

On interferometric duality in multibeam experiments

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2003 J. Phys. A: Math. Gen. 36 11481

(<http://iopscience.iop.org/0305-4470/36/45/009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.89

The article was downloaded on 02/06/2010 at 17:15

Please note that [terms and conditions apply](#).

On interferometric duality in multibeam experiments

G Bimonte^{1,2} and R Musto^{1,2}

¹ Dipartimento di Scienze Fisiche, Università di Napoli, Federico II, Complesso Universitario MSA, Via Cintia, I-80126, Napoli, Italy

² INFN, Sezione di Napoli, Napoli, Italy

E-mail: bimonte@napoli.infn.it and musto@napoli.infn.it

Received 29 May 2003

Published 29 October 2003

Online at stacks.iop.org/JPhysA/36/11481

Abstract

We critically analyse the problem of formulating duality between fringe visibility and which-way information, in multibeam interference experiments. We show that the traditional notion of visibility is incompatible with any intuitive idea of complementarity, but for the two-beam case. We derive a number of new inequalities, not present in the two-beam case, one of them coinciding with a recently proposed multibeam generalization of the inequality found by Greenberger and YaSin. We show, by an explicit procedure of optimization in a three-beam case, that suggested generalizations of Englert's inequality do not convey, differently from the two-beam case, the idea of complementarity, according to which an increase of visibility is at the cost of a loss in path information, and vice versa.

PACS numbers: 03.65.Ta, 03.67.—a

1. Introduction

Interferometric duality, as complementarity between fringe visibility and which-way information is called today, has a long, perhaps a surprisingly long history (for a recent review, see [1]). It was the central issue of the famous debate between Einstein and Bohr, on complementarity. Even if, already at that time, in defending complementarity against Einstein's criticism, Bohr pointed out that not only the system under observation, but also the measuring apparatus should be regarded as a quantum object [2], the discussion was essentially semiclassical in nature. As it was based essentially on the position–momentum Heisenberg uncertainty principle, it considered only the two extreme cases, of either purely particle-like or purely wave-like behaviour of the system. It was only in 1979 that Wootters and Zurek, [3] gave the first full quantum mechanical treatment of Young interference, in the presence of a which-way detector. They recognized that ‘in Einstein's version of the double-slit experiment,

one can retain a surprisingly strong interference pattern by not insisting on a 100% reliable determination of the slit through which each photon passes’.

By now, a consistent and simple formulation of interferometric duality has been achieved in the case of *two* interfering beams. In the absence of a which-way detector, Greenberger and YaSin [4], showed that it was possible to convert the basic quantum mechanical inequality $\text{Tr } \rho^2 \leq 1$, into one connecting the fringe visibility to the predictability of the path, based on unequal beam populations. This is an experimentally testable inequality, as it involves physically measurable quantities. For pure states, when the inequality is saturated, this statement becomes a formulation of interferometric duality; any increase in predictability is at the cost of a decrease in visibility, and vice versa.

In the case of an interference experiment performed in the presence of a which-way detector, in order to gain information on the path, one needs to carry out a measurement on the detector, after the passage of each quanton. Since, in general, no measurement ensures an unambiguous path reconstruction, the determination of the best possible measurement is a matter of statistical decision theory, that requires an *a priori* choice of an evaluation criterion. In their pioneering work, Wootters and Zurek, [3], used Shannon’s definition of information entropy [5] in order to evaluate the which-way information gained after the measurement. Following this suggestion, Englert [6], by using a different criterion for evaluating the available information, was able to establish an inequality, stating that the sum of the square of the *distinguishability*, that gives a quantitative estimate of the way³, and the *visibility* squared, is bound by one. Again, the inequality is saturated for pure states, turning into a statement of interferometric duality: any gain in distinguishability is paid for by a loss in visibility and vice versa.

In the present paper, we discuss the issue of formulating interferometric duality in the case of multibeam experiments. As an example of the problems arising, we may refer to an experiment [7] with four beams, in which the surprising result is found that scattering of a photon by one of the beams, may lead to an increase of visibility, rather than to an attenuation (for a comment see [8]). To get a better understanding of this experiment, we build an analytical three-beam example, which shows that, differently from the two-beam case, the traditional visibility may increase, after an interaction of the beams with another quantum system. This points towards the need for a different notion of visibility, and one possibility is offered in [9], where the visibility is defined as the properly normalized, rms deviation of the fringe intensity from its mean value. We briefly review Dürr’s [9] derivation, for the multibeam case, of an inequality similar to that of Greenberger and YaSin [4], that relates this new notion of visibility to a corresponding newly defined predictability. Again, in the case of a pure state the inequality is saturated and, then, in analogy with the two-beam case, may be taken as a formal definition of interferometric duality. However, as we will discuss later, this is at the cost of using a definition of predictability that has somehow lost contact with the ability of guessing the way right. Furthermore we show how, in the multibeam case, it is possible to construct new inequalities, resulting, like the one of Greenberger and YaSin, from basic quantum mechanical properties of the density matrix. Each of them can be written in terms of quantities that, in principle, may be measured in interference experiments, such as higher momenta of fringe intensity. The new inequalities then provide, exactly as the original one, independent tests on the validity of quantum mechanics in multibeam interference experiments. They also are saturated for pure states, but, at least at first sight, they do not seem to convey any simple relation with the principle of complementarity.

³ In [6] the distinguishability is expressed in terms of the optimum likelihood \mathcal{L}_{opt} for ‘guessing the way right’. This optimum likelihood is one minus the optimum average Bayes cost \bar{C}_{opt} .

Then we turn to the more interesting problem of complementarity in the presence of a which-way detector. By introducing two alternative definitions of distinguishability, Dürr constructed a generalization of Englert's inequality to the multibeam case, proposing to look at it as a formal definition of interferometric duality. We show that, for equally populated beams apart from the two-beam case, the new inequality holds as an equality only for the extreme cases where either the visibility or the distinguishability vanishes, even when the beams and the detector are both prepared in pure states. Then, there may be cases in which the distinguishability and the visibility both increase or decrease at the same time. This is in sharp contrast with the idea of complementarity, according to which '...the more clearly we wish to observe the wave nature ...the more information we must give up about ...particle properties' [3]. In a recent paper [10], an example in which this situation occurs was constructed. However, we considered there an extremely simplified model for the detector, having a two-dimensional space of states. A realistic model requires an infinite Hilbert space of states, and we analyse it in this paper. This is a much harder problem, because the task of determining the path distinguishability implies the solution of an optimization problem, that has to be performed now in an infinite-dimensional space. We report the full proof in this paper, not only for the sake of completeness, but also because it provides an example in quantum decision theory, which is a subject where few general results are known, and few cases can actually be treated. Surprisingly, in the case we examined, the distinguishability of the infinite-dimensional problem coincides with that found in [10], for the simplified model. This shows that the conclusions drawn in [10] have full generality, showing that the notion of interferometric duality in the multibeam case has not yet been properly formulated.

The paper is organized as follows: in section 2 we discuss interferometric-duality schemes, not involving which-way detectors. In section 3 we derive a new set of inequalities, not present in the two-beam case, and we comment on them. In section 4, which-way detection schemes are treated, while in section 5, we discuss the optimization problem for a three-beam example. Section 6 is devoted to our concluding remarks. The most technical proofs of the results quoted in the paper are presented in two Appendices that close the paper.

2. Visibility and predictability

We consider an n -beam interferometer, in which a beam splitter splits first a beam of quantum objects ('quantons', in brief) into n beams, that afterwards converge on a second beam splitter, where they interfere, giving rise to n output beams. We imagine that, at some instant of time, the (normalized) wavefunctions $|\psi_i\rangle$, $i = 1, \dots, n$, for the individual beams are fully localized in the region between the two beam splitters, and are spatially well separated from each other, so that $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. The state of the quanton, in front of the second beam splitter, is then described by a density matrix ρ of the form

$$\rho = \sum_{ij} \rho_{ij} |\psi_i\rangle\langle\psi_j|. \quad (2.1)$$

The diagonal elements ρ_{ii} represent the populations ζ_i of the beams, and obviously they satisfy the condition:

$$\sum_i \zeta_i = \text{Tr } \rho = 1. \quad (2.2)$$

The off-diagonal elements of ρ , that we shall denote as I_{ij} , are instead related to the probability I of finding a quanton in one of the n output beams, according to the following equation:

$$I = \frac{1}{n} \left(1 + \sum_i \sum_{j \neq i} e^{i(\phi_i - \phi_j)} I_{ij} \right). \quad (2.3)$$

Here, $\phi_i - \phi_j$ is the relative phase between beams i and j . In this paper we consider experimental settings, such that all these relative phases can be adjusted at will. However, this is not the case in a number of experimental settings, where the features of the apparatus may lead to relations among the relative phases of the beams. When this happens, the output beam intensity equation (2.3) may be rewritten, by expressing the relative phases in terms of the independently adjustable ones. An analysis of complementarity tailored on specific experimental settings, involving definite relations among the phases, may turn out to be interesting and useful. However, the purpose of the present paper is to study the problems arising when the full freedom allowed by an n -beam setting is taken into account.

Going back to equation (2.3), one notes that I does not depend at all on the populations ζ_i . In the standard case of an interferometer with two beams of interfering quantons, a typical measure of the fringe contrast is the traditional visibility \mathcal{V} , defined as

$$\mathcal{V} \equiv \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad (2.4)$$

where I_{\max} and I_{\min} are, respectively, the maximum and minimum of I . It is easy to verify, using equation (2.3) with $n = 2$, that

$$\mathcal{V} = 2|I_{12}|. \quad (2.5)$$

A few years ago, Greenberger and YaSin [4] noted that the general rules of quantum mechanics imply the existence of a simple relation connecting the visibility \mathcal{V} , to the populations ζ_i of the beams. They considered the so-called predictability

$$\mathcal{P} := |\zeta_1 - \zeta_2| \quad (2.6)$$

which can be interpreted as the *a priori* probability for ‘guessing the way right’, when one has unequal populations of the beams. It is easy to verify that the general condition

$$\text{Tr } \rho^2 \leq 1 \quad (2.7)$$

turns into the following inequality:

$$\mathcal{V}^2 + \mathcal{P}^2 \leq 1. \quad (2.8)$$

When it is saturated, namely for pure states, one can recognize in equation (2.8) a statement of wave–particle duality, because then a large predictability of the way followed by the quantons, implies a small visibility of the interference fringes, and vice versa.

