Weak measurement takes a simple form for cumulants

Graeme Mitchison*
Centre for Quantum Computation, DAMTP, University of Cambridge, Cambridge CB3 0WA, United Kingdom
(Received 18 December 2007; published 2 May 2008)

A weak measurement on a system is made by coupling a pointer weakly to the system and then measuring the position of the pointer. If the initial wave function for the pointer is real, the mean displacement of the pointer is proportional to the so-called weak value of the observable being measured. This gives an intuitively direct way of understanding weak measurement. However, if the initial pointer wave function takes complex values, the relationship between pointer displacement and weak value is not quite so simple, as pointed out recently by Jozsa [R. Jozsa, Phys. Rev. A **76**, 044103 (2007)]. This is even more striking in the case of sequential weak measurements [G. Mitchison, R. Jozsa, and S. Popescu, Phys. Rev. A **76**, 062105 (2007)]. These are carried out by coupling several pointers at different stages of the evolution of the system, and the relationship between the products of the measured pointer positions and the sequential weak values can become extremely complicated for an arbitrary initial pointer wave function. Surprisingly, all this complication vanishes when one calculates the cumulants of pointer positions. These are directly proportional to the cumulants of sequential weak values. This suggests that cumulants have a fundamental physical significance for weak measurement.

DOI: 10.1103/PhysRevA.77.052102 PACS number(s): 03.65.Ta, 03.67.−a

I. INTRODUCTION

In physics, formal simplicity is often a reliable guide to the significance of a result. The concept of weak measurement, due to Aharonov and co-workers [1,2], derives some of its appeal from the formal simplicity of its basic formulas. One can extend the basic concept to a sequence of weak measurements carried out at a succession of points during the evolution of a system [3], but then the formula relating pointer positions to weak values turns out to be not quite so simple, particularly if one allows arbitrary initial conditions for the measuring system. I show here that the complications largely disappear if one takes the cumulants of expected values of pointer positions; these are related in a formally satisfying way to weak values, and this form is preserved under all measurement conditions.

The goal of weak measurement is to obtain information about a quantum system given both an initial state $|\psi_i\rangle$ and a final, post-selected state $|\psi_f\rangle$. Since weak measurement causes only a small disturbance to the system, the measurement result can reflect both the initial and final states. It can therefore give richer information than a conventional (strong) measurement, including in particular the results of all possible strong measurements [4,5]. To carry out the measurement, a measuring device is coupled to the system in such a way that the system is only slightly perturbed; this can be achieved by having a small coupling constant g. After the interaction, the pointer's position q is measured (or possibly some other pointer observable—e.g., its momentum p). Suppose that, following the standard von Neumann paradigm [6], the interaction between measuring device and system is taken to be $H_{int}=g\delta(t)pA$, where p is the momentum of a pointer and the delta function indicates an impulsive interaction at time t. It can be shown [2] that the expectation of the pointer position, ignoring terms of order g^2 or higher, is

$$\langle q\rangle = g \text{ Re } A_w, \tag{1}$$

where A_w is the *weak value* of the observable A given by

$$A_w = \frac{\langle\psi_f|A|\psi_i\rangle}{\langle\psi_f|\psi_i\rangle}. \tag{2}$$

As can be seen, (1) has an appealing simplicity, relating the pointer shift directly to the weak value. However, this formula only holds under the rather special assumption that the initial pointer wave function ϕ is a Gaussian or, more generally, is real and has zero mean. When ϕ is a completely general wave function—i.e., is allowed to take complex values and have any mean value [3,7]—Eq. (1) is replaced by

$$\langle q\rangle = \langle q\rangle_i + g \text{ Re } A_w + g \text{ Im } A_w(\langle pq+qp\rangle_i - 2\langle q\rangle_i\langle p\rangle_i), \tag{3}$$

where, for any pointer variable x, $\langle x\rangle_i$ denotes the initial expected value $\langle\phi|x|\phi\rangle$ of x; so, for instance, $\langle q\rangle_i$ and $\langle p\rangle_i$ are the means of the initial pointer position and momentum, respectively. (Again, this formula ignores terms of order g^2 or higher.)

Equation (3) seems to have lost the simplicity of (1), but we can rewrite it as

$$\langle q\rangle = \langle q\rangle_i + g \text{ Re}(\xi A_w), \tag{4}$$

where

$$\xi = -2i(\langle qp\rangle_i - \langle q\rangle_i\langle p\rangle_i), \tag{5}$$

and Eq. (4) is then closer to the form of (1). As will become clear, this is part of a general pattern.

One can also weakly measure several observables $A_1,\ldots,A_n$ in succession [3]. Here one couples pointers at several locations and times during the evolution of the system, taking the coupling constant g_k at site k to be small. One then measures each pointer and takes the product of the po-

*g.j.mitchison@damtp.cam.ac.uk

sitions q_k of the pointers. For two observables and in the special case where the initial pointer distributions are real and have zero mean—e.g., a Gaussian—one finds [3]

$$\langle q_1 q_2\rangle = \frac{g_1 g_2}{2}\mathrm{Re}[(A_2,A_1)_w + (A_1)_w\overline{(A_2)}_w], \tag{6}$$

ignoring terms in higher powers of g_1 and g_2. Here $(A_2,A_1)_w$ is the *sequential weak value* defined by

$$(A_2,A_1)_w = \frac{\langle\psi_f|WA_2VA_1U|\psi_i\rangle}{\langle\psi_f|WVU|\psi_i\rangle}, \tag{7}$$

where U is a unitary taking the system from the initial state $|\psi_i\rangle$ to the first weak measurement, V describes the evolution between the two measurements, and W takes the system to the final state. [Note the reverse order of operators in (A_2,A_1), which reflects the order in which they are applied.] If we drop the assumption about the special initial form of the pointer distribution and allow an arbitrary ϕ, then the counterpart of (6) becomes extremely complicated: see the Appendix, Eq. (A1).

Even the comparatively simple formula (6) is not quite ideal. By analogy with (1) we would hope for a formula of the form $\langle q_1 q_2\rangle \propto \mathrm{Re}(A_2,A_1)_w$, but there is an extra term $(A_1)_w\overline{(A_2)}_w$. What we seek, therefore, is a relationship that has some of the formal simplicity of (1) and furthermore preserves its form for all measurement conditions. It turns out that this is possible if we take the *cumulant* of the expectations of pointer positions. As we shall see in the next section, this is a certain sum of products of joint expectations of subsets of the q_i, which we denote by $\langle q_1\cdots q_n\rangle^c$. For a set of observables, we can define a formally equivalent expression using sequential weak values, which we denote by $(A_n,\ldots,A_1)^c_w$. Then the claim is that, up to order n in the coupling constants g_k (assumed to be all of the same approximate order of magnitude),

$$\langle q_1\cdots q_n\rangle^c = g_1\cdots g_n\,\mathrm{Re}\{\xi(A_n,\ldots,A_1)^c_w\}, \tag{8}$$

where ξ is a factor dependent on the initial wave functions for each pointer. Equation (8) holds for any initial pointer wave function, though different wave functions produce different values of ξ. The remarkable thing is that all the complexity is packed into this one number, rather than exploding into a multiplicity of terms, as in (A1).

Note also that (4) has essentially the same form as (8) since, in the case $n=1$, $\langle A\rangle^c_w = A_w$. However, there is an extra term $\langle q\rangle_i$ in (4); this arises because the cumulant for $n=1$ is anomalous in that its terms do not sum to zero.

II. CUMULANTS

Given a collection of random variables, such as the pointer positions q_i, the cumulant $\langle q_1\cdots q_n\rangle^c$ is a polynomial in the expectations of subsets of these variables [8,9]; it has the property that it vanishes whenever the set of variables q_i can be divided into two independent subsets. One can say that the cumulant, in a certain sense, picks out the maximal correlation involving all of the variables.

We introduce some notation to define the cumulant. Let x be a subset of the integers $\{1,\ldots,n\}$. We write $\Pi_x q$ for $\Pi_{i=1}^{|x|} q_{x(i)}$, where $|x|$ is the size of x and the indices of the q's in the product run over all the integers $x(i)$ in x. Then the cumulant is given by

$$\langle q_1\cdots q_n\rangle^c = \sum_{b=\{b_1,\ldots,b_k\}} a_k \prod_{j=1}^{k}\left\langle \prod_{b_j} q\right\rangle, \tag{9}$$

where $b=\{b_1,\ldots,b_k\}$ runs over all partitions of the integers $\{1,\ldots,n\}$ and the coefficient a_k is given by

$$a_k = (k-1)!\,(-1)^{k-1}. \tag{10}$$

For $n=1$ we have $\langle q\rangle^c = \langle q\rangle$, and for $n=2$,

$$\langle q_1 q_2\rangle^c = \langle q_1 q_2\rangle - \langle q_1\rangle\langle q_2\rangle. \tag{11}$$

There is an inverse operation for the cumulant [9,10]: *Proposition II.1*:

$$\langle q_1\cdots q_n\rangle = \sum_{b=\{b_1,\ldots,b_k\}} \prod_{j=1}^{k}\left\langle \prod_{b_j} q\right\rangle^c. \tag{12}$$

