ECOLE
POLYTECHNIQUE
DE BRUXELLES

ULB

UNIVERSITÉ LIBRE DE BRUXELLES

Photonic Quantum Interferences in Linear Optical Circuits:

Perturbating few photons may lead to anomalous bunching

Master's thesis submitted in order to be awarded
the Master's Degree in Physical Engineering

Kamil Klaudiusz Pietrasz

Supervisor
Prof. Nicolas Cerf

Advisor
Ph.D. Researcher Léo Pioge

Service
Quantum Information and Communication (QuIC)

Academic Year
2024 - 2025

Abstract

In quantum mechanics, particles are classified as either bosons or fermions. These two families exhibit fundamentally different behaviors: fermions, such as electrons, are governed by the Pauli exclusion principle, which prevents them from occupying the same quantum state. Bosons, such as photons, show the opposite tendency — they are more likely to "bunch" together in the same state, a behavior that becomes particularly striking when particles are indistinguishable.

A famous example of boson bunching is the Hong-Ou-Mandel effect, where two identical photons entering a beam splitter always exit together in the same output mode. This striking effect is a hallmark of bosonic interference and forms the foundation of quantum optical technologies. But what happens when more than two photons are involved or when their indistinguishability is imperfect because of some discrepancy in their internal degrees of freedom, e.g., a polarization mismatch? This question is at the heart of ongoing research in quantum optics.

The present master's thesis is set in the context of this line of research, aiming at better understanding how partial distinguishability in the internal states of the photons affects their tendency to bunch when sent through a linear optical interferometer — a device that splits and recombines light using beam splitters and phase shifters. Understanding this behavior is crucial not only for fundamental physics but also for practical applications such as quantum computing, quantum metrology, and quantum communication, where photon interference plays a central role.

It is common knowledge that two partially distinguishable photons have a lower bunching probability than two perfectly indistinguishable photons; this is called the Hong-Ou-Mandel dip. In an attempt to generalize this effect to scenarios involving many photons in many modes, a series of conjectures have emerged in the scientific literature over time, formalizing the intuition that "indistinguishable bosons always bunch the most". It was suggested, for example, that the probability of all photons ending up in a subset of output modes of an arbitrary linear interferometer is maximum when all photons are perfectly indistinguishable.

Quite unexpectedly, however, counterexamples were discovered as the complexity of the systems increases. It was found that, in some rare cases, photons that are made distinguishable — by modifying their internal state, such as their polarization or arrival time — could actually bunch more than if they were fully indistinguishable. This refuted the above (strong) conjecture, which was dubbed the "global maximum conjecture". Subsequent work attempted to salvage this idea by weakening the conjecture: if global optimality is false, perhaps local optimality remains true near indistinguishable photons, meaning that perturbing indistinguishable photons would necessarily decrease the bunching probability in an infinitesimal neighborhood of indistinguishable photons. Yet this, too, was shown to be false through carefully constructed counterexamples.

This master's thesis further investigates this research avenue by focusing on another, even weaker conjecture — referred to as Pate's conjecture in this work. It emerged as a purely mathematical consequence of the now-disproven "local maximum conjecture". A first outcome of the thesis was to give Pate's conjecture a clear physical interpretation, leading to the understanding of its

connection with the previous bunching conjectures. As it turned out, Pate's conjecture addresses bunching in an extremely constrained setting: the internal states of only two photons are perturbed, with perturbations that are real-valued, equal in magnitude, and opposite in sign. This can be thought of as the weakest possible bunching conjecture (since the considered perturbations that might possibly lead to a violation are highly constrained), hence the conjecture that is the most likely to hold.

To explore the validity of Pate's conjecture, we have constructed a hierarchy of "intermediate conjectures" by incrementally relaxing the constraints. These new conjectures span the logical space between the (stronger) local maximum conjecture and (weaker) Pate's conjecture. These intermediate conjectures pertain to different numbers of perturbed photons (from all photons down to just two) and distinguish between real-valued and complex-valued perturbations. We have then attempted a combination of analytical and numerical methods to test these conjectures. While full analytical proofs or disproofs remained out of reach due to the computational complexity of the matrix permanent (a quantity that is central to calculating bosonic interference probabilities), our numerical simulations uncovered explicit counterexamples for several intermediate cases. Specifically, we found violations of 2-mode bunching conjecture for complex perturbations affecting as little as 5 photons out of 10 photons entering a specific linear interferometer. When additionally enforcing perturbations to be real-valued, a counterexample was found for 17 photons, all being perturbed. It must be stressed that the matrix permanent is known to be exponentially hard to compute, even with efficient algorithms such as Ryser's or the "incomplete rank" method. We developed a two-step numerical strategy in order to find such counterexamples. First, global optimization techniques were used to explore the space of photon perturbations; then, local methods were used to refine these counterexamples to higher violations. In spite of an exhaustive search, no violations were found for Pate's conjecture or for complex-valued intermediate conjectures with less than 5 perturbed photons, even when extending our search up to 22-photon systems. Our results leave open Pate's conjecture as well as a few other low-photon-number scenarios, but significantly expand the family of photon bunching conjectures that are now proven to be false.

In conclusion, this master's thesis signals potential implications for interferometric experiments where achieving perfect photon indistinguishability is very challenging. It suggests that with careful control over the photons' internal states, one may deliberately engineer anomalous bunching — producing interference effects even stronger than those arising from identical photons. This opens a pathway toward obtaining higher bunching violations, potentially large enough to be experimentally detectable. Such results may be of particular relevance in quantum optical setups used in technologies like boson sampling, quantum communication, quantum computing, or quantum-enhanced metrology, where fine-tuning photon interactions is essential. More broadly, this master's thesis points toward a novel approach in which distinguishability is not merely a noise factor to suppress, but a parameter to exploit when designing quantum interference-based devices. This should hopefully contribute to a deeper conceptual framework and better practical tools for probing the complex interplay between distinguishability and interference in multi-photon systems — a challenge at the heart of quantum optics and quantum technology.

Acknowledgments

It has been a great honor and pleasure to work alongside Professor Nicolas Cerf and researcher Léo Pioge. I am deeply grateful for their generous guidance, time, and insightful discussions. Collaborating with them gave me a meaningful glimpse into the academic world — an experience I will always cherish. I feel incredibly fortunate to have been surrounded by people who created such a welcoming and understanding atmosphere. I was treated with warmth, openness, and genuine appreciation for the effort I put in. That sense of being valued and included meant more to me than I can fully express.

I would like to thank my family: my parents, for not only making this journey possible but also supporting me unconditionally; and my friends, for believing in me, especially during difficult times, and for giving me strength and encouragement. A special thanks goes to my mother, whose persistence led me to take the special admission exam — a turning point without which none of this would have happened. I am also grateful to my father, who, through trust and quiet guidance, gave me the freedom to make my own choices — and in doing so, helped me grow.

I am also thankful to all the teachers I've had over the years. This journey has been worth every effort, even for this culminating moment alone. In particular, I want to express my gratitude to my former math teacher, Yves Delhaye — not only for encouraging me early on and giving me the confidence to aim higher, but also for recommending that I pursue the same path he once followed himself: the École Polytechnique de Bruxelles.

Beyond those I have named, I am thankful for every person and influence that helped lead me here in ways I may not even fully realize.

Contents

Nomenclature

$\text{perm}(A)$	Permanent of matrix A
$\det(A)$	Determinant of matrix A
$\overline{A}$	Complex conjugate of A
$A^\dagger$	Hermitian adjoint (conjugate transpose) of matrix A
A^t	Transpose of matrix A
$A \geq 0$	Hermitian positive semidefinite matrix A
$\mathcal{H}_n$	Set of $(n \times n)$ Hermitian positive semidefinite matrices
$\mathcal{S}_n$	Symmetric group of n elements
$[n]$	The set $\{1, \ldots, n\}$ (first n positive integers)
$A_{i,j}$	Element (i, j) of matrix A
$A_{[i_1:i_2],[j_1:j_2]}$	Submatrix of A from rows i_1 to i_2 and columns j_1 to j_2
$A_{\setminus i,\setminus j}$	Submatrix of A without row i and column j
$\otimes$	Tensor product
$\odot$	Hadamard (element-wise) product
$\mathbf{1}$	Constant vector i.e. $(1, \ldots, 1)^t$
$\mathbb{I}$	Identity matrix i.e. $\mathbb{I}_{ij} = \delta_{ij} \quad \forall i, j$
$\mathbb{E}$	Constant matrix i.e. $\mathbb{E}_{ij} = 1 \quad \forall i, j$
$\mathbb{P}$	Probability
n	Total number of input photons
m	Total number of output modes
$\mathcal{K}$	Subset of output modes where all n photon bunch
r	Size of the subset $\mathcal{K}$, i.e. $r = \lvert\mathcal{K}\rvert \leq m$
k	Number of perturbed photons, i.e. $k \leq n$
$f^{(k)}$	Leading principal submatrix of F, i.e. $f^{(k)} = F_{[1:k],[1:k]}$
$\hat{U}$	$(m \times m)$ Unitary operator representing an interferometer
M	Restricted submatrix of U i.e. $M = U_{[1:r],[1:n]}$
S	Distinguishability matrix
$\boldsymbol{v}$	Perturbation vector
$\lvert\phi_0\rangle$	Internal (unperturbed) state of the photon
$\lvert\eta_i\rangle$	Perturbative state of the photon i
$\lvert\phi_i\rangle$	Perturbed internal state of photon i
R	Violation ratio

Introduction

The Pauli exclusion principle states that identical fermions cannot occupy the same quantum state. In contrast, bosons exhibit the opposite behavior: their tendency to bunch together increases as they become more indistinguishable. This phenomenon arises from quantum superposition, where the non-deterministic paths of particles in a linear interferometer — analogous to the double-slit experiment — lead to both destructive and constructive interference. Crucially, this gives rise to boson bunching (or fermion antibunching), which is further accentuated for perfectly indistinguishable particles.

A striking manifestation of this is the Hong-Ou-Mandel effect [1], where two identical photons entering a 50/50 beam splitter always exit in the same output mode. This one-mode bunching effect lays the foundation for the broader concept of *multimode bunching*: the tendency of all photons to cluster in a subset of the output modes of an interferometer. Experimentally, this effect diminishes as photons become distinguishable — through differing polarizations, energy, time bins, etc. In an attempt to generalize the Hong-Ou-Mandel interference to multi-photon and multimode bunching, this sensitivity to distinguishability forms the basis of a fundamental conjecture in quantum optics: that the probability of multimode bunching is *maximized* for perfectly indistinguishable photons.

Photon interference lies at the heart of modern quantum optics and is key resource in the development of emerging quantum technologies. In quantum computing, it enables quantum bits to be encoded in various physical systems. One optical encoding scheme involves using two optical modes containing one photon, where quantum information is processed using only linear optical elements — beam splitters and phase shifters. Remarkably, even without nonlinear interactions (that is, no nonlinear coupling between optical modes is required), such a setup can achieve universal quantum computation [2] (paired with single photon sources and photo-detectors). Algorithms like Shor's prime factorization algorithm [3] exemplify the computational advantage that quantum systems can offer over classical ones

Beyond computation, multi-photon interference underpins protocols like Boson Sampling [4], where the output distributions of indistinguishable photons passing through a linear interferometer cannot be efficiently simulated on classical machines. In quantum communication, it enables Bell-state measurements — critical for entanglement swapping, and quantum teleportation [5]. Photon interference also plays a vital role in quantum-enhanced metrology, enabling measurements of time, position, and phase with precision beyond classical limits such as sub-femtosecond timing resolution [6], or minimal detectable phase enhanced by entangled photons [7].

Hence, understanding how distinguishability affects bunching is not merely an academic pursuit — it has practical consequences for the scalability and robustness of quantum devices whose design relies on multi-photon interference. This highlights the need to deepen our understanding of multi-photon interference, and critically assess the assumption of the conjecture that indistinguishability always optimizes interference-based bunching behavior.

However, prior work [8] has shown that this conjecture does not always hold: photons made partially distinguishable through individual perturbations of their internal states may — in rare configurations — exhibit a higher bunching probability than fully indistinguishable photons. This refuted the notion of a global maximum (in the sense of probability as a function of perturbations). Later results went further, disproving even a local maximum [9]: increased bunching can still occur for photons that are nearly — but not perfectly — indistinguishable. On a related note, the topic of interference between non-identical photons remains an active area of research: for instance, recent work [10] demonstrated a Hong-Ou-Mandel-like interference effect between two photons of different colors.

In this work, we push this line of inquiry further. We aim to demonstrate whether a partial perturbation, affecting only a subset of the photons, is sufficient to violate the local maximum conjecture. This type of conjecture will naturally emerge as we study the intricate link between the above-mentioned physical conjectures and some of the long-standing mathematical conjectures related to permanents. The results would allow us thus not only to deepen our theoretical understanding of interference and distinguishability but also signal potential implications for practical systems where perfect indistinguishability is difficult to achieve. It also suggests how careful engineering may lead to anomalous bunching, enhanced beyond that caused by identical photons, and how one could possibly obtain higher violations, making them experimentally detectable, which is yet not so common in practice.

Part I

Theoretical Background

Chapter 1

Prerequisites

We begin by reviewing the prerequisites for this work — first the necessary mathematics, then the key physical ideas, and finally a few optical concepts.

1.1 Matrix Properties

The adjoint † of a matrix A is defined as:

$$A^\dagger = \overline{A}^t \tag{1.1}$$

Where $\overline{A}$ denotes the complex conjugate of A, and t represents the transpose. A Hermitian matrix is self-adjoint, that is:

$$A^\dagger = A \tag{1.2}$$

A unitary matrix is defined as:

$$A^\dagger = A^{-1} \tag{1.3}$$

It can be seen as an extension from the real $\mathbb{R}$ to the complex domain $\mathbb{C}$ of an orthogonal matrix (which includes rotations, but also more general isometries such as reflections) which in turn is defined as:

$$A^t = A^{-1} \tag{1.4}$$

1.1.1 Positive Semi-Definiteness

We will define positive semi-definiteness (PSD) but any kind of definiteness (positive-definite, negative-definite, negative semi-definite) can be defined analogously. Throughout this work, multiple equivalent definitions of a positive semi-definite Hermitian matrix A denoted $A \geq 0$ (distinct from positive elementwise $a_{ij} \geq 0$), will be used, depending on the context. Let $\mathcal{H}_n$ denote the set of $n \times n$ positive semi-definite Hermitian matrices. A matrix $A \in \mathcal{H}_n$ is Hermitian positive semi-definite if and only if:

1. A is Hermitian $A^\dagger = A$, and at least one of the following conditions is satisfied:
   - its eigenvalues are non-negative

     $$\lambda_i \geq 0 \quad \forall i \in [n] \tag{1.5}$$

     where $[n] \equiv \{1, \ldots, n\}$

- its quadratic form is non-negative

$$\boldsymbol{x}^\dagger A\boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \in \mathbb{C}^n \tag{1.6}$$

- its leading principal minors are non-negative (Sylvester's criterion)

$$\det A_{[1:i],[1:i]} \geq 0 \quad \forall i \in [n] \tag{1.7}$$

  Where $A_{[i_1:i_2],[j_1:j_2]}$ is read as matrix A from row i_1 to i_2, and from column j_1 to j_2.

  A principal submatrix a of a matrix A shares the same diagonal elements, more precisely: a matrix from which the deleted rows and columns share the same index. "Leading" enforces that each submatrix, from the lowest dimension (1×1) until the highest dimensions $(n \times n)$, is created starting from the first row and column: $A_{[1:i],[1:i]} \quad \forall i \in [n]$. Minor corresponds to the determinant of the matrix in question.

  It is worth noting that, non-negative leading principal minors guarantee the non-negativity of all the remaining (non-leading) principal minors.

2. A admits a decomposition, for some $B \in \mathbb{C}^{n \times n}$ as:

$$A = B^\dagger B \tag{1.8}$$

The first definition is arguably the most commonly used. The second and third definitions are among the quickest ways to show that a matrix A is positive semi-definite without constructing explicit vectors or computing its eigenvalues λ_i, Which will appear later in our proof. The fourth definition is helpful for generating positive semi-definite Hermitian matrices from any given matrix B

1.1.2 Rayleigh Quotient

We will be working with eigenvalues of matrices thus motivating the introduction of the *Rayleigh quotient* $R_A(\boldsymbol{x})$ for a given complex matrix $A \in \mathbb{C}^{n \times n}$ and a nonzero vector $\boldsymbol{x} \in \mathbb{C}^n$, is defined as:

$$R_A(\boldsymbol{x}) = \frac{\boldsymbol{x}^\dagger A\boldsymbol{x}}{\|\boldsymbol{x}\|^2} \tag{1.9}$$

The Rayleigh quotient $R_A(\boldsymbol{x})$ is majorized by the largest $\lambda_1(A)$ and minorized by the smallest eigenvalue $\lambda_n(A)$ of the corresponding matrix A.

$$\lambda_1(A) \geq R_A(\boldsymbol{x}) \geq \lambda_n(A) \tag{1.10}$$

1.2 Group Theory

This section introduces essential concepts from group theory that are necessary to define the matrix functions that arise in the conjectures we investigate.

1.2.1 Symmetric Group

A set $\{1, \ldots, n\}$ can undergo a given permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ of its elements, which can be viewed as a bijection mapping $i \to \sigma(i)$ written as:

$$\sigma = \begin{pmatrix} 1 & \ldots & n \\ \sigma(1) & \ldots & \sigma(n) \end{pmatrix} \tag{1.11}$$

The identity permutation of a set $\{1, \ldots, n\}$, written as ϵ, maps every element to itself.

$$\epsilon = \begin{pmatrix} 1 & \ldots & n \\ 1 & \ldots & n \end{pmatrix} \tag{1.12}$$

Take the following permutation σ as an example

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \tag{1.13}$$

This is a cyclic permutation (each element sending onto the next one) more compactly written as $\sigma = (132)$, it reads as:

$$1 \to \sigma(1) = 3, \quad 3 \to \sigma(3) = 2, \quad 2 \to \sigma(2) = 1$$

A transposition $\tau = (i, j)$ is a 2-cycle permutation (1-cycle permutations (i), which are identities ϵ, are often omitted). The action of a composition $(\sigma\tau)$ of permutations σ and τ acting on an element i is:

$$(\sigma\tau)(i) = (\sigma(\tau(i))) \tag{1.14}$$

Any permutation σ can be decomposed as a composition of transpositions τ:

$$\sigma = \tau_1 \ldots \tau_k \tag{1.15}$$

One can verify that $(132) = (12)(23)$. The sign of such permutation σ is defined as

$$\operatorname{sgn}(\sigma) = (-1)^k \tag{1.16}$$

The symmetric group $\mathcal{S}_n$ is the set of all $n!$ possible permutations σ of $\{1, \ldots, n\}$.

1.2.2 Immanents

For a subgroup $G \subseteq \mathcal{S}_n$ of the symmetric group, the character $\chi : G \to \mathbb{C}$ is defined as

$$\chi(\sigma) = \operatorname{tr}(\rho(\sigma)) \tag{1.17}$$

For a matrix $A \in \mathbb{C}^{n \times n}$, the permutation product $d_A(\sigma) : \mathcal{S}_n \to \mathbb{C}$ is defined as

$$d_A(\sigma) = \prod_i^n a_{i,\sigma(i)} \tag{1.18}$$

The (normalized) generalized matrix function $f_{\chi,G}$, also known as an immanent, extends both the determinant and the permanent. It is defined for a matrix $A \in \mathbb{C}^{n \times n}$, a subgroup $G \subseteq \mathcal{S}_n$ of the symmetric group, and a character $\chi : G \to \mathbb{C}$, and is written as:

$$f_{\chi,G}(A) = \frac{1}{\chi(\epsilon)} \sum_{\sigma \in G} \chi(\sigma) d_A(\sigma) \tag{1.19}$$

As a reminder, ϵ is the identity permutation. When $G = \mathcal{S}_n$, this formula recovers classical matrix functions for specific χ:

- if χ is the principal character of G, that is when $\rho = \rho_{\text{triv}}$, then $f_{\chi,\mathcal{S}_n}(A)$ is the permanent

$$\operatorname{perm}(A) = \sum_{\sigma \in \mathcal{S}_n} d_A(\sigma) \tag{1.20}$$

- if χ is the alternating character of G, that is when $\rho = \rho_{\text{alt}}$, then $f_{\chi,\mathcal{S}_n}(A)$ becomes the determinant

$$\det A = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) d_A(\sigma) \tag{1.21}$$

Computational complexity of the permanent

Unlike determinant of a $(n \times n)$ matrix, which alternates signs and can be reduced to triangular form in $\mathcal{O}(n^3)$ time by Gaussian elimination, the permanent keeps all $n!$ terms *positive*, so those cancellations never occur. A brute-force sum over all $n!$ permutations would take $\mathcal{O}(n!n)$ time. The best general-purpose deterministic approach we currently have is *Ryser's algorithm*, which evaluates it in $\mathcal{O}(n2^n)$ time with $\mathcal{O}(n)$ memory use. Hence, even with Ryser's algorithm, permanents remain exponentially harder to compute than determinants.

1.3 Quantum states Formalisms

Having covered the necessary mathematical notions, we now turn to the key physical concepts.

1.3.1 Basis vectors

A set of vectors $\{|\varphi_i\rangle\} = \{|\varphi_1\rangle, \ldots, |\varphi_n\rangle\}$ forms a basis of a Hilbert space if any vector in that space can be written as a unique linear combination of that set $\{|\varphi_i\rangle\}$.

In quantum mechanics, one almost always assumes the basis is orthonormal (which is not required), which not only simplifies the calculations but also makes the physical interpretation of physical quantities more straightforward.

$$\langle\varphi_i|\varphi_j\rangle = \delta_{ij} \tag{1.22}$$

Where δ_{ij} refers to the Kronecker delta defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases} \tag{1.23}$$

The basis kets $|\varphi_i\rangle$ are then called basis states. They are the building blocks used to specify quantum states. Common examples include:

1. Computational basis for a qubit[1] $\{|0\rangle, |1\rangle\}$.
2. Electron spin states $\{|\uparrow\rangle, |\downarrow\rangle\}$.
3. Photon polarization states $\{|H\rangle, |V\rangle\}$.

1.3.2 First Quantization

A quantum system could be in a superposition, that is, a linear combination of those basis states $|\varphi_i\rangle$ weighted by their probability amplitudes $c_i \in \mathbb{C}$.

$$|\psi\rangle = \sum_i c_i |\varphi_i\rangle \tag{1.24}$$

The probability $\mathbb{P}(|\varphi_i\rangle)$ of measuring the system in a particular state $|\varphi_i\rangle$ is given by projecting the state onto the current state $|\psi\rangle$ then taking the norm squared. Using orthonormality, this yields the norm squared $|c_i|^2$ of the corresponding probability amplitude

$$\mathbb{P}(|\varphi_i\rangle) = |\langle\varphi_i|\psi\rangle|^2 = |c_i|^2 \tag{1.25}$$

[1] In practice we implement a logical qubit by choosing one particular physical degree of freedom (spin, polarization, energy levels, ...) and *mapping* $|0\rangle$ to one eigenstate, and $|1\rangle$ to another.

The following thus ensures that the total probability across all possible states adds up to 100%

$$\sum_i |c_i|^2 = 1 \tag{1.26}$$

Such a state $|\psi\rangle$ is referred to as a pure state, as opposed to a mixed state, which is described by a density operator ρ representing a convex combination of pure states $|\psi_i\rangle$.

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|, \quad p_i \in [0,1] : \sum_i p_i = 1 \tag{1.27}$$

For a system of two distinguishable particles A and B in *non-entangled pure states* $|\psi\rangle_A$ and $|\phi\rangle_B$, the joint state is given by the tensor product $\otimes$:

$$|\psi\rangle_A \otimes |\phi\rangle_B = |\psi\rangle_A |\phi\rangle_B \tag{1.28}$$

Note, however, that a tensor product does *not* always imply two different particles. It can equally well describe two distinct degrees of freedom of *one and the same particle* — for instance, the spatial mode $|j\rangle$ and the polarization $|\phi_j\rangle$ of a single photon. In such cases we still write $|j\rangle \otimes |\phi_j\rangle$.

Only states that factorize as a single tensor product (i.e. $|\psi\rangle_A \otimes |\phi\rangle_B$) are called separable (or non-entangled). When the state *cannot* be written in such form, the subsystems exhibit *entanglement*: measurement outcomes on A are statistically correlated with those on B in a way that exceeds all classical explanations. The four Bell's states are a textbook example, which maximize the entanglement between two-qubits. One of them is:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B) \tag{1.29}$$

We introduced entangled states here only to highlight the contrast with separable (non-entangled) states. From Chapter 2 onward, we will work exclusively with photons in separable states, unless explicitly stated otherwise.

This achieves a short description of first quantization: each particle is described by a wave-function (state vector $|\psi\rangle$) in a Hilbert space, and the number of particles is fixed.

1.3.3 Second Quantization

The first quantization could quickly become cumbersome for a significant number of particles. When multiple indistinguishable particles occupy the same quantum state, it becomes more natural to switch to the occupation number basis. In this formalism, contrary to the first quantization, the particle number can vary with the introduction of creation/annihilation operators. A Fock state with k different quantum states $|i\rangle$ is written as

$$|n_1, \ldots, n_k\rangle \tag{1.30}$$

with $n_i \in \mathbb{N}$ particles occupying the state $|i\rangle$. While Fock states are pure states, they belong to a different formalism called second quantization. When no particles are present, the system is said to be in the vacuum state denoted as

$$|0\rangle := |0, \ldots, 0\rangle \tag{1.31}$$

Annihilation $\hat{a}_i$ and creation operators $\hat{a}_i^\dagger$ act on Fock states by annihilating or creating a particle in the state $|i\rangle$. The action of $\hat{a}$ and $\hat{a}^\dagger$ are:

$$\frac{1}{\sqrt{n+1}} \hat{a} |n+1\rangle = |n\rangle \tag{1.32}$$

$$\frac{1}{\sqrt{n+1}}\hat{a}^\dagger |n\rangle = |n+1\rangle \tag{1.33}$$

Where $\frac{1}{\sqrt{n+1}}$ ensures that the states remain normalized due to the changing number of particles. Thus a particle in a spatial mode $|j\rangle$ of internal state $|\phi\rangle_j$ can be written in both first and second quantization (respectively the left and right side of the equality):

$$|j\rangle \otimes |\phi_j\rangle = \hat{a}^\dagger_{j,\phi_j} |0\rangle \tag{1.34}$$

Lastly, we introduce the commutator $[,]$:

$$[A, B] = AB - BA \tag{1.35}$$

and the anticommutator $\{,\}$:

$$\{A, B\} = AB + BA \tag{1.36}$$

So that now we can elegantly express the symmetry of bosons under particle exchange (with the help of commutators):

$$[\hat{a}^\dagger_i, \hat{a}^\dagger_j] = 0 \quad \forall i, j \tag{1.37}$$

While the antisymmetry of fermions under particle exchange (using anticommutators) gives:

$$\{\hat{a}^\dagger_i, \hat{a}^\dagger_j\} = 0 \quad \forall i, j \tag{1.38}$$

The anticommutator for fermions gives when $i = j$:

$$(\hat{a}^\dagger_i)^2 = 0 \tag{1.39}$$

Which reveals exactly the Pauli exclusion principle: forbidding the occupation of the same state by identical fermions.

