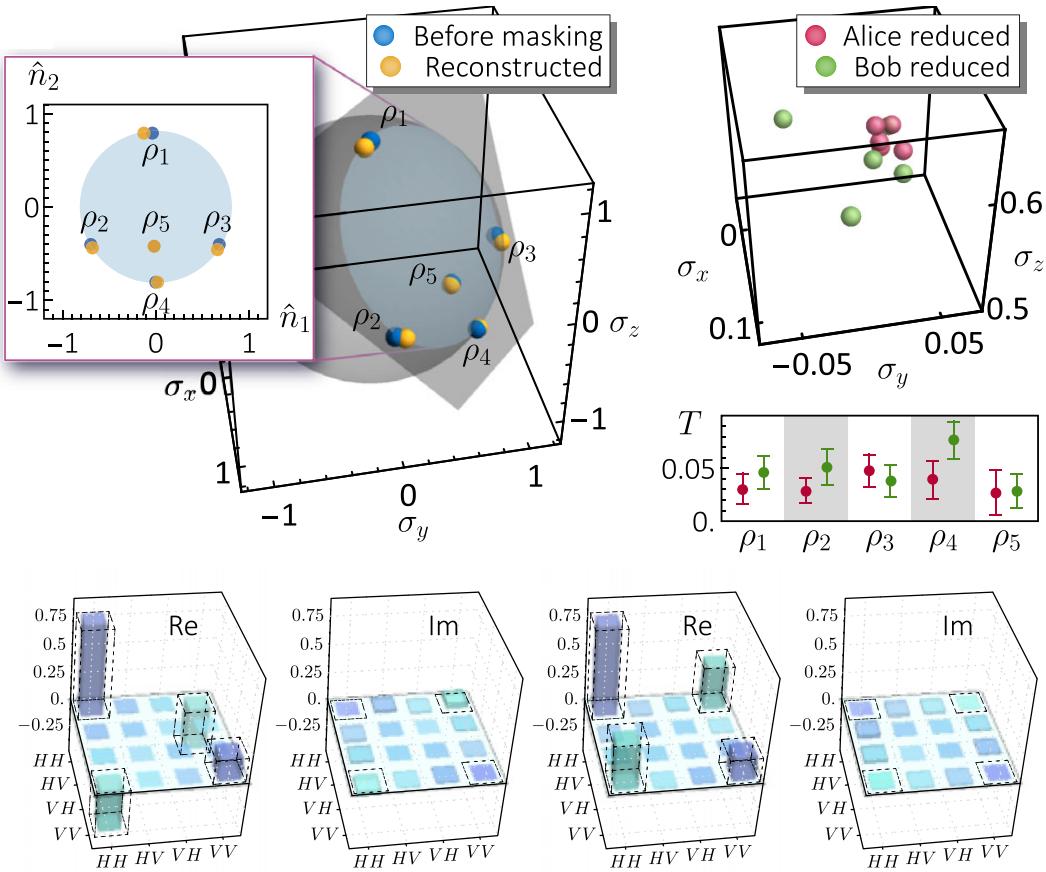


Fig. A.2 Evolution of quantum states in the masking operation. *Top left:* Pure states $\rho_1 \sim \rho_4$ and mixed state ρ_5 on the masked disk. Blue and orange dots represent the reconstructed initial and final states from quantum state tomography. The insets show the projection of these quantum states on the masked disk, with the horizontal and vertical coordinates given by $(-\hat{x} + \hat{y})/\sqrt{2}$ and $(-\hat{x} - \hat{y} + 2\hat{z})/\sqrt{6}$, respectively. *Top right:* Experimentally measured reduced single-particle states after masking, and the trace distance from their theoretical values. *Bottom:* Two-body density matrices obtained from masking ρ_1 and ρ_4 . Solid and dashed lines represent the experimental values and theoretical predictions, respectively.

that can be intuitively reproduced and are critical in illustrating the geometric form of the masked set and the fact that the individual states after masking do not carry any information from the subspace. "Quantitative" refers to providing specific performance of the experimental apparatus. In this case, we can study the fidelity of the masking operation, the form of the resulting two-party states after masking, and the robustness of the masking operation's effectiveness.

After analyzing the above two aspects, we can roughly divide the required data into two parts:

1. Evolution of quantum states in the masking operation. Specifically, this includes the density matrices of the masked quantum states before and after the masking operation. For the quantum states before masking, a comparison is made with the theoretical predictions of the masked set from a geometric