Independently of any interpretation, the inequality (2.8) represents a testable relation between measurable quantities, that follows from the first principles of quantum mechanics. Indeed, the experiments with asymmetric beams of neutrons made by Rauch *et al* [11] are compatible with it. It is interesting to observe that equation (2.8) also provides an operative, quantitative way to determine how far the beam is from being pure.

One may ask whether an inequality analogous to equation (2.8) holds in the multibeam case. Here, one’s first attitude would be to keep the definition of visibility, equation (2.5), unaltered. However, this choice has a severe fault, as we now explain. Suppose that the beams are made to interact with another system, that we call environment, and assume that the interaction does not alter the populations of the beams. If the interaction is described

as a scattering process, its effect is to give rise to an entanglement of the beams with the environment, such that

$$|\chi_0\rangle\langle\chi_0| \otimes \rho \rightarrow \rho_{b\&e} = \sum_{ij} \rho_{ij} |\chi_i\rangle\langle\chi_j| \otimes |\psi_i\rangle\langle\psi_j|. \tag{2.9}$$

Here, $|\chi_0\rangle$ and $|\chi_i\rangle$ are normalized environment states (we have assumed for simplicity that the initial state $|\chi_0\rangle$ of the environment is pure, but taking a mixture would not change the result). The entanglement with the environment alters the probability of finding a quanton in the chosen output beam. Indeed, the state ρ' of the beams, after the interaction with the environment, is obtained by tracing out the environment's degree of freedom from equation (2.9):

$$\rho' = \sum_{ij} \rho_{ij} \langle\chi_j|\chi_i\rangle |\psi_i\rangle\langle\psi_j|. \tag{2.10}$$

By plugging ρ' into equation (2.3), we obtain the new expression for the probability I' of finding a quanton in the selected output beam:

$$I' = \frac{1}{n} \left(1 + \sum_i \sum_{j \neq i} e^{i(\phi_i - \phi_j)} I_{ij} \langle\chi_i|\chi_j\rangle \right). \tag{2.11}$$

If we agree that the visibility V should be fully determined by the intensity of the output beam I' , we require that it should be defined in such a way that, for any choice of the environment states $|\chi_i\rangle$, $V' \leq V$. It is easy to convince oneself that the standard visibility \mathcal{V} fulfils this requirement for two beams, while it does not for a larger number of beams. Indeed, for two beams, $V' \leq V$ is a direct consequence of equation (2.5). Things are different already with three beams. Consider for example the three-beam state, described by the following density matrix ρ :

$$\rho = \frac{1}{3} \begin{pmatrix} 1 & -\lambda & \lambda \\ -\lambda & 1 & -\lambda \\ \lambda & -\lambda & 1 \end{pmatrix}. \tag{2.12}$$

It can be checked that ρ is positive definite if $0 \leq \lambda < 1$. A direct computation of the visibility \mathcal{V} , for $\lambda > 0$, gives the result:

$$\mathcal{V} = \frac{3\lambda}{2 + \lambda}. \tag{2.13}$$

Suppose now that the interaction with the environment is such that the environment's states in equation (2.9) satisfy the conditions: $|\chi_1\rangle = |\chi_2\rangle$ and $\langle\chi_1|\chi_3\rangle = \langle\chi_2|\chi_3\rangle = 0$. This condition is typically realized if the environment interacts only with the third beam, as happens, for example, if one scatters light off the third beam only. This is precisely the type of situation that is realized, in a four-beam context, in the experiment of [7]. With this choice for the states $|\chi_i\rangle$, the density matrix ρ' in equation (2.10) becomes

$$\rho' = \frac{1}{3} \begin{pmatrix} 1 & -\lambda & 0 \\ -\lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.14}$$

It can be verified that the new value of the visibility \mathcal{V}' is

$$\mathcal{V}' = \frac{4}{3}\lambda. \tag{2.15}$$

We see that, for $1/4 < \lambda < 1$, $\mathcal{V}' > \mathcal{V}$. We believe that these considerations lead one to abandon \mathcal{V} as a good measure of the visibility, in the multibeam case, and to search for a different definition.

Thus we need multibeam generalizations of the above definitions for the visibility and the predictability. Of course, this is a matter of choice, but it is clear that the choices for the definitions of the two quantities are tied to each other, if they are eventually to satisfy an inequality such as equation (2.8). Indeed, a simple reasoning provides us with a possible answer. One observes that, for any number of beams, it is still true that $\text{Tr } \rho^2 \leq 1$. Upon expanding the trace, one can rewrite this condition as

$$\sum_i \zeta_i^2 + \sum_i \sum_{j \neq i} |I_{ij}|^2 \leq 1. \quad (2.16)$$

One observes now that the first sum depends only on the populations ζ_i of the beams, which should determine the predictability, while the second sum depends only on the non-diagonal elements of ρ , which are those that appear in the expression of the intensity I of the output beam, equation (2.3), and thus determine the features of the interference pattern. Equation (2.16) suggests that we define the generalized visibility V as

$$V^2 = C \sum_i \sum_{j \neq i} |I_{ij}|^2 \quad (2.17)$$

where C is a constant, chosen such that the range of values of V is the interval $[0, 1]$. One finds $C = n/(n-1)$, and so we get

$$V = \sqrt{\frac{n}{n-1} \sum_i \sum_{j \neq i} |I_{ij}|^2} \quad (2.18)$$

which is the choice made in [9]. It is clear that this definition of V satisfies the above requirement, that any interaction with the environment should make V decrease, because, according to equation (2.9), the moduli $|I_{ij}|^2$ can never get larger, as a result of the interaction with the environment. Moreover, we see that for two beams $V = 2|I_{12}|$, which coincides with equation (2.5), and so $V = \mathcal{V}$. It is easy to check that V can be expressed also as a rms average, over all possible values of the phases ϕ_i , of the deviation of the intensity I of the output beam from its mean value:

$$V = \sqrt{\frac{n^3}{n-1} \langle (\Delta I)^2 \rangle_\phi}. \quad (2.19)$$

Here, the bracket $\langle \rangle_\phi$ denotes an average with respect to the phases ϕ_i and $\Delta I = I - \langle I \rangle_\phi$.

One proceeds in a similar manner with the generalized predictability P . Equation (2.16) suggests that we define P as

$$P^2 = A \sum_i \zeta_i^2 + B \quad (2.20)$$

where the constants A and B should be chosen such that the range of values of P^2 coincides with the interval $[0, 1]$. It is easy to convince oneself that this requirement uniquely fixes $A = n/(n-1)$, $B = -1/n$, and so we obtain

$$P = \sqrt{\frac{n}{n-1} \left(-\frac{1}{n} + \sum_i \zeta_i^2 \right)} \quad (2.21)$$

which is the choice of [9]. It is easy to check that this expression coincides with \mathcal{P} , equation (2.6), when $n = 2$. One may observe that this definition enjoys the following nice features:

- (i) P reaches its maximum value if and only if either one of the populations ζ_i is equal to one, and the others are zero, which corresponds to full predictability of the path;

- (ii) P reaches its minimum if and only if all the populations are equal to each other, which means total absence of predictability;
- (iii) P and P^2 are strictly convex functions. This means that for any choice of two sets of populations $\vec{\zeta}' = (\zeta'_1, \dots, \zeta'_n)$, and $\vec{\zeta}'' = (\zeta''_1, \dots, \zeta''_n)$ and for any $\lambda \in [0, 1]$ one has

$$P(\lambda\vec{\zeta}' + (1 - \lambda)\vec{\zeta}'') \leq \lambda P(\vec{\zeta}') + (1 - \lambda)P(\vec{\zeta}'') \tag{2.22}$$

where the equality sign holds if and only if the vectors $\vec{\zeta}'$ and $\vec{\zeta}''$ coincide. A similar equation holds for P^2 . This is an important property, because it means that the predictability (or its square) of any convex combination of states is never larger than the convex sum of the corresponding predictabilities (or their squares).

One can check now that P^2 and V^2 satisfy an inequality analogous to equation (2.8):

$$V^2 + P^2 \leq 1 \tag{2.23}$$

where the equal sign holds if and only if the state is pure. This result deserves a number of comments:

- (1) As in the two-beam case, the above inequality provides a testable relation between measurable quantities, and it would be interesting to verify it.
- (2) On the level of interpretations, when saturated, equation (2.23) can be regarded as a statement of wave–particle duality, in analogy with the two-beam relation, equation (2.8). In fact, since the quantity P depends only on the populations ζ_i , P may be interpreted as a particle-like attribute of the quantons. On the other hand, since the quantity V depends only on the numbers I_{ij} , that determine the interference terms in the expression of I , it is legitimate to regard V as a measure of the wave-like attributes of the quanton.
- (3) However, the quantity P does not carry the same meaning as the quantity \mathcal{P} used in the two-beam case, and the name ‘predictability’ given to it in [9] is not the most appropriate. Indeed, from the point of view of statistical decision theory [12], the natural definition of predictability would not be that in equation (2.21), but rather the following. If one interprets the number ζ_i as the probability for a quanton to be in the beam i , and if one decides to bet every time on the most populated beam \bar{i} , the sum $\sum_{i \neq \bar{i}} \zeta_i$ represents the probability of losing the bet. Then, it is natural to define the predictability \mathcal{P}_n as

$$\mathcal{P}_n = 1 - \frac{n}{n-1} \sum_{i \neq \bar{i}} \zeta_i \tag{2.24}$$

where the normalization is fixed by the requirement that $\mathcal{P}_n = 0$, if the beams are equally populated, and $\mathcal{P}_n = 1$, if any of the populations is equal to one. For $n = 2$, this definition reduces to that used by Greenberger and YaSin, in equation (2.6), and in fact it was proposed as a generalization of it in [13]. It is surely possible to write inequalities involving \mathcal{P}_n and V , but, as far as we know, none of them is saturated by arbitrary pure states, differently from equation (2.23). So, one is faced with a situation in which the less intuitive notion of ‘predictability’, given by equation (2.21), enters a sharp relation with the visibility, while the most intuitive one, given by equation (2.24), enters a relation with the visibility, that is not saturated even for pure states.