Furthermore, the commutator for fermions can be simplified using equation 1.38, thus becomes:

$$[\hat{a}^\dagger_i, \hat{a}^\dagger_j] = 2\hat{a}^\dagger_i \hat{a}^\dagger_j \tag{1.40}$$

1.4 Fundamental Interferometer Components

We are now ready to introduce *beam splitters* — a fundamental optical element that will recur throughout this work.

For completeness we will first define phase shifters, because any interferometer (and, in fact, any linear-optical circuit) can be built solely from beam splitters and phase shifters. Together they constitute a universal set for realizing arbitrary linear optical transformations [11].

Phase shifter

A photon acquires a phase $\delta \in \mathbb{R}$ when passing through a phase shifter (indicated by the arrow $\rightarrow$):

$$\hat{a}^\dagger \rightarrow \exp(i\delta)\hat{a}^\dagger \tag{1.41}$$

Beam splitter

A beam splitter is an optical device that splits an incident light beam into two (or more) output beams. It is described by a unitary matrix $\hat{U}$ and acts exclusively on the spatial mode $|j\rangle$, without affecting the internal state $|\phi_j\rangle$ of the photon.

$$\hat{U}\left(|j\rangle \otimes |\phi_j\rangle\right) = \hat{U}\,|j\rangle \otimes \mathbb{I}\,|\phi_j\rangle \tag{1.42}$$

For an m-mode beam splitter (with the same number of input and output ports); see Figure 1.1, a single photon creation operator $\hat{a}^\dagger_{j,\phi_j}$ — which creates one photon in spatial mode j with internal state $|\phi_j\rangle$ — transforms according to the $(m \times m)$ beam splitter unitary $\hat{U}$ as follows (with input $\rightarrow$ output):

$$\hat{a}^\dagger_{j,\phi_j} \rightarrow \hat{U}\hat{a}^\dagger_{j,\phi_j}\hat{U}^\dagger = \sum_{k=1}^{m} U_{kj}\hat{a}^\dagger_{j,\phi_j} \tag{1.43}$$

This will be explained in details later in Section 2.1.1. We choose the convention where the first

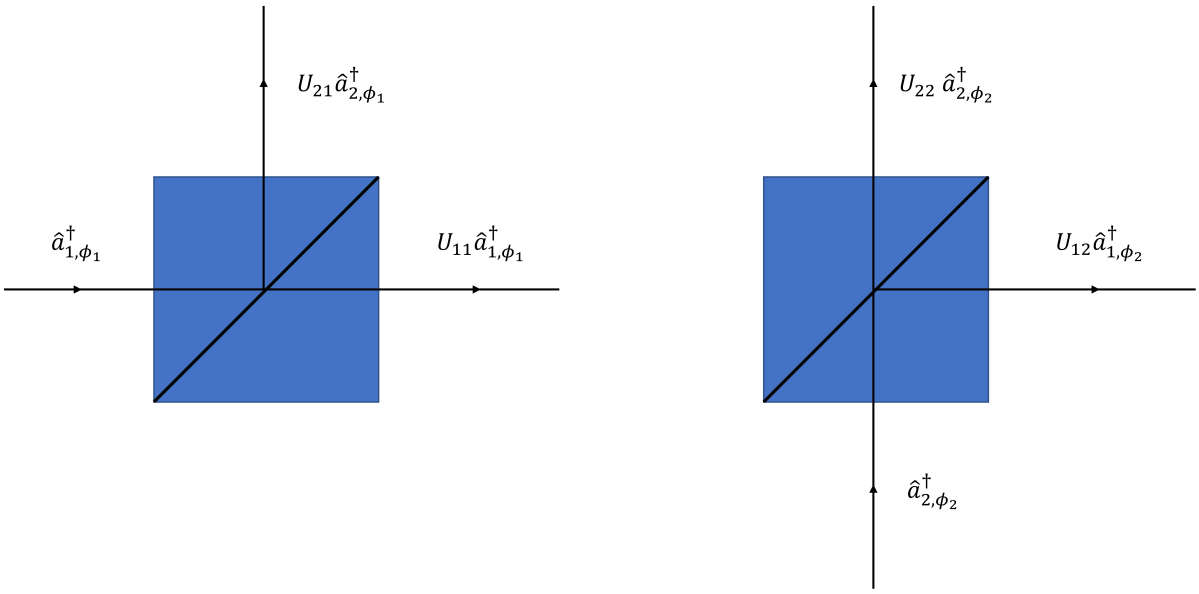


Figure 1.1: Evolution of the creation operator $\hat{a}^\dagger_{j,\phi_j}$ under the action of a two-mode beam-splitter

input mode continues along the same spatial path as the first output mode, and similarly for other modes. That is, the labeling (such as i in $\hat{a}^\dagger_i$) preserves the physical propagation direction through the beam splitter. Under this convention, the most general form of the beam splitter unitary matrix $\hat{U}$ is:

$$\hat{U} = \begin{pmatrix} \sqrt{T} & e^{i\phi}\sqrt{R} \\ -e^{-i\phi}\sqrt{R} & \sqrt{T} \end{pmatrix} \tag{1.44}$$

Where $T = |t|^2$ is the transmittance and $R = |r|^2$ is the reflectance, that is, the probability of transmission and reflection respectively. Both can be expressed using the transmission and reflection amplitudes $t, r \in \mathbb{C}$. Since the particle is inevitably either transmitted or reflected, these probabilities must satisfy

$$T + R = 1 \tag{1.45}$$

One of the possible (but not unique) beam splitter matrix $\hat{U}$ is

$$\hat{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

The reflected beam acquires a $e^{i\frac{\pi}{2}} = i$ phase shift in the reflection process, and the states must be normalized by $\frac{1}{\sqrt{2}}$ due to the beam splitter being 50/50. Both come from the satisfaction of the unitary matrix condition:

$$\hat{U}\hat{U}^\dagger = \mathbb{I} \tag{1.46}$$

Where $\mathbb{I}$ designates the identity matrix.

1.5 Two-Particle Interference

Suppose a 50/50 beam splitter ($T = R = \frac{1}{2}$) with two input and two output ports. We will explore a range of interactions, beginning with distinguishable particles, followed by indistinguishable particles. Specifically, we will consider the case of fermions, then bosons (which includes the well-known Hong-Ou-Mandel effect), and conclude with partially distinguishable particles.

We define two characteristic output patterns:

Definition 1. *bunching : the tendency of identical bosons to occupy the same spatial mode.*

Definition 2. *antibunching : the complementary behavior to bosonic bunching, for which identical fermions occupy distinct spatial modes.*

1.5.1 Distinguishable Particles

First let us consider two completely distinguishable particles, such as two photons with different polarizations: one horizontal $|H\rangle$ one vertical $|V\rangle$; or two electrons with different spins: one up $|\uparrow\rangle$, and the other down $|\downarrow\rangle$. In this case, the particles are non-interacting and remain independent and uncorrelated. If we count the number of particles obtained in the output modes, then either we find:

- one particle in each output mode $\mathbb{P}(|1,1\rangle)$ — which is exactly antibunching — occuring when both particles are transmitted T^2 or reflected R^2,

$$\mathbb{P}(|1,1\rangle) = T^2 + R^2 = \frac{1}{2} \tag{1.47}$$

- both particles in the same output mode $\mathbb{P}(|2,0\rangle)$ or $\mathbb{P}(|0,2\rangle)$ — corresponding to bunching — occuring when one of the particles is transmitted and the other reflected:

$$\mathbb{P}(|2,0\rangle) = \mathbb{P}(|0,2\rangle) = TR = \frac{1}{4} \tag{1.48}$$

1.5.2 Indistinguishable Particles

The creation operator for a single photon entering the first input mode $\hat{a}_1^\dagger$, using (1.43), evolves as:

$$\hat{a}_1^\dagger \to \frac{1}{\sqrt{2}}(\hat{a}_1^\dagger + i\hat{a}_2^\dagger) \tag{1.49}$$

When two particles enter the beam splitter, again using (1.43), the operators become

$$\hat{a}_1^\dagger \hat{a}_2^\dagger \to \frac{1}{2}(\hat{a}_1^\dagger + i\hat{a}_2^\dagger)(i\hat{a}_1^\dagger + \hat{a}_2^\dagger) \tag{1.50}$$

$$= \frac{i}{2}\left[(\hat{a}_1^\dagger)^2 + (\hat{a}_2^\dagger)^2\right] + \frac{1}{2}[\hat{a}_1^\dagger, \hat{a}_2^\dagger] \tag{1.51}$$

Fermions

In the case of fermions, we simplify using equation 1.38 and equation 1.39. Thus, equation (1.51) becomes:

$$\hat{a}_1^\dagger \hat{a}_2^\dagger \to \hat{a}_1^\dagger a_2^\dagger \tag{1.52}$$

We can then deduce the final state $|\phi\rangle_{\text{out}}$

$$|\phi\rangle_{\text{out}} = \hat{a}_1^\dagger \hat{a}_2^\dagger |0\rangle = |1,1\rangle \tag{1.53}$$

It follows that the probability of finding one fermion in each output mode is

$$\mathbb{P}(|1,1\rangle) = |\langle 1,1 |\phi\rangle_{\text{out}}|^2 = 1 \tag{1.54}$$

It is worth noting that particles are correlated: obtaining information about one of them immediately reveals the state of the other. That is, if one fermion is found in a given output mode, the other must inevtiably be in the other output mode. We exclusively have an antibunching phenomenon for identical fermions.

Bosons

As for bosons, by applying recursively (1.33) we get $(\hat{a}_1)^2 |0\rangle = \sqrt{2}\,|2,0\rangle$, similarly for $(\hat{a}_2^\dagger)^2 |0\rangle = \sqrt{2}\,|0,2\rangle$. Additionally we can use equation 1.37 to get the final state $|\psi\rangle_{\text{out}}$ becomes

$$|\psi\rangle_{\text{out}} = \frac{i}{2}((\hat{a}_1^\dagger)^2 + (\hat{a}_2^\dagger)^2)|0\rangle = \frac{i}{\sqrt{2}}(|2,0\rangle + |0,2\rangle) \tag{1.55}$$

Therefore, both bosons are guaranteed to bunch together in one of either output modes

$$\mathbb{P}(|2,0\rangle) = |\langle 2,0 |\psi\rangle_{\text{out}}|^2 = \frac{1}{2} \tag{1.56}$$

One similarly obtains $\mathbb{P}(|0,2\rangle) = \frac{1}{2}$. The particles are again correlated: if one boson is detected in a particular output mode, the other is necessarily found in the same mode. This is known as the Hong-Ou-Mandel effect.

1.5.3 Partially and Near Indistinguishable Particles

Let the initial state be described by one photon in the input mode 1 of internal state φ, different from the second photon in the input mode 2 of internal state ψ. Such state evolves through the beam-splitter as:

$$\hat{a}_{1\varphi}^\dagger \hat{a}_{2\psi}^\dagger |0\rangle \to \hat{U} \hat{a}_{1\varphi}^\dagger \hat{a}_{2\psi}^\dagger |0\rangle \tag{1.57}$$

Let us expand the second photon's state $|\psi\rangle$ in the subspace spanned by $|\varphi\rangle$ and a state $|\varphi^\perp\rangle$ orthogonal to it:

$$\begin{aligned}
\hat{a}_{2\psi}|0\rangle &= |\psi\rangle && (1.58)\\
&= \alpha|\varphi\rangle + \beta|\varphi^\perp\rangle && (1.59)\\
&= \alpha|\varphi\rangle + \exp(i\theta)\sqrt{1-|\alpha|^2}\,|\varphi^\perp\rangle && (1.60)\\
&= (\alpha\hat{a}_{2\varphi}^\dagger + \sqrt{1-|\alpha|^2}\hat{a}_{2\varphi^\perp}^\dagger)|0\rangle && (1.61)
\end{aligned}$$

Using the orthonormality between $|\varphi\rangle$ and $|\varphi^\perp\rangle$, we expressed β in terms of α up to a global phase $\exp(i\theta)$ for some $\theta \in [0, 2\pi[$.

$$|\beta|^2 = 1 - |\alpha|^2 \tag{1.62}$$

$$\beta = \exp(i\theta)\sqrt{1-|\alpha|^2} \tag{1.63}$$

The overall global phase of any state $|\psi\rangle$ is physically irrelevant. In particular, we got rid of the global phase for $|\varphi^\perp\rangle$

$$\exp(i\theta)\,|\psi\rangle = |\psi\rangle \quad \forall\,|\psi\rangle \tag{1.64}$$

The final state becomes, again using equation 1.43:

$$\begin{aligned}
\hat{U}\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{2\psi}\,|0\rangle &= \hat{U}\hat{a}^\dagger_{1\varphi}(\alpha\hat{a}^\dagger_{2\varphi} + \sqrt{1-|\alpha|^2}\hat{a}^\dagger_{2\varphi^\perp})\,|0\rangle && (1.65)\\
&= \alpha\hat{U}\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{2\varphi}\,|0\rangle + \sqrt{1-|\alpha|^2}\hat{U}\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{2\varphi^\perp}\,|0\rangle && (1.66)\\
&= \alpha(\hat{U}\hat{a}^\dagger_{1\varphi}\hat{U}^\dagger)(\hat{U}\hat{a}^\dagger_{2\varphi}\hat{U}^\dagger)\,|0\rangle + \sqrt{1-|\alpha|^2}(\hat{U}\hat{a}^\dagger_{1\varphi}\hat{U}^\dagger)(\hat{U}\hat{a}^\dagger_{2\varphi^\perp}\hat{U}^\dagger)\,|0\rangle && (1.67)\\
&= \frac{\alpha}{2}(\hat{a}^\dagger_{1\varphi} + i\hat{a}^\dagger_{2\varphi})(i\hat{a}^\dagger_{1\varphi} + \hat{a}^\dagger_{2\varphi})\,|0\rangle + \frac{\sqrt{1-|\alpha|^2}}{2}(\hat{a}^\dagger_{1\varphi} + i\hat{a}^\dagger_{2\varphi})(i\hat{a}^\dagger_{1\varphi^\perp} + \hat{a}^\dagger_{2\varphi^\perp})\,|0\rangle && (1.68)\\
&= \frac{\alpha}{2}\left([\hat{a}^\dagger_{1\varphi}, \hat{a}^\dagger_{2\varphi}] + i\left[(\hat{a}^\dagger_{2\varphi})^2 + (\hat{a}^\dagger_{1\varphi})^2\right]\right)|0\rangle && (1.69)\\
&+ \frac{\sqrt{1-|\alpha|^2}}{2}\left(\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{2\varphi^\perp} - \hat{a}^\dagger_{2\varphi}\hat{a}^\dagger_{1\varphi^\perp} + i\left[\hat{a}^\dagger_{2\varphi}\hat{a}^\dagger_{2\varphi^\perp} + \hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{1\varphi^\perp}\right]\right)|0\rangle && (1.70)
\end{aligned}$$

The first term is exactly the same as for two indistinguishable particles, scaled by a factor α. The second term corresponds to two distinguishable particles weighted by a factor $\sqrt{1-|\alpha|^2}$.

Let us now restrict to the case of (not identical) bosons, we can use equation 1.37 to get:

$$\begin{aligned}
\hat{U}\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{2\psi}\,|0\rangle &= \frac{i\alpha}{2}\left[(\hat{a}^\dagger_{2\varphi})^2 + (\hat{a}^\dagger_{1\varphi})^2\right]|0\rangle && (1.71)\\
&+ \frac{\sqrt{1-|\alpha|^2}}{2}\left(\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{2\varphi^\perp} - \hat{a}^\dagger_{2\varphi}\hat{a}^\dagger_{1\varphi^\perp} + i\left[\hat{a}^\dagger_{2\varphi}\hat{a}^\dagger_{2\varphi^\perp} + \hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{1\varphi^\perp}\right]\right)|0\rangle && (1.72)
\end{aligned}$$

Using equation 1.33, the only terms that contribute to bunching into mode 1 are:

$$\frac{i\alpha}{2}\sqrt{2}\,|2,0\rangle + i\frac{\sqrt{1-|\alpha|^2}}{2}\hat{a}^\dagger_{1\varphi}\hat{a}^\dagger_{1\varphi^\perp}\,|0\rangle \tag{1.73}$$

Thus the probability of bunching into mode 1 is:

$$\mathbb{P} = \left|\frac{i\alpha}{2}\sqrt{2}\right|^2 + \left|\frac{\sqrt{1-|\alpha|^2}}{2}\right|^2 = \frac{1+|\alpha|^2}{4} \tag{1.74}$$

Here $|\alpha|$ qualifies the indistinguishability: as $|\alpha|$ increases, photons bunch more. Therefore, the more indistinguishable the photons are, the more they bunch.

Notice that — similarly for bunching into mode 2 — this probability of bunching into mode 1 is maximized when $\alpha \to 1$ (i.e. for identical bosons):

$$\arg\max_\alpha \mathbb{P} = 1 \tag{1.75}$$

The probability of bunching into either mode is thus guaranteed

$$\mathbb{P} = \frac{1+|\alpha|^2}{2} \xrightarrow{\alpha\to 1} 1 \tag{1.76}$$

After studying two-particle interactions, we conclude that the bunching probability is maximized for identical bosons. Additionally, as the photons become more distinguishable, the bunching probability decreases. A natural question arises: if the bunching probability is maximized for *two* indistinguishable photons, could this behavior be generalized to n indistinguishable photons?

Chapter 2

Multimode boson bunching

In practice, physical systems become more interesting when they involve more than two photons and spatial modes. We therefore extend the concepts introduced earlier to larger number of photons and modes. First, we lay out the necessary fundamentals; we then survey the present-day state of the art on multiple bosons bunching into a subset of output modes.

This chapter is greatly inspired by the works Refs [8] and [9]. The ideas developed there are essential for understanding the second part of this thesis, which can be viewed as a direct continuation of those studies.

2.1 Preliminaries

The photon j enters the m-mode interferometer (cf. Figure 2.1[1]) in the state:

$$|\psi_j\rangle = |j\rangle \otimes |\phi_j\rangle = \hat{a}^\dagger_{j,\phi_j} |0\rangle \tag{2.1}$$

Where:

- $|j\rangle$ is the spatial state of the photon (corresponding to the spatial modes of the interferometer)
- $|\phi_j\rangle$ is the internal state which excludes the spatial mode (e.g. polarization, time-bin, frequency, or other degrees of freedom)

It is important to note that $\hat{U}$ acts only on spatial modes

$$\hat{U}\left(|j\rangle \otimes |\phi_j\rangle\right) = \hat{U}\,|j\rangle \otimes \mathbb{I}\,|\phi_j\rangle \tag{2.2}$$

It will be shown in the following section 2.1.1 that each photon exits in a superposition of spatial modes while its internal state $|\phi_j\rangle$ remains unchanged.

Although the interferometer has m input and m output modes, we inject n photons with only one photon per occupied input mode and allow $n \leq m$. Hence, m is best viewed as the total number of output modes (since the remaining unoccupied input modes are irrelevant). Throughout this work the input state is a product of pure, non-entangled single-photon states.

[1]The interferometer in this figure is represented as an integrated photonic circuit rather than a free-space (bulk-optic) system. The former integrates optical components (waveguides, beam splitters, phase shifters) as a monolithic (single-substrate) chip which offers robustness and scalability — ideal for practical applications in technologies (telecommunications, quantum computing etc...). In contrast, free-space systems use discrete optical elements components (mirrors, beam splitters...) aligned in air (or sometimes in vacuum) on an optical table, making them reconfigurable thus well-suited for experiments.

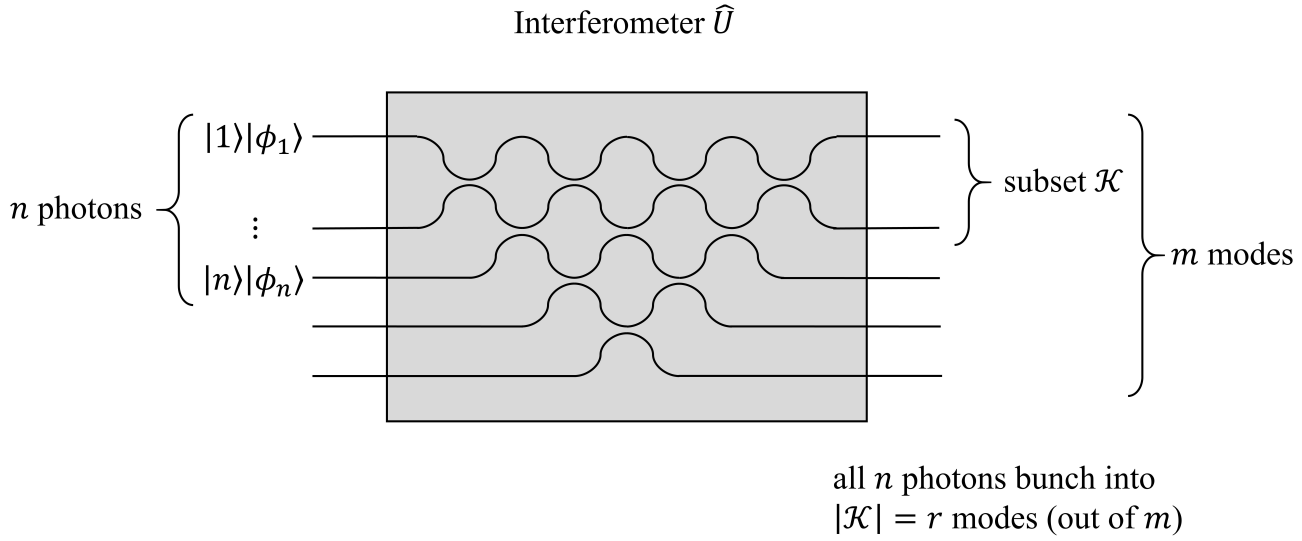


Figure 2.1: Interferometer $\hat{U}$ in which n photons enter the first $n \leq m$ input modes then bunch into a subset $\mathcal{K}$ of the m output modes, with $|\mathcal{K}| = r \leq m$, and $|j\rangle\,|\phi_j\rangle$ describing the spatial mode $|j\rangle$ and internal state $|\phi_j\rangle$ of the photon j.

Our focus is the probability P_n of all n photons bunching into a subset $\mathcal{K}$ of the output modes[2], where $|\mathcal{K}| = r \leq m$.

2.1.1 Multimode interferometer

We will derive the relation between the input and output states of a multimode interferometer.

Consider n input photons ($n \leq m$), each in a distinct spatial mode $|j\rangle$ with internal state $|\phi_j\rangle$. The beam-splitter acts only on the spatial mode $|j\rangle$ (and not on the internal state $|\phi_j\rangle$) via the $(m \times m)$ unitary $\hat{U}$ with matrix elements $U_{kj} = \langle k|\,\hat{U}\,|j\rangle$.

Beam splitter matrix

First, let us express the beam splitter matrix $\hat{U}$. Using the closure relation $\mathbb{I} = \sum_j |j\rangle\langle j|$, we can write the operator $\hat{U}$ in terms of its matrix elements $U_{k,j} = \langle k|\,\hat{U}\,|j\rangle$. Where we have m input spatial modes $|j\rangle$ and m output spatial modes $|k\rangle$:

$$\hat{U} = \mathbb{I}\cdot\hat{U}\cdot\mathbb{I} = \sum_{k=1}^{m}|k\rangle\langle k|\,\hat{U}\sum_{j=1}^{m}|j\rangle\langle j| = \sum_{k=1}^{m}\sum_{j=1}^{m}U_{kj}\,|k\rangle\langle j| \tag{2.3}$$

[2]Although from this point onward the specific convention for labeling spatial modes will no longer play a crucial role, it is helpful to clarify the difference in approach. In the case of a simple two-mode beam splitter (Figure 1.1), we previously used a convention where the output mode index was determined by the physical outcome — that is, a particle transmitted through the beam splitter retained the same mode number as its input. However, in more complex interferometric setups such as the multimode interferometer depicted in Figure 2.1, it is mathematically more convenient to adopt a different convention: we label the input and output modes according to their visual continuity in the diagram. That is, the i-th input mode connects directly to the i-th output mode via an uninterrupted path depicted as a line. In short, in Figure 2.1, the input/output are numbered in the same order(e.g. from top to bottom). This indexing simplifies calculations in practice.

Let us consider how the m-mode interferometer acts on one spatial mode $|j\rangle$ occupied by one input photon j:

$$|j\rangle \to \hat{U}\,|j\rangle = \sum_{k=1}^{m}\sum_{l=1}^{m} U_{kl}\,|k\rangle \underbrace{\langle l|j\rangle}_{\delta_{lj}} = \sum_{k=1}^{m} U_{kj}\,|k\rangle \tag{2.4}$$

This indicates that the term U_{kj} could be interpreted as the contribution (probability amplitude) of the input spatial mode $|j\rangle$ to the output spatial mode $|k\rangle$. Additionally, we notice that the photon exits in a superposition of output modes $|k\rangle$.