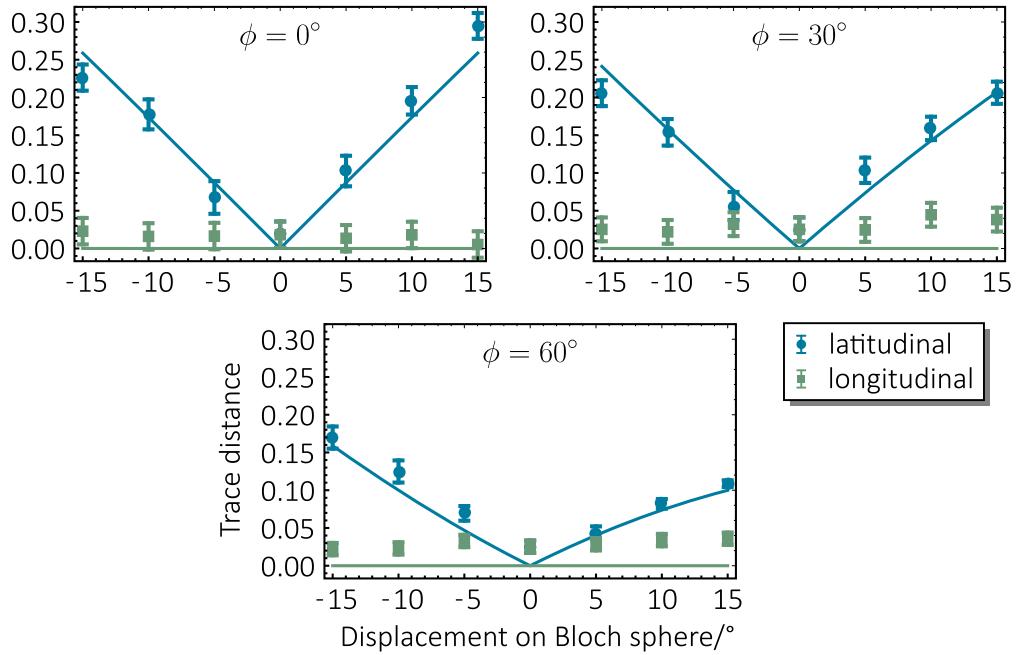


Fig. A.3 Response of the masking operation to quantum state perturbation. *Masked disk:* Intersection of the plane $x+y+1 = 1$ and the unit sphere. Each plot represents the trace distance between the reduced single-particle density matrix and the reference state when the initial state is slightly moved away from the reference point. Cases corresponding to displacement along the meridian and parallel to the meridian on the Bloch sphere are marked as green and cyan dots (experimental data) and curves (theoretical values), respectively. The error bars correspond to 1σ standard deviation estimated from Poissonian counting statistics.

perspective. For the quantum states after masking, the following is provided: (1) The form of the reduced single-particle density matrices $\tilde{\rho}_s^{A/B}$ and the deviation from their respective theoretical results $\tilde{\rho}_0^{A/B}$. To complement the geometric representation, the discrepancy is characterized using the trace distance, defined as:

$$T(\rho, \rho') := \frac{1}{2} \|\rho - \rho'\|_1 = \frac{1}{2} \text{Tr} \left[\sqrt{(\rho - \rho')^\dagger (\rho - \rho')} \right],$$

so that $T(\tilde{\rho}_s^{A/B}, \tilde{\rho}_0^{A/B}) \rightarrow 0$ indicates that the single-particle quantum states do not carry any information from the subspace. (2) The form of the two-body density matrices and the information of the initial states recovered by numerical calculation based on the inverse operation of masking, demonstrating that the information is completely preserved in the two-body correlations.

2. Response of the masking operation to quantum state perturbation. Specifically, this includes the changes in the single-particle density matrix compared to the reference value when the masked quantum states are moved within and outside the masked subset. The theoretical curves are computed and compared with experimental data points, corresponding to the geometric properties of the masked set. For clarity,

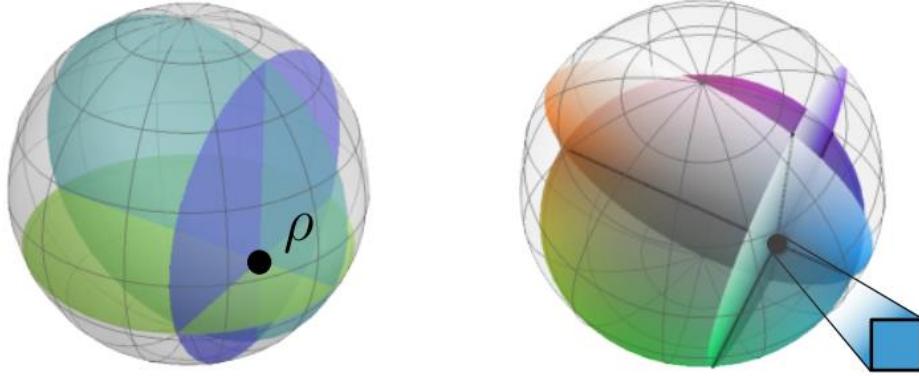


Fig. A.4 Isomorphism between the Bloch sphere and color space. In the experiment of secret sharing using quantum information masking, a partially reduced state after masking is sent to three different receivers, who can then make the three disks in the geometric representation of the masked set intersect at one point, thus recovering the original quantum state.

we choose the implemented masking operation as \mathcal{U}_0^0 , where the masked set is the meridian plane $\phi = \text{const.}$ on the Bloch sphere, and the direction away from the masked set is the parallel direction, facilitating intuitive understanding and data representation.