Proof. To see that this equation holds, we must show that the term $\Pi_{j=1}^k\langle\Pi_{b_j} q\rangle$ obtained by expanding the right-hand side is zero unless b is the partition consisting of the single set $\{1,\ldots,n\}$. Replacing each subset b_j by the integer j, this is equivalent to $\Sigma a_{k_1}\cdots a_{k_r}=0$, where the sum is over all partitions of $\{1,\ldots,k\}$ by subsets of sizes $k_1,\ldots,k_r$ and the a_k's are given by (10). In this sum we distinguish partitions with distinct integers—e.g., {1,2}, {3,4} and {1,3}, {2,4}. There are $\binom{k}{k_1,\ldots,k_r}(l_1!,\ldots,l_k!)^{-1}$ such distinct partitions with subset sizes $k_1,\ldots,k_r$, where l_i is the number of k's equal to i, so our sum may be rewritten as $k!\Sigma(-1)^{k_1-1}\cdots(-1)^{k_r-1}(l_1!\cdots l_k!k_1\cdots k_r)^{-1}$, where the sum is now over partitions in the standard sense [11]. This is $k!$ times the coefficient of x^k in

$$\left(1+x+\frac{x^2}{2!}+\cdots\right)\left(1+(-x^2/2)+\frac{(-x^2/2)^2}{2!}+\cdots\right) \times\left(1+(x^3/3)+\frac{(x^3/3)^2}{2!}+\cdots\right)\cdots \tag{13}$$

$$= e^{x-x^2/2+x^3/3\cdots} = e^{\log_e(1+x)} = 1+x. \tag{14}$$

Thus the sum is zero except for $k=1$, which corresponds to the single-set partition b. ■

Definition II.2. If $\{1,\ldots,n\}$ can be written as the disjoint union of two subsets S_1 and S_2, we say the variables corresponding to these subsets are independent if

$$\left\langle \prod_{S_1'} q \prod_{S_2'} q\right\rangle = \left\langle \prod_{S_1'} q\right\rangle\left\langle \prod_{S_2'} q\right\rangle, \tag{15}$$

for any subsets $S_i'\subseteq S_i$.

We now prove the characteristic property of cumulants.

Proposition II.3. The cumulant vanishes if its arguments can be divided into two independent subsets.

Proof. For $n=2$ this follows at once from (11) and (15), and we continue by induction. From (12) and the inductive assumption for $n-1$, we have

$$\langle q_1 \cdots q_n \rangle = \langle q_1 \cdots q_n \rangle^c + \sum_{b=\{b_1,\ldots,b_k\}\subset S_1} \prod_{j=1}^{k} \left\langle \prod_{b_j} q \right\rangle^c \times \sum_{c=\{c_1,\ldots,c_l\}\subset S_2} \prod_{j=1}^{l} \left\langle \prod_{c_j} q \right\rangle^c . \quad (16)$$

This holds because any term on the right-hand side of (12) vanishes when any subset of the partition b includes elements of both S_1 and S_2. Using (12) again, this implies

$$\langle q_1 \cdots q_n \rangle = \langle q_1 \cdots q_n \rangle^c + \left\langle \prod_{S_1} q \right\rangle \left\langle \prod_{S_2} q \right\rangle , \quad (17)$$

and by independence, $\langle q_1 \cdots q_n \rangle^c = 0$. Thus the inductive assumption holds for n. ■

In fact, the coefficients a_k in (9) are uniquely determined to have the form (10) by the requirement that the cumulant vanish when the variables form two independent subsets [12,13].

For $n=2$, the cumulant (11) is just the covariance $\langle q_1 q_2 \rangle^c = \langle (q_1 - \langle q_1 \rangle)(q_2 - \langle q_2 \rangle)\rangle$ and the same is true for $n=3$: namely, $\langle q_1 q_2 q_3 \rangle^c = \langle (q_1 - \langle q_1 \rangle)(q_2 - \langle q_2 \rangle)(q_3 - \langle q_3 \rangle)\rangle$. For $n=4$, however, there is a surprise. The covariance is given by

$$\left\langle \prod_{i=1}^{4} (q_i - \langle q_i \rangle) \right\rangle = \langle q_1 q_2 q_3 q_4 \rangle - \sum \langle q_i q_j q_k \rangle \langle q_l \rangle + \sum \langle q_i q_j \rangle \langle q_k \rangle \langle q_l \rangle - 3 \langle q_1 \rangle \langle q_2 \rangle \langle q_3 \rangle \langle q_4 \rangle , \quad (18)$$

where the sums include all distinct combinations of indices, but the cumulant is

$$\langle q_1 q_2 q_3 q_4 \rangle^c = \langle q_1 q_2 q_3 q_4 \rangle - \sum \langle q_i q_j q_k \rangle \langle q_l \rangle - \sum \langle q_i q_j \rangle \langle q_k q_l \rangle + 2 \sum \langle q_i q_j \rangle \langle q_k \rangle \langle q_l \rangle - 6 \langle q_1 \rangle \langle q_2 \rangle \langle q_3 \rangle \langle q_4 \rangle , \quad (19)$$

which includes terms like $\langle q_1 q_2 \rangle \langle q_3 q_4 \rangle$ that do not occur in the covariance. Note that, if the subsets {1,2} and {3,4} are independent, the covariance does not vanish, since independence implies we can write the first term in (18) as $\langle q_1 q_2 q_3 q_4 \rangle = \langle q_1 q_2 \rangle \langle q_3 q_4 \rangle$ and there is no canceling term. However, as we have seen, the cumulant does contain such a term, and it is a pleasant exercise to check that the whole cumulant vanishes.

III. SEQUENTIAL WEAK VALUES AND CUMULANTS

To carry out a sequential weak measurement, one starts a system in an initial state $|\psi_i\rangle$, then weakly couples pointers at several times t_k during the evolution of the system, and finally post-selects the system state $|\psi_f\rangle$. One then measures the pointers and finally takes the product of the values obtained from these pointer measurements. It is assumed that one can repeat the whole process many times to obtain the expectation of the product of pointer values. If one measures pointer positions q_k, for instance, one can estimate $\langle q_1 \cdots q_n \rangle$, but one could also measure the momenta of the pointers to estimate $\langle p_1 \cdots p_n \rangle$.

If the coupling for the kth pointer is given by $H_{int} = \delta(t - t_k) r_k p$, and if the individual initial pointer wave

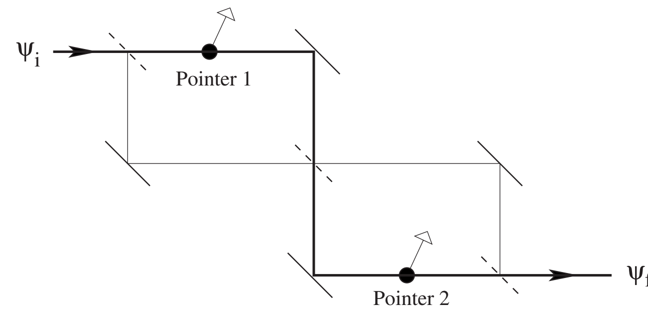

FIG. 1. The double interferometer. A sequential weak measurement is made by weakly coupling the pointers marked 1 and 2, then measuring them and finally multiplying the values so obtained.

functions are Gaussian, or, more generally, are real with zero mean, then it turns out [3] that these expectations can be expressed in terms of sequential weak values of order n or less. Here the sequential weak value of order n, $(A_n, \ldots, A_1)_w$, is defined by

$$(A_n, \ldots, A_1)_w = \frac{\langle \psi_f | U_{n+1} A_n U_n \cdots A_1 U_1 | \psi_i \rangle}{\langle \psi_f | U_{n+1} \cdots U_1 | \psi_i \rangle} , \quad (20)$$

where U_i defines the evolution of the system between the measurements of A_{i-1} and A_i.

When the A_k are projectors, $A_k = |x_k\rangle\langle x_k|$, we can write the sequential weak value as [3]

$$(A_n, \ldots, A_1)_w = \frac{\langle \psi_f | U_{n+1} | x_n \rangle \langle x_n | U_n | x_{n-1} \rangle \cdots \langle x_1 | U_1 | \psi_i \rangle}{\sum_y \langle \psi_f | U_{n+1} | y_n \rangle \langle y_n | U_n | y_{n-1} \rangle \cdots \langle y_1 | U_1 | \psi_i \rangle} = \frac{\text{amplitude}(x)}{\sum_y \text{amplitude}(y)} , \quad (21)$$

which shows that, in this case, the weak values has a natural interpretation as the amplitude for following the path defined by the x_k. Figure 1 shows an example taken from [3] where the path (labeled by "1" and "2" successively) is a route taken by a photon through a pair of interferometers, starting by injecting the photon at the top left (with state $|\psi\rangle_i$) and ending with post-selection by detection at the bottom right (with final state $|\psi\rangle_f$).