One input photon

Consider m spatial output modes $|k\rangle$ but only one photon in the spatial input mode $|j\rangle$ of internal state $|\phi_j\rangle$ created by the creation operator $\hat{a}^\dagger_{j,\phi_j}$ from the vacuum state $|0\rangle$. Then, a beam splitter acts (indicated by $\to$) on such a state via the unitary operator $\hat{U}$ as:

$$\hat{a}^\dagger_{j,\phi_j}\,|0\rangle \to \hat{U}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle \tag{2.5}$$

Now the last result $\hat{U}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle$ could be simplified in two ways:

- On one hand, using equation 2.2 then equation 2.4, we get:

$$\hat{U}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle = \Big(\hat{U}\,|j\rangle \otimes |\phi_j\rangle\Big) = \left(\sum_{k=1}^{m} U_{kj}\,|k\rangle\right) \otimes |\phi_j\rangle = \sum_{k=1}^{m} U_{kj}\hat{a}^\dagger_{k,\phi_j}\,|0\rangle \tag{2.6}$$

  Where we have rewritten the output state as: $|k\rangle \otimes |\phi_j\rangle = \hat{a}^\dagger_{k,\phi_j}\,|0\rangle$.

- On the other hand we could rewrite $\hat{U}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle$ using the fact that $\hat{U}$ is unitary $\hat{U}^{-1} = \hat{U}^\dagger$ and that it does not influence the vacuum state $\hat{U}\,|0\rangle = |0\rangle$

$$\hat{U}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle = \hat{U}\hat{a}^\dagger_{j,\phi_j}\hat{U}^\dagger\hat{U}\,|0\rangle = \hat{U}\hat{a}^\dagger_{j,\phi_j}\hat{U}^\dagger\,|0\rangle \tag{2.7}$$

Equating both expressions for $\hat{U}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle$ gives:

$$\hat{U}\hat{a}^\dagger_{j,\phi_j}\hat{U}^\dagger = \sum_{k=1}^{m} U_{kj}\hat{a}^\dagger_{k,\phi_j} \quad \forall j, \phi_j \tag{2.8}$$

Multiple input photons

Finally now treating n input photons (one per mode): the input state is a pure non-entangled state $\prod_{j=1}^{n}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle$. The annihilation operator $\hat{a}^\dagger_{j,\phi_j}$ creates a photon in the spatial mode $|j\rangle$ of internal state $|\phi_j\rangle$. The state evolves under the transformation of the beam splitter $\hat{U}$ as:

$$\prod_{j=1}^{n}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle \to \hat{U}\prod_{j=1}^{n}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle$$

We can express $\hat{U}$ inside the product using again the same mathematical trick then replace with the previously obtained expression in equation 2.8:

$$\hat{U}\prod_{j=1}^{n}\hat{a}^\dagger_{j,\phi_j}\,|0\rangle = \hat{U}\hat{a}^\dagger_{1,\phi_1}\hat{U}^\dagger\hat{U}\hat{a}^\dagger_{2,\phi_2}\hat{U}^\dagger\ldots\hat{U}\hat{a}^\dagger_{n,\phi_n}\hat{U}^\dagger\hat{U}\,|0\rangle \tag{2.9}$$

$$= \prod_{j=1}^{n}\hat{U}\hat{a}^\dagger_{j,\phi_j}\hat{U}^\dagger\,|0\rangle = \prod_{j=1}^{n}\sum_{k=1}^{m} U_{kj}\hat{a}^\dagger_{k,\phi_j}\,|0\rangle \tag{2.10}$$

For a specific photon j one thus gets:

$$\hat{a}^\dagger_{j,\phi_j} \to \hat{U}\hat{a}^\dagger_{j,\phi_j}\hat{U}^\dagger = \sum_{k=1}^{m} U_{kj}\hat{a}^\dagger_{k,\phi_j} \tag{2.11}$$

Which was the original beam-splitter relation stated in equation 1.43.

2.1.2 Multimode Bunching Probability

We will now return to bunching and derive an explicit formula for the multimode bunching probability P_n. Let us first define:

Definition 3. *multimode bunching: event in which all n input photons emerge from the interferometer $\hat{U}$ exclusively in a fixed subset $\mathcal{K}$ of the m output modes, with $|\mathcal{K}| = r \leq m$; every output port outside $\mathcal{K}$ remains unoccupied.*

For a given set of vectors $\{|\phi_i\rangle : i \in [n]\}$ The overlap between internal states can be analyzed via the Gram matrix $S \in \mathcal{H}_n$ called the *distinguishability matrix*, defined as:

$$S_{ij} = \langle\phi_i|\phi_j\rangle \quad \forall i,j \in [n] \tag{2.12}$$

Note that it is always possible to choose a set of normalized vectors $\langle\phi_i|\phi_i\rangle = 1$ to simplify the Gram matrix to $S_{ii} = 1 \quad \forall i \in [n]$. The $(n \times n)$ positive semidefinite Hermitian matrix $H \in \mathcal{H}_n$, can be seen as its correspondence for the spatial modes embodied by the linear interferometer $\hat{U}$ and the subset $\mathcal{K}$ of output modes, which is defined as:

$$H_{ij} = \sum_{k\in K} \overline{U}_{k,i}U_{k,j} \quad \forall i,j \in [n] \tag{2.13}$$

In practice, for a total of m output modes, the $(m \times m)$ matrix $\hat{U}$ describing the whole intereferometer is not relevant. Only its restriction to the subset $\mathcal{K}$ of output modes is useful, and will be denoted as a $(r \times n)$ matrix M as opposed to the $(m \times m)$ operator $\hat{U}$.

$$M = \hat{U}_{[1:r],[1:n]} \tag{2.14}$$

Notice that in such case, the matrix H can be simply redefined in terms of M as:

$$H = M^\dagger M \tag{2.15}$$

The multimode boson bunching probability P_n, that is, the probability of finding all n photons in a subset $\mathcal{K}$ of the interferometer output modes, can be obtained by projecting the output state of the interferometer $\prod_{j=1}^n \hat{a}^\dagger_j |0\rangle \to \hat{U}\prod_{j=1}^n \hat{a}^\dagger_j |0\rangle$ onto the subspace where all the photons occupy the subset $\mathcal{K}$. Such probability P_n is expressed as follows:

$$P_n(S) = \text{perm}(H \odot S) \tag{2.16}$$

Where $\odot$ denotes the Hadamard product $(A \odot B)_{ij} = A_{ij}B_{ij}$

2.2 Global Maximum Conjecture

The Hong-Ou-Mandel effect described in Section 1.5.2 begs the question whether n indistinguishable bosons do maximize the bunching probability in a given subset $\mathcal{K}$ of output modes.

The opposite statement about fermions is true: indistinguishable fermions always maximize the multimode antibunching probabilities.

Finally, this physical conjecture for bosons is, as will be explained in this section, equivalent to a mathematical conjecture which in turn is implied by a conjeture *permanent-on-top* proven for $n \leq 3$ (later seen in details in section 2.4.2).

We define:

Definition 4. *partial distinguishability : characteristic of two (normalized) internal states* $|\psi\rangle$ *and* $|\phi\rangle$ *whose overlap* $\langle\cdot|\cdot\rangle$ *is*

$$0 < |\langle\phi|\psi\rangle| < 1 \tag{2.17}$$

For example, the state of two partially distinguishable photons could be: $|\psi\rangle = |H\rangle$ and $|\phi\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$

Definition 5. *anomalous bunching: phenomenon in which the probability that partially distinguishable bosons exceeds the corresponding bunching probability for completely indistinguishable bosons.*

Now we can properly introduce the physical conjecture P1 about multimode boson bunching [12], motivated by the above mentioned observations. The authors stated it in Ref [9] as follows:

Conjecture P1. *(Shchesnovich 2016) [9] Consider any linear interferometer* $\hat{U}$ *and any subset* $\mathcal{K}$ *of output modes. Among all possible separable input states of n classically correlated photons, the probability that all output photons bunch into* $\mathcal{K}$ *reaches its global maximum if the photons are perfectly indistinguishable.*

It could be verified that this conjecture holds for single-mode bunching $|\mathcal{K}| = 1$. Let us define a $(n \times n)$ matrix $\mathbb{E}$ as:

$$\mathbb{E}_{ij} = 1 \quad \forall i, j \in [n] \tag{2.18}$$

Since $S = \mathbb{E}$ if all photons are perfectly indistinguishable, and since $\mathrm{perm}(H \odot \mathbb{E}) = \mathrm{perm}(H)$, the conjecture P1 can be written using equation (2.16) as:

$$\mathrm{perm}(H \odot S) \leq \mathrm{perm}(H) \tag{2.19}$$

Coincidentally this physical Conjecture P1 can be linked by equivalence with the mathematical Conjecture M1 by Bapat and Sunder from Ref [13], which is the following:

Conjecture M1. *(Bapat, Sunder 1985) Let* $A, B \in \mathcal{H}_n$ *then*

$$\mathrm{perm}(A \odot B) \leq \mathrm{perm}(A) \prod_{i=1}^{n} B_{ii} \tag{2.20}$$

Without loss of generality [14], B can be restricted to a Gram matrix with

$$B_{ii} = 1 \quad \forall i \in [n] \tag{2.21}$$

Thus making equation (2.20) equivalent to equation (2.19). Intuitively, and given the long-standing mathematical conjecture, one would naturally expect the conjecture P2 for the global maximum to be true.

Counterexample

Against all expectations, for the first time there are partially distinguishable bosons that bunch more than indistinguishable bosons. The counterexample to Conjecture M1 was found by Drury [15] in 2016, with the lowest-dimensional (2×7) matrix, physically corresponding to $n = 7$ input photons that bunch into $|\mathcal{K}| = 2$. It is more useful to provide its decomposition instead $H = M^\dagger M$ (possible for any Hermitian positive-semi definite matrices [16]):

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & \sqrt{2} & 1 & \omega & \omega^2 & \omega^3 & \omega^4 \end{pmatrix} \tag{2.22}$$

Where $\omega = \exp(i\frac{2\pi}{q})$ with $q = n - 2$. From this mathematical counterexample, several instances of interferometers were found by the authors in Ref [8], one depicted in Figure 2.2 — the source of this Figure is also from Ref [8], to disprove conjecture P1.

Note that the currently defined H is in accordance with the definition of the matrix H in equation 2.13. Meanwhile M, as defined in equation 2.14, does not directly correspond to $\hat{U}$ (representing the whole interferometer), but rather to its restricted dimension from m total output modes down to $|\mathcal{K}|$ output modes (in which all the photons bunch). The rank r of the matrix M is equal to $r = |\mathcal{K}| = 2$. To summarize, to find a counterexample matrix $\hat{U}$, one only needs to work with its $(r \times n)$ submatrix M (thus of rank r) representing n photons, one per input mode, bunching into r output modes (out of m in total).

This explicit counterexample clearly proves the surprising fact: there is an interferometer configuration (as described by the matrix M), for which *anomalous bunching* takes place, i.e. *partially distinguishable* photons present a higher multimode bunching probability than perfectly indistinguishable photons.

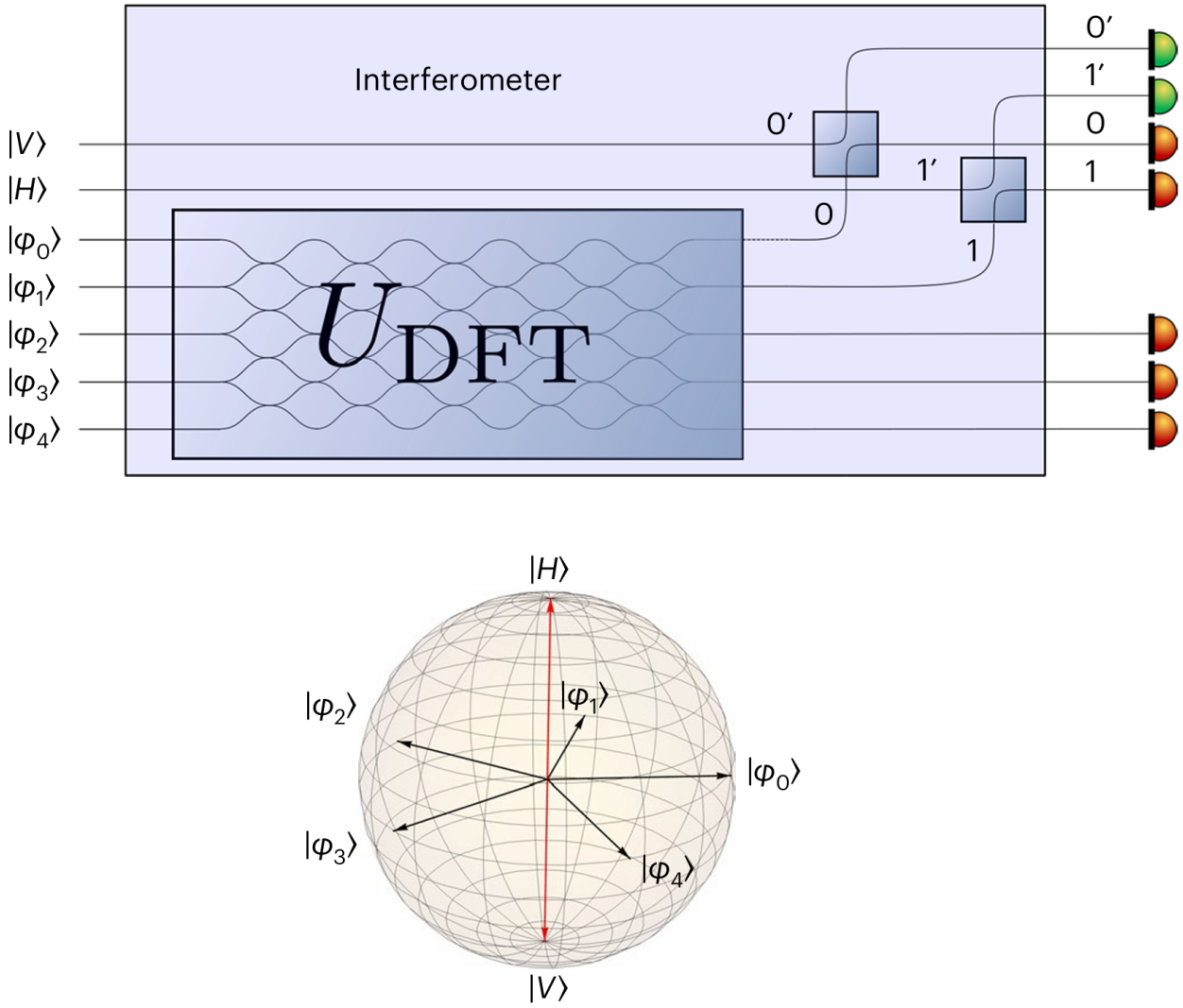


Figure 2.2: The image is from Ref [8] displaying an optical set up: 7-mode interferometer found by authors in Ref [8] corresponding to the mathematical counterexample of the Conjecture M1 from Ref [15] found by Drury. Seven photons with specifically prepared polarization states bunch into the first two output modes (depicted by green light detectors) with a higher probability than if they all shared the same polarization (which would effectively make them indistinguishable). Below, their polarizations are represented on a Bloch sphere. As described by the authors [8], the setup can be generalized to n modes with $q = n - 2$ (here $q = 5$) photons specifically prepared with polarization states $|\phi_j\rangle = \frac{1}{\sqrt{2}}(|H\rangle + \omega^j |V\rangle) \quad \forall j = 0, \ldots, q-1$ in the $q = 5$ mode discrete Fourier transform $U_{jk} = \frac{1}{\sqrt{q}}\omega^{jk} \quad \forall j, k = 0, \ldots, q-1$ whose two modes are sent respectively to two beam splitters of equal transmittance $\frac{2}{n}$ to interfere with the remaining $n - q = 2$ vertically and horizontally polarized photons.

2.3 Local Maximum Conjecture

We have concluded the previous section with a counterexample to the global maximum Conjecture P1. This naturally leads to a new question: is the multimode bunching probability at least *locally* maximized by perfectly indistinguishable bosons, even if not globally?

Moreover, the stationarity result from Theorem 1 (proven in Ref [9]), shows that the multimode

bunching probability $P_n(S)$ is unaffected by first order perturbations. The vanishing of the first derivative is a necessary condition for local maximum; the second condition is that the second derivative must be negative.

Theorem 1. *(Pioge et al. 2023). Consider any linear interferometer $\hat{U}$ and any (nontrivial) subset $\mathcal{K}$ of output modes. For an arbitrary choice of the perturbation parameters v_i and (normalized) perturbation state $|\eta_i\rangle$, the multimode bunching probability P_n into $\mathcal{K}$ is stationary i.e.*

$$\frac{\partial P_n(S)}{\partial \epsilon} = 0 \tag{2.23}$$

Finally, it has been observed out that the counterexamples to Conjecture P1 are, in a certain sense, far from indistinguishability. More precisely, any total system state $|\Psi\rangle = |\phi_1\rangle \dots |\phi_n\rangle$ contains a permutationally symmetric component, whose weight is quantified by the *indistinguishability measure* $d(S)$:

$$d(S) = \frac{\mathrm{perm}(S)}{n!} \tag{2.24}$$

This measure takes the value $d(\mathbb{E}) = 1$ for perfectly indistinguishable photons, but is significantly smaller in Drury's counterexample to Conjecture M1.

To properly introduce the local maximum conjecture. We must first define:

Definition 6. *separable infinitesimal perturbation: perturbation $|\eta_i\rangle$ proportional by a factor ϵv_i applied to the internal state $|\phi_0\rangle$ of the photon i. Its perturbed internal state reads as (with a normalization constant α_i):*

$$|\phi_i\rangle = \frac{1}{\alpha_i}(|\phi_0\rangle + \epsilon v_i |\eta_i\rangle) \tag{2.25}$$

Definition 7. *near indistinguishability: characteristic of two (normalized) internal states $|\psi\rangle$ and $|\phi\rangle$ whose overlap $\langle\cdot|\cdot\rangle$ is*

$$\langle\phi|\psi\rangle = 1 + \mathcal{O}(\epsilon^2) \tag{2.26}$$

An example of two nearly indistinguishable bosons is one in a horizontally polarized state, while the other experiences an infinitesimal perturbation in the direction of an orthogonal state: $|\psi\rangle = |H\rangle$ and $|\phi\rangle = \frac{1}{\alpha_i}(|H\rangle + \epsilon v_i |V\rangle)$. This scenario will be used explicitly in section 2.3.1). Notice that *near indistinguishability* is a special case of *partial distinguishability.*

All these observations naturally lead to a conjecture P2 regarding the local maximum as stated in Ref [9] by the authors:

Conjecture P2. *[9] (Pioge et al. 2023) Consider any linear interferometer $\hat{U}$ and any subset $\mathcal{K}$ of output modes. Starting from the state of n indistinguishable photons, the probability that all output photons bunch into $\mathcal{K}$ can only decrease if a separable infinitesimal perturbation is applied to the internal state of the photons, making them slightly distinguishable.*

The validity of the conjecture P1 would in turn, imply the validity of Conjecture P2. Interestingly, this latter conjecture appears to be coincidentally linked to another mathematical Conjecture M2, originally proposed by Bapat and Sunder in Ref [17]. More specifically, Conjecture P2 implies Conjecture M2 (cf. section 2.3.1) but the converse is not true: there is no equivalence as it was the case between Conjecture P1 and Conjecture M1.

Let $A_{\setminus i,\setminus j}$ denote the matrix A without the row i and column j. We define a new matrix F whose (i,j) entry is:

$$F_{i,j} = A_{i,j}\,\mathrm{perm}(A_{\setminus i,\setminus j}) \tag{2.27}$$

Conjecture M2, stated by Bapat and Sunder in Ref [17] in terms of the matrix F (as defined in equation 2.27), is as follows:

Conjecture M2. *(Bapat, Sunder 1986) Let F be the matrix associated to $A \in \mathcal{H}_n$ then* $\mathrm{perm}(A)$ *is the largest eigenvalue of F.*

This conjecture could alternatively be written as: let $A \in \mathcal{H}_n$ then

$$\lambda_1(F) = \mathrm{perm}(A) \tag{2.28}$$

Note that $A \in \mathcal{H}_n$ implies $F \in \mathcal{H}_n$ [17] thus its eigenvalues are non negative. Additionally, $\mathrm{perm}(A)$ is always an eigenvalue of F since by using the Laplace expansion for permanents (equation 2.98) one gets

$$F\mathbf{1} = \mathrm{perm}(A)\mathbf{1} \Leftrightarrow \sum_j F_{i,j} = \mathrm{perm}(A) \tag{2.29}$$

Where $\mathbf{1}$ is the constant vector $\mathbf{1} = (1, \ldots, 1)^t$. However, the conjecture states that the $\mathrm{perm}(A)$ *is* the largest eigenvalue. The determinant analogue to this conjecture can be proven to be true, that is: the smallest eigenvalue of F is $\det(A)$ where F this time is defined as $F_{ij} = a_{ij}\det(A_{\setminus i,\setminus j})$.

Counterexample

The counterexample to conjecture M2 was found by Drury [18] in 2018, this time of dimension (2×8) thus physically corresponding to $n = 8$ photons sent in each input mode but still bunching in $|\mathcal{K}| = 2$ output modes. The matrix is $H = M^\dagger M$ where

$$M = \begin{pmatrix} -7+4i & 9-3i & -6+2i & 3+4i & 7+6i & 4-4i & i & 5-8i \\ 4-5i & 1+4i & -8-2i & -7+4i & 1-4i & 1-8i & 8-6i & 1-3i \end{pmatrix} \tag{2.30}$$

For which the largest eigenvalue $\lambda_1(F)$ of F surpasses $\mathrm{perm}(H) = 2'977'257'622'144'118'400$ by $\lambda_1(F)/\,\mathrm{perm}(H) \approx 1.01956$

This provides another instance of *anomalous bunching* — but now with *nearly indistinguishable* photons, an even stricter condition than *partially distinguishable* photons. As in the global maximum counterexample, one can build a physical interferometer corresponding to the mathematical counterexample M. For that device, the two-mode bunching probability is *not* maximized by fully indistinguishable bosons; instead, it is still exceeded by nearly indistinguishable bosons.

2.3.1 Perturbation Analysis

We will now analyze the implication between the physical conjecture P2 and the mathematical conjecture M2 for the global maximum. In this section, the exact perturbation analysis was done in Ref [9], here we restrict ourselves to an even more specific case of photon's internal state (cf. equation 2.35) which will be explained in details in the following paragraphs.

We analyze how a small perturbation, proportional to a (normalized) state $|\eta_i\rangle$, applied to the internal state $|\phi_0\rangle$ (shared by all photons) affects the bunching probability $P_n(S)$. Let $\epsilon > 0$ denote the overall magnitude of the perturbation, and let v_i be the i-th component of a normalized complex vector $\boldsymbol{v} \in \mathbb{C}^n$.

Note that ϵ and v_i represent distinct quantities: ϵ controls the global strength of the perturbation whereas v_i scales the contribution of the perturbation for photon i. Since $\boldsymbol{v}$ is normalized, the rescaling required during normalization, namely $\boldsymbol{v} \to \boldsymbol{v}/\|\boldsymbol{v}\|$, can be absorbed into ϵ, effectively transforming it as $\epsilon \to \|\boldsymbol{v}\|\epsilon$.

In summary, each photon is initially in the internal state $|\phi_0\rangle$. A small perturbation is then applied individually to each photon, modifying its internal state by an amount $\epsilon v_i |\eta_i\rangle$. As a result, the photon i enters the interferometer in the state:

$$|\phi_i\rangle = \frac{1}{\alpha_i}(|\phi_0\rangle + \epsilon v_i |\eta_i\rangle) \quad \forall i \in [n] \tag{2.31}$$

Here α_i is a normalization factor ensuring $\langle\phi_i|\phi_i\rangle = 1$

$$\alpha_i = \sqrt{1 + 2\epsilon\Re[\overline{v}_i \langle\eta_i|\phi_0\rangle] + \epsilon^2|v_i|^2} \tag{2.32}$$

The stationarity of P_n [9], opens up the possibility of it being a potential local maximum. However, to verify if it is indeed a local maximum, we must continue with the calculation for the second order. Its second order expansion is particularly interesting to study because it will reveal how Conjecture P2 and Conjecture M2 are linked.

Second Order Perturbation

We now present a detailed account of the second order perturbation of the internal state of photons. The following derivations are taken directly from Ref [9], and are reproduced here.

Let us assume that the internal state $|\phi_0\rangle$ is orthogonal to the perturbation state $|\eta_i\rangle$ for each photon:

$$\langle\eta_i|\phi_0\rangle = 0 \quad \forall i \in [n] \tag{2.33}$$

Additionally, let us assume that the perturbed internal state of every photon is spanned by the same two orthogonal vectors: for example a horizontal polarization state $|H\rangle$ and vertical $|V\rangle$. Thus equation 2.31 becomes:

$$|\phi_i\rangle = \frac{1}{\alpha_i}(|H\rangle + \epsilon v_i |V\rangle) \quad \forall i \in [n] \tag{2.34}$$

$$\langle H|V\rangle = 0, \quad \langle H|H\rangle = \langle V|V\rangle = 1 \tag{2.35}$$

These two assumptions alone, equation 2.31 and equation 2.33, restrict our Conjecture P2 down to specific cases, which will be relevant to understand the relation to its mathematical counterpart.