The data required for the two parts described above are represented in Figures A.2 and A.3, respectively. In addition, key data and computational results are provided in the paper, such as the average fidelity of 97.7% between the masked two-body quantum states and the target states, and the average trace distance of 1.55×10^{-2} between the single-particle states and the target states, etc. In this way, not only can the experimental results correspond to the theoretical predictions, directly demonstrating the successful implementation of the required functionality and validating the theoretical expectations, but also a large number of quantitative results can be obtained, proving the high precision and effectiveness of the device.

A.4 Applications

After completing the initial data measurements, how can we explore the potential applications of the device in quantum information processing and quantum communication? We start by considering the functionalities that the device can achieve and combine them with the guidance from theoretical work to address this question. The work by Modi *et al.* [201] suggests that quantum information masking can find applications in quantum secret sharing, and Liang *et al.* [202] has proposed a scheme for secret sharing based on geometric properties. Specifically, the operation involves using two different masking maps to mask the same pure state and sending a partially reduced state after masking to two different receivers. Each receiver alone cannot obtain any information about the initial state before masking, but by comparing their received states, the two receivers can make the disks in the geometric representation of the masked set intersect at one point, thus recovering the original quantum state.

Compared to previous theoretical predictions, the device has further achieved masking of mixed states, which was not studied in previous works. It extends the range of the masked set from a circular ring on the surface of the Bloch sphere to the entire disk inside the sphere. Therefore, applications can be further developed based on this achievement. Specifically, it is worth noting that the unit sphere can be used to

encode all colors, as describing any color only requires three parameters, and the parameter space has the same topological structure as the unit sphere, as shown in [Figure A.4](#). The specific correspondence can be found on websites such as Wikipedia. Therefore, a very intuitive and conceptually validating application is to encode colors using quantum states and sequentially encrypt and recover these states, which is equivalent to transmitting an image pixel by pixel. In each step, three different masking maps are applied to the same quantum state, and by using the single-particle reduced state obtained from the measurement of one side of the detector, the position of the disk can be determined. Then, by comparing the coordinates of the three disks through classical communication, the original quantum state can be recovered. Using this method, we demonstrate an example: the sharing and secure transmission of an image of a bird in [Figure A.5](#).

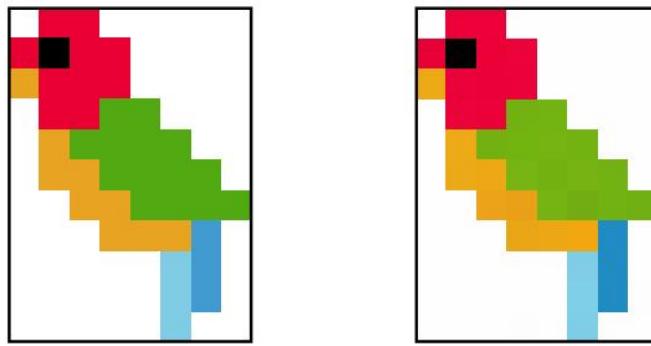


Fig. A.5 Sharing a colored image using a secret sharing protocol. The left image is the original image, and the right image is the image recovered by the receiver.

References

- [1] P. W. Shor, SIAM review **41**, 303 (1999).
- [2] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Nature **574**, 505 (2019).
- [3] J. Arrazola, V. Bergholm, K. Brádler, T. Bromley, M. Collins, I. Dhand, A. Fumagalli, T. Gerrits, A. Goussev, L. Helt, *et al.*, Nature **591**, 54 (2021).
- [4] N. D. Mermin, Reviews of Modern Physics **65**, 803 (1993).
- [5] D. Bohm, Physical Review **85**, 166 (1952).
- [6] A. Einstein, B. Podolsky, and N. Rosen, Physical Review **47**, 777 (1935).
- [7] R. W. Spekkens, Physical Review A **71**, 052108 (2005).
- [8] A. Fine, Physical Review Letters **48**, 291 (1982).
- [9] A. M. Gleason, Journal of Mathematics and Mechanics **6**, 885 (1957).
- [10] S. Kochen and E. Specker, Journal of Mathematics and Mechanics **17**, 59 (1967).
- [11] A. Peres, Journal of Physics A: Mathematical and General **24**, L175 (1991).
- [12] A. Cabello, J. Estebaranz, and G. García-Alcaine, Physics Letters A **212**, 183 (1996).
- [13] A. Cabello, Physical Review Letters **101**, 210401 (2008).
- [14] S. Yu and C. H. Oh, Physical Review Letters **108**, 030402 (2012).
- [15] A. Cabello, S. Severini, and A. Winter, Physical Review Letters **112**, 040401 (2014).
- [16] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature **510**, 351 (2014).
- [17] Z.-P. Xu, J.-L. Chen, and O. Gühne, Physical Review Letters **124**, 230401 (2020).
- [18] Y.-F. Huang, C.-F. Li, Y.-S. Zhang, J.-W. Pan, and G.-C. Guo, Physical Review Letters **90**, 250401 (2003).
- [19] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos, Nature **460**, 494 (2009).
- [20] E. Amselem, M. Rådmank, M. Bourennane, and A. Cabello, Physical Review Letters **103**, 160405 (2009).
- [21] V. D'Ambrosio, I. Herbauts, E. Amselem, E. Nagali, M. Bourennane, F. Sciarrino, and A. Cabello, Physical Review X **3**, 011012 (2013).
- [22] X.-M. Hu, J.-S. Chen, B.-H. Liu, Y. Guo, Y.-F. Huang, Z.-Q. Zhou, Y.-J. Han, C.-F. Li, and G.-C. Guo, Physical Review Letters **117**, 170403 (2016).
- [23] P. Wang, J. Zhang, C.-Y. Luan, M. Um, Y. Wang, M. Qiao, T. Xie, J.-N. Zhang, A. Cabello, and K. Kim, Science Advances **8**, eabk1660 (2021).
- [24] D. A. Meyer, Physical Review Letters **83**, 3751 (1999).
- [25] A. Kent, Physical Review Letters **83**, 3755 (1999).