In the last section, the cumulant was defined for expectations of products of variables. One can define the cumulant for other entities by formal analogy—for instance, for density matrices [10] or hypergraphs [9]. We can do the same for sequential weak values, defining the cumulant by (9) with $\langle \Pi_{b_j} q \rangle$ replaced by $(\overleftarrow{A_{b_j(|b_j|)}, \ldots, A_{b_j(1)}})_w$, where the arrow indicates that the indices, which run over the subset b_j, are arranged in ascending order from right to left. For example, for $n=1$, $(A_w)^c = A_w$, and for $n=4$,

$$(A_4, A_3, A_2, A_1)_w^c = (A_4, A_3, A_2, A_1)_w - \sum (\overleftarrow{A_i, A_j, A_k})_w (A_l)_w - \sum (\overleftarrow{A_i, A_j})_w (\overleftarrow{A_k, A_l})_w + 2 \sum (\overleftarrow{A_i, A_j})_w (A_k)_w (A_l)_w - 6 (A_1)_w (A_2)_w (A_3)_w (A_4)_w . \quad (22)$$

There is a notion of independence that parallels (15): given a disjoint partition $S_1 \cup S_2 = \{1, \ldots, n\}$ such that

$$(\overleftarrow{A_{S'_1 \cup S'_2}})_w = (\overleftarrow{A_{S'_1}})_w (\overleftarrow{A_{S'_2}})_w, \tag{23}$$

for any subsets $S'_i \subseteq S_i$; then, we say the observables labeled by the two subsets are *weakly independent*. There is then an analog of Lemma II.3.

Lemma III.1. The cumulant $(A_n, \ldots, A_1)^c_w$ vanishes if the A_k are weakly independent for some subsets S_1 and S_2.

As an example of this, if one is given a bipartite system $\mathcal{H}^A \otimes \mathcal{H}^B$ and initial and final states that factorize as $|\psi_i\rangle = |\psi_i\rangle^A \otimes |\psi_i\rangle^B$ and $|\psi_f\rangle = |\psi_f\rangle^A \otimes |\psi_f\rangle^B$, then observables on the A and B parts of the system are clearly weakly independent. Another class of examples comes from what one might describe as a "bottleneck" construction, where at some point the evolution of the system is divided into two parts by a one-dimensional projector (the bottleneck) and its complement, and the post-selection excludes the complementary part. Then, if all the measurements before the projector belong to S_1 and all those after the projector belong to S_2, the two sets are weakly independent. This follows because we can write

$$\begin{aligned}(\overleftarrow{A_{S'_1 \cup S'_2}})_w &= \frac{\langle \psi_f | U_{n+1} A_n \cdots U_{k+1} A_k W_k | \psi_b \rangle \langle \psi_b | V_k A_{k-1} \cdots A_1 U_1 | \psi_i \rangle}{\langle \psi_f | U_{n+1} \cdots U_{k+1} W_k | \psi_b \rangle \langle \psi_b | V_k \cdots U_1 | \psi_i \rangle} \\ &= \frac{\langle \psi_f | U_{n+1} A_n \cdots U_{k+1} A_k W_k | \psi_b \rangle \langle \psi_b | V_k \cdots U_1 | \psi_i \rangle}{\langle \psi_f | U_{n+1} \cdots U_{k+1} W_k | \psi_b \rangle \langle \psi_b | V_k \cdots U_1 | \psi_i \rangle} \frac{\langle \psi_f | U_{n+1} \cdots U_{k+1} W_k | \psi_b \rangle \langle \psi_b | V_k A_{k-1} \cdots A_1 U_1 | \psi_i \rangle}{\langle \psi_f | U_{n+1} \cdots U_{k+1} W_k | \psi_b \rangle \langle \psi_b | V_k \cdots U_1 | \psi_i \rangle} = (\overleftarrow{A_{S'_1}})_w (\overleftarrow{A_{S'_2}})_w,\end{aligned}$$

where $W_k|\psi_b\rangle\langle\psi_b|V_k$ is the part of U_k lying in the post-selected subspace. As an illustration of this, suppose we add a connecting link (Fig. 2, "L") between the two interferometers in Fig. 1, so $|\psi_b\rangle\langle\psi_b|$, the bottleneck, is the projection onto L, and post-selection discards the part of the wave function corresponding to the path L'. Then measurements at "1" and "2" are weakly independent; in fact, $(A_1)_w = 1/2$, $(A_2)_w = 1/2$, and $(A_2, A_1)_w = 1/4$. Note that the same measurements are *not* independent in the double interferometer of Fig. 1, where $(A_1)_w = 0$, $(A_2)_w = 0$, and yet, surprisingly, $(A_2, A_1)_w = -1/2$ [3].

IV. MAIN THEOREM

Consider n system observables $A_1, \ldots, A_n$. Suppose s_k, for $k = 1, \ldots, n$, are observables of the kth pointer—namely, Hermitian functions $s_k(q_k, p_k)$ of pointer position q_k and momentum p_k—and the interaction Hamiltonian for the weak measurement of system observable A_k is $H_k = g_k s_k A_k$, where g_k is a small coupling constant (all g_k being assumed of the same order of magnitude g). Suppose further that the pointer observables r_k are measured after the coupling. Let ϕ_k be the kth pointer's initial wave-function. For any variable x_k associated with the kth pointer, write $\langle x_k \rangle_i$ for $\langle \phi_k | x_k | \phi_k \rangle$.

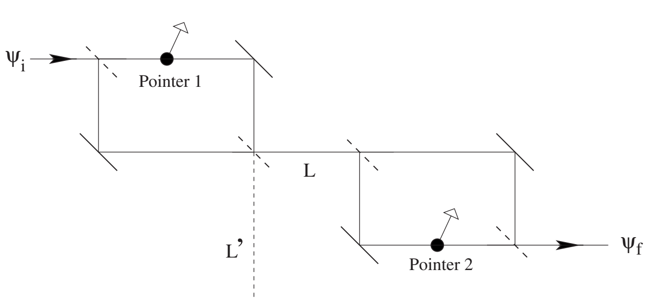

FIG. 2. A "bottleneck" (L) is added to the double interferometer of Fig. 1. This makes the measurements at 1 and 2 weakly independent.

We are now almost ready to state the main theorem, but first need to clarify the measurement procedure. When we evaluate expectations of products of the r_k for different sets of pointers—for instance, when we evaluate $\langle r_1 r_2 \rangle$—we have a choice. We could either couple the entire set of n pointers and then select the data for pointers 1 and 2 to get $\langle r_1 r_2 \rangle$. Or we could carry out an experiment in which we couple just pointers 1 and 2 to give $\langle r_1 r_2 \rangle$. These procedures give different answers. For instance, if we couple three pointers and measure pointers 1 and 2 to get $\langle r_1 r_2 \rangle$, in addition to the terms in g_1, g_2, and $g_1 g_2$ we also get terms in $g_2 g_3$ and $g_1 g_3$ involving the observable A_3. This means we get a different cumulant $\langle r_1 \cdots r_n \rangle^c$, depending on the procedure used. In what follows, we regard each expectation as being evaluated in a separate experiment, with only the relevant pointers coupled. It will be shown elsewhere that, with the alternative definition, the theorem still holds but with a different value of the constant ξ.

Theorem IV.1 (cumulant theorem). For $n \geq 2$, for any pointer observables r_k and s_k and for any initial pointer wave functions ϕ_k, up to total order n in the g_k,

$$\langle r_1 \cdots r_n \rangle^c = g_1 \cdots g_n \, \mathrm{Re}\{\xi (A_n, \ldots, A_1)^c_w\}, \tag{24}$$

where ξ (sometimes written more explicitly as $\xi_{r_1 \ldots r_n}$) is given by

$$\xi = 2(-i)^n \left(\prod_{k=1}^{n} \langle r_k s_k \rangle_i - \prod_{k=1}^{n} \langle r_k \rangle_i \langle s_k \rangle_i \right). \tag{25}$$

For $n = 1$ the same result holds, but with the extra term $\langle r \rangle_i$:

$$\langle r\rangle = \langle r\rangle_i + g\,\mathrm{Re}(\xi A_w). \tag{26}$$

Proof. We use the methods of [3] to calculate the expectations of products of pointer variables for sequential weak measurements. Let the initial and final states of the system be $|\psi_i\rangle$ and $|\psi_f\rangle$, respectively. Consider some subset $b=\{b_1,\ldots,b_\kappa\}$ of $\{1,\ldots,n\}$, with $b_1\le b_2\le\cdots\le b_\kappa$. The state of the system and the pointers $b_1,\ldots,b_\kappa$ after the coupling of those pointers is

$$\Psi_{\mathcal{S},\mathcal{M}} = U_{n+1}\cdots U_{b_\kappa+1}e^{-ig_{b_\kappa}s_{b_\kappa}A_{b_\kappa}}U_{b_\kappa}\cdots e^{-ig_{b_1}s_{b_1}A_{b_1}} \times U_{b_1}\cdots U_1|\psi_i\rangle\phi_{b_1}(r_{b_1})\cdots\phi_{b_\kappa}(r_{b_\kappa}), \tag{27}$$

and following post-selection by the system state $|\psi_f\rangle$, the state of the pointers is

$$\Psi_{\mathcal{M}} = \langle\psi_f|U_{n+1}\cdots U_{b_\kappa+1}e^{-ig_{b_\kappa}s_{b_\kappa}A_{b_\kappa}}U_{b_\kappa}\cdots e^{-ig_{b_1}s_{b_1}A_{b_1}} \times U_{b_1}\cdots U_1|\psi_i\rangle\phi_{b_1}(r_{b_1})\cdots\phi_{b_\kappa}(r_{b_\kappa}). \tag{28}$$