In the second order (the perturbation magnitude converges as $\epsilon \to 0$), using the McLaurin expansion $(1+x)^n \xrightarrow{x\to 0} 1 + nx$, the normalization constant α_i becomes:

$$\frac{1}{\alpha_i} = (1 + \epsilon^2|v_i|^2)^{-1/2} = 1 - \frac{1}{2}\epsilon^2|v_i|^2 \tag{2.36}$$

and their product can be simplified to:

$$\frac{1}{\alpha_i\alpha_j} = 1 - \frac{1}{2}\epsilon^2(|v_i|^2 + |v_j|^2) + \mathcal{O}(\epsilon^4) \tag{2.37}$$

using the orthonormality in equation 2.35, the projection of the perturbed internal state of the photon i into photon j is:

$$(\langle H| + \epsilon\overline{v}_i \langle V|)(|H\rangle + \epsilon v_j |V\rangle) = 1 + \epsilon^2\overline{v}_i v_j \tag{2.38}$$

Using the last two equations, the distinguishability matrix S as defined in equation 2.12 becomes:

$$S_{ij} = \langle\phi_i|\phi_j\rangle \tag{2.39}$$

$$= 1 + \epsilon^2[\overline{v}_i v_j - \frac{1}{2}(|v_i|^2 + |v_j|^2)] + \mathcal{O}(\epsilon^4) \tag{2.40}$$

$$= 1 + \epsilon^2 X_{ij} + \mathcal{O}(\epsilon^4) \tag{2.41}$$

Where we defined the matrix X:

$$X_{ij} = \overline{v}_i v_j - \frac{1}{2}(|v_i|^2 + |v_j|^2) \tag{2.42}$$

Let us remind the Hadamard product $(a \odot b)_{i,j} = a_{i,j} b_{i,j}$, and use the matrix $\mathbb{E}$ as previously defined in equation 2.18. The multimode bunching probability $P_n(S)$ is:

$$\begin{aligned} P_n(S) &= \text{perm}(H \odot S) && (2.43)\\ &= \text{perm}(H \odot \mathbb{E} + H \odot \epsilon^2 X) + \mathcal{O}(\epsilon^4) && (2.44)\\ &= \text{perm}(B + A) + \mathcal{O}(\epsilon^4) && (2.45) \end{aligned}$$

Where we defined $B = H \odot \mathbb{E} = H$ and $A = H \odot \epsilon^2 X$. The permanent of the sum is given by the Minc's formula:

$$\text{perm}(A + B) = \sum_{\substack{i,j \subseteq [n] \\ |i| = |j|}} \text{perm}(A_{i,j}) \, \text{perm}(B_{\setminus i, \setminus j}) \tag{2.46}$$

$$\tag{2.47}$$

Where the sum involves i, j which are all subsets of $[n]$ such that their size are the same i.e. $|i| = |j|$: the sum goes over all possibilities of zero columns and rows, then one column and row, and so on and so forth until n columns and rows.

We decompose the sum in the Minc's formula into three parts based on the choice of subsets i and j:

1. empty set: this corresponds to $i = j = \emptyset$, in that case this simplifies to

$$\sum_{\substack{i,j \\ i=j=\emptyset}} \text{perm}(A_{i,j}) \, \text{perm}(B_{\setminus i, \setminus j}) = \text{perm}(A_{\emptyset,\emptyset}) \, \text{perm}(B) = \text{perm}(H) \tag{2.48}$$

2. singleton subsets: this corresponds to subset i consisting of a single element from n, that is $|i| = 1$, likewise for $|j| = 1$. The sum over all such possible i and j subsets simplifies to:

$$\begin{aligned} \sum_{\substack{i,j \subseteq [n] \\ |i| = |j| = 1}} \text{perm}(A_{i,j}) \, \text{perm}(B_{\setminus i, \setminus j}) &= \sum_{\substack{i,j \subseteq [n] \\ |i| = |j| = 1}} \text{perm}((H \odot \epsilon^2 X)_{i,j}) \, \text{perm}(H_{\setminus i, \setminus j}) && (2.49)\\ &= \sum_{\substack{i,j \subseteq [n] \\ |i| = |j| = 1}} (H \odot \epsilon^2 X)_{i,j} \, \text{perm}(H_{\setminus i, \setminus j}) && (2.50)\\ &= \epsilon^2 \sum_{\substack{i,j \subseteq [n] \\ |i| = |j| = 1}} (F \odot X)_{i,j} && (2.51) \end{aligned}$$

3. remaining sets: all the remaining subsets consisting of more than one element from $[n]$ create a matrix $A_{i,j}$ of two or more rows and column whose permanent result in terms containing $\mathcal{O}(\epsilon^4)$

Therefore the bunching probability $P_n(S)$ becomes:

$$P_n(S) = \text{perm}(H) + \epsilon^2 \sum_{i,j} (F \odot X)_{i,j} + \mathcal{O}(\epsilon^4) \tag{2.52}$$

The increase of bunching probability $\delta P_n(S)$ in slightly perturbed bosons compared to indistinguishable ones is:

$$\begin{aligned}\delta P_n &= \mathrm{perm}(H \odot S) - \mathrm{perm}(H) && (2.53)\\ &= \epsilon^2 \sum_{i,j} (F \odot X)_{i,j} + \mathcal{O}(\epsilon^4) && (2.54)\\ &= \epsilon^2 \sum_{i,j} [\overline{v}_i F_{i,j} v_j - \frac{1}{2}(|v_i|^2 F_{i,j} + |v_j|^2 F_{i,j})] + \mathcal{O}(\epsilon^4) && (2.55)\\ &= \epsilon^2 [\boldsymbol{v}^\dagger F \boldsymbol{v} - \mathrm{perm}(H)] + \mathcal{O}(\epsilon^4) && (2.56)\end{aligned}$$

Where for the first term, we used

$$\sum_{i,j} \overline{v}_i F_{i,j} v_j = \boldsymbol{v}^\dagger F \boldsymbol{v} \tag{2.57}$$

Additionally, using Laplace expansion formula for the permanent

$$\sum_i F_{i,j} = \sum_j F_{i,j} = \mathrm{perm}(H) \tag{2.58}$$

and for a normalized vector $\|\boldsymbol{v}\|^2 = 1$, the remaining terms simplify as:

$$\sum_{i,j} |v_i|^2 F_{i,j} = \sum_i |v_i|^2 \sum_j F_{i,j} = \mathrm{perm}(H) \sum_i |v_i|^2 = \mathrm{perm}(H) \tag{2.59}$$

Conjecture 2 claims that $\delta P_n \leq 0$ for all perturbations $\boldsymbol{v}$. In particular, for $\boldsymbol{v}_{\max}$ eigenvector of the maximum eigenvalue $\lambda_1(F)$. Now we can majorize $\delta P_n(S)$ using equation 1.10

$$\delta P_n \leq \max_{\boldsymbol{v}} \delta P_n = \epsilon^2 [\lambda_1(F) - \mathrm{perm}(H)] + \mathcal{O}(\epsilon^4) \tag{2.60}$$

Or equivalently:

$$\lambda_1(F) \leq \mathrm{perm}(H) \tag{2.61}$$

But by definition of the maximum eigenvalue, and since $\mathrm{perm}(H)$ is also an eigenvalue, we have:

$$\mathrm{perm}(H) \leq \lambda_1(F) \tag{2.62}$$

In other words, equation 2.61 implies exactly Conjecture M2:

$$\lambda_1(F) = \mathrm{perm}(H) \tag{2.63}$$

As mentioned previously, Conjecture M2 is a special case of conjecture P2 because of assuming a specific perturbation described in equation 2.31, paired with the condition in equation 2.33. Therefore it is only true that the validity of Conjecture P2 implies the validity of Conjecture M2, and not the converse.

2.4 Permanent-on-Top Conjecture

The *Permanent-on-Top* conjecture M3 serves as a foundation from which emerged the main conjectures M1 and M2 investigated in this work. The goal of this section is to demonstrate that there is no possible *anomalous bunching* for $n = 3$ photons (or less) by proving the validity of the *permanent-on-top* conjecture for the dimension $n = 3$.

2.4.1 Schur Power Matrix

First let us define the $(n! \times n!)$ Schur power matrix $\pi(A)$ associated to the $(n \times n)$ matrix A, indexed by permutations $\sigma, \tau \in \mathcal{S}_n$ defined as follows:

$$(\pi(A))_{\sigma,\tau} = \prod_{i=1}^{n} a_{\tau(i),\sigma(i)} \tag{2.64}$$

Positive semi-definite Hermitianity conservation

It is worth noting that it conserves positve definite Hermitianity, that is, $A \in \mathcal{H}_n$ implies $\pi(A) \in \mathcal{H}_{n!}$ since it is the principal submatrix of the n-fold Kronecker product $\otimes^n A$.

Equivalency in left regular representation

The Schur power matrix $\pi(A)$ can also be described using the left regular representation ρ_{reg} on an $n!$-dimensional inner product space V spanned by the orthonormal basis $V = \{e_\sigma : \sigma \in \mathcal{S}_n\}$

$$\pi(A) = \sum_{\sigma \in S_n} d_A(\sigma) \rho_{\text{reg}}(\sigma) \tag{2.65}$$

One can verify this by computing its matrix element (σ, τ) using equation (A.2)

$$(\pi(A))_{\sigma,\tau} = \sum_{\alpha \in S_n} d_A(\alpha) \langle e_\sigma | \underbrace{\rho_{\text{reg}}(\alpha) | e_\tau \rangle}_{|e_{\alpha\tau}\rangle} \tag{2.66}$$

since the scalar product is nonzero only when $\sigma = \alpha\tau \Leftrightarrow \sigma\tau^{-1} = \alpha$

$$(\pi(A))_{\sigma,\tau} = d_A(\sigma\tau^{-1}) = \prod_i a_{i,\sigma\tau^{-1}(i)} \tag{2.67}$$

Since $\sigma\tau^{-1}(i) = \sigma(\tau^{-1}(i))$, by substituting $j = \tau^{-1}(i) \Leftrightarrow \tau(j) = i$ one finally gets

$$(\pi(A))_{\sigma,\tau} = \prod_j a_{\tau(j),\sigma(j)} \tag{2.68}$$

Generalized matrix function as eigenvalues

It can be shown [17] that the Schur power matrix $\pi(A)$ admits the generalized matrix function $f_{\chi,G}(A)$ as eigenvalue for any character $\chi(\sigma)$ of a subgroup $G \subseteq \mathcal{S}_n$ of the symmetric group. In particular, for $\pi(A)$ operator on $V = \{e_\sigma : \sigma \in \mathcal{S}_n\}$ associated to the $n \times n$ matrix A, one can verify that indeed it admits both $\det(A)$ and $\text{perm}(A)$ as eigenvalues but only the former $\det(A)$ has been proven (implicitly) by Schur to be the lowest.

$$\begin{cases} \pi(A)v_{\text{perm}} = \text{perm}(A)v_{\text{perm}}, & v_{\text{perm}} := \sum_\sigma e_\sigma \\ \pi(A)v_{\det} = \det(A)v_{\det}, & v_{\det} := \sum_\sigma \text{sgn}(\sigma) e_\sigma \end{cases} \tag{2.69}$$

2.4.2 Conjecture

Schur in Ref [19] implicitly showed that the smallest eigenvalue of the operator $\pi(A)$ associated to the matrix $A \geq 0$ is $\det(A)$. This motivated the introduction of a conjecture counterpart for permanents, introduced by Soules [20], referred as *permanent-on-top* in this work, and was stated in Ref [17] as follows:

Conjecture M3. *(Soules 1983) Let $A \geq 0$, then is the largest eigenvalue of the Schur power matrix $\pi(A)$ associated to the matrix A equal to the permanent of A.*

Equivalent inequality

One way to describe the Conjecture M3 is the following:

$$\pi(A) \leq \text{perm}(A)\mathbb{I} \tag{2.70}$$

This inequality could be rewritten by explicitly writing the eigenvalues of $\pi(A)$ as being majorized by perm(A). For $\pi(A) \in \mathcal{H}_{n!}$, it is Hermitian and positive semi-definite, and hence admits a spectral decomposition:

$$\pi(A) = V\Lambda V^\dagger \tag{2.71}$$

Where V is a matrix whose columns are orthonormal eigenvectors of $\pi(A)$. The matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{n!})$ containing the eigenvalues is diagonal. This implies

$$V^\dagger \pi(A) V = \Lambda \tag{2.72}$$

Moreoever, since the permanent is invariant under unitary transformation

$$V^\dagger(\text{perm}(A)\mathbb{I})V = \text{perm}(A)\mathbb{I} \tag{2.73}$$

therefore the inequality (2.70) becomes

$$\Lambda \leq \text{perm}(A)\mathbb{I} \tag{2.74}$$

This holds if and only if perm(A) is the greatest eigenvalue of $\pi(A)$ and thus is equivalent to equation (2.70). One could write it even more directly:

$$\lambda_1(\pi(A)) = \text{perm}(A) \tag{2.75}$$

Where we denote the i-th greatest eigenvalue $\lambda_i(X)$ of the $(n \times n)$ matrix X sorted in descending order

$$\lambda_1 \geq \cdots \geq \lambda_n \tag{2.76}$$

Proof

The conjecture M3 has been proven by Bapat and Sunder for $n \leq 3$ in Ref [17]. The proof is the following.

Proof. For $n = 1$ it is trivial. For $n = 2$, $\pi(A) \in \mathcal{H}_n$ if $A \in \mathcal{H}_n$ therefore it can be decomposed as $\pi(A) = V\Lambda V^\dagger$ as explained in equation (2.69), that is:

$$V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}; \quad \Lambda = \begin{pmatrix} \text{perm}(A) & 0 \\ 0 & \det(A) \end{pmatrix} \tag{2.77}$$

Additionally, $\forall A \geq 0 : \text{perm}(A) \geq \det A$ which concludes the proof for n=2.

For $n = 3$, since the Schur power matrix $\pi(A)$ can be written using equation (2.65) in the left regular representation ρ_{reg}, and since $A \geq 0$, the inequality (2.70) is equivalent to: $\forall x \in V = \{e_\sigma : \sigma \in \mathcal{S}_n\}$

$$\sum_{\sigma \in \mathcal{S}_n} d_A(\sigma) \langle x| \rho_{\text{reg}}(\sigma) |x\rangle \leq \text{perm}(A) \langle x|x\rangle \tag{2.78}$$

As mentioned in equation (A.14), since the left regular representation ρ_{reg} can be decomposed as a direct sum of irreducible representations $\rho(\sigma)$ of S_n, inequality (2.78) is equivalent to:

$$\alpha := \sum_{\sigma \in \mathcal{S}_n} d_A(\sigma)\rho(\sigma) \leq \text{perm}(A)\mathbb{I} \tag{2.79}$$

S_3 only has 3 irreducible representations: ρ_{triv}, ρ_{alt}, and ρ_{std}. The latter being a restriction of the natural representation ρ_{nat} to the subspace $V = \{v \in \mathbb{C}^3\} \perp |1\rangle$ orthogonal to the constant vector $|1\rangle = (1,1,1)^t$ as described in equation (A.11). It suffices to show the particular case of the representation $\rho(\sigma) = \rho_{\text{nat}}(\sigma)$. The (i,j) entry of the operator α becomes:

$$\alpha_{i,j} = \sum_{\sigma \in \mathcal{S}_n} d_A(\sigma) \langle e_j | \underbrace{\rho_{\text{nat}}(\sigma) |e_i\rangle}_{|e_{\sigma(i)}\rangle} \tag{2.80}$$

Using equation (A.4), and noting that the scalar product is nonzero only when $j = \sigma(i)$, this reduces down to

$$\alpha_{i,j} = \sum_{j=\sigma(i)} d_A(\sigma) = F_{i,j} \tag{2.81}$$

Where the entry (i,j) of the newly defined matrix F is given by:

$$F_{i,j} := A_{i,j} \operatorname{perm}(A_{\setminus i,\setminus j}) \tag{2.82}$$

Here, $A_{\setminus i,\setminus j}$ denotes the matrix A without the row i and column j. Note again that $A \in \mathcal{H}_n$ implies $F \in \mathcal{H}_n$. The inequality to prove becomes

$$F \leq \operatorname{perm}(A)\mathbb{I} \tag{2.83}$$

Which is equivalent to checking the semi definite-positiveness of the matrix G defined as:

$$G = \operatorname{perm}(A)\mathbb{I} - F \geq 0 \tag{2.84}$$

In an attempt to further simplify the matrix A by reducing the number of working entries $a_{i,j}$. Observe that, for a specific diagonal matrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ the inequality (2.70) remains invariant when A undergoes the following transformation

$$A \to DAD^\dagger \tag{2.85}$$

Thus for any matrix A it is always possible to choose D such that after transformation, A has a normalized diagonal. Additionally, since $A \geq 0$ we get:

$$A = \begin{pmatrix} 1 & d & e \\ \overline{d} & 1 & f \\ \overline{e} & \overline{f} & 1 \end{pmatrix} \tag{2.86}$$

The diagonally normalized matrix $A \to DAD^\dagger$ can be without loss of generality further simplified [17]: its element f remains complex, while it is sufficient for $d, e \geq 0$ to be real non-negative. Intuitively, this matrix A can be treated similarly to the indistuighisbality matrix $S_{i,j} = \langle \phi_i | \phi_j \rangle$ whose elements are also equivalent up to global phase. Indeed, the latter is also normalized diagonally because of orthonormal states $|\phi_i\rangle$. Since the quantum state $|\phi_i\rangle$ is physically equivalent to $\exp(i\theta_i) |\phi_i\rangle$ we can redefine

$$|\phi_i\rangle \to \exp(i\theta_i) |\phi_i\rangle \tag{2.87}$$

Under this transformation, the entries are equivalent up to a global phase $\exp(i(\theta_j - \theta_i))$

$$\langle \phi_i | \phi_j \rangle \to \exp(i(\theta_j - \theta_i)) \langle \phi_i | \phi_j \rangle \tag{2.88}$$

Let us eliminate the phase for d and e

$$d = \exp(i\alpha)|d| \to \exp(i(\theta_2 - \theta_1))d = |d| \tag{2.89}$$

$$e = \exp(i\beta)|e| \to \exp(i(\theta_3 - \theta_1))e = |e| \tag{2.90}$$

This fixes $(\theta_3 - \theta_2) = (\alpha - \beta)$ thus f transforms as:

$$f = \exp(i\gamma)|f| \to \exp(i(\alpha - \beta))f = \exp(i(\alpha - \beta + \gamma))|f| \tag{2.91}$$

Thus f remains the only complex entry after cancelling the global phase for d and e effectively making them real entries. Let us rename:

$$|d| \to d \geq 0 \tag{2.92}$$

$$|e| \to e \geq 0 \tag{2.93}$$

$$\exp(i(\alpha - \beta + \gamma))|f| \to f \in \mathbb{C} \tag{2.94}$$

The matrix A becomes:

$$A = \begin{pmatrix} 1 & d & e \\ d & 1 & f \\ e & \overline{f} & 1 \end{pmatrix} \tag{2.95}$$

From this one can construct the matrix G from equation (2.84)

$$G = \begin{pmatrix} \alpha + \beta & -\alpha & -\beta \\ -\overline{\alpha} & \overline{\alpha} + \gamma & -\gamma \\ -\overline{\beta} & -\overline{\gamma} & \overline{\beta} + \overline{\gamma} \end{pmatrix} \tag{2.96}$$

Where $\alpha = d^2 + def$, $\beta = e^2 + de\overline{f}$, $\gamma = |f|^2 + def$. Now we will use equation (1.7) to prove that $G \geq 0$. First notice: $\det G = 0$ because $G|1\rangle = 0$. The two remaining leading principal minors are: $G_{1,1}$ and $\det(G_{[1;2],[1:2]})$. The first one is positive, using the definition (2.84)

$$G_{1,1} = \text{perm}(A) - F_{1,1} = \sum_{j \neq 1} F_{1j} \geq 0 \tag{2.97}$$

Where we used the Laplace expansion formula for the permanent equation 2.58:

$$\text{perm}(A) = \sum_i F_{i,j} = \sum_j F_{i,j} = F_{jj} + \sum_{j \neq i} F_{i,j} \tag{2.98}$$

and the fact that both $\text{perm}(A) \geq 0$ and $F_{jj} \geq 0$ (since $F \in \mathcal{H}_n$ is implied by $A \in \mathcal{H}_n$) are non negative real, implying that $\sum_{j \neq i} F_{i,j} \geq 0$ must be as well. The second leading principal minor is:

$$\det(G_{[1;2],[1:2]}) = \begin{vmatrix} \alpha + \beta & -\alpha \\ -\overline{\alpha} & \overline{\alpha} + \gamma \end{vmatrix} = \alpha\gamma + \beta\gamma + \overline{\alpha}\beta \tag{2.99}$$

After substiting α, β, γ then for $f = a + ib$ this becomes:

$$\alpha\gamma + \beta\gamma + \overline{\alpha}\beta = (de)^2(1 + |f|^2) + (d^2 + e^2)|f|^2 + de(\Re[f])^2 + 2de\Re[f](d^2 + e^2 + |f|^2) \tag{2.100}$$

$$= 3(dea)^2 + (d^2 + e^2 + 2dea)b^2 \tag{2.101}$$

$$+ 2dea(d^2 + e^2 + a^2) \tag{2.102}$$

The first term is positive

$$3(dea)^2 \geq 0 \tag{2.103}$$

The second term can be simplified using $A \geq 0 \Rightarrow \det(A_{[2:3],[2:3]}) \geq 0 \Leftrightarrow |f|^2 < 1$ and $|a| \leq |f|$

$$d^2 + e^2 + 2dea \geq d^2 + e^2 - 2|dea| \tag{2.104}$$

$$\geq d^2 + e^2 - 2de = (d - e)^2 \geq 0 \tag{2.105}$$

The third term can be expressed as a quadratic form of a newly defined matrix C evaluated at a specific vector $\boldsymbol{v}$

$$2dea(d^2 + e^2 + a^2) = \boldsymbol{v}^\dagger C \boldsymbol{v} \tag{2.106}$$

Where

$$C = \begin{pmatrix} 1 & a & e \\ a & 1 & d \\ e & d & 1 - b^2 \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} ea \\ da \\ de \end{pmatrix} \tag{2.107}$$

Note that C is Hermitian thus it suffices to prove $C \geq 0$ using equation (1.7). The first principal minor is trivial. As for the second, since $a^2 \leq |f|^2 < 1$

$$\det(C_{[1:2],[1:2]}) = 1 - a^2 \geq 1 - |f|^2 > 0 \tag{2.108}$$

Finally, the last leading principal minor after simplification becomes

$$\det C = \det A + (ab)^2 \geq \det A > 0 \tag{2.109}$$

This proves $C \geq 0$ which is equivalent to equation (1.6). Now since

$$\boldsymbol{x}^\dagger C \boldsymbol{x} \geq 0 \quad \forall \boldsymbol{x} \in \mathbb{C}^3 \tag{2.110}$$

In particular it is true for $\boldsymbol{x} = \boldsymbol{v}$ which concludes the proof. □

To conclude, this proof has the following physical implication: there is no anomalous bunching for $n \leq 3$.

An attempt was made to extend the proof for $n = 4$ (instead of $n = 3$) by computing all terms explicitly starting from a matrix A with this time $(n-1)(n-2)/2 = 3$ instead of 1 complex entries. Although this approach is, in principle, analytically feasible, the sheer number of terms soon became unmanageable and the attempt was ultimately unsuccessful. For example, this time the matrix F would have as entries, a permanent of dimension $(n-1) = 3$ with $(n-1)! = 3!$ terms instead of 2!. This would lead to the second principal minor of G (to be proven $\det(G_{[1:2],[1:2]}) \geq 0$) to have $((n-1)^2 - 1)((n-1)!)^2 = 288$ terms[3] instead of 12. The third principal minor and the determinant of G would naturally have even more terms.

Counterexample

Shchesnovich [21] found a $n = 5$ dimensional counterexample $A = M^\dagger M$ to conjecture M3 using numerical search where:

$$M = \begin{pmatrix} 4 - 2i & 2 + 3i & -4 + 4i & -3 - 4i & 1 \\ 2 + 4i & 3i & 2 + 4i & 3i & -5 + 7i \end{pmatrix} \tag{2.111}$$

Indeed, the largest eigenvalue $\lambda_1(\pi(A)) = 320(2'185'775 + \sqrt{160'600'333'345})$ of $\pi(A)$ is bigger than its permanent $\text{perm}(A) = 814'016'640$.

Soules showed that disproving conjecture M3 for real matrices implies that the smallest order for which it fails must be a singular counterexample. It was proven by Drury that conjecture M3 did fail for real matrices.

Despite the conjecture M3 being proven for $n \leq 3$ but failing for $n \geq 5$, the case of $n = 4$ still remains unkown.

[3] $(n-1)^2$ comes from the product of diagonal entries of G, one diagonal entry having $(n-1)$ entries of F, The substraction of -1 from $(n-1)^2$ comes from a simplification that happens similarly as in equation 2.99

As for the physical implication [12], the *permanent-on-top* conjecture M3 differs from the global maximum conjecture M1 by the constraint on the internal states of the input photons: the former allows for the internal states to be entangled (non-separable) while the latter restricts to only non-entangled states. Therefore, this counterexample proves that *anomalous bunching* is possible for $n = 5$ entangled (internal state) particles.

Part II

Results

Chapter 3

Analytical Investigations

Previously, in section 2.3, the local maximum conjecture M2 has been falsified. However, there exists another conjecture M4 implied by conjecture M2. Hence, the main objective of this master thesis is to find a physical interpretation, that is, to construct a physical conjecture P3 corresponding to its mathematical counterpart — conjecture M4. A secondary objective would be to verify the validity of conjecture M4.

3.1 Pate Conjecture

The local maximum Conjecture M2 implies a new conjecture M4 in Ref [22], which, unlike the former, has not yet been disproven. This conjecture is from Ref [23] and is the following:

Conjecture M4.

$$\frac{F_{11} - F_{12} - F_{21} + F_{22}}{2} \leq \mathrm{perm}(A) \tag{3.1}$$

This conjecture was the original focus of our investigation, which later shifted toward a broader set of conjectures that lie between Conjecture M2 and Conjecture M4. To understand this shift, it is necessary to explain why Conjecture M4 turns out to be a special case of Conjecture M2, a relationship that will be discussed in detail in this section.