- [26] R. Clifton and A. Kent, Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **456**, 2101 (2000).
- [27] Y. Hasegawa, R. Loidl, G. Badurek, M. Baron, and H. Rauch, Nature **425**, 45 (2003).
- [28] H. Bartosik, J. Klepp, C. Schmitzer, S. Sponar, A. Cabello, H. Rauch, and Y. Hasegawa, Physical Review Letters **103**, 040403 (2009).
- [29] A. Cabello, M. Kleinmann, and C. Budroni, Physical Review Letters **114**, 250402 (2015).
- [30] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, Reviews of modern physics **79**, 135 (2007).
- [31] B. Bartlett, A. Dutt, and S. Fan, Optica **8**, 1515 (2021).
- [32] S. Yokoyama, R. Ukai, S. C. Armstrong, C. Sornphiphatphong, T. Kaji, S. Suzuki, J.-i. Yoshikawa, H. Yonezawa, N. C. Menicucci, and A. Furusawa, Nature Photonics **7**, 982 (2013).
- [33] M. V. Larsen, X. Guo, C. R. Breum, J. S. Neergaard-Nielsen, and U. L. Andersen, Science **366**, 369 (2019).
- [34] R. Raussendorf and H. J. Briegel, Physical Review Letters **86**, 5188 (2001).
- [35] E. Schrödinger, Naturwissenschaften **23**, 844 (1935).
- [36] M. A. Nielsen and I. Chuang, *Quantum computation and quantum information* (Cambridge University Press, 2000).
- [37] E. J. Bergholtz, J. C. Budich, and F. K. Kunst, Reviews of Modern Physics **93**, 015005 (2021).
- [38] G.-C. Wick, Physical Review **96**, 1124 (1954).
- [39] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Physical Review Letters **101**, 020403 (2008).
- [40] B. Liu, Y. Huang, Y. Gong, F. Sun, Y. Zhang, C. Li, and G. Guo, Physical Review A **80**, 044101 (2009).
- [41] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson, arXiv preprint arXiv:2102.13036 (2021).
- [42] K. Bharti, M. Ray, A. Varvitsiotis, N. A. Warsi, A. Cabello, and L.-C. Kwek, Physical Review Letters **122**, 250403 (2019).
- [43] L. Lovász, M. Saks, and A. Schrijver, Linear Algebra and its applications **114**, 439 (1989).
- [44] L. Lovász, IEEE Transactions on Information theory **25**, 1 (1979).
- [45] A. Cabello, Physical Review A **93**, 032102 (2016).
- [46] Y. Xiao, Z.-P. Xu, Q. Li, J.-S. Xu, K. Sun, J.-M. Cui, Z.-Q. Zhou, H.-Y. Su, A. Cabello, J.-L. Chen, C.-F. Li, and G.-C. Guo, Optics Express **26**, 32 (2018).
- [47] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Physical review letters **102**, 010401 (2009).
- [48] D. Schmid and R. W. Spekkens, Physical Review X **8**, 011015 (2018).
- [49] O. Gühne, C. Budroni, A. Cabello, M. Kleinmann, and J.-Å. Larsson, Physical Review A **89**, 062107 (2014).
- [50] M. Um, Q. Zhao, J. Zhang, P. Wang, Y. Wang, M. Qiao, H. Zhou, X. Ma, and K. Kim, Physical Review Applied **13**, 034077 (2020).
- [51] M.-H. Li, X. Zhang, W.-Z. Liu, S.-R. Zhao, B. Bai, Y. Liu, Q. Zhao, Y. Peng, J. Zhang, Y. Zhang, et al., Physical Review Letters **126**, 050503 (2021).
- [52] N. Harrigan and R. W. Spekkens, Foundations of Physics **40**, 125 (2010).
- [53] M. S. Leifer and O. J. Maroney, Physical Review Letters **110**, 120401 (2013).
- [54] M. F. Pusey, J. Barrett, and T. Rudolph, Nature Physics **8**, 475 (2012).
- [55] J. Barrett, E. G. Cavalcanti, R. Lal, and O. J. Maroney, Physical Review Letters **112**, 250403 (2014).
- [56] M. Ringbauer, B. Duffus, C. Branciard, E. G. Cavalcanti, A. G. White, and A. Fedrizzi, Nature Physics **11**, 249 (2015).