Expanding each exponential, we have

$$\langle r_{b_1}\cdots r_{b_\kappa}\rangle = \frac{\int \bar{\Psi}_{\mathcal{M}} r_{b_1}\cdots r_{b_\kappa}\Psi_{\mathcal{M}}\,dr_{b_1}\cdots dr_{b_\kappa}}{\int |\Psi_{\mathcal{M}}|^2 dr_{b_1}\cdots dr_{b_\kappa}} \tag{29}$$

$$= \frac{\sum_{i_1,\ldots,i_n\in b;j_1,\ldots,j_n\in b}\alpha_{i_1,\ldots,i_n}\bar{\alpha}_{j_1,\ldots,j_n}u^{b_1}_{i_{b_1}j_{b_1}}\cdots u^{b_\kappa}_{i_{b_\kappa}j_{b_\kappa}}}{\sum_{i_1,\ldots,i_n\in b;j_1,\ldots,j_n\in b}\alpha_{i_1,\ldots,i_n}\bar{\alpha}_{j_1,\ldots,j_n}v^{b_1}_{i_{b_1}j_{b_1}}\cdots v^{b_\kappa}_{i_{b_\kappa}j_{b_\kappa}}}, \tag{30}$$

where $i_k\ge 0$ are integers, $i_1,\ldots,i_n\in b$ means that $i_l=0$ for $l\notin b$, and

$$\alpha_{i_1,\ldots,i_n} = \left(\prod_{k=1}^{n} g_k^{i_k}\right)(A_n^{i_n},\ldots,A_1^{i_1})_w, \tag{31}$$

$$u^k_{lm} = \int (m!)^{-1}\overline{(-is_k)^m\phi_k(r_k)}\,r_k(l!)^{-1}(-is_k)^l\phi_k(r_k)dr_k, \tag{32}$$

$$v^k_{lm} = \int (m!)^{-1}\overline{(-is_k)^m\phi_k(r_k)}(l!)^{-1}(-is_k)^l\phi_k(r_k)dr_k. \tag{33}$$

Let us write (30) as

$$\langle r_{b_1}\cdots r_{b_\kappa}\rangle = \frac{\sum_{\mathbf{i}\in b,\mathbf{j}\in b}x_{\mathbf{i};\mathbf{j}}}{\sum_{\mathbf{i}\in b,\mathbf{j}\in b}y_{\mathbf{i};\mathbf{j}}}, \tag{34}$$

where

$$x_{\mathbf{i};\mathbf{j}} = \alpha_{i_1,\ldots,i_n}\bar{\alpha}_{j_1,\ldots,j_n}u^{b_1}_{i_{b_1}j_{b_1}}\cdots u^{b_\kappa}_{i_{b_\kappa}j_{b_\kappa}}, \tag{35}$$

$$y_{\mathbf{k};\mathbf{l}} = \alpha_{k_1,\ldots,k_n}\bar{\alpha}_{l_1,\ldots,l_n}v^{b_1}_{k_{b_1}l_{b_1}}\cdots v^{b_\kappa}_{k_{b_\kappa}l_{b_\kappa}}, \tag{36}$$

and $\mathbf{i}$ denotes the index set $\{i_1,\ldots,i_n\}$, etc. Define

$$X_b = \sum_{\mathbf{i}\in b,\mathbf{j}\in b}x_{\mathbf{i};\mathbf{j}},\quad Y_b = \sum_{\mathbf{k}\in b,\mathbf{l}\in b}y_{\mathbf{k};\mathbf{l}}. \tag{37}$$

Then

$$\langle r_1\cdots r_n\rangle^c = \sum_{b_1,\ldots,b_k}(k-1)!(-1)^{k-1}\prod_{l=1}^{k}\langle r_{b_l(1)}\cdots r_{b_l(|b_l|)}\rangle \tag{38}$$

$$= \sum_{b_1,\ldots,b_k}(k-1)!(-1)^{k-1}\prod_{l=1}^{k}\frac{X_{b_l}}{Y_{b_l}}. \tag{39}$$

Set $\mathcal{Y}=\Pi_{b\subset\{1,\ldots,n\}}Y_b$, where b in the product ranges over all distinct subsets of the integers $\{1,\ldots,n\}$. Then $\mathcal{Y}\langle r_1\cdots r_n\rangle^c$ is an (infinite) weighted sum of terms

$$z_{\mathcal{I}} = (x_{\mathbf{i}(1);\mathbf{j}(1)}\cdots x_{\mathbf{i}(m);\mathbf{j}(m)})(y_{\mathbf{k}(1);\mathbf{l}(1)}\cdots y_{\mathbf{k}(m');\mathbf{l}(m')}), \tag{40}$$

where

$$\mathcal{I} = \mathcal{I}_i\cup\mathcal{I}_j\cup\mathcal{I}_k\cup\mathcal{I}_l = \{\mathbf{i}(1),\ldots,\mathbf{i}(m)\}\cup\{\mathbf{j}(1),\ldots,\mathbf{j}(m)\}\cup\{\mathbf{k}(1),\ldots,\mathbf{k}(m')\}\cup\{\mathbf{l}(1),\ldots,\mathbf{l}(m')\} \tag{41}$$

denotes the set of all the index sets that occur in $z_{\mathcal{I}}$. The strategy is to show that, when the size of the index set $\mathcal{I}$ is less than n, the coefficient of $z_{\mathcal{I}}$ vanishes; by (31), this implies that all coefficients of order less than n in g vanish. We then look at the index sets of size n, corresponding to terms of order g^n, and show that the relevant terms sum up to the right-hand side of (24). But if $\mathcal{Y}\langle r_1\cdots r_n\rangle^c=g^n x+O(g^{n+1})$ for some x, then we also have $\langle r_1\cdots r_n\rangle^c=g^n x+O(g^{n+1})$, since $\mathcal{Y}=1+O(g)$.

Let $b=\{b_1,\ldots,b_s\}$ be a partition of $\{1,\ldots,n\}$. We say that b is a *valid* partition for $\mathcal{I}$ if the following is true.

(i) For each r with $1\le r\le m$, $\mathbf{i}(r)+\mathbf{j}(r)\in b_l$, for some b_l, and we can associate a distinct b_l to each r. (Here $\mathbf{i}+\mathbf{j}$ means the index set $\{i_1+j_1,\ldots,i_n+j_n\}$.)

(ii) For each r with $1\le r\le m'$, $\mathbf{k}(r)+\mathbf{l}(r)\in S$, for some subset $S\subset\{1,\ldots,n\}$ that is not in the partition b—i.e., for which $S\ne b_l$ for any l—and we can associate a distinct S to each r. Let $\gamma(\mathcal{I},b)$ be the number of ways of associating a subset S to each r.

Lemma IV.2. The coefficient of $z_{\mathcal{I}}$ in $\mathcal{Y}\langle r_1\cdots r_n\rangle^c$ is zero if all the index sets in $\mathcal{I}$ have a zero at some position r.

Proof. If we expand $\mathcal{Y}\langle r_1\cdots r_n\rangle^c$ using (39), each term in this expansion is associated with a partition b of $\{1,\ldots,n\}$. Let b be a valid partition for $\mathcal{I}$, and let $c=\{c_1,\ldots,c_s\}$ denote the partition derived from b by removing r from the subset b_l that contains it and deleting that subset if it contains only r. Then the following partitions include b and are all valid:

$$c^{(1)}=\{(rc_1),c_2,\ldots,c_s\},$$

$$c^{(2)}=\{c_1,(rc_2),\ldots,c_s\},$$

$$,\ldots,$$

$$c^{(s)}=\{c_1,c_2,\ldots,(rc_s)\},$$

$$c^{(s+1)}=\{r,c_1,c_2,\ldots,c_s\}. \tag{42}$$

Each partition $c^{(i)}$, for $1\leq i\leq s+1$, contributes $\gamma(\mathcal{I},b)$ to the coefficient of $z_{\mathcal{I}}$ in $\mathcal{Y}\prod_{l=1}^{k}X_{c^{(i)}}/Y_{c^{(i)}}$, and since this term has coefficient $(s-1)!(-1)^{(s-1)}$ in (39) for partitions $c^{(1)},c^{(2)},\ldots,c^{(s)}$, and $s!(-1)^s$ for $c^{(s+1)}$, the sum of all contributions is zero. ■

From Eqs. (31) and (41), the power of g in the term $z_{\mathcal{I}}$ is $|I|=|I_i|+|I_j|+|I_k|+|I_l|$. This, together with the preceding lemma, implies that the lowest-order nonvanishing terms in $\mathcal{Y}\langle r_1\cdots r_n\rangle^c$ are $z_{\mathcal{I}}$'s that have a "1" occurring once and once only in each position; we call these *complete lowest-degree* terms.

Lemma IV.3. The coefficient of a complete lowest-degree term $z_{\mathcal{I}}$ in $\mathcal{Y}\langle r_1\cdots r_n\rangle^c$ is zero unless only one of the four classes of indices in $\mathcal{I}$—viz., $\mathcal{I}_i$, $\mathcal{I}_j$, $\mathcal{I}_k$,. or $\mathcal{I}_l$—has nonzero terms.