3.1.1 Conjecture Derivation

We start with Conjecture M2 that is equation 2.28, where $A \in \mathcal{H}_n$

$$\lambda_1(F) = \mathrm{perm}(A) \tag{3.2}$$

Using equation 1.10, it majorizes the Rayleigh Quotient $R_F(\boldsymbol{v})$ for any vector $\boldsymbol{v} \in \mathbb{C}^n$

$$R_F(\boldsymbol{v}) = \frac{\boldsymbol{v}^\dagger F \boldsymbol{v}}{\boldsymbol{v}^\dagger \boldsymbol{v}} \leq \max_{\boldsymbol{v}} R_F(\boldsymbol{v}) = \lambda_1(F) = \mathrm{perm}(A) \tag{3.3}$$

Since the only entries of F appearing in equation 3.1 are the first two columns and rows, equation 3.3 is a fortiori true for redefined $\boldsymbol{v}$ with only first two nonzero components:

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{3.4}$$

Thus the numerator of the left hand side inequality simplifies to

$$\boldsymbol{v}^\dagger F \boldsymbol{v} = |v_1|^2 F_{11} + \overline{v}_1 v_2 F_{12} + v_1 \overline{v}_2 F_{21} + |v_2|^2 F_{22} \tag{3.5}$$

Which satisfies equation 3.1 when $|v_1|^2 = |v_2|^2 = 1$ and $\overline{v}_1 v_2 = v_1 \overline{v}_2 = -1$ and $\boldsymbol{v}^\dagger \boldsymbol{v} = |v_1|^2 + |v_2|^2 = 2$, that is:

$$v_1 = \exp(i\theta) \quad \forall \theta \in [0, 2\pi[\tag{3.6}$$

$$v_2 = -v_1 \tag{3.7}$$

This can be interpreted as follows: let the internal state $|H\rangle$ of the boson i be perturbed by $\epsilon v_i |V\rangle$ so that:

$$|\phi_i\rangle = \frac{1}{\alpha_i}(|H\rangle + \epsilon v_i |V\rangle) \tag{3.8}$$

This setup suggests that Conjecture M4 is a special case of Conjecture M2. Specifically, it imposes the following additional conditions:

1. Only two bosons are perturbed: $v_i = 0 \quad \forall i \in [n] \backslash \{1, 2\}$
2. The perturbation of one boson is exactly opposite in phase to that of the other: $v_2 = -v_1$, with both having the same magnitude: $|v_1| = |v_2|$, regardless of the perturbation's orientation: $v_1 = \exp(i\theta) \quad \forall \theta \in [0, 2\pi[$

We can associate a physical Conjecture P3 to the mathematical conjecture M4:

Conjecture P3. *Consider any linear interferometer $\hat{U}$ and any subset $\mathcal{K}$ of output modes. Starting from the state of n indistinguishable photons, the probability that all output photons bunch into $\mathcal{K}$ can only decrease if a separable infinitesimal perturbation,* ***equal in magnitude but opposite in phase****, is applied to the internal state of two photons.*

3.2 Intermediate Conjectures

Regarding the first previously stated requirement (that only two photons are perturbed): the first two columns and rows are clearly chosen arbitrarily, since the matrix F can be permuted by rearranging its columns and rows.

A more precise and general formulation would refer not the submatrix $F_{[1:2],[1:2]}$, but rather to an arbitrary (2×2) *principal submatrix* $f^{(2)}$ of F. This is equivalent to having a perturbation vector $\boldsymbol{v}$ with exactly two nonzero components, corresponding to two perturbed photons, while the remaining $(n-2)$ photons remain unperturbed.

Let us now generalize the first requirement from $k = 2$ perturbed photons to any k perturbed photons among n. Moreover, we will drop the second requirement, which imposes a specific structure on the perturbation vector $\boldsymbol{v}$. This allows us to define a family of conjectures $\forall k \in [n] \backslash \{1\}$ (excluding the case of a single perturbed photon) which interpolate between Conjecture M4 and Conjecture M2. More specifically, each member of this family is described as follows:

Conjecture M5. *Let $f^{(k)}$ be any $(k \times k)$ principal submatrix of F associated to $A \in \mathcal{H}_n$ then* $\mathrm{perm}(A)$ *is the largest eigenvalue of $f^{(k)}$*

Again, one of the possible inequalities for this conjecture (using the notation of eigenvalues in descending order) would be:

$$\lambda_1(f^{(k)}) = \mathrm{perm}(A) \tag{3.9}$$

It is worth noting that the validity of the Conjecture M5 for any given k, starting with $k = n$ (which corresponds exactly to Conjecture M2), implies the validity of the conjecture for $(k-1)$,

and so on, down to $k = 2$. Indeed, Conjecture M2 implies Conjecture M5 because the former applies to any perturbation vector $\boldsymbol{v}$, and in particular, one may choose $\boldsymbol{v}$ with only k nonzero components, yielding a matrix F whose relevant block is precisely $f^{(k)}$.

As previously discussed, the case $k = 2$ in this family of conjectures implies conjecture M4. The latter, however, adds further constraint: the perturbation vector $\boldsymbol{v}$ must satisfy the specific condition described in the second requirement above. Note that such perturbation vector is also a real vector.

The physical Conjecture P4 that can be attributed to the mathematical Conjecture M5 is:

Conjecture P4. *Consider any linear interferometer $\hat{U}$ and any subset $\mathcal{K}$ of output modes. Starting from the state of n indistinguishable photons, the probability that all output photons bunch into $\mathcal{K}$ can only decrease if a separable infinitesimal perturbation is applied to the internal state of k* ***photons*** *(**among** n).*

3.3 Real Intermediate Conjectures

It is important to note that conjecture M5 does not impose the eigenvector $\boldsymbol{v}_{\max} \in \mathbb{C}^n$, associated with the largest eigenvalue $\lambda_1(f^{(k)})$, to be real. This contrasts with Conjecture M4, which does require real perturbation vectors.

Therefore, we can define a related set of Conjectures M6, which are real-valued counterparts of Conjectures M5. In these versions, we restrict to *real eigenvectors* $\boldsymbol{v} \in \mathbb{R}^n$.

Restricting to real eigenvectors of a Hermitian positive semi-definite matrix $F \in \mathcal{H}_n$, is in the context of Rayleigh quotient $R_F(\boldsymbol{v})$, equivalent to considering its real part $\Re[F]$: if F is decomposed into its real $\Re[F]$ and imaginary $\Im[F]$ part then:

$$\boldsymbol{v}^\dagger F \boldsymbol{v} = \boldsymbol{v}^\dagger \Re[F] \boldsymbol{v} + i \boldsymbol{v}^\dagger \Im[F] \boldsymbol{v} \tag{3.10}$$

However, since $F \geq 0$ is positive semi-definite and we restrict ourselves to real vectors $\boldsymbol{v} \in \mathbb{R}^n$, it implies that:

$$\boldsymbol{v}^\dagger F \boldsymbol{v} = \boldsymbol{v}^\dagger \Re[F] \boldsymbol{v} \quad \forall \boldsymbol{v} \in \mathbb{R}^n \tag{3.11}$$

Conjecture M6. *Let $f^{(k)}$ be any $(k \times k)$ principal submatrix of $\Re[F]$ associated to $A \in \mathcal{H}_n$ then* $\mathrm{perm}(A)$ *is the largest eigenvalue of $f^{(k)}$*

The physical Conjectures P5 implying the mathematical Conjectures M6 are as follows:

Conjecture P5. *Consider any linear interferometer $\hat{U}$ and any subset $\mathcal{K}$ of output modes. Starting from the state of n indistinguishable photons, the probability that all output photons bunch into $\mathcal{K}$ can only decrease if a separable infinitesimal* ***real perturbation*** *is applied to the internal state of k photons (among n).*

The only known counterexample to the local maximum conjecture M2 has a perturbation vector $\boldsymbol{v}$ with:

- no nonzero components — all photons are infinitesimally perturbed (and not just their subset) $v_i \neq 0 \quad \forall i$, thus we don't know if the set of conjectures M5 is true.
- no complex components, hence the validity of the real counterpart of conjectures M5, that is conjecture M6, is still undetermined.

Every previously presented physical conjecture, along with its relationship to its mathematical counterpart, is summarized in Table 3.1.

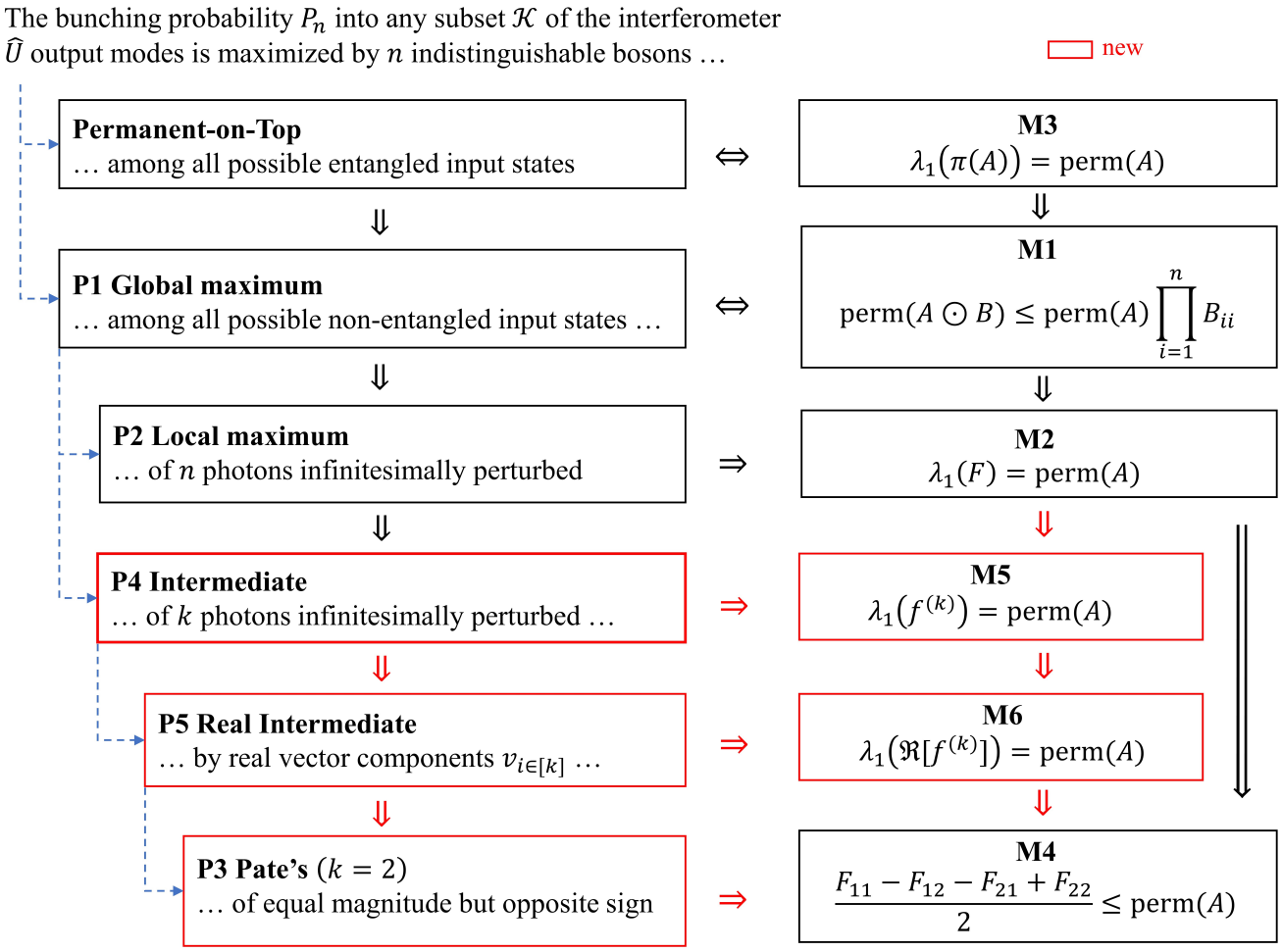


Figure 3.1: Tree of conjectures representing the relationship between physical (left) and mathematical (right) conjectures. The dotted blue lines represent a continuation (initiated by three dots '...') of a statement (ended with three dots '...'). For example, statement found in P4 is a continuation of P1 not P2. New contributions from this work, such as conjectures and the mathematical implications, are represented by the color red. Note that some implications were already known, such as M2 implying M4.

Chapter 4

Numerical Experiments

Analytically proving the conjectures is unfortunately quite challenging. As illustrated by the example discussed at the end of the proof in Section 2.4.2. Therefore, we now turn to a numerical approach, aiming to explicitly construct counterexamples that could potentially falsify the intermediate conjectures M5 (where only a subset of photons is perturbed), their real counterpart — conjectures M6, or even ultimately the special case of two oppositely yet equal in magnitude perturbed bosons, that is conjecture M4.

We will be studying specific (n, r, k) cases where:

- n is the number of photons
- r is the size $|\mathcal{K}| = r$ of the subset $\mathcal{K}$ of the m output modes into which all n photons bunch, with $r \leq m$
- k is the number of *perturbed* photons among n, with $k \leq n$

Those parameters play the following role in the numerical search: n and r define the size $(r \times n)$ of the matrix M from which not F but its principal matrix $f^{(k)}$ is calculated, which as the notation suggests, is determined by the parameter k.

4.1 Problem Statement

Let us first properly define the problem to be solved numerically. The numerical search consists of maximizing a function to optimize $\mathrm{fct}(\boldsymbol{x})$ by varying a vector $\boldsymbol{x}$ (initiated as $\boldsymbol{x}_0$ for the algorithms that would require initiation) which ends up as $\boldsymbol{x}'$ afterwards:

$$\boldsymbol{x}' = \arg\max_{\boldsymbol{x}} \mathrm{fct}(\boldsymbol{x}) \tag{4.1}$$

In our particular case, the function to be maximized is

$$\mathrm{fct}(\boldsymbol{x}) = \frac{\lambda_1(f^{(k)}(\boldsymbol{x}))}{\mathrm{perm}(H(\boldsymbol{x}))} \tag{4.2}$$

The principal submatrix $f^{(k)}$ is constructed from F as:

$$f^{(k)} = F_{[1:k],[1:k]} \tag{4.3}$$

Of course F is being created from H as:

$$F_{ij} = H_{ij}\,\mathrm{perm}(H_{\setminus i,\setminus j}) \tag{4.4}$$

Which in turn A is guaranteed to be positive semi-definite since it is generated from U as:

$$A = H = M^\dagger M \tag{4.5}$$

Where the $(r \times n)$ matrix M is the $(n \times n)$ interferometer $\hat{U}$ limited to a subset $\mathcal{K}$ output modes corresponding to a total number of r modes into which all n photons bunch.

$$M = U_{[1:r],[1:n]} \tag{4.6}$$

To integrate with the code, the $(r \times n)$ matrix M must instead be represented as a $(2rn)$-dimensional real vector $\boldsymbol{x} \in \mathbb{R}^{2rn}$, where the factor 2 accounts for the complex entries of M being decomposed into two real components in $\boldsymbol{x}$.

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_{2rn} \end{pmatrix} \Rightarrow M = \begin{pmatrix} x_1 + ix_2 & \dots & x_{2r(n-1)+1} + ix_{2r(n-1)+2} \\ \vdots & \ddots & \vdots \\ x_{2r-1} + ix_{2r} & \dots & x_{2rn-1} + ix_{2rn} \end{pmatrix} \tag{4.7}$$

The code then operates by varying each component of $\boldsymbol{x}$ to maximize the function $\mathrm{fct}(\boldsymbol{x})$. A counterexample to the intermediate Conjecture M5 is found if $\mathrm{fct}(\boldsymbol{x}) > 1$ for a given $\boldsymbol{x}$. By contrapositive reasoning, since Conjecture P4 implies M4, such counterexample would also refute Conjecture P4.

4.2 Methods

4.2.1 Optimizers type

Optimizers can be roughly categorized in 2: local and global optimizers. The former is generally preferred since it aims to "exploit" already given information and thus is more certain to converge. However local optimizers generally get trapped in a local minimum. Exactness in local optimizers is sacrificed for flexibility: by making use first of exploration before exploitation, which is what global optimizers do. Their state of the art, relies on a sort of metaheuristic algorithm which can be broken down to a "guided randomness":

1. the initial solution is generated often at random (but could be set manually)
2. the solution space is explored through randomness or some sort of probabilistic decision to escape local minima
3. the search is kept broad to avoid premature convergence again via randomness

4.2.2 Convergence failsafe

In order to limit the convergence temporally in the case of an optimizer being stuck, or simply taking too much time for any reason, generally a time limit t_{max} is put.

An additional stop method is a limit on the maximum number of calls $fcall_{max}$ on the function to be optimized, which in essence, is more suitable when working with varying matrix sizes (reusing optimizer for $n = 8$, then $n = 9$...) since for an increased dimension, the time limit t_{max} would have to bet increased accordingly as well contrary to $fcall_{max}$ which would remain constant.

However, one should consider the following as a disclaimer: the maximum function calls $fcall_{max}$ should be adjusted to each algorithm accordingly, and is not at all representative of the time or computational effort, that is, it serves only as a metric measure within the same algorithm, but

should not be used at all to compare two different algorithms. For example, one algorithm may take much more time than another despite having a lower $fcall_{max}$. The former, being only one of the hard coded output (of the algorithm) by the programmer.

Another constraint required by some optimizers consists of an additional predefined domain in which entries are allowed to work: $\boldsymbol{x}_i \in [a, b] \quad \forall i$. Whenever such search range was required, for any optimizer, the given input range was always chosen the same, which is $[-50, +50]$.

4.2.3 Differentiation

As mentioned, some optimizers are derivative based. Differentiations compute explicitly the derivatives at a given point, just like evaluating a function. By default, the method of finite differences is used since it doesn't require any explicit formula for the derivative. However, the drawback is that it is very slow paced because of function evaluations. For example, one typr of finite difference is central difference, where the error term is proportional to the square of the step $\mathcal{O}(h^2)$

$$\partial_x f(x) = \frac{f(a_+) - f(a_-)}{a_+ - a_-} + \mathcal{O}(h^2), \quad a_\pm = x \pm h \tag{4.8}$$

However a much more preferred way is Automatic Differentiation. It is used within measures, that is, if the analytical derivatives of the function are known. However, the derivative of our optimizing function is not supported. In such case, to use Automatic Differentiation, one would need to manually program it.

There are two kinds of automatic differentiation:

- forward: $\mathbb{R}^m \to \mathbb{R}^n$ where each input i is evaluated separately giving Jacobian column J_i thus requiring m passes (1 per input) which is especially efficient when $m < n$

$$J_i = \begin{pmatrix} \partial_{x_i} f_1 \\ \dots \\ \partial_{x_i} f_n \end{pmatrix}, \quad J = \begin{pmatrix} J_1 & \dots & J_m \end{pmatrix} \tag{4.9}$$

- backward: $\mathbb{R}^m \to \mathbb{R}$ where all inputs are evaluated together. It is only applicable for scalar function for which it is much more efficient than forward differentiation. Which is the case of our optimizing function (under the condition that its derivative would be explicitly manually programmed)

$$J_i = \begin{pmatrix} \partial_{x_i} f \end{pmatrix}, \quad J = \begin{pmatrix} J_1 & \dots & J_m \end{pmatrix} \tag{4.10}$$

4.2.4 Optimizers testing

In the first place, different optimization algorithms have been tested from mainly two julia packages: Optim.jl [1] and Optimization.jl[2]. Both serve as a framework to keep a uniform syntax. The latter may require additional packages such as:

- Metaheuristics.jl[3]
- NOMAD.jl[4]
- Evolutionary.jl[5]

[1] https://julianlsolvers.github.io/Optim.jl/stable/
[2] https://docs.sciml.ai/Optimization/stable/
[3] https://github.com/jmejia8/Metaheuristics.jl
[4] https://github.com/bbopt/NOMAD.jl
[5] https://github.com/wildart/Evolutionary.jl

- CMAEvolutionStrategy.jl[6]
- GCMAE.jl[7]

The former, Optim.jl, does not require any additional packages. An additional third package BlackBoxOptim.jl[8] has been used exclusively for BlackBoxOptimization (BBO). Finally, one exclusive package has been used, that is also external to Optim.jl and Optimization.jl : NaturalES.jl[9].

The list of optimizers is summarized in the following Table 4.1. The same named algorithm could be used from different packages, such as CMAES[10]. Thus to avoid confusion, by default the algorithms will be instead referred by their acronym (when mentioned in the tables) unless not explicitly written. The table 4.2 and table 4.3 are examples for the specific case $(n, r, k) = (8, 2, 8)$ as defined in the beginning of this chapter 4. Note that these tables are not representative and only serve as an example for one specific (n, r, k) case.

Optimizer	**Acronym**	**Framework**	**Original Package**
Black Box Optimization	BBO	BlackBoxOptim.jl	
LBFGS	LBFGS	Optim.jl	
Nelder Mead		Optim.jl	
GradientDescent	GD	Optim.jl	
ConjugateGradient	CG	Optim.jl	
Particle Swarm	PS	Optim.jl	
Newton		Optim.jl	
IPNewton		Optim.jl	
SAMIN		Optim.jl	
Evolutionary Centers Algorithm	ECA	Optimization.jl	Metaheuristics.jl
Simulated Annealing	SA	Optimization.jl	Metaheuristics.jl
NOMADOpt		Optimization.jl	NOMAD.jl
Whale Optimization Algorithm	WOA	Optimization.jl	Metaheuristics.jl
Gravitational Search Algorithm	CGSA	Optimization.jl	Metaheuristics.jl
Artificial Bee Colony	ABC	Optimization.jl	Metaheuristics.jl
Particle Swarm Optimization	PSO	Optimization.jl	Metaheuristics.jl
Differential Evolution	DE	Optimization.jl	Evolutionary.jl
Genetic Algorithm	GA	Optimization.jl	Evolutionary.jl
Evolutionary Strategy Algorithm	ES	Optimization.jl	Evolutionary.jl
CMAES	CMAES	Optimization.jl	Evolutionary.jl
CMAEvolutionStrategyOpt	CMAE	Optimization.jl	CMAEvolutionStrategy.jl
GCMAES	GCMAES	Optimization.jl	GCMAES.jl
NaturalES			NaturalES.jl

Table 4.1: Listed optimizers used for testing

To keep numerical search within limits of realizable computations, we will limit ourselves to a constant $r = 2$ (except explicit mentions) corresponding to the matrix M of rank 2, that is, limited to bunching in $r = 2$ output modes. More specifically, in order to save up time. The testings were performed on the smallest dimension $n = 8$, which was the only known counterexample, proven by Drury, to the local maximum Conjecture M2 — that is $(n, r, k) = (8, 2, 8)$.

[6]https://github.com/jbrea/CMAEvolutionStrategy.jl
[7]https://github.com/AStupidBear/GCMAES.jl
[8]https://github.com/robertfeldt/BlackBoxOptim.jl
[9]https://github.com/francescoalemanno/NaturalES.jl
[10]CMAES stands for Covariance Matrix Adaptation Evolution Strategy

Optimizer	Time [s]	Ratio
Nelder Mead	0.261	1.019 31
LBFGS	2.759	1.019 56
PS	6.179	1.019 56
CG	6.281	1.019 56
GD	9.928	1.019 56
IPNewton	23.148	1.019 56
BBO	26.737	1.019 56
GD	31.673	1.019 54
SAMIN	44.854	1.019 56
Newton	454.187	1.019 56

Table 4.2: Example of $(n, r, k) = (8, 2, 8)$ with tested optimizers from Optim.jl with Ratio $\lambda_1(f^{(k)})(\text{perm}(H))^{-1}$. The best time-wise local optimizer is Nelder Mead, if one wants to have a more refined ratio, LBFGS emerges at the top, although taking significantly more time. The first two global optimizers are PS (particle swarm) and BBO (black box optimization) however the former has a lower success rate of finding counterexamples (not explicitly seen in this table).

Optimizer	Time [s]	Ratio
SA	77.143	1.019 45
BBO	78.239	1.019 56
ECA	78.322	1.019 56
PS	120.223	1.019 48
DE	76.483	1.007 72
ABC	150.226	1.002 76
ES	0.25	1.0
GA	0.283	1.0
CMAES	2.94	1.0
CMAE	4.918	1.0
NOMADOpt	8.023	1.0
GCMAESOpt	10.729	1.0
NaturalES	20.294	1.0
PSO	77.27	1.0
CGSA	120.808	1.0
WOA	120.906	1.0

Table 4.3: Example for $(n, r, k) = (8, 2, 8)$ of tested optimizers from Optimization.jl with Ratio $\lambda_1(f^{(k)})(\text{perm}(H))^{-1}$ with time limit set to 120s. The emerging global optimizers are SA (simulated annealing) and BBO (black box optimization), however the former — just like PS (particle swarm) — presents a lower success rate of finding counterexamples (not seen explicitly in this table).

Local Optimizers

It has been noted that, gradient based optimizers (i.e. LBFGS, GD) but not exclusive (IPNewton uses second derivative as well) find a counterexample $\boldsymbol{x}$ given an initial guess $\boldsymbol{x}_0$ only if $\boldsymbol{x}_0$ is relatively close to $\boldsymbol{x}$. That is, each entry i of the initial guess $(\boldsymbol{x}_0)_i$ is updated to $\boldsymbol{x}_i$ within a predefined maximum function calls $fcall$.

The Table 4.2 and Table 4.3 display what is typically found when running $(8, 2, 8)$ with random initial guesses (to avoid bias). Note that the result might sometimes differ for some initial guess. For example, LBFGS could be stuck for a specific x_0 unlike Nelder Mead (or vice versa). However,

in most randomly selected x_0, the first local optimizer that came out on top was Nelder Mead time wise, and LBFGS, if one values accuracy in highest achieved optimization instead (cf. Table 4.2).

Although not being derivative based, Nelder Mead, also known as Simplex Search Algorithm, is best known for multidimensional unconstrained optimization. Its unreliance on derivatives greatly speeds up the process of optimization which could be put to great use. Although it was considerably faster than LBFGS, it did not manage to optimize the function any better.

This comparison between LBFGS and Nelder Mead can be summarized from the results in Table 4.4, and was done as follows. First BBO solves for random initial guess $\boldsymbol{x}_0$ and returns $\boldsymbol{x}'$:

$$\boldsymbol{x}' = \arg\max_x \mathrm{fct}(\boldsymbol{x}) \quad (\text{initial guess } \boldsymbol{x}_0) \tag{4.11}$$

We repeat the process until we have 20 different $\boldsymbol{x}'$ satisfying

$$\mathrm{fct}(\boldsymbol{x}') = \lambda_1(\boldsymbol{x}')/\operatorname{perm}(H(\boldsymbol{x}')) > 1 \tag{4.12}$$

The first $\boldsymbol{x}'$ is given to both LBFGS and Nelder Mead. They return an optimized $\boldsymbol{x}''$:

$$\boldsymbol{x}'' = \arg\max_x \mathrm{fct}(\boldsymbol{x}) \quad (\text{initial guess } \boldsymbol{x}') \tag{4.13}$$

The time, ratio $\mathrm{fct}(\boldsymbol{x}'')$, and number of function calls $fcall$ are measured. This is repeated for all $\boldsymbol{x}'$ (cf. Table 4.4).