- [57] J. Pearl, *Causality* (Cambridge university press, 2009).
- [58] E. G. Cavalcanti, Physical Review X **8**, 021018 (2018).
- [59] J. Pearl and E. Cavalcanti, Quantum **5**, 518 (2021).
- [60] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Physical Review X **7**, 031021 (2017).
- [61] L. Vaidman, Y. Aharonov, and D. Z. Albert, Physical Review Letters **58**, 1385 (1987).
- [62] Y. Aharonov, F. Colombo, S. Popescu, I. Sabadini, D. C. Struppa, and J. Tollaksen, Proceedings of the National Academy of Sciences **113**, 532 (2016).
- [63] Y. Aharonov, S. Popescu, D. Rohrlich, and P. Skrzypczyk, New Journal of Physics **15**, 113015 (2013).
- [64] N. D. Mermin, Physical Review Letters **74**, 831 (1995).
- [65] M. S. Leifer and R. W. Spekkens, Physical Review Letters **95**, 200405 (2005).
- [66] S. Yu and C. Oh, arXiv preprint arXiv:1408.2477 (2014).
- [67] M. Waegell and J. Tollaksen, Quantum Studies: Mathematics and Foundations **5**, 325 (2018).
- [68] J. S. Bell, Physics Physique Fizika **1**, 195 (1964).
- [69] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, *et al.*, Nature **526**, 682 (2015).
- [70] M. Giustina, M. A. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, *et al.*, Physical Review Letters **115**, 250401 (2015).
- [71] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, *et al.*, Physical Review Letters **115**, 250402 (2015).
- [72] B. Bell Test Collaboration *et al.*, Nature **557**, 212 (2018).
- [73] A. Cabello, Physical Review Letters **86**, 1911 (2001).
- [74] A. Cabello, Physical Review Letters **127**, 070401 (2021).
- [75] M. Sadiq, P. Badziąg, M. Bourennane, and A. Cabello, Physical Review A **87**, 012128 (2013).
- [76] D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's theorem, quantum theory and conceptions of the universe* (Springer, 1989) pp. 69–72.
- [77] L. Hardy, Physical Review Letters **71**, 1665 (1993).
- [78] S. L. Braunstein and P. Van Loock, Reviews of modern physics **77**, 513 (2005).
- [79] R. W. Spekkens, Physical Review Letters **101**, 020401 (2008).
- [80] H. J. Briegel and R. Raussendorf, Physical Review Letters **86**, 910 (2001).
- [81] R. Raussendorf, Physical Review A **88**, 022322 (2013).
- [82] S. Bravyi, D. Gosset, and R. König, Science **362**, 308 (2018).
- [83] S. Bravyi, D. Gosset, R. König, and M. Tomamichel, Nature Physics **16**, 1040 (2020).
- [84] S. Bravyi and A. Kitaev, Physical Review A **71**, 022316 (2005).
- [85] D. Gottesman, arXiv preprint quant-ph/9807006 (1998).
- [86] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Physical Review Letters **119**, 170501 (2017).
- [87] J. B. Altepeter, E. R. Jeffrey, and P. G. Kwiat, Advances in Atomic, Molecular, and Optical Physics **52**, 105 (2005).
- [88] J. L. O'Brien, G. Pryde, A. Gilchrist, D. James, N. K. Langford, T. Ralph, and A. White, Physical Review Letters **93**, 080502 (2004).
- [89] F. Flamini, N. Spagnolo, and F. Sciarrino, Reports on Progress in Physics **82**, 016001 (2018).
- [90] S. Takeuchi, Optics Letters **26**, 843 (2001).
- [91] C. Zhang, Y.-F. Huang, C.-J. Zhang, J. Wang, B.-H. Liu, C.-F. Li, and G.-C. Guo, Optics Express **24**, 27059 (2016).
- [92] R. Drever, J. L. Hall, F. Kowalski, J. Hough, G. Ford, A. Munley, and H. Ward, Applied Physics B **31**, 97 (1983).