Proof. Consider first the case where the indices in $\mathcal{I}_j$ and $\mathcal{I}_l$ are zero and where both $\mathcal{I}_i$ and $\mathcal{I}_k$ have some nonzero indices. Let $b=\{b_1,\ldots,b_r\}$ be the partition whose subsets consists of the nonzero positions in index sets $\mathbf{i}(t)$ in $\mathcal{I}_i$, and let $c=\{c_1,\ldots,c_s\}$ be some partition of the remaining integers in $\{1,\ldots,n\}$. Suppose $s\leq r$. Then we can construct a set of partitions by mixing b and c; these have the form

$$d^{(w)}=\{c_{i_1},\ldots,c_{i_t},(x_1b_1),\ldots,(x_rb_r)\}, \tag{43}$$

where each x_i is either empty or consists of some c_i and all the subsets c_i are present once only in the partition. If any $d^{(w)}$ is valid, all the other mixtures will also be valid. Furthermore, the set of all valid partitions can be decomposed into nonoverlapping subsets of mixtures obtained in this way.

Any mixture $d^{(w)}$ gives the same value of $\gamma(\mathcal{I},d^{(w)})$, which we denote simply by γ; so to show that all the contributions to the coefficient of $z_{\mathcal{I}}$ cancel, we have only to sum over all the mixtures, weighting a partition with t subsets by $(t-1)!(-1)^{t-1}$. This gives

$$\text{coefficient of } z_{\mathcal{I}}=\gamma\sum_{i=0}^{s}(s+r-1)!(-1)^{s+r-i}\binom{s}{i}\binom{r}{i}i!$$

$$=\gamma(-1)^{s+r-1}s!\sum(s+r-i-1)\cdots$$

$$\times(s-i+1)\binom{r}{i}(-1)^i$$

$$=\gamma(-1)^{s+r-1}s!\frac{\partial^{r-1}}{\partial x^{r-1}}\{x^{s-1}(x-1)^r\}|_{x=1}=0.$$

The above argument applies equally well to the situation where $\mathcal{I}_i$ and $\mathcal{I}_l$ both have some nonzero indices and indices in $\mathcal{I}_j$ and $\mathcal{I}_k$ are zero. If the nonzero indices are present in $\mathcal{I}_i$ and $\mathcal{I}_j$, we can take any valid partition $a=\{a_1,\ldots,a_r\}$ and divide each subset a_k into two subsets b_k and c_k with the indices from $\mathcal{I}_i$ in b_k and those from $\mathcal{I}_j$ in c_k. All the mixtures of type (43) are valid, and they include the original partition a. By the above argument, the coefficients of $z_{\mathcal{I}}$ arising from them sum to zero. Other combinations of indices are dealt with similarly.

Note that, for $n=4$ and for the index sets $(1,1,0,0)\in\mathcal{I}_i$ and $(0,0,1,1)\in\mathcal{I}_j$, the "mixture" argument shows that coefficient of $z_{\mathcal{I}}$ coming from $\langle r_1r_2r_3r_4\rangle$ cancels that coming from $\langle r_1r_2\rangle\langle r_3r_4\rangle$ to give zero. This cancellation occurs with the cumulant (19), but not with the covariance (18), where the term $\langle r_1r_2\rangle\langle r_3r_4\rangle$ is absent. ■

The only terms that need to be considered, therefore, are complete lowest-degree terms with nonzero indices only in one of the sets $\mathcal{I}_i$, $\mathcal{I}_j$, $\mathcal{I}_k$, and $\mathcal{I}_l$. It is easy to calculate the coefficients one gets for such terms. Consider the case of $\mathcal{I}_i$. We only need to consider the single partition b whose subsets are the index sets of $\mathcal{I}_i$. For this partition, by (40), (35), and (36):

$$z_{\mathcal{I}}=\prod_{e=1}^{t}\alpha_{\mathbf{i}(e)}\prod_{k=1}^{n}u_{1,0}^{k}v_{0,0}^{k}$$

$$=g_1\cdots g_n\prod_{e=1}^{t}(A_{\mathbf{i}(e)(|\mathbf{i}(e)|)},\ldots,A_{\mathbf{i}(e)(1)})_w\prod_{k=1}^{n}\langle r_ks_k\rangle_i. \tag{44}$$

From (39), $z_{\mathcal{I}}$ appears in $\mathcal{Y}\langle r_1\cdots r_n\rangle^c$ with a coefficient $(t-1)!(-1)^{t-1}$. So summing over all $z_{\mathcal{I}}$ with indices in $\mathcal{I}_i$, one obtains $g_1\cdots g_n(A_n,\ldots,A_1)_w^c\prod_{k=1}^{n}(-i\langle r_ks_k\rangle_i)$. Similarly, from (31)–(33), summing over the $z_{\mathcal{I}}$ with indices in $\mathcal{I}_j$ gives the complex conjugate of $g_1\cdots g_n(A_n,\ldots,A_1)_w^c\prod_{k=1}^{n}(-i\langle r_ks_k\rangle_i)$. Thus $\mathcal{I}_i$ and $\mathcal{I}_j$ together give $g_1\cdots g_n[2\prod_{k=1}^{n}(-i\langle r_ks_k\rangle_i)]\text{Re}\{(A_n,\ldots,A_1)_w^c\}$.

This corresponds to (24), but with only the first half of ξ as defined by (25). The rest of ξ comes from the index sets $\mathcal{I}_k$ and $\mathcal{I}_l$. However, the sum of the coefficients of $z_{\mathcal{I}}$ for the same index set in $\mathcal{I}_i$ and $\mathcal{I}_k$ is zero. This is true because, for any complete lowest degree index set, the sum of coefficients for all $z_{\mathcal{I}}$ with the indices divided in any manner between $\mathcal{I}_i$ and $\mathcal{I}_k$ is zero, being the number ways of obtaining that index set from $\mathcal{Y}$ times $\Sigma_{t=1}^{n}(t-1)(-1)^{t-1}$. But by Lemma IV.3, the coefficient of $z_{\mathcal{I}}$ is zero unless the index set comes wholly from $\mathcal{I}_i$ or $\mathcal{I}_k$. Now (40), (35), and (36) tell us that, for an index set in $\mathcal{I}_k$,

$$z_{\mathcal{I}} = \prod_{e=1}^{t} \alpha_{\mathbf{i}(e)} \prod_{k=1}^{n} u_{0,0}^{k} v_{1,0}^{k}$$
$$= g_1 \cdots g_n \prod_{e=1}^{t} (A_{\mathbf{i}(e)(|\mathbf{i}(e)|)}, \ldots, A_{\mathbf{i}(e)(1)})_w \prod_{k=1}^{n} \langle r_k \rangle_i \langle s_k \rangle_i, \tag{45}$$

and from the above argument, this appears in $\mathcal{Y}\langle r_1 \cdots r_n \rangle^c$ with coefficient $-(t-1)!(-1)^{t-1}$. Again, the index sets in $\mathcal{I}_l$ give the complex conjugate of those in $\mathcal{I}_k$. Thus we obtain the remaining half of ξ, which proves (24) for $n \geq 2$. For $n=1$ the constant terms (of order zero in g) in $\mathcal{Y}\langle r \rangle$ do not vanish, but the proof goes through if we consider $\mathcal{Y}(\langle r \rangle - \langle r \rangle_i)$ instead. ∎

V. EXPLORING THE THEOREM

Consider first the simplest case, where $n=1$ and $r=q$. We take $H_{int} = g\delta(t)pA$ throughout this section, so $s=p$. Then (26) and (25) give

$$\langle q \rangle = \langle q \rangle_i + g \, \mathrm{Re}(\xi_q A_w), \tag{46}$$

with

$$\xi_q = -2i(\langle qp \rangle_i - \langle q \rangle_i \langle p \rangle_i),$$

which we have already seen as Eqs. (4) and (5). If we measure the pointer momentum, so $r=p$, we find

$$\langle p \rangle = \langle p \rangle_i + g \, \mathrm{Re}(\xi_p A_w), \tag{47}$$

with

$$\xi_p = -2i(\langle p^2 \rangle_i - \langle p \rangle_i^2),$$

which is equivalent to the result obtained in [7].

For two variables, our theorem for $r_1 = q_1, r_2 = q_2$, is

$$\langle q_1 q_2 \rangle^c = g_1 g_2 \, \mathrm{Re}[\xi_{qq}(A_2, A_1)_w^c], \tag{48}$$

with

$$\xi_{qq} = 2(\langle q_1 \rangle_i \langle p_1 \rangle_i \langle q_2 \rangle_i \langle p_2 \rangle_i - \langle q_1 p_1 \rangle_i \langle q_2 p_2 \rangle_i). \tag{49}$$

The calculations in the Appendix allow one to check (48) and (49) by explicit evaluation; see (A3). Note in passing that, if one writes $\Delta q = \sqrt{\langle (q_1 - \langle q_1 \rangle)^2 \rangle}$, the Cauchy-Schwarz inequality

$$\{\langle q_1 q_2 \rangle^c\}^2 = \{\langle (q_1 - \langle q_1 \rangle)(q_2 - \langle q_2 \rangle) \rangle\}^2$$
$$\leq \langle (q_1 - \langle q_1 \rangle)^2 \rangle \langle (q_2 - \langle q_2 \rangle)^2 \rangle$$

implies a Heisenberg-type inequality

$$\Delta q_1 \Delta q_2 \geq g_1 g_2 \, \mathrm{Re}\{\xi_{qq}(A_2, A_1)_w^c\},$$

relating the pointer noise distributions of two weak measurements carried out at different times during the evolution of the system.