$\mu \pm \sigma$	**LBFGS**	**Nelder Mead**
time[s]	2.113 ± 0.464	0.338 ± 0.025
ratio	1.01956 ± 0	1.01924 ± 0.00048
$fcall$	166.21 ± 36.05	1745.74 ± 132.54

Table 4.4: Comparison of LBFGS vs Nelder Mead in terms of the average μ and standard deviation σ for time of the maximization, their achieved ratio $\lambda_1/\operatorname{perm}(H)$, and the required total number of function calls $fcall$ for 20 different samples (given identical to both LBFGS and Nelder Mead) initial guesses that already satisfied the ratio surpassing 1

However, local optimizers have a lower success rate starting with a random initial guess $\boldsymbol{x}_0$ than global optimizers. They are especially useful if one prioritizes the optimization of the ratio $\lambda_1(f^{(k)})(\operatorname{perm}(H))^{-1}$, given $\boldsymbol{x_0}$ for which the ratio is already surpassing 1 (thus is guaranteed to be a counterexample). For such cases, another variation has been also tested where, first, the faster solver has been used — Nelder Mead — then to increase the ratio even further, the slower but more precise optimizer has been used — LBFGS. However, this did not significantly speed up the time of the process, compared to solely using LBFGS.

Global Optimizers

To find such correct initial guess — which satisfies a ratio surpassing 1 without necessarily being further optimized — a higher success rate is achieved with global optimizers.

Although the top global optimizer was PS (particle swarm) in Table 4.2, its success rate on average was lower than the second top optimizer in that table — BBO. However, SA (simulated annealing) was found to have an even lower success rate than PS, despite seeming to be faster than BBO in Table 4.3.

The optimizer that managed to find counterexamples globally with the highest success rate was BBO (black box optimization), which requires no finetunning of hyperparameters. Particle Swarm required generally (with some $\boldsymbol{x}_0$ exceptions) a bigger time whose result greatly varied with its hyperparameter (number of particles generated) but still performed on average more poorly than BBO. The only global optimizer that managed to have a similar success rate as BBO was ECA. However, it also did require parameters tuning (contrary to BBO).

4.2.5 Chosen Optimizers

Since our goal was to find counterexamples to conjectures, attempting to maximize the ratio beyond 1 is optional (unless that is explicitly the given objective): a counterexample is immediately found for a given $\boldsymbol{x}$ which results with the ratio $\lambda_1/\operatorname{perm}(H) > 1$.

For that purpose BBO was chosen: the global optimizer, with the highest success rate found for $(8, 2, 8)$. It will be the mainly used optimizer, unless explicitly stated otherwise. The stop condition will simply consist of either achieving $fcall_{\max}$ or a ratio $\lambda_1(f^{(k)})(\operatorname{perm}(H))^{-1} > 1$.

If the objective is to further attempt to find a higher ratio: we will choose a specific counterexample $\boldsymbol{x}'$ resulting from BBO, and use it with LBFGS as initial guess to converge further possibly leading to a higher ratio. It turns out that no matter which initial guess $\boldsymbol{x}'$ is chosen for LBFGS (such that $\lambda_1/\operatorname{perm}(H) > 1$), the maximum attained ratio (at least by LBFGS) always resulted in the same one. This will be convenient for later, when comparing different (n, r, k) cases, since each one of them will only have one found maximal ratio attained by LBFGS, no matter its initial guess.

4.2.6 Incrementations and Trace Method

As the search time increases exponentially with n, in an attempt to reduce it, it was suggested to use the resulting $\boldsymbol{x}'$ already solved by the optimizer for $(n, r, k) = (8, 2, 8)$ as *seed*. From this *seed* one would try to make use of it by further propagating to other dimensions such as $(n + 1, r, k)$ or $(n, r, k - 1)$ or even $(n, r + 1, k)$.

An example of the n incrementing is as follows: for $(n, r, k) = (8, 2, 8)$ let the optimizer solve the problem starting from initial guess $\boldsymbol{x}_0$, and as its best fit, it returns $\boldsymbol{x}'$:

$$\boldsymbol{x}' = \arg\max_{\boldsymbol{x}} \operatorname{fct}(\boldsymbol{x}) \quad (\text{initial guess } \boldsymbol{x}_0) \tag{4.14}$$

This $\boldsymbol{x}'$ counts as seed, that is, as initial guess to solve for $(n + 1, r, k) = (9, 2, 8)$ and returns $\boldsymbol{x}''$

$$\boldsymbol{x}'' = \arg\max_{\boldsymbol{x}} \operatorname{fct}(\boldsymbol{x}) \quad (\text{initial guess } \boldsymbol{x}') \tag{4.15}$$

To be more precise, the incrementation methods work by either: incrementing $n \to n + 1$ or decrementing $k \to k - 1$

$$f^{(k)} \to f^{(k-1)} \tag{4.16}$$

In the former case $n \to n + 1$, the matrix M gets an additional column $\epsilon \to 0$

$$M = \begin{pmatrix} M_{11} & \dots & M_{1n} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ M_{r1} & \dots & M_{rn} & 0 \end{pmatrix} \tag{4.17}$$

The same could be done with rows if this time we increment $r \to r + 1$.

A final attempt consisted of changing into a potentially simpler function to optimize, where instead of maximizing only the biggest eigenvalue $\lambda_1(f^{(k)})$, the two biggest eigenvalues are maximized but normalized by the sum of eigenvalues. The latter being exactly equal to the trace of the matrix

$$\mathrm{tr}(f^{(k)}) = \sum_i \lambda_i(f^{(k)}) \tag{4.18}$$

thus the name trace method $\mathrm{fct}_{\mathrm{tr}}$:

$$\mathrm{fct}_{\mathrm{tr}} = \frac{\lambda_1(f^{(k)}) + \lambda_2(f^{(k)})}{\mathrm{tr}(f^{(k)})} \tag{4.19}$$

The point would be to first optimize using this new function, starting with initial guess $\boldsymbol{x}_0$

$$\boldsymbol{x}' = \arg\max_x \mathrm{fct}_{\mathrm{tr}}(\boldsymbol{x}) \quad (\text{initial guess } \boldsymbol{x}_0) \tag{4.20}$$

Then to follow this up with BBO:

$$\boldsymbol{x}'' = \arg\max_x \mathrm{fct}(\boldsymbol{x}) \quad (\text{initial guess } \boldsymbol{x}') \tag{4.21}$$

4.2.7 Sample Evaluation

For each (n, r, k) search, a sample size of 60 trials is planned, with a predefined maximum number of function calls $fcall_{\max} = 3 \cdot 10^4$ which both of course should be preferably maximized within the possible limits to increase the chance of discovery of new counterexamples. However, one such search for big dimensions such as $(n, r, k) = (22, 2, 8)$ takes about 2.5 days. A subsequent search with LBFGS for just only 1 among 60 optimized $\boldsymbol{x}'$ (outputs of BBO), would also take in the order of days. Note that the time taken could greatly vary depending on the technical equipment.

Additionally, to monitor each (n, r, k) cases, a distribution of the total $fcall$ of each sample is plotted, that is, it either converges to a stopping condition which is to terminate once $\mathrm{fct} = \lambda_1 / \mathrm{perm}(H) > 1$, in that case $fcall < fcall_{max}$, or the algorithm simply exits after achieving the maximum number of function calls threshold $fcall = fcall_{\max}$ which usually ends up with a ratio very close but just below 1 that is $\mathrm{fct} < 1$. Note that this only serves the purpose to monitor how well the current algorithm BBO is doing for each (n, r, k). The results of the $fcall$ distributions are not perfectably replicable since:

- the seed initial guess is chosen randomly
- the algorithm is a meta heuristic optimization thus partially probabilistic
- the results also vary depending on the total number of taken samples (in this case 60)
- the limit in the search of each sample that is, the bigger the threshold $fcall_{max} = 3 \cdot 10^4$, is chosen arbitrarily

The $fcall$ distribution also allows to verify if the $fcall_{\max}$ limit is not set too low, in which case, all the function calls of each sample would end up thresholding at $fcall = fcall_{\max}$ However, as shall be discusedd in section **Relaxation and Restriction**, the fact of seeing an accumulation in the distribution near the $fcall_{\max}$ limit could be due to other reasons, just as it is the case for $(17, 2, 5)$ depicted in Figure 4.1.

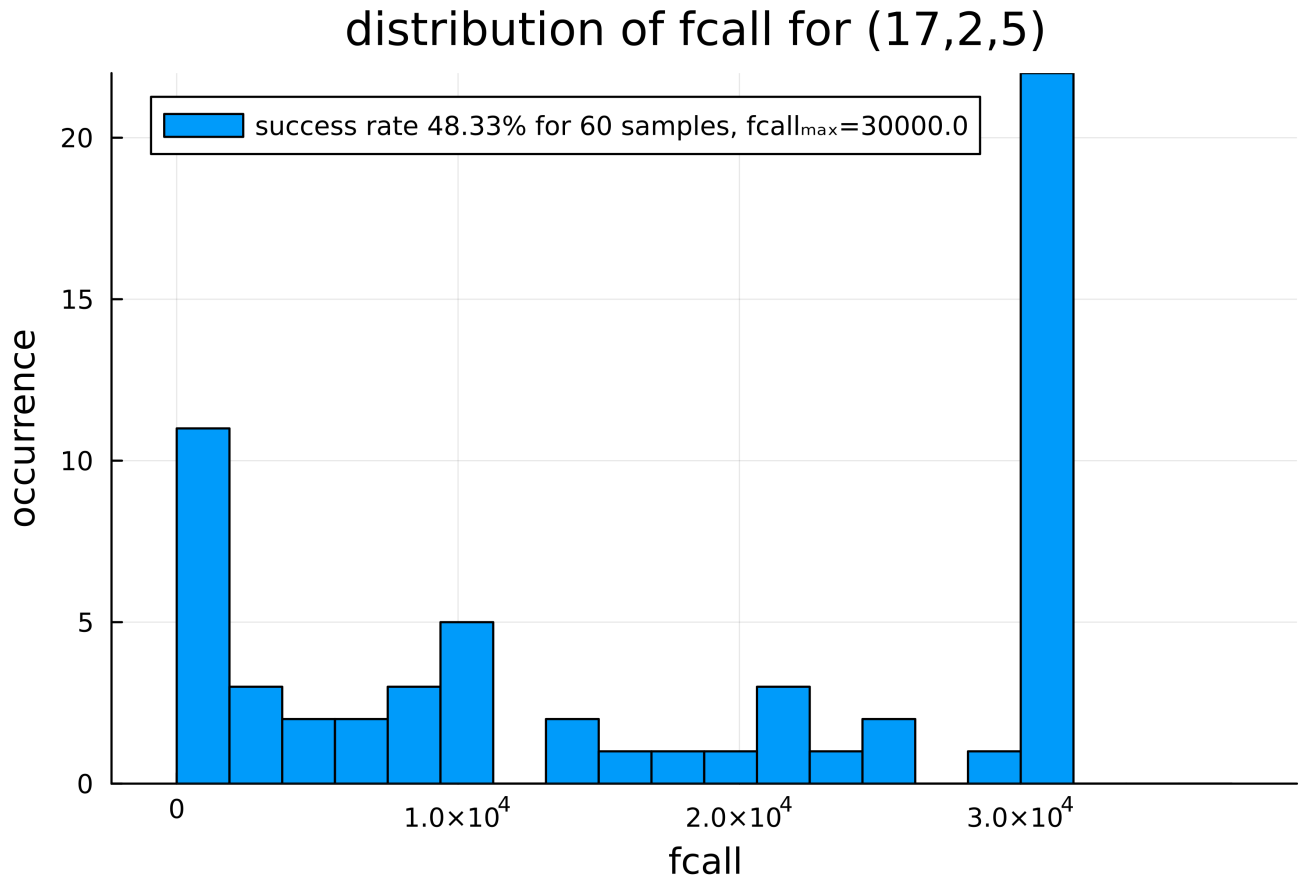


Figure 4.1: Distribution of $fcall$ for 60 samples in the case $(n, r, k) = (17, 2, 5)$ may not directly result from a too low set limit $fcall_{\max}$. The success hit rate of reaching fct $= \lambda_1 / \operatorname{perm}(H) > 1$ is 48.33%

Coming back to the counterexamples for intermediate Conjectures M5. To avoid spending time on futile calculations. A safe approach would consist of finding counterexamples to the set of intermediate Conjectures M5 by progressively approaching $k = 2$ starting from $k = n - 1 = 7$. The idea is to look for k first with the lowest possible n, which is $n = 8$ bounded by the previously found counterexample $(8, 2, 8)$. Only if this fails, then n would be incremented by 1, which of course would be best to avoid since this considerably increases the computational cost (mainly due to permanent).

4.2.8 Permanent Computation Method

Two methods when calculating the permanent were used based on tests with varying matrix dimension n: the *incomplete rank* algorithm [11] — introduced in Ref [**barvinok1994**] — outpaced Ryser's algorithm [24] from $n \geq 17$ (cf. Figure 4.2). This is to be expected since the former, is specifically adapted to work with matrices of low rank, which was our case $r = 2$. Thus the method of calculating the permanent was adapted accordingly for each (n, r, k) cases.

[11] https://github.com/benoitseron/Permanents.jl/tree/incomplete_rank

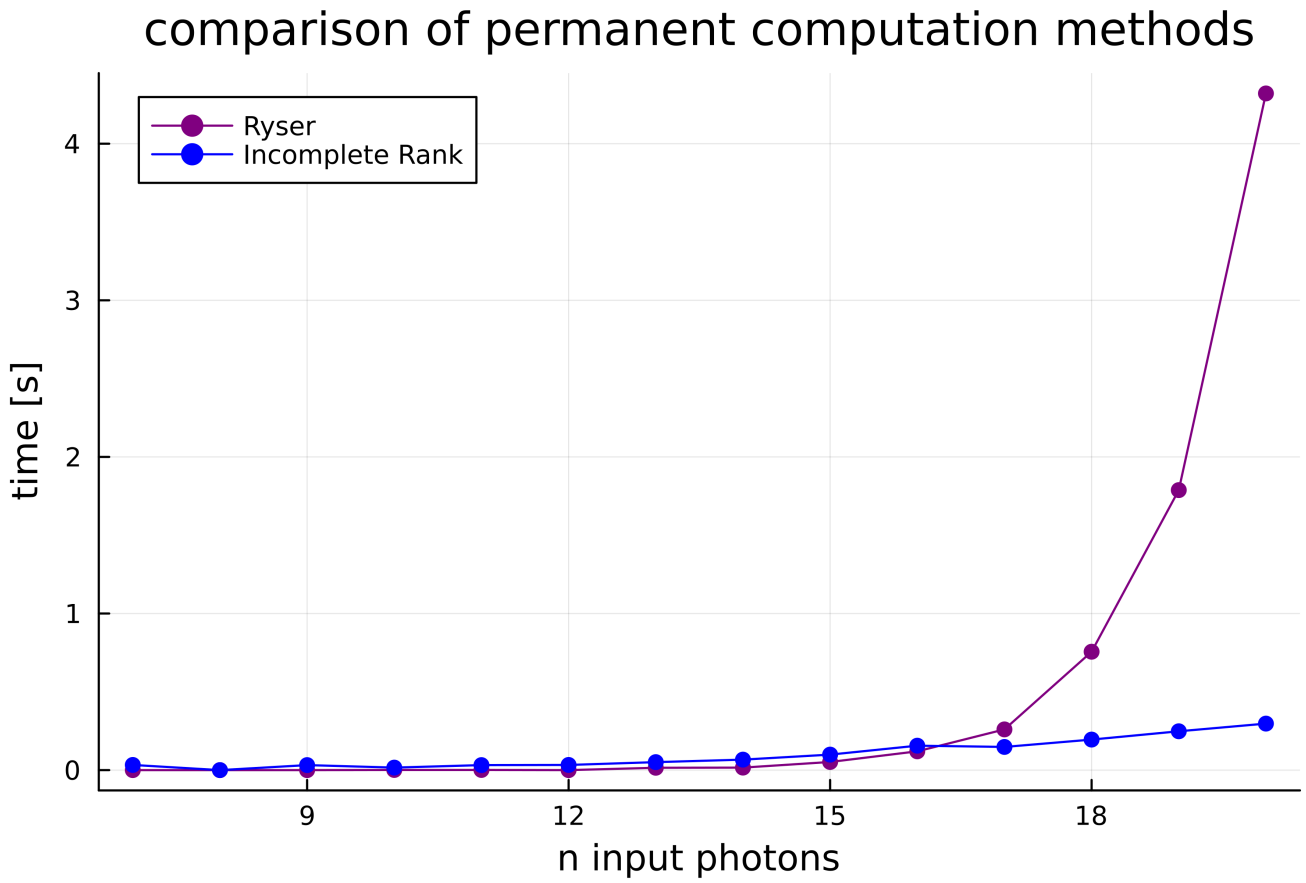


Figure 4.2: Comparison of two methods to compute the permanent: Ryser's versus the *incomplete rank* method, by measuring the time [s] taken to calculate the ratio $\lambda_1(F)/\operatorname{perm}(H)$ of varying matrix dimension n

4.3 Numerical search

4.3.1 Counterexamples

Local maximum conjecture

After having chosen an optimal optimizer by testing out on the solved case $k = 8 = n$, it was attempted to find a lower dimensional case, that is with lower n, that would satisfy $n = k$ Conjecture M2. Unfortunately, no such counterexamples have been found for $4 \leq n < 8$ for 60 samples.

Intermediate conjectures

Coincidentally, for the same $n = 8$, a counterexample has been found for $k = 7$ and $k = 6$, although the latter with very low success rate 3.33% for 60 samples cf. Table 4.5. As for $k = 5$, the lowest dimensional counterexample found was $n = 10$ with an even lower success rate 1.67%. No further counterexamples have been found for $k \leq 4$, where the search range for n was $n \in [16, 22]$.

$n\backslash k$	**8**	**7**	**6**	**5**	**4**
8	31.67	93.33	3.33	0	
9	100	100	78.33	0	
10	100	100	100	1.67	
11	100	100	100	11.67	
12	100	100	100	18.33	
13	100	100	100	31.66	
14	100	100	100	45	
15	100	100	100	50	
16	100	100	100	58.33	0
17				48.33	0
18				51.67	0
19				53.33	0
20				45	0
21				50	0
22	100			46.67	0

Table 4.5: Success rates (in percent, without %) of reaching fct $= \lambda_1/\operatorname{perm}(H) > 1$ for 60 samples for each (n, k) pair with $r = 2$ specifically (using BBO $fcall_{\max} = 3{\cdot}10^4$). It is clear that restricting to lower k and lower n has a restrictive effect numerically (for the optimizer to find a violation). Interestingly, it stagnates for the column $k = 5$ perturbed photons despite increasing to higher dimensions up to $n = 22$ photons in total. It could be that, despite the number of counterexamples increasing (with increasing n or k), and thus should be leading to higher success rates, the search space (i.e. elements of matrix M to maximize the ratio) could be increasing even "faster" (due to increased matrix M dimension of $n \times r$) which would explain the stagnation around 50% success rate for $k = 5 : 14 \leq n \leq 22$.

As mentioned previously in section 4.2.4 (subsection Local optimizers), interestingly once BBO has found a counterexample, that is, returning $\boldsymbol{x}'$ such that the ratio fct $= \lambda_1/\operatorname{perm}(H)$ is bigger than 1, then for any such tested counterexamples, the ratio further maximized by LBFGS (by taking $\boldsymbol{x}'$ as initial guess) always ended up the same. This is why all the (n, r, k) cases could be summarized in the following Table 4.6 with ratio expressed in per mille[12] as: $10^{-3}(\lambda_1(f^{(k)})/\operatorname{perm}(H) - 1)$. Note that, LBFGS was not used to maximize ratio in column $k = 4$ of Table 4.6 since BBO did not find counterexamples for $k = 4$ (cf. Table 4.5). The maximal ratio achieved by LBFGS is also illustrated in Figure 4.3 where each row of the Table 4.6 — corresponding to a constant dimension $n = c^{te}$ — is plotted as a separate line.

For the specific $k = 5$ (see column where $k = 5$ in Table 4.6), there was a search of 60 samples for $n \in [8, 22]$ using BBO. This search was continued with LBFGS to maximize the ratio only for $n \leq 16$ and for $n = 22$, the skip from $n = 16$ to $n = 22$ was indicated by vertical dots $\vdots$ in the Table 4.6. The highest achieved ratio pursued with LBFGS (after BBO) for $(n, r, k) = (22, 2, 5)$ is:

$$10^{-3}\left(\frac{\lambda_1(f^{(5)})}{\operatorname{perm}(H)} - 1\right) = 106.261 \quad \text{for } (n, r, k) = (22, 2, 5) \tag{4.22}$$

[12]Since the Table 4.6 only displays maximum ratios, they are always above 1. Thus it can be omitted by substracting 1 and displaying in per mille to save more space for decimals.

$n\backslash k$	8	7	6	5	4
8	19.561	17.983	9.660		
9	66.601	53.894	36.564		
10	104.119	83.158	58.056	18.969	
11	134.581	107.847	76.052	34.257	
12	159.914	129.023	91.931	47.218	
13	181.354	147.039	107.119	57.895	
14	199.806	162.414	119.897	66.803	
15	215.921	175.775	131.069	74.353	
16	228.181	187.582	140.918	80.835	
			⋮		
22	290.426			106.261	

Table 4.6: Maximized ratio $(\lambda_1(f^{(k)})/\operatorname{perm}(H) - 1)‰$ (in permil without ‰) by LBFGS, with initial guess optimized by BBO for each (n, k) pair with $r = 2$ specifically. Note that LBFGS was only used for (n, k) pairs that had at least one found counterexample via BBO which was not the case for $k = 5 : n = 8, 9$ and $k = 4 : 16 \leq n \leq 22$ thus left as blank. The only known counterexample before this thesis was $(n, r, k) = (8, 2, 8)$ found by Drury [18] corresponding to ratio 19.561 (cf. eq 2.30). Every other (n, k) pair is newly found in this thesis. Thus, after this work, it is still unkown whether $k < 5$ (less than 5 perturbed photons) could lead to a violation of intermediate conjectures M5, or Pate's conjecture M4.

It is also interesting to see what is the highest possible ratio to achieve overall. Of course the highest corresponds to the most relaxed case which, in the range $n \leq 22$, is $(n, r, k) = (22, 2, 8)$

$$10^{-3}\left(\frac{\lambda_1(f^{(5)})}{\operatorname{perm}(H)} - 1\right) = 290.426 \quad \text{for } (n, r, k) = (22, 2, 8) \tag{4.23}$$

The lowest n dimensional counterexamples for each k — considering having a low ratio close to 1 — could turn out to be a numerical error due to the accumulation of the floating point errors. To check if this is not the case, it is relevant, beyond just checking the ratio $\text{fct} = \lambda_1/\operatorname{perm}(H)$, to also check individually either λ_1 or $\operatorname{perm}(H)$. This is represented in the Table 4.7 together with their corresponding matrices M found in Appendix B.

In Table 4.7, the row $(20, 20)$ serves as a more relaxed numerically counterexample to maximize the ratio which yielded 487.51. The last row $(17, 17)$ is a counterexample for real intermediate conjectures M6 with lowest $n = k$ dimension found discussed in more details later in section "**Real intermediate conjectures**", all the other (n, k) are counterexamples to complex intermediate conjectures M5 (less tight than their real counterpart).