3. Higher order inequalities

In a multibeam interferometer a new interesting feature is present, which is absent in the two-beam case, and puts equation (2.8) in a new perspective. In fact, equation (2.23), that relates the populations of the beams ζ_i to the features of the interference fringes, is only the

first of a collection of inequalities that we now discuss. The new inequalities, exactly like equation (2.23), rest on the first principles of quantum mechanics and can be derived along similar lines, by considering higher powers of the density matrix ρ . Indeed, for n beams, one has the following $n - 1$ independent inequalities:

$$\text{Tr } \rho^m \leq 1 \quad m = 2, \dots, n. \quad (3.1)$$

For example, with three beams, if we take $m = 3$ we obtain

$$0 < \sum_i \zeta_i^3 + 3 \sum_i \zeta_i \sum_{j \neq i} |I_{ij}|^2 + 3(I_{12}I_{23}I_{31} + \text{h.c.}) \leq 1. \quad (3.2)$$

This inequality, like equation (2.23), may be translated in terms of physically measurable quantities, although in a more elaborate way. First, we note that the combination of non-diagonal elements of the density matrix, that appears in the last term of the rhs of the above equation represents the third moment of the intensity I of the output beam:

$$(I_{12}I_{23}I_{31} + \text{h.c.}) = \frac{\langle (\Delta I)^3 \rangle_\phi}{\langle I \rangle_\phi^3}. \quad (3.3)$$

On the other hand, the quantities $|I_{ij}|^2$ that appear in the middle terms are related, as in equation (2.5), to the visibilities V_{ij} of the three interference patterns, that are obtained by letting the beams i and j interfere with each other, after intercepting the remaining beam. Therefore, we may rewrite equation (3.2) as

$$0 < \sum_i \zeta_i^3 + \frac{3}{4} \sum_i \zeta_i \sum_{j \neq i} V_{ij}^2 + 3 \frac{\langle (\Delta I)^3 \rangle_\phi}{\langle I \rangle_\phi^3} \leq 1 \quad (3.4)$$

which shows clearly that the novel inequality is a testable relation, to be checked by experiment.

This example illustrates the general structure of the new higher order inequalities. As the number n of beams and the power of m in equation (3.1) increase, higher and higher moments of the intensity I will appear. Furthermore, data related to the interference patterns formed by all possible subsets of beams that can be sorted out of the n beams, will appear.

A few comments are in order. On one hand, the higher order inequalities are similar to equation (2.23), in that they are all testable in principle, and become equalities for beams in a pure state. On the other hand, differently from equation (2.23), they do not exhibit a natural splitting of the particle-like quantities ζ_i from the wave-like quantities I_{ij} into two separate, positive definite terms.

The existence of this sequence of inequalities suggests that, from the point of view of complementarity, the two-beam and the multibeam cases are different. For two beams, the basic properties of the density matrix are completely expressed in terms of a single duality relation, such as equation (2.8). In the multibeam case, a whole sequence of independent inequalities is needed, if one is to fully express the basic properties of the density matrix. Except for the first one, none of these inequalities seems to be related in any simple way to the intuitive concept of wave-particle duality. It seems then that the lowest order inequality, equation (2.23), still carries an idea of wave-particle duality, but only at the cost of averaging out the effects related to higher order moments.

4. Which-way detection

The notion of predictability, introduced in section 2, does not express any real knowledge of the path followed by individual quantons, but at most our *a priori* ability of predicting it. A more interesting situation arises if the experimenter actually tries to gain which-way

information on individual quantons, by letting them interact with a detector placed in the region where the beams are still spatially separated. The analysis proceeds assuming that the detector can also be treated as a quantum system, and that the particle–detector interaction is described by some unitary process. A detector can be considered as a part of the environment, whose state and whose interaction with the beams can, to some extent, be controlled by the experimenter. If we let $|\chi_0\rangle$ be the initial state of the detector (which we assume to be pure, for simplicity), the interaction with the particle will give rise to an entangled density matrix $\rho_{\text{b\&e}}$, of the form considered earlier, in equation (2.9). This time, however, we interpret the states $|\chi_i\rangle$ as n normalized (but not necessarily orthogonal!) states of the which-way detectors. The existence of a correlation between the detector state $|\chi_i\rangle$ and the beam $|\psi_i\rangle$, in equation (2.9), is at the basis of the detector’s ability to store which-way information. We observed earlier that the very interaction of the quantons with the detector causes, as a rule, a decrease in the visibility. According to the intuitive idea of the wave–particle duality, one would like to explain this decrease of the visibility as a consequence of the fact that one is trying to gain which-way information on the quantons. In order to see if this is the case, we need to read out the which-way information stored in the detector. We thus consider the final detector state ρ_D , obtained by taking a trace of equation (2.9) over the particle’s degrees of freedom:

$$\rho_D = \sum_i \zeta_i |\chi_i\rangle\langle\chi_i|. \quad (4.1)$$

As we see, ρ_D is a mixture of the n final states $|\chi_i\rangle$, corresponding to the n possible paths, weighted by the fractions ζ_i of quantons taking the respective paths. Thus the problem of determining the trajectory of the particle reduces to the following one: after the passage of each particle, is there a way to decide in which of the n states $|\chi_i\rangle$ the detector was left? If the states $|\chi_i\rangle$ are orthogonal to each other, the answer is obviously yes. If, however, the states $|\chi_i\rangle$ are not orthogonal to each other, there is no way to unambiguously infer the path: whichever detector observable W one picks, there will be at least one eigenvector of W having a non-zero projection onto more than one state $|\chi_i\rangle$. Therefore, when the corresponding eigenvalue is obtained as the result of a measurement, no unique detector state can be inferred, and only probabilistic judgments can be made. Under such circumstances, the best the experimenter can do is to select the observable that provides as much information as possible *on average*, namely after many repetitions of the experiment. Of course, this presupposes the choice of a definite criterion to measure the average amount of which-way information delivered by a certain observable W .

Let us see in detail how this is done. Consider an observable W , and let Π_μ be the projector onto the subspace of the detector’s Hilbert space \mathcal{H}_D , associated with the eigenvalue w_μ . The *a priori* probability p_μ of getting the result w_μ is

$$p_\mu = \text{Tr}_D(\Pi_\mu \rho_D) = \sum_i \zeta_i P_{i\mu} \quad (4.2)$$

where Tr_D denotes a trace over the detector’s Hilbert space \mathcal{H}_D and $P_{i\mu} = |\langle\chi_i|\Pi_\mu|\chi_i\rangle|^2$. The quantity $\zeta_i P_{i\mu}$ coincides with the probability of getting the value w_μ , when all the beams, except the i th one, are intercepted before reaching the detector, and indeed this provides us with a way to measure the numbers $\zeta_i P_{i\mu}$. When the interferometer is operated with n beams, one may interpret the normalized probability $Q_{i\mu}$:

$$Q_{i\mu} = \frac{\zeta_i P_{i\mu}}{p_\mu} \quad (4.3)$$

as the *a posteriori* relative probability, for a particle to be in the i th beam, provided that the measurement of W gave the outcome w_μ .

On the other hand, if W is measured after the passage of each quanton, one can sort the quantons in the output beam into distinct subensembles, according to the result w_μ of the measurement. The subensembles of quantons are described by density matrices $\rho_{(\mu)}$ of the form

$$\rho_{(\mu)} = \frac{1}{p_\mu} \text{Tr}_D(\Pi_\mu \rho_{b\&c}) := \sum_{ij} \rho_{(\mu)ij} |\psi_i\rangle \langle \psi_j| \quad (4.4)$$

where we defined

$$\rho_{(\mu)ij} = \frac{1}{p_\mu} \langle \chi_j | \Pi_\mu | \chi_i \rangle \rho_{ij}. \quad (4.5)$$

We see that the *a posteriori* probabilities $Q_{i\mu}$ coincide with the diagonal elements of the density matrices $\rho_{(\mu)ij}$, and thus represent also the populations of the beams, for the sorted subensembles of quantons.