When one or both of the q_k in (48) is replaced by the pointer momentum p_k, we get

$$\langle q_1 p_2 \rangle^c = g_1 g_2 \, \mathrm{Re}[\xi_{qp}(A_2, A_1)_w^c], \tag{50}$$

$$\langle p_1 p_2 \rangle^c = g_1 g_2 \, \mathrm{Re}[\xi_{pp}(A_2, A_1)_w^c], \tag{51}$$

with

$$\xi_{qp} = -2(\langle q_1 p_1 \rangle_i \langle p_2^2 \rangle_i - \langle q_1 \rangle_i \langle p_1 \rangle_i \langle p_2 \rangle_i^2), \tag{52}$$

$$\xi_{pp} = -2(\langle p_1^2 \rangle_i \langle p_2^2 \rangle_i - \langle p_1 \rangle_i^2 \langle p_2 \rangle_i^2). \tag{53}$$

Consider now the special case where ϕ is real with zero mean. Then the very complicated expression for $\langle q_1 q_2 \rangle$ in (A1) reduces to

$$\langle q_1 q_2 \rangle = \frac{g_1 g_2}{2} \mathrm{Re}[(A_2, A_1)_w + (A_1)_w (\bar{A}_2)_w], \tag{54}$$

as shown in [3]. Two further examples from [3] are

$$\langle q_1 q_2 q_3 \rangle = \frac{g_1 g_2 g_3}{4} \mathrm{Re}[(A_3, A_2, A_1)_w + (A_3, A_2)_w (\bar{A}_1)_w$$
$$+ (A_3, A_1)_w (\bar{A}_2)_w + (A_2, A_1)_w (\bar{A}_3)_w], \tag{55}$$

$$\langle q_1 q_2 q_3 q_4 \rangle = \frac{g_1 g_2 g_3 g_4}{8} \mathrm{Re}[(A_4, A_3, A_2, A_1)_w$$
$$+ (A_4, A_3, A_2)_w (\bar{A}_1)_w + \cdots + (A_4, A_3)_w \overline{(A_2, A_1)}_w$$
$$+ \cdots]. \tag{56}$$

We can use these formulas to calculate the cumulant $\langle q_1 \cdots q_n \rangle$ and thus check Theorem IV.1 for this special class of wave functions ϕ. Each formula contains on the right-hand side a leading sequential weak value, but there are also extra terms, such as $(A_1)_w (\bar{A}_2)_w$ in (54) and $(A_2, A_1)_w (\bar{A}_3)_w$ in (55). All these extra terms are eliminated when the cumulant is calculated, and we are left with (24) with $\xi_{q_1, \ldots, q_n} = (1/2)^{n-1}$.

This gratifying simplification depends on the fact that the cumulant is a sum over all partitions. For instance, it does not occur if one uses the covariance instead of the cumulant. To see this, look at the case $n=4$: The term $\langle q_1 q_2 q_3 q_4 \rangle$ in $\mathrm{Cov}(q_1, q_2, q_3, q_4)$, the covariance of pointer positions, gives rise via (56) to weak value terms like $(A_4, A_3)_w \overline{(A_2, A_1)}_w$. However, (18) together with (54)–(56) show that $\mathrm{Cov}(q_1, q_2, q_3, q_4)$ has no other terms that generate any multiple of $(A_4, A_3)_w \overline{(A_2, A_1)}_w$, and consequently this weak value expression cannot be canceled and must be present in $\mathrm{Cov}(q_1, q_2, q_3, q_4)$. This means that there cannot be any equation relating $\mathrm{Cov}(q_1, q_2, q_3, q_4)$ and $\mathrm{Cov}(A_4, A_3, A_2, A_1)_w$. This negative conclusion does not apply to the cumulant $\langle q_1 q_2 q_3 q_4 \rangle^c$, as this includes terms such as $\langle q_1 q_2 \rangle \langle q_3 q_4 \rangle$; see (19).

VI. SIMULTANEOUS WEAK MEASUREMENT

We have treated the interactions between each pointer and the system individually, the Hamiltonian for the kth pointer and system being $H_k = g_k \delta(t - t_k) s_k A_k$, but of course we can equivalently describe the interaction between all the pointers and the system by $H = \Sigma_k g_k \delta(t - t_k) s_k A_k$. For sequential measurements we implicitly assume that all the times t_k are distinct. However, the limiting case where there is no evolution between coupling of the pointers and all the t_k's are equal is of interest and is the *simultaneous* weak measurement considered in [14–16]. In this case, the state of the pointers after post-selection is given by

$$\Psi_{\mathcal{M}} = \langle\psi_f|e^{-i(g_1 s_1 A_1 + \cdots + g_n s_n A_n)}|\psi_i\rangle \phi_1(r_1)\cdots\phi_n(r_n). \quad (57)$$

The exponential $e^{-i(g_1 s_1 A_1 + \cdots + g_n s_n A_n)}$ here differs from the sequential expression $e^{-i g_n s_n A_n} \cdots e^{-i g_1 s_1 A_1}$ in (28) in that each term in the expansion of the latter appears with the operators in a specific order—viz., the arrow order ← as in (22)—whereas in the expansion of the former the same term is replaced by a symmetrized sum over all orderings of operators. For instance, for arbitrary operators X, Y, and Z, the third degree terms in $e^X e^Y e^Z$ include $X^3/3!$, $X^2 Y/2!$, and XYZ, whose counterparts in $e^{(X+Y+Z)}$ are, respectively, $X^3/3!$, $\{X^2Y + XYX + YX^2\}/3!$, and $\{XYZ + XZY + YXZ + YZX + ZXY + ZYX\}/3!$. Apart from this symmetrization, the calculations in Section IV can be carried through unchanged for simultaneous measurement. Thus if we replace the sequential weak value by the *simultaneous weak value* [14–16]

$$(A_{i_k}, \ldots, A_{i_1})_{ws} = \frac{1}{k!}\sum_{\pi\in S_k} (A_{i_{\pi(k)}}, \ldots, A_{i_{\pi(1)}})_w, \quad (58)$$

where the sum on the right-hand side includes all possible orders of applying the operators, we obtain a version of Theorem IV.1 for simultaneous weak measurement:

$$\langle r_1 \cdots r_n\rangle^c = g_1 \cdots g_n \,\mathrm{Re}\{\xi(A_n, \ldots, A_1)^c_{ws}\}. \quad (59)$$

Likewise, relations such (54) and (55), etc., hold with simultaneous weak values in place of the sequential weak values; indeed, these relations were first proved for simultaneous measurement [14,15].

From (58) we see that, when the operators A_k all commute, the sequential and simultaneous weak values coincide. One important instance of this arises when the operators A_k are applied to distinct subsystems, as in the case of the simultaneous weak measurements of the electron and positron in Hardy's paradox [17,18].

When the operators do not commute, the meaning of simultaneous weak measurement is not so obvious. One possible physical interpretation follows from the well-known formula

$$e^{X+Y} = \lim_{N\to\infty} (e^{X/N} e^{Y/N})^N \quad (60)$$

and its analogs for more operators. Suppose two pointers, one for A_1 and one for A_2, are coupled alternately in a sequence of N short intervals (Fig. 3) with coupling strength g_k/N for each interval. This is an enlarged sense of sequential weak measurement [3] in which the same pointer is used repeatedly, coherently preserving its state between couplings. The state after post-selection is

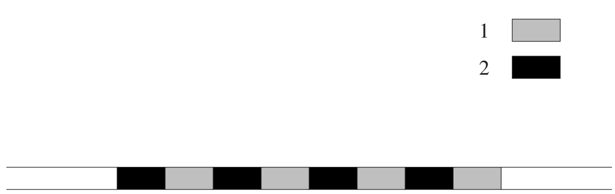

FIG. 3. An approximation to simultaneous weak measurement is obtained by alternately coupling two pointers, 1 and 2 , weakly and for short intervals, many times in succession.

$$\Psi_{\mathcal{M}} = \langle\psi_f|(e^{-i(g_2/N)s_2 A_2} e^{-i(g_1/N)s_1 A_1})^N|\psi_i\rangle \phi_1(r_1)\phi_2(r_2). \quad (61)$$

From (60) we deduce that

$$\Psi_{\mathcal{M}} \approx \langle\psi_f|e^{-i(g_2 s_2 A_2 + g_1 s_1 A_1)}|\psi_i\rangle \phi_1(r_1)\phi_2(r_2). \quad (62)$$

This picture readily extends to more operators A_k.

One can also simulate a simultaneous measurement by averaging the results of a set of sequential measurements with the operators in all orders; in effect, one carries out a set of experiments that implement the averaging in (58). There is then no single act that counts as simultaneous measurement, but weak measurement in any case relies on averaging many repeats of experiments in order to extract the signal from the noise. In a certain sense, therefore, sequential measurement includes and extends the concept of simultaneous measurement. However, if we wish to accomplish simultaneous measurement in a single act, then we need a broader concept of weak measurement where pointers can be reused; indeed, we can go further and consider generalized weak coupling between one time-evolving system and another, followed by measurement of the second system. However, even in this case, the measurement results can be expressed algebraically in terms of the sequential weak values of the first system [3].

VII. LOWERING OPERATORS

Lundeen and Resch [16] showed that, for a Gaussian initial pointer wave function, if one defines an operator a by

$$a_{LR} = \langle p^2\rangle_i^{1/2}\left(q + \frac{ip}{2\langle p^2\rangle_i}\right),$$

then the relationship

$$\langle a_{LR}\rangle = g\langle p^2\rangle_i^{1/2} A_w$$

holds. They argued that a_{LR} can be interpreted physically as a lowering operator, carrying the pointer from its first excited state $|1\rangle$, in number-state notation, to the Gaussian state $|0\rangle$ (despite the fact that the pointer is not actually in a harmonic potential). Although a_{LR} is not an observable, $\langle a_{LR}\rangle$ can be regarded as a prescription for combining expectations of pointer position and momentum to get the weak value.