Let us compare the eigenvalues in Table 4.7. Indeed we have $\lambda_1(f^{(k)}) > \operatorname{perm}(H)$. Notice that $\operatorname{perm}(H) = \lambda_2(F)$. The eigenvalue interlacing theorem (for leading principal submatrix $f^{(k)}$ of Hermitian $F = f^{(n)}$) — a corollary of Courant-Fischer theorem — states that

$$\lambda_i(f^{(n)}) \geq \lambda_i(f^{(k)}) \geq \lambda_{i+(n-k)}(f^{(n)}) \quad \forall i \in [m] \tag{4.24}$$

Which indeed is verified here for all (n, k) pairs we have:

$$\lambda_1(f^{(n)}) = \lambda_1(F) \geq \lambda_1(f^{(k)}) \geq \lambda_{1+(n-k)}(f^{(n)}) \tag{4.25}$$

We even have a stronger condition met here:

$$\lambda_1(f^{(k)}) \geq \lambda_2(F) = \operatorname{perm}(H) \geq \lambda_{1+(n-k)}(f^{(k)}) \tag{4.26}$$

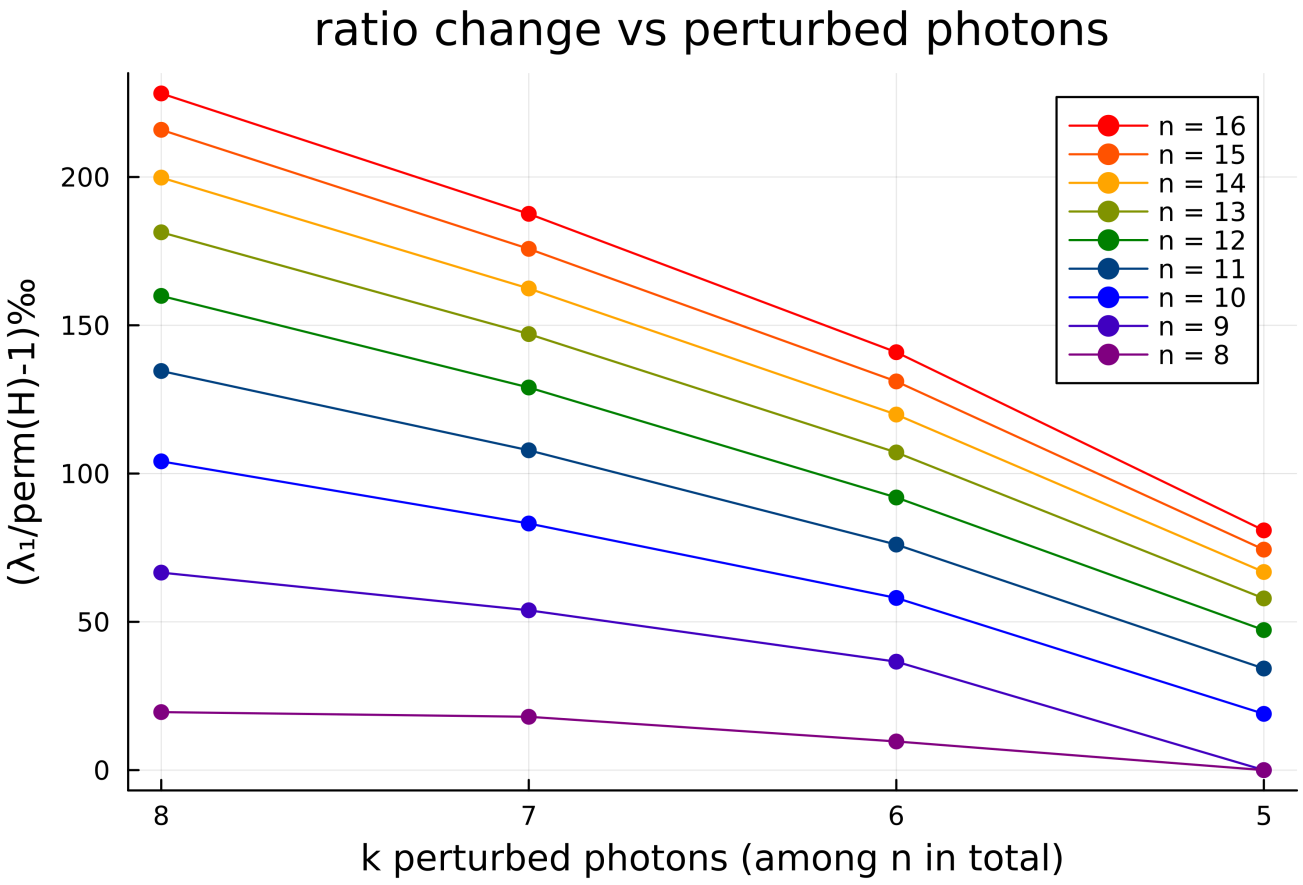


Figure 4.3: Graphical representation of the ratio $\left(\frac{\lambda_1}{\text{perm}(H)} - 1\right)$ ‰ (in per mille) versus k perturbed photons (among n input photons). The dots are real data from Table 4.6 but the continuous line has been added only for visibility reasons: the maximum ratio for each (n, r, k) has not been continuously computed.

(n,k)	$\boldsymbol{v}$ type	ratio [‰]	$\lambda_1(f^{(k)})$	$\lambda_2(f^{(k)})$	$\lambda_1(F)$	$\lambda_2(F)$	$\text{perm}(H)$
(8,8)	$\mathbb{C}^8$	19.56	276.08	270.79	276.08	270.79	270.79
(8,7)	$\mathbb{C}^7$	17.98	277.09	240.32	277.25	272.20	272.20
(8,6)	$\mathbb{C}^6$	9.66	294.28	224.12	294.79	291.46	291.46
(10,5)	$\mathbb{C}^5$	18.97	9188.69	5428.30	9438.16	9017.63	9017.63
(20,20)	$\mathbb{C}^{20}$	487.51	8.80×10^{11}	5.92×10^{11}	8.80×10^{11}	5.92×10^{11}	5.92×10^{11}
(17,17)	$\mathbb{R}^{17}$	37.36	1.34×10^{9}	1.29×10^{9}	1.34×10^{9}	1.29×10^{9}	1.29×10^{9}

Table 4.7: Maximized ratio $= (\lambda_1(f^{(k)})/\,\text{perm}(H) - 1)$‰ (with $r = 2$ specifically) by LBFGS (exceptionally $(17, 17)$ only by Nelder Mead) with initial guess optimized by BBO. The first 5 rows are counterexamples for each k with lowest dimension n, paired with the first two largest eigenvalues λ_1, λ_2 of matrix $F = f^{(n)}$ and its submatrix $f^{(k)}$. A ratio consisting of very similar but small numerator and denominator could indicate a false positive, just like in the case of big values (possibly due to accumulated floating-point errors). This could be additionally verified with the violation ratio (discussed in future sections). Note that the eigenvalue interlacing theorem is verified. Notice that $\lambda_2(F) = \text{perm}(H)$.

Time constraint

The used algorithm — BBO — being nondeterministic, the results are not definitive but only reflect the amount of search that has been carried out (60 samples for each (n, k) pair with

$r = 2$) for which the hit rate of finding counterexamples to $k \leq 4$ has been 0%. One could potentially find such counterexamples (for $k \leq 4$) by: increasing the sample size past 60, increasing the maximum number of calls beyond the current $fcall_{\max} = 3 \cdot 10^4$ for BBO (unique for each optimizer), increasing the domain search beyond $\boldsymbol{x} \in [-50, 50]$, increasing n, increasing r. The main constraint was the time complexity of permanents, especially as n increased. For example, for $n = 22$, a 60 sample search takes around 2.5 days while for $n = 8$ it only takes several seconds. To further maximize the ratio using LBFGS for only one counterexample (among 60), it will take another 2.5 days for dimension $n = 22$

Methods Comparison

When it comes to comparing the incremental methods with direct methods, see Table 4.8: the incrementation of n, and decrementation of k did see slight not significant benefit compared the direct method when it comes to increasing the success rate of finding counterexamples. Note that this is not certain and could be due simply to randomness. A more indepth statement whether n incrementation or k decrementation does present benefits relative to the direct method, could require a much bigger sample size than 60, and the behavior might vary with n or/and k.

For a search of 30 samples, trying to find counterexamples by direct method for $(8, 3, 7)$ leads to 0% success rate. This was drastically improved by the r incrementation method to 93.33% success rate, this could be seen as a *transfer* of the solutions from the seed $(8, 2, 7)$ having the same success rate. However, using this for a case where there are no initially found counterexamples (for a r before being incremented), or for cases with little to no counterexamples, such as $(8, 2, 6)$, did not increase the success rate. In short, it is easier to numerically find anomalous bunching into two modes $r = 2$ than three.

Finally, when it comes to choosing another function to optimize, that is the trace method: this one delivered no result hitting a 0% success rate for the simplest case $(8, 2, 8)$. This could indicate that counterexamples are not nearby the point in space for which both first two eigenvalues are high relative to the rest, but rather only the first one.

The results are displayed in Table 4.8 by comparing the success rate of achieving a higher ratio $\lambda_1(F)/\operatorname{perm}(H)$ than 1 (using BBO and 60 samples). Note that the last column "seed" represents the success ratio of the method's corresponding seed, e.g. the fourth line in Table 4.8 reads as k decrementation for $(8, 2, 6)$, the seed thus being $(8, 2, 6 - 1)$.

Relaxation and Restriction

With help of the $fcall$ distribution for each (n, r, k) case, one can see the success rate, but also compute their average and standard deviation. Obviously, the standard deviation would be biased towards 0 when, say the $fcall_{max}$ is too low, resulting in not enough given $fcall$ (thus time) for the search to converge, but this also results in a guaranteed 0% success rate.

It is worth noticing that the success rate decreases for constant n as k decreases as illustrated in Table 4.5.

This is natural since decreasing k restricts to a more particular case of lower number of perturbed photons, and thus has a restricting effect. Similarly, for constant k, as n increases, the success rate increases as well. It has a relaxing effect making the counterexamples being easier to find hence when the relaxation effect on the interferometer configuration outweights the restricting effects of the increased k, this is when another counterexample is found.

Interestingly it can be noticed for $k = 5$ that the success rate has trouble surpassing 60%. This could be again explained similarly as with the argument of increased r: even though it has

(n,r,k)	**direct**	**n**	**r**	**k**	**seed**
$(8,3,6)$	0		1.67		1.67
$(8,3,7)$	0		93.33		93.33
$(8,3,8)$	0		33.33		33.33
$(8,2,6)$	1.67			0	93.33
$(8,2,7)$	86.67			93.33	31.67
$(10,2,5)$	1.67			5	100
$(13,2,6)$	100			100	100
$(15,2,6)$	98.33			98.33	100
$(14,2,7)$	100	100			100
$(16,2,7)$	98.33	100			100
$(20,2,5)$	40	45			53.33

Table 4.8: Comparison of methods: direct; n, r incrementation, and k decrementation (resp. "n", "r", "k") by the success rate of ratio $\lambda_1(f^{(k)})/\operatorname{perm}(H)$ surpassing 1, for 60 samples $fcall_{\max} = 3 \cdot 10^4$ using BBO. The column "seed" stands for the success ratio of the seed (n,r,k) case corresponding to their respective incrementation method

analytically a relaxing effect, the number of variables has been increased considerably expanding the search domain to beyond the capabilities of the used optimizer making it in practice more restrictive numerically due to the total number of possible choices.

Real intermediate conjectures

Finally adressing the intermediate conjectures but their real counterpart that is, Conjectures M6, maximizing $\lambda_1/\operatorname{perm}(H)$ this time while restricting $\boldsymbol{v}$ to only real vectors. This is equivalent to taking the real part of the matrix F before calculating the eigenvalue λ_1. For this, no counterexamples have been found for the easiest case that is $(8,2,8)$. However, by taking a much more relaxed case $n = k$, a counterexample has been found for $(17,2,17)$ with a ratio $(\lambda_1/\operatorname{perm}(H)-1)‰ = 37.36‰$ found by BBO and only optimized by Nelder Mead (non-optimized by LBFGS).

No thorough search was conducted for real counterexamples. A search of 60 samples was executed for only $n = k \leq 12$ A separate 1-2 samples per (n,r,k) case was initiated starting with $n = k = 20$ until $n = k = 17$, which were fruitful, a counterexample has been found. No counterexamples were found for the subsequent search $n = k = 16$ among 15 samples. Although a quick search found a real perturbation counterexample for $n = k = 17$ (the explicit M matrix can be found in Appendix B), a more detailed search could prove to be useful in hopes of finding an even lower dimensional counterexample. The search was not continued due to the limited time. As mentioned before, due to the low ratio (non optimized by LBFGS), it could be due to accumulated floating point errors. In the next section, a new *violation ratio* will be introduced that will additionally help us verify such sensitive counterexamples.

4.3.2 Violation Ratio

It is interesting to point out that even for the global maximum conjecture P1, there were previously no known counterexamples of anomalous bunching where only k photons are partially perturbed and the remaining $(n-k)$ photons are perfectly indistinguishable.

One can study such violations with respect to the magnitude of such perturbations. Those coun-

terexamples where the function to be maximized was

$$\text{fct} = \frac{\lambda_1(f^{(k)})}{\text{perm}(H)} \tag{4.27}$$

are only valid to approximate the curvature of the bunching probability near indistinguishability. The found counterexample M (which maximizes fct), not only provides the direct description of the interferometer but also the exact perturbation vector $\boldsymbol{v}_{\max}$, corresponding to the eigenvector associated to the highest eigenvalue λ_1 as explained in Section 4.1

$$f^{(k)}\boldsymbol{v}_{\max} = \lambda_1 \boldsymbol{v}_{\max} \tag{4.28}$$

whose i-th component $(\boldsymbol{v}_{\max})_i$ indicates by how much each photon i should be perturbed. As a reminder, Conjecture M2 was developped as a specific case of Conjecture P2 for a two dimensional space spanned by orthogonal vectors $\{|\psi_0\rangle, |\eta_i\rangle\}$ ($|\eta_i\rangle$ being the same for every photon i). We can naturally choose, just as in equation 2.34, the following state for each photon i:

$$|\psi_i\rangle = \frac{1}{\alpha_i}(|H\rangle + \epsilon(\boldsymbol{v}_{\max})_i |V\rangle) \tag{4.29}$$

Since this is known for every photon, one can compute the distinguishability matrix $S_{i,j} = \langle\psi_i|\psi_j\rangle$. Finally we define a new ratio, called the *violation ratio* $R(\epsilon)$, that quantifies by how much the bunching probability overshoots for a specific counterexample found $\text{perm}(H \odot S)$, compared to the one of fully indistinguishable bosons $\text{perm}(H)$.

$$R(\epsilon) = \frac{\text{perm}(H \odot S)}{\text{perm}(H)} \tag{4.30}$$

Note that this ratio $R(\epsilon)$ is not to be confused with the previously used ratio $\lambda_1(f^{(k)})(\text{perm}(H))^{-1}$ which is, in the local approximation, the curvature (second order) of $R(\epsilon)$ around $\epsilon \to 0$ for $\boldsymbol{v} = \boldsymbol{v}_{\max}$. Using the calculations found in equation 2.60 we get:

$$\begin{aligned} P_n(S) = \text{perm}(H \odot S) &= \text{perm}(H) + \epsilon^2[\boldsymbol{v}_{\max}^\dagger F \boldsymbol{v}_{\max} - \text{perm}(H)] + \mathcal{O}(\epsilon^4) && (4.31)\\ &= \text{perm}(H) + \epsilon^2[\lambda_1 - \text{perm}(H)] + \mathcal{O}(\epsilon^4) && (4.32)\\ R(\epsilon) = \frac{\text{perm}(H \odot S)}{\text{perm}(H)} &\overset{\epsilon\to 0}{\longrightarrow} 1 + \epsilon^2\left[\frac{\lambda_1}{\text{perm}(H)} - 1\right] + \mathcal{O}(\epsilon^4) && (4.33) \end{aligned}$$

One can extend the violation ratio $R(\epsilon)$ beyond the local analytical approximation by evaluating it numerically for a given perturbation vector $\boldsymbol{v}$ as a function of ϵ. In the case $(n, r, k) = (8, 2, 8)$, the eigenvctor $\boldsymbol{v}_{\max}$ that maximizes $\lambda_1(\text{perm}(H))^{-1}$ yields the violation ratio $R(\epsilon)$ shown in Figure 4.4a. Its second order approximation near $\epsilon \to 0$ is:

$$R(\epsilon) \overset{\epsilon\to 0}{\longrightarrow} 1 + \epsilon^2\left[\frac{\lambda_1}{\text{perm}(H)} - 1\right] + \mathcal{O}(\epsilon^4) = 1 + 0.0196\epsilon^2 + \mathcal{O}(\epsilon^4) \tag{4.34}$$

The ratio $R(\epsilon)$ reaches its maximum value $R_{\max} = 1.0231$ at some finite $\epsilon_{\max} = 1.2416$, where:

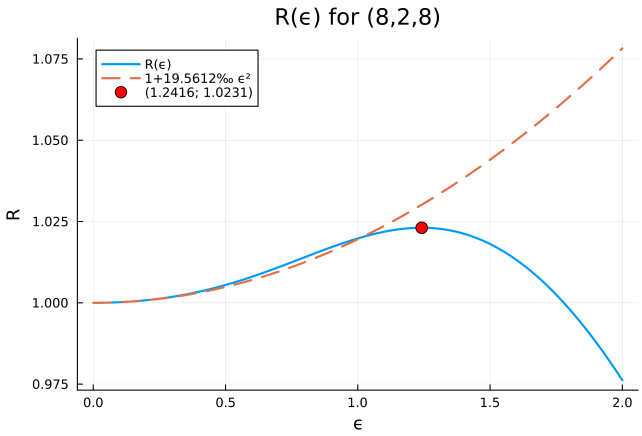
$$R_{\max} \equiv \max_{\epsilon} R(\epsilon) = R(\epsilon_{\max} = 1.2416) = 1.0231 \tag{4.35}$$

Similarly the violation ratio is depicted for the cases $(n, r, k) = (8, 2, 7)$ in Figure 4.4b, $(8, 2, 6)$ in Figure 4.4c, $(12, 2, 5)$ in Figure 4.4d, $(16, 2, 5)$ in Figure 4.4e and $(22, 2, 8)$ in Figure 4.4f.

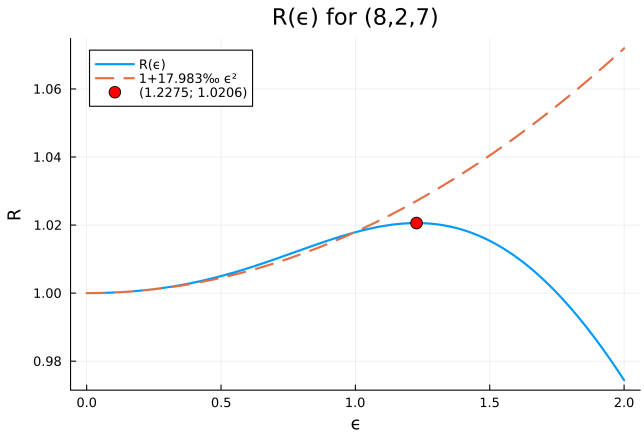
Coming back to our real perturbation counterexample $(17, 2, 17)$, the exact plot in Figure 4.5a confirms that it is not due to accumulated floating point errors: indeed the exact violation ratio

depicted in blue, does surpass 1 and even goes beyond the non maximized curvature of the red line (which is only an approximation for small ϵ).

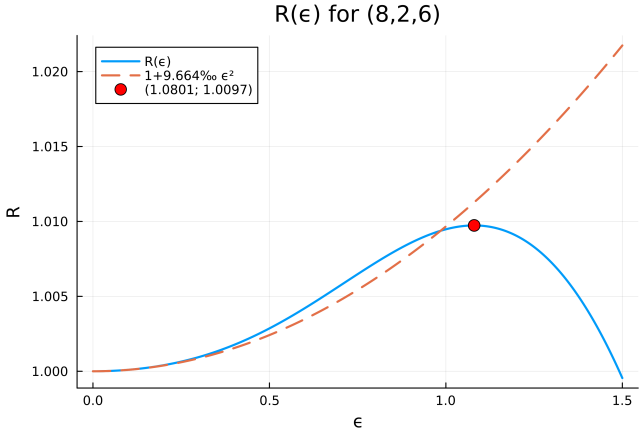
It is also interesting to obtain a highest possible violation ratio, by choosing a more relaxed (n, r, k) case, i.e. easier to numerically optimize, which was chosen here to be $(20, 2, 20)$ for a complex perturbation vector (cf. Figure 4.5b) whose ratio $\lambda_1 / \operatorname{perm}(H)$ was optimized by LBFGS, we get a violation ratio of $R(\epsilon) = 6.0174$.



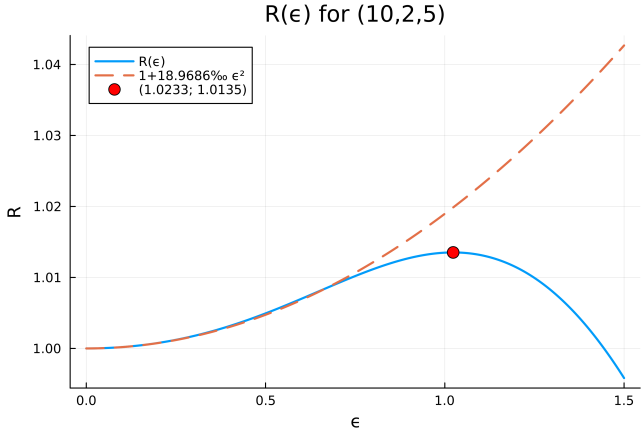
(a) $(n, r, k) = (8, 2, 8)$; $R(\epsilon) \approx 1 + 0.0196\epsilon^2$
$(\epsilon_{\max}; R_{\max}) = (1.2416; 1.0231)$



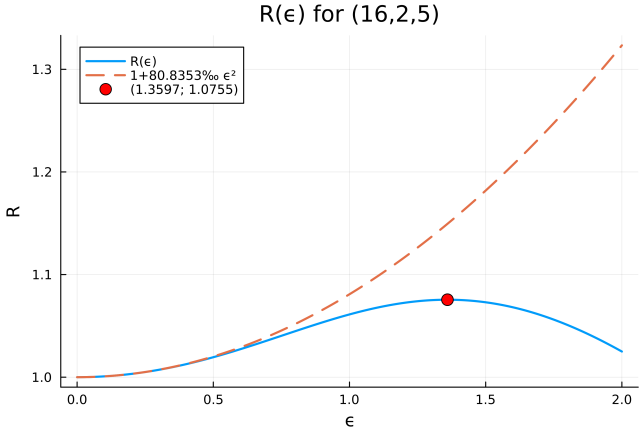
(b) $(n, r, k) = (8, 2, 7)$; $R(\epsilon) \approx 1 + 0.018\epsilon^2$
$(\epsilon_{\max}; R_{\max}) = (1.2275; 1.0206)$



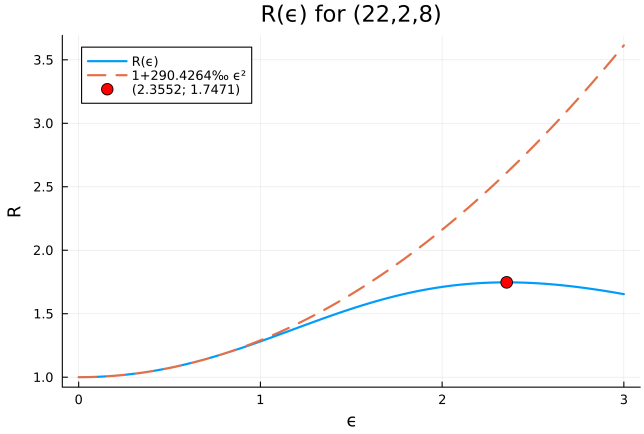
(c) $(n, r, k) = (8, 2, 6)$; $R(\epsilon) \approx 1 + \epsilon^2 9.66‰$
$(\epsilon_{\max}; R_{\max}) = (1.0801; 1.0097)$



(d) $(n, r, k) = (10, 2, 5)$; $R(\epsilon) \approx 1 + \epsilon^2 18.969‰$
$(\epsilon_{\max}; R_{\max}) = (1.0233; 1.0135)$

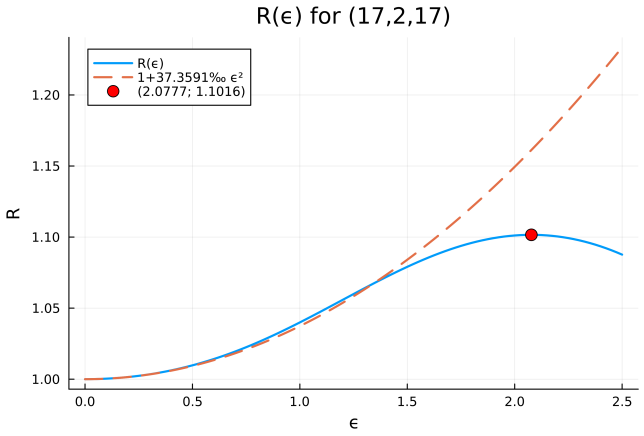


(e) $(n, r, k) = (16, 2, 5)$; $R(\epsilon) \approx 1 + \epsilon^2 80.83‰$
$(\epsilon_{\max}; R_{\max}) = (1.3597; 1.0755)$

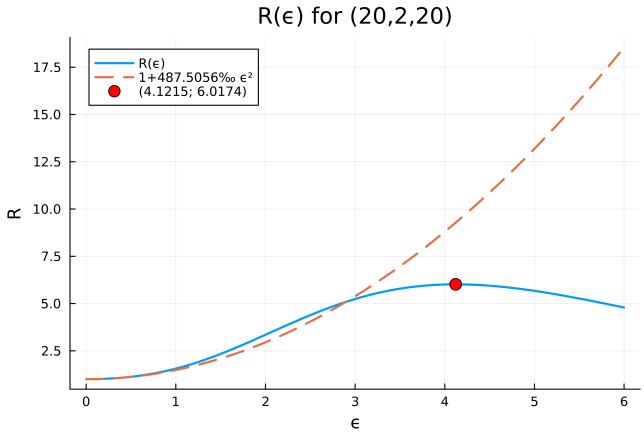


(f) $(n, r, k) = (22, 2, 8)$; $R(\epsilon) \approx 1 + \epsilon^2 290.43‰$
$(\epsilon_{\max}; R_{\max}) = (2.3552; 1.7471)$

Figure 4.4: Numerically computed violation ratio $R(\epsilon)$ (blue line) and its second-order approximation in ϵ^2 (dotted red line) for various (n, r, k) parameter sets.



(a) Numerically computed violation ratio $R(\epsilon)$ for the real perturbation counterexample $(n, r, k) = (17, 2, 17)$; $R(\epsilon) \approx 1 + \epsilon^2 37.36‰$ $(\epsilon_{\max}; R_{\max}) = (2.0777; 1.1016)$



(b) Numerically computed violation ratio $R(\epsilon)$ for the complex perturbation counterexample $(n, r, k) = (20, 2, 20)$; $R(\epsilon) \approx 1 + \epsilon^2 487.51‰$ $(\epsilon_{\max}; R_{\max}) = (4.1215; 6.0174)$

Conclusion

In the first part of this work, we introduced the notion of *bunching*, a characteristic property of bosonic particles, which becomes particularly pronounced as they approach *indistinguishability*. We then explored a sequence of physical conjectures that naturally arise when attempting to describe this phenomenon. These were subsequently linked to mathematical conjectures, revealing a deep connection to the theory of matrix *permanents*.

As one conjecture was disproved — typically by constructing explicit counterexamples — it became necessary to refine the scope of these statements. Each new conjecture was crafted by restricting the previously failed one to more specific or constrained cases.

For instance, the initial conjecture claimed that the *bunching probability* is *globally maximized* when photons are perfectly indistinguishable (and possibly entangled), meaning it exceeds that of any partially distinguishable photons. This evolved first to the *non-entangled* case, which was later restricted to *near indistinguishable* photons (cf. Figure 3.1), where the conjecture became one of a local maximum: indistinguishable photons outperform only those subject to *separable infinitesimal perturbations* to each photon. From there, the conjectures evolved further by limiting the number of perturbed photons, considering only a *perturbed subset* rather than all. Among these, one final refinement involved *real-valued* perturbations, which, when applied specifically to *two photons* with equal magnitude but opposite direction, yielded the most constrained version studied, hence the loosest conjecture (the one that is the most likely to hold).