Let us now consider the case of two beams. For each outcome w_μ , one can consider the predictability $\mathcal{P}_\mu(W)$ and the visibility $\mathcal{V}_\mu(W)$ associated with the corresponding subensemble of quantons:

$$\mathcal{P}_\mu(W) = |\rho_{(\mu)11} - \rho_{(\mu)22}| = |Q_{1\mu} - Q_{2\mu}| \quad (4.6)$$

$$\mathcal{V}_\mu(W) = 2|\rho_{(\mu)12}|. \quad (4.7)$$

Note that both quantities depend, of course, on the observable W . It is clear that an inequality like equation (2.6) holds for each subensemble separately:

$$\mathcal{P}_\mu^2(W) + \mathcal{V}_\mu^2(W) \leq 1. \quad (4.8)$$

The equality sign holds if and only if the subensemble is a pure state, which is surely the case if the beams and the detector are separately prepared in pure states, before they interact. When the eigenvalue w_μ is observed, it is natural to define the average amount of which-way knowledge delivered by W as the predictability $\mathcal{P}_\mu(W)$ of the corresponding subensemble of quantons. In order to measure the overall ability of the observable W to discriminate the paths, one defines a quantity $\mathcal{K}(W)$,⁴ which is some average of the partial predictabilities $\mathcal{P}_\mu(W)$. The procedure implicitly adopted by Englert in [6] is to define $\mathcal{K}(W)$ as the weighted average of the numbers $\mathcal{P}_\mu(W)$, with weights provided by the *a priori* probabilities p_μ :

$$\mathcal{K}(W) = \sum_{\mu} p_\mu \mathcal{P}_\mu(W). \quad (4.9)$$

One can also introduce the ‘erasure visibility’ [14], relative to W , as the weighted average of the partial visibilities:

$$\mathcal{V}(W) = \sum_{\mu} p_\mu \mathcal{V}_\mu(W). \quad (4.10)$$

For any W , these quantities can be shown to satisfy the following inequality, that is a direct consequence of equation (4.8):

$$\mathcal{K}^2(W) + \mathcal{V}^2(W) \leq 1. \quad (4.11)$$

Moreover, one can prove that

$$\mathcal{P}^2 \leq \mathcal{K}^2(W) \quad (4.12)$$

⁴ Indeed, Englert considers the *likelihood* \mathcal{L}_W for guessing the way right. In our notation, $\mathcal{L}_W = (1 + \mathcal{K}(W))/2$.

which gives expression to the intuitive idea that any observable W that we decide to measure provides us with a better knowledge of the path than that available on the basis of a mere *a priori* judgement. One also has the other inequality

$$\mathcal{V}^2 \leq \mathcal{V}^2(W). \tag{4.13}$$

For the proofs of these inequalities, we refer the reader to [1], where they are derived in a number of independent ways. In the so-called which-way sorting schemes, it is natural to select the observable W so as to maximize $\mathcal{K}(W)$, and one then defines the distinguishability \mathcal{D} of the paths as the maximum value of $\mathcal{K}(W)$:

$$\mathcal{D} = \max_W \{\mathcal{K}(W)\}. \tag{4.14}$$

It is easy to see that equations (4.11), (4.13) and (4.14) together imply the following inequality, analogous to equation (2.8), first derived by Englert in [6]:

$$\mathcal{D}^2 + \mathcal{V}^2 \leq 1. \tag{4.15}$$

Thus, given the visibility \mathcal{V} , there is an upper bound for the distinguishability, set by the above relation. But Englert in fact proves much more than this: he shows that equation (4.15) becomes an *identity* when both the beams and the detector are in a pure state. In our opinion, this fact is essential to justify the interpretation of equation (4.15) as a statement of the complementary character of the wave and particle attributes of a quanton. In fact, this implies that when the beam of quantons and the detector are as noiseless as they can possibly be in quantum mechanics, namely when they are in pure states, an increase in any of the two terms is necessarily accompanied by an exactly quantifiable corresponding decrease of the other.

A possible generalization of the above considerations to the multibeam case is as follows [9]. One again sorts the quantons into subensembles, depending on the outcome of the measurement of W . For each outcome w_μ , one uses the generalized predictability P in equation (2.21) and the generalized visibility V in equation (2.18) to define the ‘conditioned which-way knowledge’ $K_\mu(W)$:

$$K_\mu(W) = \sqrt{\frac{n}{n-1} \left(-\frac{1}{n} + \sum_i Q_{i\mu}^2 \right)} \tag{4.16}$$

and the ‘partial erasure visibility’ $V_\mu(W)$:

$$V_\mu(W) = \sqrt{\frac{n}{n-1} \sum_i \sum_{j \neq i} |\rho_{(\mu)ij}|^2}. \tag{4.17}$$

In view of equation (2.23), they satisfy an inequality analogous to equation (4.8):

$$K_\mu^2(W) + V_\mu^2(W) \leq 1. \tag{4.18}$$

Again, as in the two-beam case, the equality sign holds if the subensembles are pure. The author of [9] now considers two different definitions for the ‘which-way knowledge’ and the ‘erasure visibility’, associated with W , as a whole. The first one is closer to equation (4.9):

$$K(W) := \sum_\mu p_\mu K_\mu(W) \quad V(W) := \sum_\mu p_\mu V_\mu(W). \tag{4.19}$$

The second one, inspired by the work of Brukner and Zeilinger [15], is⁵:

$$\tilde{K}^2(W) := \sum_\mu p_\mu K_\mu^2(W) \quad \tilde{V}^2(W) := \sum_\mu p_\mu V_\mu^2(W). \tag{4.20}$$

⁵ We use here a notation different from that of [9]. Our $\tilde{K}^2(W)$ and $\tilde{V}^2(W)$ correspond, respectively, to $n/(n-1)I_{KW}$ and $n/(n-1)I_{VW}$, in [9].

The quantities introduced above are related by the following chains of inequalities, the proofs of which can be found in [9]:

$$V \leq V(W) \leq \tilde{V}(W) \quad P \leq K(W) \leq \tilde{K}(W). \quad (4.21)$$

These inequalities show that $\tilde{K}(W)$ and $\tilde{V}(W)$ provide more efficient measures for the average which-way information and for the erasure visibility, respectively. However, the author of [9] observes that the quantities $K(W)$ and $V(W)$ are preferable to $\tilde{K}(W)$, and $\tilde{V}(W)$, respectively, because they are the ones that reduce, for $n = 2$, to the definitions used in the two-beam case. We would like to point out that, since $K_\mu(W)$ and $V_\mu(W)$ are essentially variances of the diagonal and non-diagonal elements, respectively, of the density matrices for the subensembles of quantons, it appears more natural, from a statistical point of view, to combine them in quadrature, as done in equation (4.20). This suggests that one should adopt the definition with the quadrature also in the two-beam case.

By taking the suprema of all the quantities defined above, over all possible observables W , one can define a set of four quantities that characterize the state ρ of the beams. For example, upon taking the maxima of $K(W)$ and $\tilde{K}(W)$, we end up with two possible definitions for the which-way distinguishability, D and \tilde{D} , respectively:

$$D = \max_W \{K(W)\} \quad \tilde{D} = \max_W \{\tilde{K}(W)\}. \quad (4.22)$$

Similarly, by taking the suprema of $V(W)$ and $\tilde{V}(W)$, we obtain two definitions of the so-called ‘coherence’ of the beams [16]:

$$C = \sup_W \{V(W)\} \quad \tilde{C} = \sup_W \{\tilde{V}(W)\}. \quad (4.23)$$

(The reader may find in [1] an explanation of why one has maxima in the definition of distinguishability and only suprema in that of coherence.)

The quantities introduced above satisfy a set of inequalities that all follow from the chains of inequalities of equations (4.21), and from the following inequality, that can be obtained from equation (4.18), on averaging over all possible outcomes w_μ :

$$\tilde{K}^2(W) + \tilde{V}^2(W) \leq 1. \quad (4.24)$$

It is clear that this inequality is saturated, regardless of the observable W , when the state of the combined detector–beam system is pure. This is an immediate consequence of equation (4.18).

One of the central results of [9] is the following inequality, generalizing equation (4.15):

$$\tilde{D}^2 + V^2 \leq 1. \quad (4.25)$$

Since $\tilde{D} \geq D$, this also implies

$$D^2 + V^2 \leq 1. \quad (4.26)$$

Thus we see that also in the multibeam case, the visibility V sets an upper limit for the amount of which-way information irrespective of how one measures it, via D or \tilde{D} . In [9] it is suggested that the above two inequalities provide multibeam generalizations of the two-beam wave–particle duality relation equation (4.15).

Even if equations (4.26) and (4.25) represent correct inequalities that can be tested in an experiment, in our opinion, their interpretation as an expression of wave–particle duality appears disputable. The root of the problem is that the above inequalities, differently from the two-beam case, cannot be saturated, in general, even if the beams and the detector are prepared in pure states (in appendix A, we actually prove that equation (4.26), for example, can be saturated only if $D = P$, which means that the detector does not provide any information).

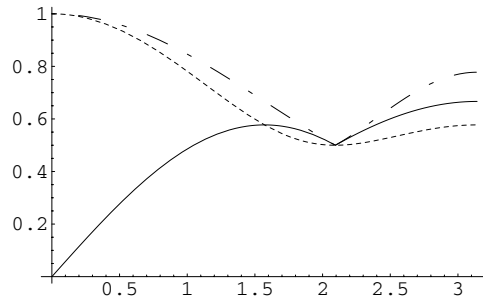


Figure 1. Plots of the quantities D (solid line), V (dotted line) and $D^2 + V^2$ (dashed line), as functions of θ in a three-beam situation.

Therefore, one may conceive the possibility of designing two which-way detectors D_1 and D_2 such that $V_1 > V_2$, while, at the same time, $D_1 > D_2$. This possibility, which conflicts with the intuitive idea of complementarity, actually occurs, as we anticipated in [10], and as we report in the next section.