- [93] X.-M. Hu, W.-B. Xing, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, P. Erker, and M. Huber, Physical Review Letters **125**, 090503 (2020).
- [94] L. Li, Z. Liu, X. Ren, S. Wang, V.-C. Su, M.-K. Chen, C. H. Chu, H. Y. Kuo, B. Liu, W. Zang, G.-C. Guo, L. Zhang, Z. Wang, S. Zhu, and D. P. Tsai, Science **368**, 1487 (2020).
- [95] X.-Y. Xu, Q.-Q. Wang, W.-W. Pan, K. Sun, J.-S. Xu, G. Chen, J.-S. Tang, M. Gong, Y.-J. Han, C.-F. Li, and G.-C. Guo, Physical Review Letters **120**, 260501 (2018).
- [96] S. Ecker, F. Bouchard, L. Bulla, F. Brandt, O. Kohout, F. Steinlechner, R. Fickler, M. Malik, Y. Guryanova, R. Ursin, and M. Huber, Physical Review X **9**, 041042 (2019).
- [97] L. Allen, M. W. Beijersbergen, R. Spreeuw, and J. Woerdman, Physical Review A **45**, 8185 (1992).
- [98] L. Marrucci, C. Manzo, and D. Paparo, Physical Review Letters **96**, 163905 (2006).
- [99] B.-G. Englert, C. Kurtsiefer, and H. Weinfurter, Physical Review A **63**, 032303 (2001).
- [100] D. E. Browne and T. Rudolph, Physical Review Letters **95**, 010501 (2005).
- [101] T. Bodiya and L.-M. Duan, Physical Review Letters **97**, 143601 (2006).
- [102] K. Wang, X. Zhan, Z. Bian, J. Li, Y. Zhang, and P. Xue, Physical Review A **93**, 052108 (2016).
- [103] E. Bolduc, N. Bent, E. Santamato, E. Karimi, and R. W. Boyd, Optics letters **38**, 3546 (2013).
- [104] W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
- [105] J.-S. Xu, M.-H. Yung, X.-Y. Xu, S. Boixo, Z.-W. Zhou, C.-F. Li, A. Aspuru-Guzik, and G.-C. Guo, Nature Photonics **8**, 113 (2014).
- [106] J.-S. Xu, K. Sun, Y.-J. Han, C.-F. Li, J. K. Pachos, and G.-C. Guo, Nature Communications **7**, 13194 (2016).
- [107] J.-S. Xu, K. Sun, J. K. Pachos, Y.-J. Han, C.-F. Li, and G.-C. Guo, Science advances **4**, eaat6533 (2018).
- [108] N. K. Langford, T. Weinhold, R. Prevedel, K. Resch, A. Gilchrist, J. O'Brien, G. Pryde, and A. White, Physical Review Letters **95**, 210504 (2005).
- [109] N. Kiesel, C. Schmid, U. Weber, R. Ursin, and H. Weinfurter, Physical Review Letters **95**, 210505 (2005).
- [110] R. Okamoto, H. F. Hofmann, S. Takeuchi, and K. Sasaki, Physical Review Letters **95**, 210506 (2005).
- [111] Y. Aharonov, D. Z. Albert, and L. Vaidman, Physical Review Letters **60**, 1351 (1988).
- [112] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd, Reviews of Modern Physics **86**, 307 (2014).
- [113] J.-S. Tang, Y.-T. Wang, S. Yu, D.-Y. He, J.-S. Xu, B.-H. Liu, G. Chen, Y.-N. Sun, K. Sun, Y.-J. Han, C.-F. Li, and G.-C. Guo, Nature Photonics **10**, 642 (2016).
- [114] L. Xiao, X. Zhan, Z. Bian, K. Wang, X. Zhang, X. Wang, J. Li, K. Mochizuki, D. Kim, N. Kawakami, *et al.*, Nature Physics **13**, 1117 (2017).
- [115] L. Xiao, T. Deng, K. Wang, G. Zhu, Z. Wang, W. Yi, and P. Xue, Nature Physics **16**, 761 (2020).
- [116] C.-K. Hong, Z.-Y. Ou, and L. Mandel, Physical Review Letters **59**, 2044 (1987).
- [117] N. Bent, H. Qassim, A. Tahir, D. Sych, G. Leuchs, L. L. Sánchez-Soto, E. Karimi, and R. Boyd, Physical Review X **5**, 041006 (2015).
- [118] M. Navascués, S. Pironio, and A. Acín, Physical Review Letters **98**, 010401 (2007).
- [119] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, Nature Communications **6**, 6288 (2015).
- [120] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Physical Review Letters **23**, 880 (1969).
- [121] N. D. Mermin, American Journal of Physics **62**, 880 (1994).
- [122] Y. Xiao, Z.-P. Xu, Q. Li, H.-Y. Su, K. Sun, A. Cabello, J.-S. Xu, J.-L. Chen, C.-F. Li, and G.-C. Guo, Optica **5**, 718 (2018).
- [123] K. F. Pál and T. Vértesi, Physical Review A **82**, 022116 (2010).
- [124] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Physical Review Letters **114**, 140403 (2015).