Since all conjectures up to and including the *local maximum* case were disproven, the second part of this work focused on the remaining conjectures. These were tested numerically for specific cases, but have not yet been (dis)proven analytically for every case. In particular when only considering two-mode bunching $r = 2$, the cases involving perturbation subsets with $k = 8, 7, 6, 5$ photons were numerically disproven (cf. Table 4.6). They required a system of $n = 8, 8, 8, 10$ respectively. However, the search yielded no counterexamples for a perturbed subset of $k = 4$ photons, despite searching for as high as $n = 22$. Additionally, a real perturbation counterexample has been found for a relaxed case $n = k = 17$.

The current state of knowledge about anomalous bunching can be summarized as follows (see also Table 4.9 for a schematic view):

- There is no anomalous bunching for a single perturbed photon, i.e. it is always holds that

$$\lambda_1(f^{(1)}) = F_{11} = H_{11}\,\mathrm{perm}(H_{\setminus 1,\setminus 1}) \leq \mathrm{perm}(H) \quad \forall H \in \mathcal{H}_n \quad \forall H \in \mathcal{H}_n \tag{4.36}$$

  This result follows from Corollary 1 in Ref [22].

- Complex perturbation vectors:
  - It is unknown whether anomalous bunching can occur when perturbing only a subset of 2 to 4 photons.

  - Anomalous bunching has been found when perturbing a subset of 5 or more photons, up to all photons involved. In systems with 8 photons, anomalous bunching has been found for subsets of 6,7 or 8 perturbed photons. When perturbing as few as 5 photons, at least a 10-photon system was required to obtain anomalous bunching.

- Real perturbation vectors:
  - It remains open whether anomalous bunching can occur when perturbing only a subset of photons.
  - If all photons are perturbed, anomalous bunching has been found in systems with 17 photons. The case of 16 photons or fewer remains open.

conjecture type	$\boldsymbol{v} \in \mathbb{C}^k$	$\boldsymbol{v} \in \mathbb{R}^k$
$k = n : n \geq 17$	$\times$	$\times$
$k < n : 17 > k \geq 5$	$\times$	?
$k < n : 5 > k \geq 2$	?	?
$k < n : k = 2, v_2 = -v_1$	?	?
$k < n : k = 1$	✓	✓

Table 4.9: Entries indicate the status of conjectures: "$\times$" indicates that the conjecture is falsified, "✓" indicates that the conjecture has been verified, "?" is used when no counterexample has been found so far. In the first column, $n = k$ refers to the local maximum conjecture M2 while intermediate conjectures M5 are expressed by $k < n$. The case of $k = 2 : v_2 = -v_1$ refers to Pate's conjecture M4, notice that both columns $\mathbb{C}^k$ and $\mathbb{R}^k$ refer to the same conjecture (since when looking for $v_2 = -v_1$ for $\boldsymbol{v} \in \mathbb{C}^2$ it becomes inevitably real).

These numerical limits were largely due to the exponential complexity of computing permanents, even when using the most well-known general-purpose algorithm — Ryser's method. Beyond $n = 17$, we found that the incomplete-rank method — tailored to matrices of rank $r = 2$ (corresponding to two-mode bunching subset) — outperformed Ryser's algorithm. Our most effective numerical strategy involved first applying a global optimizer (such as BBO) to find potential counterexamples, followed by a local optimizer to refine the solution and potentially increase the violation. Different optimizers trade off precision and runtime, so chaining them — first fast (like Nelder Mead), then accurate (such as LBFGS) — proved beneficial, although the gains were not significant.

A natural direction for further research would be to improve optimization performance by explicitly deriving the gradient (and possibly the hessian) of the function to be optimized. This would accelerate numerical counterexample discovery. On the analytical side, it would be valuable to verify the *permanent-on-top* conjecture for the only remaining unkown dimension $n = 4$, or for the yet-to-be-(dis)proven intermediate conjectures with perturbed subsets $k = 4$ or smaller.

Another avenue for exploration stems from the observation that the matrix $L = \text{perm}(A)\mathbb{I} - F$ resembles a graph Laplacian. While F does not fit the standard definition of an adjacency matrix — since it may have nonzero diagonal elements, and off-diagonal elements differing from 1 and 0 — it can be interpreted as a weighted adjacency matrix of a generalized graph with self loops. Moreover, since F is Hermitian (and not just symmetric), it may represent a bidirectional structure. In this analogy, $\text{perm}(A)\mathbb{I}$ plays the role of the degree matrix, as it satisfies $(\text{perm}(A)\mathbb{I})_{ii} = \sum_j F_{ij}$. This viewpoint, although nonstandard, might illuminate structural properties of the permanent and help import tools from graph theory into the study of bosonic interference.

Finally, a broader shift in focus might also prove fruitful. Rather than seeking ever more restrictive conjectures with decreasing perturbed subsets of k photons, one could instead look for

higher *violations* — counterexamples involving the larger number of photons n, with all photons perturbed (i.e. $k = n$), thus much easier to find numerically. Such was the objective for a case involving $n = k = 20$ photons, all perturbed, whose violation ratio reached $R(\epsilon) = 6$ (cf. Figure 4.5b). This could guide the design of real experimental setups capable of detecting *anomalous bunching*, otherwise difficult to observe in practice due to the small violations.

References

[1]C. K. Hong, Z. Y. Ou, and L. Mandel, "Measurement of subpicosecond time intervals between two photons by interference", Physical Review Letters **59**, 2044–2046 (1987).

[2]E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics", Nature **409**, 46–52 (2001).

[3]P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", SIAM Journal on Computing **26**, 1484–1509 (1997).

[4]S. Aaronson and A. Arkhipov, "The computational complexity of linear optics", Theory of Computing **9**, 143–152 (2013).

[5]J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, "Multiphoton entanglement and interferometry", Reviews of Modern Physics **84**, 777–838 (2012).

[6]A. Lyons, G. C. Knee, E. Bolduc, T. Roger, J. Leach, E. M. Gauger, and D. Faccio, "Attosecond-resolution Hong-Ou-Mandel interferometry", Science Advances **4** (2018).

[7]P. Kok, H. Lee, and J. P. Dowling, "Creation of large-photon-number path entanglement conditioned on photodetection", Physical Review A **65** (2002).

[8]B. Seron, L. Novo, and N. J. Cerf, "Boson bunching is not maximized by indistinguishable particles", Nature Photonics **17**, 702–709 (2023).

[9]L. Pioge, B. Seron, L. Novo, and N. J. Cerf, "Anomalous bunching of nearly indistinguishable bosons", arXiv:2308.12226v2 (2024).

[10]F. Mann, H. M. Chrzanowski, F. Gewers, M. Placke, and S. Ramelow, "Hong-Ou-Mandel effect with two frequency-entangled photons of vastly different color", arXiv:2504.03304v1 (2025).

[11]M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator", Physical Review Letters **73**, 58–61 (1994).

[12]V. S. Shchesnovich, "Universality of generalized bunching and efficient assessment of boson sampling", Physical Review Letters **116** (2016).

[13]R. Bapat and V. Sunder, "On majorization and schur products", Linear Algebra and its Applications **72**, 107–117 (1985).

[14]F. Zhang, "Notes on hadamard products of matrices", Linear and Multilinear Algebra **25**, 237–242 (1989).

[15]S. Drury, "A counterexample to a question of bapat and sunder", The Electronic Journal of Linear Algebra **31**, 69–70 (2016).

[16]G. H. Gloub and C. F. VanLoan, *Matrix computations*, 3rd (Johns Hopkins University Press, 1996).

[17]R. Bapat and V. Sunder, "An extremal property of the permanent and the determinant", Linear Algebra and its Applications **76**, 153–163 (1986).

[18]S. Drury, "A counterexample to a question of bapat and sunder", Mathematical Inequalities & Applications **21**, 517–520 (2018).

[19]I. Schur, "Über endliche gruppen und hermitesche formen", Mathematische Zeitschrift **1**, 184–207 (1918).

[20]H. Minc, "Theory of permanents 1978-1981", Linear and Multilinear Algebra **12**, 227–263 (1983).

[21]V. S. Shchesnovich, "The permanent-on-top conjecture is false", Linear Algebra and its Applications **490**, 196–201 (2016).

[22]I. M. Wanless, "Lieb's permanental dominance conjecture", arXiv:2202.01867v1 (2022).

[23]T. H. Pate, "On permanental compounds", Linear Algebra and its Applications **429**, 1093–1101 (2008).

[24]B. Seron and A. Restivo, "Bosonsampling. jl: a julia package for quantum multi-photon interferometry", https://arxiv.org/abs/2212.09537v2 (2022).

[25]A. I. Barvinok, "Two algorithmic results for the traveling salesman problem", Mathematics of Operations Research **21**, 65–84 (1996).

Appendix A

Additional Definitions

Those definitions are relevant for the proof of the Permanent-on-Top conjecture for $n \leq 3$ in section 2.4.2.

A.1 Representations

A representation ρ of a group G is a homomorphism from G to the group of invertible matrices acting on a vector space V. To ensure the group operation being mirrored by the matrices, the homomorphism condition must be satisfied:

$$\rho(\sigma\tau) = \rho(\sigma)\rho(\tau) \quad \forall \sigma, \tau \in G \tag{A.1}$$

A.1.1 Left regular representation

We will introduce the left regular representation for the a specific group being the symmetric group $\mathcal{S}_n$. To each permutation $\sigma \in \mathcal{S}_n$ in the symmetric group $\mathcal{S}_n$ we associate a vector $\boldsymbol{e}_\sigma$ indexed by the corresponding permutation σ. Let V be a complex vector space of dimension $n!$ with basis $\{\boldsymbol{e}_\sigma : \sigma \in \mathcal{S}_n\}$. The left regular representation ρ_{reg} of the symmetric group $\mathcal{S}_n$ is a map from $\mathcal{S}_n$ to the space of linear operators on V that acts on the basis vectors as follows:

$$\rho_{\text{reg}}(\sigma)\boldsymbol{e}_\tau = \boldsymbol{e}_{\sigma\tau} \quad \forall \sigma, \tau \in \mathcal{S}_n \tag{A.2}$$

This means that each element $\sigma \in \mathcal{S}_n$ is associated with a linear operator $\rho_{\text{reg}}(\sigma)$ on V, which permutes the basis vector $\boldsymbol{e}_\tau$ by left multiplication $\boldsymbol{e}_{\sigma\tau}$.

Let us denote by $\{\tau_i : i \in [n!]\}$ the set of all possible permutations in the symmetric group $\mathcal{S}_n$. The entry (i, j) of $\rho_{\text{reg}}(\sigma)$ is

$$(\rho_{\text{reg}}(\sigma))_{i,j} = \begin{cases} 1 & \text{if } \tau_i = \sigma\tau_j \\ 0 & \text{else} \end{cases} \tag{A.3}$$

A.1.2 Natural representation

Let V be a complex vector space this time of dimension n with standard basis $\{\boldsymbol{e}_i : i \in [n]\}$. The natural representation ρ_{nat} is a map from $\mathcal{S}_n$ to the space of linear operators on V and acts on the basis vectors by permuting them as:

$$\rho_{\text{nat}}(\sigma)\boldsymbol{e}_i = \boldsymbol{e}_{\sigma(i)} \quad \forall i = [n] \tag{A.4}$$

For any vector $\boldsymbol{v} \in V$, it is often abbreviated as $\rho_{\text{nat}}(\sigma)\boldsymbol{v} = \sigma \cdot \boldsymbol{v}$, and is understood as the permutation of the components of $\boldsymbol{v}$. More specifically, the action of σ on the component i of the vector $\boldsymbol{v}$ is:

$$\sigma \cdot \boldsymbol{v}_i = \boldsymbol{v}_{\sigma^{-1}(i)} \quad \forall i = [n] \tag{A.5}$$

As an example, the representation for $\sigma = (132)$ in S_3 is

$$\rho_{\text{nat}}((132)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \tag{A.6}$$

A.1.3 Irreducible Representation

However, the natural representation ρ_{nat} can be decomposed into a direct sum of irreducible representations. More precisely, a representation ρ is called irreducible if the only subspaces that remain invariant under the action of all group elements (i.e. any vector in the subspace is mapped to another vector within that same subspace) are the trivial ones: the zero vector space $\{0\}$ and the whole space V itself. That is,

$$\forall v \in V, \forall \sigma \in G : \rho(\sigma)v \in V \tag{A.7}$$

Since each permutation σ can be expressed as a composition of cycles, it can also be categorized by a partition λ, which records the length of each cycle in descending order. For S_3, the $n! = 3!$ permutations σ fall into 3 such partitions as depicted in Table A.1. The number of irreducible

Partition λ	**Permutations** σ
$(1)(1)(1)$	$\epsilon = (1)(2)(3)$
$(2)(1)$	$(12), (13), (23)$
(3)	$(123), (132)$

Table A.1: Partitions λ of permutations σ in S_3

representations for S_n equals the number of partition $p(n)$ of n. For S_3, we have $p(3) = 3$ irreducible representations, which can be defined generally for any n dimensional symmetric group S_n as follows:

1. the trivial representation ρ_{triv}: every group element acts as the identity transformation on a 1 dimensional subspace V

   $$\rho_{\text{triv}}(\sigma)\boldsymbol{v} = \boldsymbol{v} \quad \forall \boldsymbol{v} \in V, \forall \sigma \in \mathcal{S}_n \tag{A.8}$$

   This vector space V is typically spanned by the constant vector $\mathbf{1} \in \mathbb{C}^n$ defined as:

   $$\mathbf{1} \equiv \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \tag{A.9}$$

   Thus, $\rho_{\text{triv}}(\sigma) = 1$ is a (1×1) identity matrix $\mathbb{I}$.

2. the alternating representation ρ_{alt} also acts on a 1 dimensional subspace but distinguishes permutations by their sign

   $$\rho_{\text{alt}}(\sigma)\boldsymbol{v} = \text{sgn}(\sigma)\boldsymbol{v} \tag{A.10}$$

   The representation $\rho_{\text{alt}}(\sigma) = \text{sgn}(\sigma)$ is also a (1×1) but signed matrix.

3. the standard representation ρ_{std} is the restriction of the natural representation ρ_{nat} to the $(n-1)$ dimensional subspace V which is orthogonal to the trivial representation ρ_{triv} vector subspace $\mathbf{1}$

$$V = \{\boldsymbol{v} \in \mathbb{C}^n : \mathbf{1}^\dagger \boldsymbol{v} = 0\} \tag{A.11}$$

For example in $\mathcal{S}_3$ the basis $\{\boldsymbol{v}_1 = (1,-1,0)^t, \boldsymbol{v}_2 = (1,1,-2)^t\}$ satisfies condition (A.11), then a given permutation $\sigma = (12)$ acts as the following:

$$\sigma \cdot \boldsymbol{v}_1 = (-1,1,0)^t = -\boldsymbol{v}_1, \quad \sigma \cdot \boldsymbol{v}_2 = (1,1,-2)^t = \boldsymbol{v}_2 \tag{A.12}$$

Note that we have used the natural representation $\rho_{\text{nat}}(\sigma)\boldsymbol{v} = \sigma \cdot \boldsymbol{v}$ acting on $\boldsymbol{v} \in \mathbb{C}^3$ however, the representation itself $\rho_{\text{std}}((12))$

$$\rho_{\text{std}}((12)) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{A.13}$$

is a (2×2) matrix acting on the 2 dimensional subspace $V \perp \mathbf{1}$ not on $\mathbb{C}^3$ itself.

A.1.4 Direct Sum Decomposition

The natural representation ρ_{nat} decomposes as a direct sum of ρ_{triv} and ρ_{std}. The left regular representation ρ_{reg} decomposes into a direct sum of all irreducible representations ρ_i of the group, each appearing with multiplicity d_i equal to its dimension:

$$\rho_{\text{reg}} \cong \bigoplus_{i=1}^{p(n)} d_i \rho_i \tag{A.14}$$

Since the irreducible components act on invariant and mutually orthogonal subspaces, an inequality, or operator property involving ρ_{reg} can be verified separately on each irreducible component.

Appendix B

Counterexample Matrices

The following matrices are presented with values specified to at least 14 decimal places, consistent with 64-bit double precision floating-point representation (which provides approximately 15-17 significant decimal digits), reflecting the raw output of our numerical optimization. Note that this formatting does not imply 16-digit accuracy; floating-point error accumulation typically yields fewer significant digits in practice. However, truncating to an insufficient number of digits could potentially degrade the violation ratio, justifying this representation.

B.1 Intermediate Conjectures

For complex perturbation vectors, the $(n \times r)$ matrix $M^\dagger = \Re[M^\dagger] + i\Im[M^\dagger]$ of the counterexamples for each k (starting with $k = 8$), with the lowest dimensional n, as mentioned in Table 4.7 and yielding violation ratios displayed in Figure 4.4, is written here explicitly.

For $(n, r, k) = (8, 2, 8)$:

$$\Re[M^\dagger] = \begin{pmatrix} -19.27973998905716 & 24.512422508995286 \\ 22.40138556143034 & -10.02868763374688 \\ -13.30844385583821 & -29.248767684153986 \\ 23.407372605925325 & -34.454504121262424 \\ 31.778411040349255 & 16.906151001676 \\ -36.30859114520126 & -5.87958203044978 \\ -6.902277478759458 & 48.95580487502791 \\ 21.994108273071195 & -10.76957920872963 \end{pmatrix} \tag{B.1}$$

$$\Im[M^\dagger] = \begin{pmatrix} -15.383247811712959 & 1.3550781218063954 \\ 13.061567317606574 & 29.669737394624317 \\ -53.463247986301724 & 3.802685571345312 \\ -27.476897157061707 & 4.944864320806435 \\ -12.630027482869336 & -50.79651437617374 \\ -42.18655345610969 & -21.746101859795033 \\ 13.907436379945949 & -34.45090195317378 \\ 42.13199141347784 & -6.348960299447917 \end{pmatrix} \tag{B.2}$$

For $(n, r, k) = (8, 2, 7)$:

$$\Re[M^\dagger] = \begin{pmatrix} -33.026763960672774 & -50.16684213066005 \\ -40.6755966142037 & 5.630973827465017 \\ -55.93331391031506 & 6.463978613725208 \\ -1.0682981491067614 & -28.52471679613835 \\ 51.2843422261179 & -47.89679849529764 \\ 14.925717757971404 & 43.33760267859043 \\ -21.982309516842857 & 54.210827809067275 \\ -0.22400618997379143 & 39.79910184184562 \end{pmatrix} \tag{B.3}$$

$$\Im[M^\dagger] = \begin{pmatrix} -31.155126752434967 & -40.83043012475833 \\ -37.91136985466681 & 28.8111644863626 \\ -22.67788780222025 & -9.062272050970382 \\ -0.1092117821575023 & -2.6106434665392415 \\ 2.7603853854041076 & 22.226853624520707 \\ 40.93018229233945 & -0.5213681553916506 \\ -40.84463715229701 & -28.929234979928836 \\ 31.108881506343657 & -47.993580965427704 \end{pmatrix} \tag{B.4}$$

For $(n, r, k) = (8, 2, 6)$:

$$\Re[M^\dagger] = \begin{pmatrix} -20.340960180844508 & 16.388260643704754 \\ 28.14403065215319 & -4.1677202619718114 \\ -35.00529737592378 & -4.5903395182711 \\ -15.305518480718558 & -16.187014292861218 \\ -32.702675456929 & 11.331494673390504 \\ -30.252760086663383 & 25.61803860566624 \\ 0.7359880628821807 & 30.014430279874585 \\ 0.7084379341529433 & 13.093502236308575 \end{pmatrix} \tag{B.5}$$

$$\Im[M^\dagger] = \begin{pmatrix} -21.300418940954856 & -22.37247773777849 \\ -21.686790752035627 & -7.6688596377204705 \\ 12.539694053350313 & 39.63532280239014 \\ 42.66834880750554 & 36.768332457804775 \\ -16.815858390448295 & 41.467576417004395 \\ 7.531856340005444 & -23.752369070799226 \\ -6.64009153217154 & -38.47240277208332 \\ -0.037813115291300416 & 34.88496564263725 \end{pmatrix} \tag{B.6}$$

For $(n, r, k) = (10, 2, 5)$:

$$\Re[M^\dagger] = \begin{pmatrix} 12.724087157697031 & 11.019125503365544 \\ 23.934098806525675 & -29.471170065235818 \\ 7.058545937333791 & -4.019375150031794 \\ -20.0541661502745 & 18.900716379844614 \\ 27.400182150392745 & 18.79358932278734 \\ -34.41482917956147 & 3.9399148920108593 \\ 12.908497963524834 & -16.75948865379304 \\ 3.7846573878857805 & -46.73885326709621 \\ 20.155187144667256 & 44.36384399841827 \\ -8.147853826544226 & -21.823305673824606 \end{pmatrix} \tag{B.7}$$

$$\Im[M^{\dagger}] = \begin{pmatrix} -46.33990512010141 & -15.622266404370444 \\ 9.587928139512249 & 24.448073230810582 \\ -29.811804634545183 & 4.031145355193529 \\ -29.529461316940885 & 28.735917681804352 \\ 6.332360986428548 & 34.21883824818084 \\ 17.807842406886017 & 44.37172970647607 \\ -28.187539406980658 & -41.37945060722293 \\ -37.92599226080596 & -3.299300410320387 \\ 12.754362081226537 & -48.45265967706525 \\ -5.046010530617657 & 43.95393795310394 \end{pmatrix} \tag{B.8}$$

The counterexample $M^{\dagger}$ with a more numerically relaxed case $(n, r, k) : (20, 2, 20)$ to obtain a much bigger violation ratio as displayed in Figure 4.5b, is:

$$\Re[M^{\dagger}] = \begin{pmatrix} -11.608084974123337 & -50.884401995545254 \\ 19.327659784763746 & 42.28501364727814 \\ 16.025167376900818 & -47.98468848715327 \\ 20.400591585453018 & 21.359479580831938 \\ 32.87359928098447 & -27.222094831723584 \\ -46.57288856350903 & -37.0131786899834 \\ 35.2244243120546 & 10.007624649941382 \\ -43.64195339807457 & 12.462027262653208 \\ 1.410700364745222 & -30.869516826973317 \\ 22.794141406792075 & 38.711366990881565 \\ -48.512230214443186 & -24.184715567805753 \\ 55.26349358414746 & -0.42551957705115195 \\ -60.82380788850235 & 20.920719208460184 \\ 23.04721345984038 & -10.459118845986376 \\ -11.977190514630701 & -53.221893222026324 \\ -6.6745726747647245 & 30.461256430032215 \\ 39.38354212419131 & 31.251054838153273 \\ 35.09352300750899 & 38.22425656460953 \\ -42.45410360044216 & 18.73422069007815 \\ -45.048652653322215 & 38.47170095963702 \end{pmatrix} \tag{B.9}$$

$$\Im[M^\dagger] = \begin{pmatrix} -16.28931526741547 & 36.69360614445305 \\ 16.951064583887426 & -1.43405829928676 \\ 7.817389074503644 & 42.56284804360472 \\ -22.830216244984086 & 25.246243820345097 \\ 38.593960415099815 & -47.4022337226595 \\ 13.667698324153438 & -19.166499316880753 \\ 49.206372049899855 & 23.49689980177754 \\ -35.02159970662092 & -23.218910291418872 \\ 34.10099661409676 & -38.43235746407254 \\ -30.75393701633848 & -37.14817232871657 \\ -35.95906627610786 & 9.144856732938317 \\ 9.82492539721687 & 0.9872977654142935 \\ 11.310051465351735 & 29.345062589090013 \\ 18.68122536229227 & 51.30542989200656 \\ -9.090609499679013 & -48.8399822190716 \\ 20.229125082773464 & -50.151507208772365 \\ 23.08845998872662 & -33.312576042468244 \\ 36.80373572515839 & 17.06170560874865 \\ -0.3332365510084254 & -18.819692898117236 \\ 5.337010621925767 & 48.28155628847052 \end{pmatrix} \tag{B.10}$$

B.2 Real Intermediate Conjectures

For $(n, r, k) = (17, 2, 17)$ the counterexample for a real perturbation vector — yielding a violation ratio shown in Figure 4.5a — is:

$$\Re[M^\dagger] = \begin{pmatrix} 44.737878390903546 & -18.423381189686705 \\ -18.738937868577334 & -8.676902970401017 \\ 7.813233814523968 & 26.92957857212379 \\ -28.573502691903556 & 10.59110315437202 \\ 32.214666059533386 & 20.176038986476378 \\ -6.263218656723517 & 1.7045294342386288 \\ 34.1721422479508 & -21.779445882978173 \\ -12.159162379384409 & -44.61334706422904 \\ -35.83427663822305 & -2.810639451740424 \\ 51.58005762024606 & -12.930080685999464 \\ 14.353998255447413 & -23.753905443206815 \\ 11.71495690935856 & 48.22942255273919 \\ -32.1575616388388 & -12.241669154823242 \\ -19.419832256334704 & -40.13714771194084 \\ -21.836522611358326 & 1.431427900055775 \\ -19.223130870028804 & -23.246458096866313 \\ -11.208488130958438 & -44.73638926415173 \end{pmatrix} \tag{B.11}$$

$$\Im[M^{\dagger}] = \left(\begin{array}{cc} 31.92076032033974 & 3.9446466351508875 \\ 28.75331090425896 & -21.195989993958893 \\ 25.805070508006942 & 3.6376442146807415 \\ -3.622674046200489 & 19.86419014832973 \\ 5.399979995071319 & 46.52620590670907 \\ -9.032636256679115 & 40.5492567809099 \\ 17.53129971019158 & -46.77584375181678 \\ 3.909942771446728 & 35.121842846046874 \\ -3.81496956482659 & 24.076563488705705 \\ 31.76935692267677 & 45.25567856034303 \\ 11.161690903630472 & 0.6533520799479308 \\ -29.524111616369147 & 53.93153453940263 \\ -12.768738633535053 & -14.470849910369743 \\ -6.102844820279533 & -29.885677184494334 \\ -23.827680560414215 & -6.1039489967498195 \\ -25.286443647139336 & -24.667814625288074 \\ -23.507062716646683 & 18.829897840191418 \end{array}\right) \tag{B.12}$$