5. A three-beam example

The example discussed in [10], was based on a three-beam interferometer with equally populated beams, described by the pure state:

$$\rho = \frac{1}{3} \sum_{ij} |\psi_i\rangle\langle\psi_j|. \tag{5.1}$$

For the sake of simplicity, it was assumed there that the detector’s Hilbert space was a two-dimensional space \mathcal{H}_2 . Its rays were described via the Bloch parametrization, such that

$$\frac{1 + \hat{n} \cdot \vec{\sigma}}{2} = |\chi\rangle\langle\chi| \tag{5.2}$$

where \hat{n} is a unit three-vector and $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is any representation of the Pauli matrices. We denoted by $|\hat{n}\rangle\langle\hat{n}|$ the ray corresponding to the vector \hat{n} . We required that the directions $\hat{n}_+, \hat{n}_-, \hat{n}_0$ associated with the states $|\chi_i\rangle$ were coplanar, and such that \hat{n}_+ and \hat{n}_- both formed an angle θ with \hat{n}_0 . We imagined that θ could be varied at will, by acting on the detector, and in [10] we obtained the following expressions for the visibility V and the distinguishability D , as functions of θ :

$$V(\theta) = \sqrt{\frac{1 + \cos \theta + \cos^2 \theta}{3}} \tag{5.3}$$

$$D(\theta) = \frac{1}{\sqrt{3}} \sin \theta \quad \text{for } 0 \leq \theta \leq 2/3\pi \tag{5.4}$$

$$D(\theta) = \frac{2}{3} \sin^2 \left(\frac{\theta}{2} \right) \quad \text{for } 2/3\pi \leq \theta \leq \pi. \tag{5.5}$$

The values of V and D are plotted in figure 1. By looking at it, one realizes that something unexpected happens: while in the interval $0 \leq \theta < \pi/2$, V decreases and D increases, as expected from the wave–particle duality, we see that in the interval $\pi/2 \leq \theta \leq \pi$, V and D decrease and increase simultaneously! We see that if we pick two values θ_1 and θ_2 in this

region, we obtain two which-way detectors that precisely realize the situation described at the end of the previous section.

The analysis of [10], that we have summarized here, is not realistic though, because of the simplifying assumption of a detector with a two-dimensional Hilbert space of states. Even assuming that the detector's final states $|\chi_i\rangle$ span a two-dimensional subspace \mathcal{H}_2 , still one has to take into account that the full Hilbert space \mathcal{H}_D of a realistic device is infinite dimensional. Now, it is known from the theory of quantum detection [12, 17] that the optimum discrimination among an assigned set of quantum states is not always achieved by an observable that leaves invariant the subspace spanned by them. However, the value of D quoted above corresponds to maximizing the which-way knowledge over the restricted set of detector observables W , that leave the subspace \mathcal{H}_2 invariant. Then, in order to complete the proof, we need to show that no observable in \mathcal{H}_D can perform better than the one determined in [10], by considering only operators that live in \mathcal{H}_2 . Filling this gap is by no means an easy job, because it is a matter of solving an optimization problem in an infinite-dimensional Hilbert space. There is no general strategy for solving this sort of problem, and we can rely only on a few known general results [12, 17, 18]. The interested reader can find the lengthy procedure to compute D in appendix B. Here, we content ourselves with sketching the method followed, and presenting the results.

For the sake of definiteness, let us agree to use $K(W)$ as our measure of the which-way information. At the end of this section, we shall discuss what changes if one instead uses $\tilde{K}(W)$. The determination of the optimal observable W_{opt} is facilitated by the observation that, even when \mathcal{H}_D is infinite dimensional, the problem can be formulated entirely in the subspace \mathcal{H}_2 , as we now explain. One observes that the probabilities $P_{i\mu}$ that enter the definition of $K(W)$ can also be written as

$$P_{i\mu} = \langle \chi_i | \Pi_\mu | \chi_i \rangle = \langle \chi_i | \Pi \Pi_\mu \Pi | \chi_i \rangle \equiv \langle \chi_i | A_\mu | \chi_i \rangle \quad (5.6)$$

where Π is the orthogonal projector onto \mathcal{H}_2 , and $A_\mu = \Pi \Pi_\mu \Pi$ is a positive (Hermitian) operator on the subspace \mathcal{H}_2 . Thus we see that the operators A_μ contain all the information we need about W in order to compute the which-way knowledge. It is to be noted that A_μ are not projection operators, in general. However, they must provide a decomposition of the identity onto \mathcal{H}_2 , since

$$\sum_\mu A_\mu = \sum_\mu \Pi \Pi_\mu \Pi = \Pi \left(\sum_\mu \Pi_\mu \right) \Pi = \Pi. \quad (5.7)$$

Such a collection of operators on \mathcal{H}_2 provides an example of what is known in mathematics as a positive operator valued measure (POVM in short). Note though that, while any Hermitian operator in \mathcal{H}_D gives rise, by projection, to a POVM in \mathcal{H}_2 , the converse may not be true⁶. Our strategy to determine W_{opt} is then to search first for the optimal POVM A_{opt} in \mathcal{H}_2 (the notion of which-way knowledge is obviously defined for an arbitrary POVM, as well), and to check at the end if A_{opt} can be realized by projecting onto \mathcal{H}_2 an operator W in \mathcal{H}_D , as in equation (5.6). If this is the case, W is guaranteed to be optimal, and we can say that $D = K(A_{\text{opt}})$. The determination of A_{opt} is facilitated by a general theorem [18], that states that for any measure of the which-way knowledge that is a weighted average of a convex function, the optimal POVM consists of rank-one operators. This is the case for the which-way knowledge K , which is a weighted average of the predictability P , which indeed is a convex

⁶ In effect, this problem arises only if \mathcal{H}_D is finite dimensional. If \mathcal{H}_D is infinite dimensional, all POVMs are acceptable, because a general theorem due to Neumark [19] ensures that all POVMs of any Hilbert space can be realized as projections of self-adjoint operators from a larger Hilbert space.

function. The A_μ being rank-one operators, we are assured that there exist non-negative numbers $2\alpha_\mu \leq 1$ and unit vectors \hat{m}_μ such that

$$A_\mu = 2\alpha_\mu |\hat{m}_\mu\rangle\langle\hat{m}_\mu| = \alpha_\mu (1 + \hat{m}_\mu \cdot \vec{\sigma}). \quad (5.8)$$

The condition for a POVM, equation (5.7), is equivalent to the following conditions for the numbers α_μ and the vectors \hat{m}_μ :

$$\sum_\mu \alpha_\mu = 1 \quad \sum_\mu \alpha_\mu \hat{m}_\mu = 0. \quad (5.9)$$

The interested reader may find in appendix B how the optimal POVM can be determined. Here we just report the result: for all values of θ , A_{opt} turns out to have only two non-vanishing elements, A_\pm , such that

$$A_\pm = \frac{1 \pm \sigma_x}{2} \quad \text{for } 0 \leq \theta < 2\pi/3 \quad (5.10)$$

$$A_\pm = \frac{1 \pm \sigma_z}{2} \quad \text{for } 2\pi/3 < \theta \leq \pi. \quad (5.11)$$

It is clear that the operators A_\pm coincide with the projectors found in [10], showing that it was indeed sufficient to carry out the optimization procedure in \mathcal{H}_2 .

It should be appreciated that this coincidence is by no means trivial, and strictly depends on the choice of $K(W)$ as a measure of which-way knowledge. For example, for $\theta = 2\pi/3$, it is known [12, 17], that, with either Shannon's entropy or Bayes' cost function as measures of information, the optimal POVM actually consists of three elements, and thus it is not associated with an operator in \mathcal{H}_2 .

Then, our observation that the inequality equation (2.23) fails to carry the physical picture associated with the idea of complementarity is now fully demonstrated. We have checked that a similar conclusion can be drawn if, rather than K , one uses the alternative definition of distinguishability \bar{D} provided by equation (4.22). In fact, it turns out that the optimal POVM for \bar{K} coincides with the one found earlier, in the interval $0 \leq \theta < 2/3\pi$, and so $\bar{D} = D$. The proof of this can be found in appendix B.

6. Conclusions

The intuitive concept of complementarity has found, in the case of two-beam interference experiments, a satisfactory, fully quantum mechanical formulation as interferometric duality. In this paper, we critically analysed the difficulties encountered in the attempt of generalizing this concept to multibeam experiments, and discussed the shortcomings that are present, in our opinion, in recent proposals. It seems to us fair to say that interferometric duality has not yet found a proper formulation in the multibeam case. To justify this conclusion, let us recall the different points we have elaborated in the paper.

In the two-beam case, general quantum mechanical requirements on the density matrix imply the Greenberger–YaSin inequality, that, when saturated, expresses interferometric duality. This inequality has been generalized to the multibeam case [9], leading to a formal definition of interferometric duality for more than two beams. The price paid is that the corresponding generalized concept of predictability has lost the intuitive connection with minimizing the error in guessing the way right. The traditional concept of predictability may enter, together with the generalized visibility, in an inequality that is not saturated, and then cannot convey the idea of complementarity, which requires that better visibility is necessarily related to a loss in information.

We have shown that general requirements of quantum mechanics imply new inequalities, that are not present in the two-beam case. These inequalities are again experimentally testable. They deserve further study but, at present, they do not seem to exhibit a direct relation with the idea of complementarity.

Interferometric duality may be fully analysed only in the presence of which-way detectors. In the two-beam case, Englert has shown that the visibility enters, with the distinguishability, into an inequality that is saturated for pure states. As maximizing the distinguishability minimizes the error in guessing the way right by performing a measurement, this relation fully expresses interferometric duality. In deriving an analogous inequality for the multibeam case, Dürr has introduced two alternative notions of distinguishability. However, we have shown that this inequality is never saturated, apart from trivial cases. Then, a pure inequality may be consistent with a situation in which an increase (decrease) in visibility goes together with an increase (decrease) in distinguishability, in contrast to the intuitive idea of interferometric duality. We have given a full proof that this possibility actually occurs in a realistic example. The inequalities proposed by Dürr, in terms of generalized visibility and distinguishability, are then correct quantum mechanical relations, testable in principle, but they fail to convey the idea of interferometric duality.

It then seems fair to conclude that interference duality in multibeam experiments has not yet been properly formulated. We leave the problem open, but we note it is by no means necessary that quantum mechanics should provide us with an exact formulation of this concept in the multibeam case. Maybe one should content oneself with its formulation in the two-beam case, where the semiclassical intuitive idea of complementarity was first introduced. Maybe quantum mechanics just provides us with the values of observable quantities and experimentally testable inequalities. The analysis we have performed may hint in this direction, but further investigation is required.