- [125] M. Ray, N. G. Boddu, K. Bharti, L.-C. Kwek, and A. Cabello, New Journal of Physics **23**, 033006 (2021).
- [126] N. D. Mermin, Physical Review Letters **65**, 1838 (1990).
- [127] M. Ardehali, Physical Review A **46**, 5375 (1992).
- [128] T. Vidick and S. Wehner, Physical Review Letters **107**, 030402 (2011).
- [129] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Reviews of Modern Physics **86**, 419 (2014).
- [130] E. A. Aguilar, M. Farkas, D. Martínez, M. Alvarado, J. Cariñe, G. B. Xavier, J. F. Barra, G. Cañas, M. Pawłowski, and G. Lima, Physical Review Letters **120**, 230503 (2018).
- [131] D. Pierangeli, G. Marcucci, and C. Conti, Physical Review Letters **122**, 213902 (2019).
- [132] J. Lee Rodgers and W. A. Nicewander, The American Statistician **42**, 59 (1988).
- [133] A. Cabello, P. Badziag, M. T. Cunha, and M. Bourennane, Physical Review Letters **111**, 180404 (2013).
- [134] W. Asavanant, Y. Shiozawa, S. Yokoyama, B. Charoensombutamon, H. Emura, R. N. Alexander, S. Takeda, J.-i. Yoshikawa, N. C. Menicucci, H. Yonezawa, *et al.*, Science **366**, 373 (2019).
- [135] M. V. Larsen, X. Guo, C. R. Breum, J. S. Neergaard-Nielsen, and U. L. Andersen, Nature Physics **17**, 1018 (2021).
- [136] N. C. Menicucci, Physical Review A **83**, 062314 (2011).
- [137] O. Gühne, G. Tóth, P. Hyllus, and H. J. Briegel, Physical Review Letters **95**, 120405 (2005).
- [138] G. Tóth, O. Gühne, and H. J. Briegel, Physical Review A **73**, 022303 (2006).
- [139] A. Cabello, O. Gühne, and D. Rodríguez, Physical Review A **77**, 062106 (2008).
- [140] F. Baccari, R. Augusiak, I. Šupić, J. Tura, and A. Acín, Physical Review Letters **124**, 020402 (2020).
- [141] A. Cabello, Physical Review Letters **95**, 210401 (2005).
- [142] V. Scarani, A. Acín, E. Schenck, and M. Aspelmeyer, Physical Review A **71**, 042325 (2005).
- [143] Z. Chen, Physical Review A **68**, 052106 (2003).
- [144] M. Hein, J. Eisert, and H. J. Briegel, Physical Review A **69**, 062311 (2004).
- [145] Z.-E. Su, W.-D. Tang, D. Wu, X.-D. Cai, T. Yang, L. Li, N.-L. Liu, C.-Y. Lu, M. Żukowski, and J.-W. Pan, Physical Review A **95**, 030103 (2017).
- [146] W. Tang, S. Yu, and C. Oh, Physical Review Letters **110**, 100403 (2013).
- [147] J. C. Adcock, S. Morley-Short, A. Dahlberg, and J. W. Silverstone, Quantum **4**, 305 (2020).
- [148] D. P. DiVincenzo and P. W. Shor, Physical Review Letters **77**, 3260 (1996).
- [149] Z.-B. Chen, J.-W. Pan, Y.-D. Zhang, Č. Brukner, and A. Zeilinger, Physical Review Letters **90**, 160408 (2003).
- [150] L. C. Ryff, American Journal of Physics **65**, 1197 (1997).
- [151] J.-L. Chen, X.-J. Ye, C. Wu, H.-Y. Su, A. Cabello, L. C. Kwek, and C. H. Oh, Scientific reports **3**, 1 (2013).
- [152] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Physical Review Letters **98**, 140402 (2007).
- [153] K. Sun, J.-S. Xu, X.-J. Ye, Y.-C. Wu, J.-L. Chen, C.-F. Li, and G.-C. Guo, Physical Review Letters **113**, 140402 (2014).
- [154] L. Carroll, *Alice's adventures in Wonderland* (Maecenas Press, 1969).
- [155] T. Denkmayr, H. Geppert, S. Sponar, H. Lemmel, A. Matzkin, J. Tollaksen, and Y. Hasegawa, Nature Communications **5**, 4492 (2014).
- [156] D. P. Atherton, G. Ranjit, A. A. Geraci, and J. D. Weinstein, Optics letters **40**, 879 (2015).
- [157] J. M. Ashby, P. D. Schwarz, and M. Schlosshauer, Physical Review A **94**, 012102 (2016).
- [158] R. Corrêa, M. F. Santos, C. Monken, and P. L. Saldanha, New Journal of Physics **17**, 053042 (2015).
- [159] D. Das and A. K. Pati, New Journal of Physics **22**, 063032 (2020).