If instead of a_{LR} one takes

$$a = q + \frac{ip}{2\langle p^2\rangle_i}, \quad (63)$$

then the even simpler relationship

$$\langle a\rangle = gA_w \quad (64)$$

holds. We refer to a as a generalized lowering operator.

Lundeen and Resch also extended their lowering operator concept to simultaneous weak measurement of several observables A_k. Rephrased in terms of our generalized lowering operators a_k defined by (63), their finding [16] can be stated as

$$\langle a_1 \cdots a_n \rangle = g_1 \cdots g_n (A_1, \dots, A_n)_{ws}. \tag{65}$$

This is of interest for two reasons. First, the entire simultaneous weak value appears on the right-hand side, not just its real part; and second, the "extra terms" in the simultaneous analogs of (54)–(56) have disappeared. The lowering operator seems to relate directly to weak values.

We can generalize these ideas in two ways. First, we extend them from simultaneous to sequential weak measurements. Second, instead of assuming the initial pointer wave function is a Gaussian, we allow it be arbitrary; we do this by defining a generalized lowering operator

$$a = q + i\frac{p}{\eta}, \tag{66}$$

with

$$\eta = -i\bar{\xi}_p/\bar{\xi}_q.$$

For a Gaussian ϕ, $\eta = 2\langle p^2 \rangle_i$, so the above definition reduces to (63) in this case. In general, however, ϕ will not be annihilated by a and is therefore not the number state $|0\rangle$ (this state is a Gaussian with complex variance η^{-1}). Nonetheless, there is an analog of Theorem IV.1 in which the whole sequential weak value, rather than its real part, appears:

Theorem VII.1 (cumulant theorem for lowering operators). For $n > 1$,

$$\langle a_1 \cdots a_n \rangle^c = g_1 \cdots g_n \vartheta (A_n, \dots, A_1)^c_w, \tag{67}$$

where ϑ is given by

$$\vartheta = \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \frac{(-1)^{\Sigma i_j}\xi_{r_{i_1},\dots,r_{i_n}}(\bar{\xi}_{r_{1-i_1}} \cdots \bar{\xi}_{r_{1-i_n}})}{2(\bar{\xi}_{p_1} \cdots \bar{\xi}_{p_n})}. \tag{68}$$

For $n=1$ the same result holds, but with the extra term $\langle a \rangle_i$:

$$\langle a \rangle = \langle a \rangle_i + \vartheta g A_w. \tag{69}$$

Proof. Put $r_0 = q$ and $r_1 = p$. Then

$$\langle a_1 \cdots a_n \rangle^c = \langle (q_1 + ip_1/\eta_1) \cdots (q_n + ip_n/\eta_n) \rangle^c$$

$$= \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \frac{(-1)^{\Sigma i_j}\langle r_{i_1} \cdots r_{i_n} \rangle^c (\bar{\xi}_{r_{1-i_1}} \cdots \bar{\xi}_{r_{1-i_n}})}{(\bar{\xi}_{p_1} \cdots \bar{\xi}_{p_n})}$$

$$= g_1 \cdots g_n [\vartheta (A_n, \dots, A_1)^c_w + \varpi \overline{(A_n, \dots, A_1)}^c_w],$$

where we used Theorem IV.1 to get the last line, and where ϑ is given by (68) and ϖ by

$$\varpi = \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \frac{(-1)^{\Sigma i_j}\bar{\xi}_{r_{i_1},\dots,r_{i_n}}(\bar{\xi}_{r_{1-i_1}} \cdots \bar{\xi}_{r_{1-i_n}})}{2(\bar{\xi}_{p_1} \cdots \bar{\xi}_{p_n})}$$

[note the bar over $\bar{\xi}_{r_{i_1},\dots,r_{i_n}}$ that is absent in the definition of ϑ by (68)].

We want to prove $\varpi = 0$, and to do this it suffices to prove that the complex conjugate of the numerator is zero—i.e.,

$$\varpi' = \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} (-1)^{\Sigma i_j}\xi_{r_{i_1},\dots,r_{i_n}}(\xi_{r_{1-i_1}} \cdots \xi_{r_{1-i_n}}) = 0.$$

Let $a_k = \langle q_k s_k \rangle_i$, $b_k = \langle q_k \rangle_i \langle s_k \rangle_i$, $c_k = \langle p_k s_k \rangle_i$, and $d_k = \langle p_k \rangle_i \langle s_k \rangle_i$. Using the definition of ξ in (25), the above equation can be written

$$\varpi'/[2^{n+1}(-1)^n] = \prod_{k=1}^{n} \{a_k(c_k - d_k) - c_k(a_k - b_k)\}$$
$$- \prod_{k=1}^{n} \{b_k(c_k - d_k) - d_k(a_k - b_k)\}$$
$$= \prod (b_k c_k - a_k d_k) - \prod (b_k c_k - a_k d_k) = 0.$$

■

Suppose the interaction Hamiltonian has the standard von Neumann form $H_{int} = gpA$, so $s = p$ in the definition of ξ by Eq. (25). Then for $n = 1$, since $\bar{\xi}_p = \xi_p$ and $\overline{\langle qp \rangle}_i = \langle pq \rangle_i$, $\vartheta = (-i)(\xi_q - \bar{\xi}_q) = (-i)(\langle qp \rangle_i - \langle pq \rangle_i) = 1$, so we get the even simpler result

$$\langle a \rangle = \langle a \rangle_i + gA_w. \tag{70}$$

This is valid for all initial pointer wave functions, and therefore extends Lundeen and Resch's equation (64). It seems almost too simple: there is no factor corresponding to ξ in Eq. (46). However, a dependence on the initial pointer wave function is of course built into the definition of a through η.

For $n > 1$ it is no longer true that $\vartheta = 1$, even with the standard interaction Hamiltonian. However, if in addition $\langle p \rangle_i = 0$, then

$$\vartheta = (-i)^n \prod_{k=1}^{n} (\langle q_k p_k \rangle_i - \langle p_k q_k \rangle_i) = (-i)^n (i)^n = 1.$$

Thus $\langle a_1 \cdots a_n \rangle^c = g_1 \cdots g_n (A_n, \dots, A_1)^c_w$ for all n. Applying the inverse operation for the cumulant, given by Proposition II.1, we deduce the following.

Corollary VII.2. If $\langle p \rangle_i = 0$—e.g., if the initial pointer wave function ϕ is real—then for $n > 1$

$$\langle a_1 \cdots a_n \rangle = g_1 \cdots g_n (A_n, \dots, A_1)_w. \tag{71}$$

This is the sequential weak value version of the result for simultaneous measurements, Eq. (65), but is more general than the Gaussian case treated in [16].

We might be tempted to try to repeat the above argument for pointer positions q_k instead of the lowering operators a_k by applying the anticumulant to both sides of (24). This fails, however, because of the need to take the real part of the weak values; in fact, this is one way of seeing where the extra terms come from in (54)–(56) and their higher analogs.

Note also that (71) does not hold for general ϕ, since then different subsets of indices may have different values of ϑ.

VIII. DISCUSSION

The procedure for sequential weak measurement involves coupling pointers at several stages during the evolution of the system, measuring the position (or some other observable) of each pointer and then multiplying the measured values together. In [3] it was argued that we would really like to measure the product of the values of the operators $A_1, \ldots, A_n$ and that this corresponds to the sequential weak value $(A_n, \ldots, A_1)_w$. Multiplication of the values of pointer observables is the best we can do to achieve this goal. However, this brings along extra terms, such as $(A_1)_w(\bar{A}_2)_w$ in (54), which are an artifact of this method of extracting information. From this perspective, the cumulant extracts the information we really want.

In [3], a somewhat idealized measuring device was being considered, where the pointer position distribution is real and has zero mean. When the pointer distribution is allowed to be arbitrary, the expressions for $\langle q_1 \cdots q_n \rangle$ become wildly complicated [see, for instance, (A1)]. Yet the cumulant of these terms condenses into the succinct Eq. (24) with all the complexity hidden away in the one number ξ. Why does the cumulant have this property?

Recall that the cumulant vanishes when its variables belong to two independent sets. The product of the pointer positions $q_1, \ldots, q_n$ will include terms that come from products of disjoint subsets of these pointer positions, and the cumulant of these terms will be sent to zero, by Lemma II.3. For instance, with $n=2$, the pointers are deflected in proportion to their individual weak values, according to (4), and the cumulant subtracts this component, leaving only the component that arises from the $O(g^2)$ influence of the weak measurement of A_1 on that of A_2. The subtraction of this component corresponds to the subtraction of the term $(A_1)_w(\bar{A}_2)_w$ from (54). In general, the cumulant of pointer positions singles out the maximal correlation involving all the q_i and the theorem tells us that this is directly related to the corresponding "maximal correlation" of sequential weak values, $(A_n, \ldots, A_1)^c$, which involves all the operators.