Acknowledgments

We gratefully acknowledge an interesting discussion with Professor G Marmo and Professor E C G Sudarshan. The work of GB was partially supported by the PRIN *S.Inte.Si*. The work of RM was partially supported by EC program HPRN-EC-2000-00131, and by the PRIN '*Teoria dei campi, superstringhe e gravità*'.

Appendix A

In this appendix, we prove the following result: for any number $n > 2$ of beams in a pure state ρ , and any detector in a pure initial state, the inequality equation (4.26) is satisfied as an equality if and only if $D = P$, namely when the detector provides no information at all. The proof consists in showing that the equal sign in equation (4.26) holds only if the detector states $|\chi_i\rangle$ are proportional to each other, which obviously implies $D = P$. Consider the optimal operator W_{opt} such that $K(W_{\text{opt}}) = D$ (we assume that such an operator exists), and let $V(W_{\text{opt}})$ be the corresponding erasure visibility. It follows then from equations (4.21) and equations (4.24) that

$$D^2 + V^2 \leq D^2 + V^2(W_{\text{opt}}) = K^2(W_{\text{opt}}) + V^2(W_{\text{opt}}) \leq \tilde{K}^2(W_{\text{opt}}) + \tilde{V}^2(W_{\text{opt}}) = 1. \quad (\text{A.1})$$

We see that a necessary condition to have $D^2 + V^2 = 1$ is that

$$V = V(W_{\text{opt}}). \quad (\text{A.2})$$

In what follows, we shall not consider the trivial case $V = 0$, and we shall suppose that $V > 0$. In order to study equation (A.2), we take advantage of the fact that, $K(W)$ being convex, the

spectrum of W_{opt} can be taken to be non-degenerate [18]. If we let $|w_\mu\rangle$ be the eigenvectors of W_{opt} , with non-vanishing projection onto some of the states $|\chi_i\rangle$, by using the expressions of equation (4.17) for the partial visibilities, we can write

$$\begin{aligned} \frac{n-1}{n}V^2(W_{\text{opt}}) &= \frac{n-1}{n} \sum_{\mu} \sum_{\nu} p_{\mu} p_{\nu} V_{\mu} V_{\nu} \\ &= \sum_{\mu} \sum_{\nu} \sqrt{\sum_i \sum_{j \neq i} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle|^2} \sqrt{\sum_p \sum_{q \neq p} |\langle w_{\nu} | \tilde{\rho}_{pq} | w_{\nu} \rangle|^2} \end{aligned} \tag{A.3}$$

where

$$\tilde{\rho}_{ij} := \rho_{ij} |\chi_i\rangle \langle \chi_j|. \tag{A.4}$$

Now, the Cauchy–Schwarz inequality for real vectors implies that

$$\sqrt{\sum_i \sum_{j \neq i} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle|^2} \sqrt{\sum_p \sum_{q \neq p} |\langle w_{\nu} | \tilde{\rho}_{pq} | w_{\nu} \rangle|^2} \geq \sum_i \sum_{j \neq i} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle| \cdot |\langle w_{\nu} | \tilde{\rho}_{ij} | w_{\nu} \rangle|. \tag{A.5}$$

Upon using this relation in equation (A.3), we obtain

$$\begin{aligned} \frac{n-1}{n}V^2(W_{\text{opt}}) &\geq \sum_{\mu\nu} \sum_i \sum_{j \neq i} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle| \cdot |\langle w_{\nu} | \tilde{\rho}_{ij} | w_{\nu} \rangle| \\ &= \sum_i \sum_{j \neq i} \left(\sum_{\mu} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle| \right)^2. \end{aligned} \tag{A.6}$$

Obviously

$$\sum_{\mu} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle| \geq \left| \sum_{\mu} \langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle \right|. \tag{A.7}$$

Then, equation (A.6) becomes

$$\frac{n-1}{n}V^2(W_{\text{opt}}) \geq \sum_i \sum_{j \neq i} \left| \sum_{\mu} \langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle \right|^2 = \sum_i \sum_{j \neq i} |\text{Tr}_D(\tilde{\rho}_{ij})|^2 = \frac{n-1}{n}V^2. \tag{A.8}$$

Clearly, $V^2(W_{\text{opt}})$ becomes equal to V if and only if all the inequalities involved in the derivation of equation (A.8) become equalities. Note that the case $n = 2$ is special, for then the Cauchy–Schwarz inequalities of equation (A.5) are necessarily equalities, because the sums in equation (A.5) contain just one term. However, for $n > 2$, we have the equal sign if and only if there exist positive constants c_{μ} such that

$$c_{\mu} |\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle| = c_{\nu} |\langle w_{\nu} | \tilde{\rho}_{ij} | w_{\nu} \rangle| \quad \forall i \neq j. \tag{A.9}$$

Since $\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle = \langle w_{\mu} | \chi_i \rangle \langle \chi_j | w_{\mu} \rangle \rho_{ij}$, and we assume $\rho_{ij} \neq 0$, the above condition is equivalent to

$$c_{\mu} |\langle w_{\mu} | \chi_i \rangle \langle \chi_j | w_{\mu} \rangle| = c_{\nu} |\langle w_{\nu} | \chi_i \rangle \langle \chi_j | w_{\nu} \rangle| \quad \forall i \neq j. \tag{A.10}$$

On the other hand, the set of inequalities equation (A.7) become equalities if and only if, for all $j \neq i$, the phases of the complex numbers $\langle w_{\mu} | \tilde{\rho}_{ij} | w_{\mu} \rangle$, and then of the numbers $\langle w_{\mu} | \chi_i \rangle \langle \chi_j | w_{\mu} \rangle$, do not depend on μ :

$$\arg(\langle w_{\mu} | \chi_i \rangle \langle \chi_j | w_{\mu} \rangle) = \theta_{ij}. \tag{A.11}$$

Now, for $n > 2$ and $V > 0$, equation (A.10) implies that the matrix elements $\langle w_\mu | \chi_i \rangle$ are all different from zero. To see this, we separate the states $|\chi_i\rangle$ into two subsets, A and B . A contains the detector states which are orthogonal to some of the eigenstates $|w_\mu\rangle$. B contains the remaining states. We can prove that, for $V > 0$, A must be empty. This is done in two steps: first we prove that if A contains some detector states, then it contains all of them. In the second step, we show that the elements of A are orthogonal to each other. By combining the two facts, it follows that A must be empty, because otherwise all detector states would be orthogonal to each other, and then, by taking a W that has the detector states as eigenvectors, we would achieve $D = 1$ and $V(W_{\text{opt}}) = 0$, which is not possible, because we assumed that $V > 0$. So, let us show first that if A contains some detector states, it contains all. In fact, let $|\chi_1\rangle$ be one of its elements. Then there exists a value of μ , say $\mu = 2$, such that $\langle w_2 | \chi_1 \rangle = 0$. On the other hand, since the vectors $|w_\mu\rangle$ form a basis for the vectors $|\chi_i\rangle$, there must be some eigenvector, say $|w_1\rangle$, such that $\langle w_1 | \chi_1 \rangle \neq 0$. Suppose now that B contains an element, say $|\chi_n\rangle$, and consider equation (A.10), for $i = 1, j = n, \mu = 2$ and $\nu = 1$: $c_2 |\langle w_2 | \chi_1 \rangle \langle \chi_n | w_2 \rangle| = c_n |\langle w_1 | \chi_1 \rangle \langle \chi_n | w_1 \rangle|$. It is clear that the lhs vanishes, while the rhs does not. It follows that there cannot be such a $|\chi_n\rangle$. Then, if A contains just one detector state, it contains all.

Now we can turn to the second step. In order to prove that all elements of A are orthogonal to each other, consider for example equation (A.10) for $\mu = 2$ and $i = 1$: they imply that, for any $j \neq 1$ and any ν , the numbers $|\langle w_\nu | \chi_1 \rangle \langle \chi_j | w_\nu \rangle|$ must vanish. But this implies $\langle w_\nu | \chi_1 \rangle \langle \chi_j | w_\nu \rangle = 0$. Summing over all values of ν , we obtain

$$0 = \sum_\nu \langle w_\nu | \chi_1 \rangle \langle \chi_j | w_\nu \rangle = \langle \chi_j | \chi_1 \rangle. \quad (\text{A.12})$$

So, $|\chi_1\rangle$ is orthogonal to all other detector states $|\chi_i\rangle$. The same reasoning applies to all elements of A , and thus we conclude that all detector states are orthogonal to each other.

Having proved that all matrix elements $\langle \chi_i | w_\mu \rangle$ are different from zero, we can now show that the detector's states $|\chi_i\rangle$ are indeed proportional to each other. Since $n > 2$, for any $i \neq j$, we can find a k distinct from both i and j . Consider now equation (A.10) for the couples i, k and j, k , and divide the first by the second. This is legitimate, because all inner products $\langle w_\mu | \chi_i \rangle$ are different from zero. We get

$$\frac{|\langle w_\mu | \chi_i \rangle|}{|\langle w_\mu | \chi_j \rangle|} = \frac{|\langle w_\nu | \chi_i \rangle|}{|\langle w_\nu | \chi_j \rangle|} \quad \forall i \neq j. \quad (\text{A.13})$$

This is the same as

$$\frac{|\langle w_\mu | \chi_i \rangle|}{|\langle w_\nu | \chi_i \rangle|} = \frac{|\langle w_\mu | \chi_j \rangle|}{|\langle w_\nu | \chi_j \rangle|} \quad \forall i \neq j. \quad (\text{A.14})$$

Since $\sum_\mu |\langle w_\mu | \chi_i \rangle|^2 = 1$ for all i , it is easy to verify that the above equations imply

$$|\langle w_\mu | \chi_i \rangle| = |\langle w_\mu | \chi_j \rangle|. \quad (\text{A.15})$$

To proceed, we make use now of equation (A.11). If we set $\alpha_{\mu i} = \arg \langle w_\mu | \chi_i \rangle$, equation (A.11) implies

$$\alpha_{\mu i} - \alpha_{\mu j} = \theta_{ij} \quad (\text{A.16})$$

which obviously means that, for fixed i and j and variable μ , the phases of the complex numbers $\langle w_\mu | \chi_i \rangle$ and $\langle w_\mu | \chi_j \rangle$ differ by the overall phase θ_{ij} , and this implies:

$$|\chi_i\rangle = e^{i\theta_{ij}} |\chi_j\rangle. \quad (\text{A.17})$$

Since all detector states differ by a phase, it obviously follows that the detector provides no information at all, and thus $D = P$.