- [160] G. Pryde, J. O'Brien, A. White, T. Ralph, and H. Wiseman, Physical Review Letters **94**, 220405 (2005).
- [161] A. Danan, D. Farfurnik, S. Bar-Ad, and L. Vaidman, Physical Review Letters **111**, 240402 (2013).
- [162] L. Vaidman, Physical Review A **87**, 052104 (2013).
- [163] C. Ferrie and J. Combes, Physical Review Letters **113**, 120404 (2014).
- [164] Y. Kim, D.-G. Im, Y.-S. Kim, S.-W. Han, S. Moon, Y.-H. Kim, and Y.-W. Cho, npj Quantum Information **7**, 1 (2021).
- [165] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Physical Review Letters **70**, 1895 (1993).
- [166] P. Chowdhury, A. K. Pati, and J.-L. Chen, Photonics Research **9**, 1379 (2021).
- [167] M. Barbieri, C. Cinelli, P. Mataloni, and F. De Martini, Physical Review A **72**, 052110 (2005).
- [168] R. L. Franco and G. Compagno, Physical Review Letters **120**, 240403 (2018).
- [169] R. P. Feynman, in *Feynman and computation* (CRC Press, 2018) pp. 133–153.
- [170] J. Preskill, arXiv:9712048 [quant-ph] (1997), *Fault-tolerant quantum computation*.
- [171] A. Y. Kitaev, Physics Uspekhi **44**, 131 (2001).
- [172] A. Y. Kitaev, Annals of Physics (Amsterdam) **303**, 2 (2003).
- [173] X. G. Wen, Physical Review Letters **66**, 802 (1991).
- [174] D. Rainis and D. Loss, Physical Review B **85**, 174533 (2012).
- [175] P. Fendley, Journal of Statistical Mechanics: Theory and Experiment **2012**, 11020 (2012).
- [176] D. C. Tsui, H. L. Stormer, and A. C. Gossard, Physical Review Letters **48**, 1559 (1982).
- [177] R. B. Laughlin, Physical Review Letters **50**, 1395 (1983).
- [178] A. Dua, B. Malomed, M. Cheng, and L. Jiang, Physical Review B **100**, 144508 (2019).
- [179] H. Bartolomei, M. Kumar, R. Bisognin, A. Marguerite, J.-M. Berroir, E. Bocquillon, B. Plaçais, A. Cavanna, Q. Dong, U. Gennser, Y. Jin, and G. Fève, Science **368**, 173 (2020).
- [180] J. Nakamura, S. Liang, G. Gardner, and M. Manfra, Nature Physics **16**, 931 (2020).
- [181] A. Aspuru-Guzik and P. Walther, Nature Physics **8**, 285 (2012).
- [182] I. M. Georgescu, S. Ashhab, and F. Nori, Reviews of Modern Physics **86**, 153 (2014).
- [183] J. Noh, T. Schuster, T. Iadecola, S. Huang, M. Wang, K. P. Chen, C. Chamon, and M. C. Rechtsman, Nature Physics **16**, 985 (2020).
- [184] M. Yang, Y. Xiao, Y.-W. Liao, Z.-H. Liu, X.-Y. Xu, J.-S. Xu, C.-F. Li, and G.-C. Guo, Laser & Photonics Reviews **14**, 1900251 (2020).
- [185] E. Fradkin and L. P. Kadanoff, Nuclear Physics B **170**, 1 (1980).
- [186] A. Hutter and D. Loss, Physical Review B **93**, 125105 (2016).
- [187] S. Pancharatnam, in *Proceedings of the Indian Academy of Sciences-Section A*, Vol. 44 (Springer, 1956) pp. 398–417.
- [188] V. Bargmann, Journal of Mathematical Physics **5**, 862 (1964).
- [189] P. W. Shor, Physical Review A **52**, R2493 (1995).
- [190] D. Gottesman, A. Kitaev, and J. Preskill, Physical Review A **64**, 012310 (2001).
- [191] H.-L. Huang, M. Narożniak, F. Liang, Y. Zhao, A. D. Castellano, M. Gong, Y. Wu, S. Wang, J. Lin, Y. Xu, H. Deng, H. Rong, J. P. Dowling, C.-Z. Peng, T. Byrens, X. Zhu, and J.-W. Pan, Physical Review Letters **126**, 090502 (2021).
- [192] Y. Guo, S. Cheng, X. Hu, B.-H. Liu, E.-M. Huang, Y.-F. Huang, C.-F. Li, G.-C. Guo, and E. G. Cavalcanti, Physical Review Letters **123**, 170402 (2019).
- [193] G. Brassard, A. Broadbent, and A. Tapp, Foundations of Physics **35**, 1877 (2005).
- [194] P. Yin, W.-H. Zhang, L. Xu, Z.-G. Liu, W.-F. Zhuang, L. Chen, M. Gong, Y. Ma, X.-X. Peng, G.-C. Li, *et al.*, Light: Science & Applications **10**, 103 (2021).
- [195] R. Corrêa and P. L. Saldanha, Physical Review A **104**, 012212 (2021).
- [196] R. S. Bhati *et al.*, Physics Letters A , 127955 (2022).

- [197] A. Benhemou, T. Angkhanawin, C. S. Adams, D. E. Browne, and J. K. Pachos, arXiv preprint arXiv:2111.04132 (2021).
- [198] J. Preskill, *Quantum* **2**, 79 (2018).
- [199] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, *Nature Communications* **5**, 4213 (2014).
- [200] W. M. Kirby, A. Tranter, and P. J. Love, *Quantum* **5**, 456 (2021).
- [201] K. Modi, A. K. Pati, A. Sen, U. Sen, *et al.*, *Physical Review Letters* **120**, 230501 (2018).
- [202] X.-B. Liang, B. Li, and S.-M. Fei, *Physical Review A* **100**, 030304 (2019).