The simple relationship between the cumulants of weak values and pointer positions also suggests an experimental strategy. Suppose one wishes to estimate weak values of some system from measurements of coupled pointers. One can evaluate the cumulants of the pointer positions, deduce the cumulants of weak values, and then derive the individual weak values by applying the anticumulant (12). In this way one avoids having to unravel the immensely complicated relations that occur, for example, in (A1). This procedure is further simplified by using lowering operators, since then cumulants of pointer positions and weak values are directly proportional (67) without the intervening operation of taking the real part (24).

Finally, we emphasize an important feature of our theorem. There are many choices of pointer observable $r(p,q)$—e.g., position, momentum, or some Hermitian combination of them—and likewise many ways of coupling the pointer with the system, which can be via a Hamiltonian $H_{int}=gs(p,q)A$ with any Hermitian $s(p,q)$. Different choices of these variables r and s lead only to a different multiplicative constant ξ in front of $(A_n, \ldots, A_1)^c_w$ in (24). We always extract the same function of sequential weak values, $(A_n, \ldots, A_1)^c_w$, from the system. This argues both for the fundamental character of sequential weak values and also for the key role played by their cumulants.

ACKNOWLEDGMENTS

I am indebted to J. Åberg for many discussions and for comments on drafts of this paper; I thank him particularly for putting me on the track of cumulants. I also thank A. Botero, P. Davies, R. Jozsa, R. Koenig, and S. Popescu for helpful comments. A preliminary version of this work was presented at a workshop on "Weak Values and Weak Measurement" at Arizona State University under the aegis of the Center for Fundamental Concepts in Science, directed by P. Davies.

APPENDIX: AN EXPLICIT CALCULATION

To calculate $\langle q_1 q_2 \rangle$ for arbitrary pointer wave functions ϕ_1 and ϕ_2, we use (28) to determine the state of the two pointers after the weak interaction and then evaluate the expectation using (29), keeping only terms up to order g^2. We define

$$\mu_k = \langle q_k \rangle_i, \quad \nu_k = \langle p_k \rangle_i, \quad \zeta_k = \langle p_k^2 \rangle_i,$$

$$\rho_k = \langle q_k p_k \rangle_i, \quad \sigma_k = \langle q_k p_k^2 \rangle_i, \quad \tau_k = \langle p_k q_k p_k \rangle_i.$$

Then, expanding the exponential in (28) and substituting Ψ in (29) gives, up to order g^2:

$$\begin{aligned}\langle q_1 q_2 \rangle = {} & \mu_1\mu_2 - ig_1\{[(A_1)_w - (\bar{A}_1)_w]\mu_1\nu_1\mu_2 - (\bar{A}_1)_w\bar{\rho}_1\mu_2 + (A_1)_w\rho_1\mu_2\} - ig_2\{[(A_2)_w - (\bar{A}_2)_w]\mu_1\mu_2\nu_2 - (\bar{A}_2)_w\mu_1\bar{\rho}_2 \\ & + (A_2)_w\mu_1\rho_2\} + g_1^2\left\{|(A_1)_w|^2(\tau_1\mu_2 - \mu_1\zeta_1\mu_2) + [(A_1^2)_w + (\bar{A}_1^2)_w]\frac{\mu_1\zeta_1\mu_2}{2}\right\} - g_1^2\left\{[(A_1)_w - (\bar{A}_1)_w]^2\mu_1\nu_1^2\mu_2 + (A_1^2)_w\frac{\sigma_1\mu_2}{2}\right. \\ & \left. + (\bar{A}_1^2)_w\frac{\bar{\sigma}_1\mu_2}{2}\right\} + g_2^2\left\{|(A_2)_w|^2(\mu_1\tau_2 - \mu_1\mu_2\zeta_2) + [(A_2^2)_w + (\bar{A}_2^2)_w]\frac{\mu_1\mu_2\zeta_2}{2}\right\} - g_2^2\left\{[(A_2)_w - (\bar{A}_w)_2]^2\mu_1\mu_2\nu_2^2\right. \\ & \left. + (A_2^2)_w\frac{\mu_1\sigma_2}{2} + (\bar{A}_2^2)_w\frac{\mu_1\bar{\sigma}_2}{2}\right\} + g_1g_2\{(A_1)_w(\bar{A}_2)_w\rho_1\bar{\rho}_2 + (\bar{A}_1)_w(A_2)_w\bar{\rho}_1\rho_2 - (A_2,A_1)_w\rho_1\rho_2 - \overline{(A_2,A_1)}_w\bar{\rho}_1\bar{\rho}_2\} \\ & - g_1g_2\{2[(A_1)_w - (\bar{A}_1)_w][(A_2)_w - (\bar{A}_2)_w]\mu_1\nu_1\mu_2\nu_2\} + g_1g_2\{[(A_2,A_1)_w + \overline{(A_2,A_1)}_w - (A_1)_w(\bar{A}_2)_w\end{aligned}$$

$$- (\bar{A}_1)_w(A_2)_w]\mu_1\nu_1\mu_2\nu_2\} + g_1^2\{[(A_1)_w - (\bar{A}_1)_w](A_1)_w\nu_1\rho_1\mu_2 - [(A_1)_w - (\bar{A}_1)_w](\bar{A}_1)_w\nu_1\bar{\rho}_1\mu_2\} + g_2^2\{[(A_2)_w - (\bar{A}_2)_w] \times(\times A_2)_w\mu_1\nu_2\rho_2 - [(A_2)_w - (\bar{A}_2)_w](\bar{A}_2)_w\mu_1\nu_2\bar{\rho}_2\} + g_1g_2\{[(A_1)_w - (\bar{A}_1)_w](A_2)_w\mu_1\nu_1\rho_2 - [(A_1)_w - (\bar{A}_1)_w](\bar{A}_2)_w\mu_1\nu_1\bar{\rho}_2\} + g_1g_2\{[(A_2)_w - (\bar{A}_2)_w](A_1)_w\rho_1\mu_2\nu_2 - [(A_2)_w - (\bar{A}_2)_w](\bar{A}_1)_w\bar{\rho}_1\mu_2\nu_2\}. \tag{A1}$$

To calculate the cumulant $\langle q_1, q_2\rangle^c = \langle q_1 q_2\rangle - \langle q_1\rangle\langle q_2\rangle$ we need $\langle q\rangle$ up to order g^2:

$$\langle q\rangle = \mu + ig\{A_w(\mu\nu - \rho) - \bar{A}_w(\mu\nu - \bar{\rho})\} + g^2|A_w|^2(\tau - \mu\zeta + 2\mu\nu^2 - \nu\rho - \nu\bar{\rho}) + g^2 \times\left\{(A^2)_w\left(\frac{\mu\zeta}{2} - \frac{\sigma}{2}\right) - (\bar{A}^2)_w\left(\frac{\mu\zeta}{2} - \frac{\bar{\sigma}}{2}\right) + (A_w)^2(\nu\rho - \mu\nu^2) + (\bar{A}_w)^2(\nu\bar{\rho} - \mu\nu^2)\right\}. \tag{A2}$$

Substituting from (A1) and (A2) a radical simplification occurs:

$$\langle q_1 q_2\rangle^c = g_1 g_2\{(A_2, A_1)_w - (A_1)_w(A_2)_w\}(\mu_1\nu_1\mu_2\nu_2 - \rho_1\rho_2) + \text{complex conjugate}. \tag{A3}$$

This, of course, is what Theorem IV.1 tells us.

[1] Y. Aharonov and D. Rohrlich, *Quantum Paradoxes* (Wiley-VCH, Weinheim, Germany, 2005).
[2] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988).
[3] G. Mitchison, R. Jozsa, and S. Popescu, Phys. Rev. A **76**, 062105 (2007).
[4] O. Oreshkov and T. A. Brun, Phys. Rev. Lett. **95**, 110409 (2005).
[5] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **59**, 1070 (1999).
[6] J. von Neumann, *Mathematical Foundations of Quantum Theory* (Princeton University Press, Princeton, 1983).
[7] R. Jozsa, Phys. Rev. A **76**, 044103 (2007).
[8] M. Kendall and A. Stuart, *The Advanced Theory of Statistic* (Charles Griffin, London, 1977), Vol. 1.
[9] A. Royer, J. Math. Phys. **24**, 897 (1983).
[10] D. L. Zhou, B. Zeng, Z. Xu, and L. You, Phys. Rev. A **74**, 052110 (2006).
[11] T. M. Apostol, *Introduction to Analytic Number Theory* (Springer-Verlag, New York, 1976).
[12] J. K. Percus, Commun. Math. Phys. **40**, 283 (1975).
[13] B. Simon, *Functional Integration and Quantum Physics* (Academic Press, New York, 1979).
[14] K. J. Resch and A. M. Steinberg, Phys. Rev. Lett. **92**, 130402 (2004).
[15] K. J. Resch, J. Opt. B: Quantum Semiclassical Opt. **6**, 482 (2004).
[16] J. S. Lundeen and K. J. Resch, Phys. Lett. A **334**, 337 (2005).
[17] L. Hardy, Phys. Rev. Lett. **68**, 2981 (1992).
[18] Y. Aharonov, A. Botero, S. Popescu, B. Reznik, and J. Tollaksen, in *Proceedings of NATO ARW Mykonos 2000 Decoherence and its Implications in Quantum Computation and Information Transfer: Proceedings of the NATO advanced research workshop, Mykonos, Greece, 2000*, edited by A. Gonis and P. Turchi (IOS Press, Bristol, 1991).