Appendix B

In this appendix, we determine the rank-one POVM that maximizes the which-way knowledge, for the three-beam interferometer considered in section 5. The procedure is different, depending on whether we choose to measure the which-way knowledge by means of K or \tilde{K} . We consider first K , because it is the simplest case. We can prove then that, for any number of beams with equal populations ζ_i and any choice of the detector states $|\chi_i\rangle$ in \mathcal{H}_2 , the POVM A that maximizes K can be taken to have only two non-vanishing elements, $A = \{A_1, A_2\}$. The proof is as follows. First, we note that for any rank-one POVM consisting of only two elements, the conditions for a POVM, equation (5.9), imply

$$\alpha_1 = \alpha_2 = \frac{1}{2} \quad \hat{m}_1 + \hat{m}_2 = 0. \tag{B.1}$$

Thus, all rank-one POVM with two elements are characterized by a pair of unit vectors \hat{m}_μ , that are opposite to each other. Such a POVM clearly coincides with the projector valued measure (PVM) associated with the Hermitian operator $\hat{m}_1 \cdot \vec{\sigma}$ in \mathcal{H}_2 . We let A be the optimal PVM, that can be obtained by considering all possible directions for \hat{m}_1 . We can show that such an A represents the optimal POVM. To see this, we prove that the which way-knowledge $K(A)$ delivered by A is not less than that delivered by any other POVM C . By virtue of the theorem proved in [18], it is sufficient to consider POVMs C made of rank-one operators. In order to evaluate $K(C)$, it is convenient to rewrite the quantities $p_\mu K_\mu$, for any element $C_\mu = 2\alpha_\mu^{(C)}(1 + \hat{m}_\mu^{(C)} \cdot \vec{\sigma})$ of C , as

$$\begin{aligned} p_\mu K_\mu &= \left[\frac{n}{n-1} \left(-\frac{p_\mu^2}{n} + \sum_{i=1}^n \zeta_i^2 P_{i\mu}^2 \right) \right]^{1/2} \\ &= \alpha_\mu^{(C)} \sqrt{\frac{n}{n-1}} \left\{ -\frac{1}{n} \left[1 + \left(\hat{m}_\mu^{(C)} \cdot \sum_i \zeta_i \hat{n}_i \right)^2 \right] \right. \\ &\quad \left. + \sum_i \zeta_i^2 \left[1 + \left(\hat{m}_\mu^{(C)} \cdot \hat{n}_i \right)^2 \right] + 2\hat{m}_\mu^{(C)} \cdot \sum_i \zeta_i \left(\zeta_i - \frac{1}{n} \right) \hat{n}_i \right\}^{1/2}. \end{aligned} \tag{B.2}$$

We observe now that, for equally populated beams, $\zeta_i = 1/n$, the last sum in the above equation vanishes, and the expression for $p_\mu K_\mu$ becomes invariant under the exchange of $\hat{m}_\mu^{(C)}$ with $-\hat{m}_\mu^{(C)}$. Consider now the POVM B , such that

$$B_\mu^+ = \frac{1}{2}C_\mu \quad B_\mu^- = \frac{1}{2}\alpha_\mu^{(C)}(1 - \hat{m}_\mu^{(C)} \cdot \vec{\sigma}) \tag{B.3}$$

Of course, $p_\mu^{(+)} K_\mu^{(+)} = p_\mu K_\mu/2$, while the invariance of $p_\mu K_\mu$ implies $p_\mu^{(-)} K_\mu^{(-)} = p_\mu^{(+)} K_\mu^{(+)}$. It follows that the average information for B and C are equal, $K(B) = K(C)$. Now, for each value of μ , the pair of operators $B_\mu^\pm/\alpha_\mu^{(C)} = (1 \pm \hat{m}_\mu^{(C)} \cdot \vec{\sigma})/2$ constitutes by itself a POVM, with two elements. Thus, the POVM C can be regarded as a collection of POVMs with two elements, each taken with a non-negative weight $\alpha_\mu^{(C)}$. But then $K(C)$, being equal to the average of the amount of information provided by a number of POVM with two elements, cannot be larger than the amount of information $K(A)$ delivered by the best POVM with two elements. Thus we have shown that $K(C) = K(B) \leq K(A)$, which shows that A is the optimal POVM.

It remains to find A for the example considered in section 5, but this is easy. If we let β and γ be the polar angles that identify the vector \hat{m}_1 , one finds for the square of the which-way information the following expression:

$$K^2 = \frac{4}{9} \left[\cos^2 \beta \sin^2 \left(\frac{\theta}{2} \right) + 3 \sin^2 \beta \cos^2 \gamma \cos^2 \left(\frac{\theta}{2} \right) \right] \sin^2 \left(\frac{\theta}{2} \right). \tag{B.4}$$

For all values of θ , the which-way information is maximum if $\cos \gamma = \pm 1$, i.e. if the vector \hat{m}_1 lies in the same plane as the vectors \hat{n}_i . As for the optimal value of β , it depends on θ . For $0 \leq \theta < 2\pi/3$, the best choice is $\beta = \pm\pi/2$, and one gets the PVM in equation (5.10), which gives the path distinguishability D given in equation (5.4). For larger values of θ , one has $\beta = 0$ and then the optimal PVM is that of equation (5.11), with D given by equation (5.5).

We turn now to the case when the which-way information is measured by means of \tilde{K} . Since the square of the predictability is a convex function, we are ensured by the general theorem proved in [18] that the optimal POVM is made of rank-one operators, of the form (5.8). We split the computation of the optimal POVM into two steps. First, we prove a lemma, which actually holds for any measure of the which-way information F , which is a weighted average of a convex function of the *a posteriori* probabilities $Q_{i\mu}$.

Lemma. *Consider an interferometer with n beams and arbitrary populations ζ_i . Let the detector states $|\chi_i\rangle$ be in \mathcal{H}_2 , and have coplanar vectors \hat{n}_i . Then, the optimal POVM is necessarily such that all the vectors \hat{m}_μ in equation (5.8) lie in the same plane containing the vectors \hat{n}_i .*

The proof of the lemma is as follows. Let B be an optimal POVM. Suppose that some of the vectors $\hat{m}_\mu^{(B)}$ do not belong to the plane containing the vectors \hat{n}_i , which we assume to be the xz plane. We show below how to construct a new POVM A providing no less information than B , and such that the vectors $\hat{m}_\mu^{(A)}$ all belong to the xz plane. The first step in the construction of A consists in symmetrizing B with respect to the xz plane. The symmetrization is done by replacing each element B_μ of B , not lying in the xz plane, by the pair (B'_μ, B''_μ) , where $B'_\mu = B_\mu/2$, and B''_μ has the same weight α_μ as B'_μ , while its vector $\hat{m}_\mu^{(B)''}$ is the symmetric of $\hat{m}_\mu^{(B)}$ with respect to the xz plane. It is easy to verify that the symmetrization preserves the conditions for a POVM (equations (5.9)). Since all the vectors \hat{n}_i belong by assumption to the xz plane, the which-way knowledge actually depends only on the projections of the vectors $\hat{m}_\mu^{(B)}$ in the plane xz . This implies, as it is easy to check, that symmetrization with respect to the xz plane does not change the amount of which-way knowledge. We assume therefore that B has been preliminarily symmetrized in this way. Now we show that we can replace, one after the other, each pair of symmetric elements (B'_μ, B''_μ) by another pair of operators, whose vectors lie in the xz plane, without reducing the information provided by the POVM. Consider for example the pair (B'_κ, B''_κ) . We construct the unique pair of unit vectors \hat{u}_κ and \hat{v}_κ , lying in the xz plane, and such that

$$\hat{u}_\kappa + \hat{v}_\kappa = 2(m_\kappa^{(B)x}\hat{i} + m_\kappa^{(B)z}\hat{k}) \quad (\text{B.5})$$

where \hat{i} and \hat{j} are the directions of the x and z axes, respectively. Note that $\hat{u}_\kappa \neq \hat{v}_\kappa$. Consider now the collection of operators obtained by replacing the pair (B'_κ, B''_κ) with the pair (A'_κ, A''_κ) such that

$$A'_\kappa = \alpha_\kappa^{(B)}(1 + \hat{u}_\kappa \cdot \vec{\sigma}) \quad A''_\kappa = \alpha_\kappa^{(B)}(1 + \hat{v}_\kappa \cdot \vec{\sigma}). \quad (\text{B.6})$$

It is clear, in view of equations (B.5), that the new collection of operators still forms a resolution of the identity, and thus represents a POVM. Equations (B.5) also imply

$$\begin{aligned} P_{i\kappa}^{(B)'} &= P_{i\kappa}^{(B)''} = \alpha_\kappa(1 + m_\kappa^{(B)x}n_i^x + m_\kappa^{(B)z}n_i^z) \\ &= \frac{1}{2}\alpha_\kappa(1 + u_\kappa^x n_i^x + u_\kappa^z n_i^z) + \frac{1}{2}\alpha_\kappa(1 + v_\kappa^x n_i^x + v_\kappa^z n_i^z) \\ &= \frac{1}{2}(P_{i\kappa}^{(A)'} + P_{i\kappa}^{(A)''}). \end{aligned} \quad (\text{B.7})$$

Now, define $\lambda'_\kappa := p_\kappa^{(A)'} / (2p_\kappa^{(B)})$, and $\lambda''_\kappa := p_\kappa^{(A)''} / (2p_\kappa^{(B)})$, where $p_\kappa^{(B)} := p_\kappa^{(B)'} = p_\kappa^{(B)''}$. Since $p_\kappa^{(A)'} + p_\kappa^{(A)''} = 2p_\kappa^{(B)}$, we have $\lambda'_\kappa + \lambda''_\kappa = 1$. It is easy to verify that

$$Q_{i\kappa}^{(B)'} = Q_{i\kappa}^{(B)''} = \lambda'_\kappa Q_{i\kappa}^{(A)'} + \lambda''_\kappa Q_{i\kappa}^{(A)''}. \quad (\text{B.8})$$

But then, the convexity of F implies

$$\begin{aligned}
p_\kappa^{(B)'} F(\vec{Q}_\kappa^{(B)'}) + p_\kappa^{(B)''} F(\vec{Q}_\kappa^{(B)''}) \\
&= 2p_\kappa^{(B)} F(\vec{Q}_\kappa^{(B)}) \\
&= 2p_\kappa^{(B)} F(\lambda'_\kappa \vec{Q}_\kappa^{(A)'} + \lambda''_\kappa \vec{Q}_\kappa^{(A)''}) \leq 2p_\kappa^{(B)} [\lambda'_\kappa F(\vec{Q}_\kappa^{(A)'}) + \lambda''_\kappa F(\vec{Q}_\kappa^{(A)''})] \\
&= p_\kappa^{(A)'} F(\vec{Q}_\kappa^{(A)'}) + p_\kappa^{(A)''} F(\vec{Q}_\kappa^{(A)''}).
\end{aligned} \tag{B.9}$$

It follows that the new POVM is no worse than B . By repeating this construction, we can obviously eliminate from B all the p pairs of elements not lying in the xz plane, until we get a POVM A , which provides no less information than B , whose elements all lie in the xz plane. This concludes the proof of the lemma.

Now we can proceed as follows: we consider the POVMs consisting of two elements only, and having vectors \hat{m}_i parallel to the x axis. By direct evaluation one can check that $\tilde{K}(A)$ equals the expression in equation (5.4). We can prove that, for $0 \leq \theta < 2\pi/3$, such an A provides no less information than any other POVM, C , consisting of more than two elements. By virtue of the lemma just proved, we lose no generality if we assume that all the vectors $m_\mu^{(C)}$ of C lie in the xz plane. Our first move is to symmetrize C with respect to the z axis, by introducing a POVM B , consisting of pairs of elements (B'_μ, B''_μ) , having equal weights, and vectors \hat{m}'_μ and \hat{m}''_μ that are symmetric with respect to the z axis:

$$B'_\mu = \frac{1}{2}C_\mu \quad B''_\mu = \frac{1}{2}\alpha_\mu^{(C)}(1 - \hat{m}'_\mu \sigma_x + \hat{m}''_\mu \sigma_z). \tag{B.10}$$

B provides as much information as C . Indeed, in view of equation (B.10), we find

$$P_{\pm\mu}^{(C)} = 2P_{\pm\mu}^{(B)'} = 2P_{\mp\mu}^{(B)''}. \tag{B.11}$$

The invariance of the predictability with respect to permutations of its arguments, then ensures that $\tilde{K}(B) = \tilde{K}(C)$. Thus, we lose no information if we consider a POVM, B , that is symmetric with respect to the z axis. Now we describe a procedure of reduction that, applied to a symmetric POVM-like B , gives rise to another symmetric POVM \hat{B} , which contains two elements less than B , but nevertheless gives no less information than B . The procedure works as follows: we pick at will two pairs of elements of B , say (B'_N, B''_N) and (B'_{N-1}, B''_{N-1}) , and consider the unique pair of symmetric unit vectors $\hat{u}_\pm = \pm u^x \hat{i} + u^z \hat{k}$ such that

$$u^z = \frac{1}{\alpha_N^{(B)} + \alpha_{N-1}^{(B)}} (\alpha_N^{(B)} m_N^{(B)z} + \alpha_{N-1}^{(B)} m_{N-1}^{(B)z}). \tag{B.12}$$

Consider the symmetric collection \hat{B} , obtained from B after replacing the four elements $(B'_N, B''_N, B'_{N-1}, B''_{N-1})$ by the pair $(\hat{B}'_{N-1}, \hat{B}''_{N-1})$ such that

$$\hat{B}'_{N-1} = (\alpha_N^{(B)} + \alpha_{N-1}^{(B)})(1 + \hat{u}_+ \cdot \vec{\sigma}) \quad \hat{B}''_{N-1} = (\alpha_N^{(B)} + \alpha_{N-1}^{(B)})(1 + \hat{u}_- \cdot \vec{\sigma}). \tag{B.13}$$

\hat{B} is still a POVM, as is easy to verify. Moreover, \hat{B} provides no less information than B , as we now show. Indeed, after some algebra, one finds

$$\frac{\tilde{K}(\hat{B}) - \tilde{K}(B)}{\alpha_N^{(B)} + \alpha_{N-1}^{(B)}} = g(u^z) - \frac{\alpha_N^{(B)}}{\alpha_N^{(B)} + \alpha_{N-1}^{(B)}} g(m_N^{(B)z}) - \frac{\alpha_{N-1}^{(B)}}{\alpha_N^{(B)} + \alpha_{N-1}^{(B)}} g(m_{N-1}^{(B)z}) \tag{B.14}$$

where the function $g(x)$ has the expression:

$$g(x) = -\frac{3 + x(1 + 2 \cos \theta)}{6} + \frac{(1 + x)^2 + 2(1 + x \cos \theta)^2 + 2(1 - x^2) \sin^2 \theta}{6 + 2x(1 + 2 \cos \theta)}. \tag{B.15}$$

In view of equation (B.12), the rhs of equation (B.14) is of the form

$$g(\lambda x_1 + (1 - \lambda)x_2) - \lambda g(x_1) - (1 - \lambda)g(x_2) \tag{B.16}$$

where $\lambda = \alpha_N^{(B)} / (\alpha_N^{(B)} + \alpha_{N-1}^{(B)})$, while $x_1 = m_N^{(B)z}$ and $x_2 = m_{N-1}^{(B)z}$. It may be checked that, for all values of θ , such that $0 \leq \theta < 2\pi/3$, $g(x)$ is concave, for $x \in [-1, 1]$, and so the rhs of equation (B.16) is non-negative for any value of $\lambda \in [0, 1]$. This implies that the rhs of equation (B.14) is non-negative as well, and so $\tilde{K}(\hat{B}) \geq \tilde{K}(B)$. After enough iterations of this procedure, we end up with a symmetric POVM consisting of two pairs of elements (B'_1, B''_1) and (B'_2, B''_2) . But then, the conditions for a POVM, equations (5.9), imply that the quantity between the brackets on the rhs of equation (B.12) vanishes, and so equation (B.12) gives $u^z = 0$. This means that the last iteration gives rise precisely to the PVM A . By putting everything together, we have shown that $\tilde{K}(C) = \tilde{K}(B) \leq \tilde{K}(\hat{B}) \cdots \leq \tilde{K}(A)$, and this is the required result.

References

- [1] Englert B G and Bergou J A 2000 *Opt. Commun.* **179** 337
- [2] A collection of original papers, and an historical review on the subject can be found in Wheeler J W and Zurek W H 1983 *Quantum Theory and Measurements* (Princeton, NJ: Princeton University Press)
- [3] Wootters W K and Zurek W H 1979 *Phys. Rev. D* **19** 473
- [4] Greenberger D M and YaSin A 1988 *Phys. Lett. A* **128** 391
- [5] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
Shannon C E 1948 *Bell Syst. Tech. J.* **27** 623
- [6] Englert B G 1996 *Phys. Rev. Lett.* **77** 2154
- [7] Mei M and Weitz M 2001 *Phys. Rev. Lett.* **86** 559
- [8] Luis A 2002 *J. Phys. A: Math. Gen.* **34** 8597
- [9] Dürr S 2001 *Phys. Rev. A* **64** 042113
- [10] Bimonte G and Musto R 2003 Comment on ‘Quantitative wave–particle duality in multibeam interferometers’
Preprint quant-ph/0304004 (Phys. Rev. A at press)
- [11] Rauch H and Summhammer J 1984 *Phys. Lett. A* **104** 44
Summhammer J, Rauch H and Tuppinger D 1987 *Phys. Rev. A* **36** 4447
- [12] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
- [13] Jaeger G, Shimony A and Vaidman L 1995 *Phys. Rev. A* **51** 54
- [14] Scully M O and Drühl K 1982 *Phys. Rev. A* **25** 2208
- [15] Brukner C and Zeilinger A 1999 *Phys. Rev. Lett.* **83** 3354
Brukner C and Zeilinger A 2001 *Phys. Rev. A* **63** 022113
- [16] Meystre P and Scully M O 1983 Observation, state reduction and the quantum eraser in quantum optics *Advances in Laser Spectroscopy (NATO ASI Series vol B95)* ed F T Arecchi, F Strumia and H Walther (New York: Plenum)
- [17] Peres A 1990 *Found. Phys.* **20** 1441
Peres A 1993 *Quantum Theory: Concepts and Methods* (Dordrecht: Kluwer)
- [18] Davies E B 1978 *IEEE Trans. Inf. Theory* **24** 596
- [19] Neumark M A 1940 *Izv. Akad. Nauk SSSR Ser. Mat.* **4** 53
Neumark M A 1943 *C.R. (Dokl.) Acad. Sci. URSS* **41** 359