Invariants in copolar interferometry: An Abelian gauge theory

Nithyanandan Thyagarajan[*]

CSIRO, Space & Astronomy, P. O. Box 1130, Bentley, WA 6102,
Australia and National Radio Astronomy Observatory, 1003 Lopezville Rd, Socorro, New Mexico 87801, USA

Rajaram Nityananda[†]

Azim Premji University, Bikkanahalli Main Road, Sarjapura, Bengaluru 562125, Karnataka, India

Joseph Samuel[‡]

International Centre for Theoretical Sciences, Bengaluru 560089, Karnataka, India
and Raman Research Institute, Bengaluru 560080, Karnataka, India

An N-element interferometer measures correlations among pairs of array elements. Closure invariants associated with closed loops among array elements are immune to multiplicative, element-based ("local") corruptions that occur in these measurements. Till recently, it has been unclear how a complete set of independent invariants can be analytically determined. We view the local, element-based corruptions in copolar correlations as gauge transformations belonging to the gauge group $GL(1, \mathbb{C})$. Closure quantities are then naturally gauge invariant. We use this to provide a simple and effective formalism and identify the complete set of independent closure invariants from copolar interferometric correlations using only quantities defined on $(N-1)(N-2)/2$ elementary and independent triangular loops. The $(N-1)(N-2)/2$ closure phases and $N(N-3)/2$ closure amplitudes (totaling $N^2 - 3N + 1$ real invariants), familiar in astronomical interferometry, naturally emerge from this formalism, which unifies what has required separate treatments until now. We do not require autocorrelations but can easily include them if reliably measured. This unified view clarifies issues relating to noise and inference of object model parameters. It also allows us to extend the rule of parallel transport associated with Pancharatnam phase in optics to apply to amplitudes as well. The framework presented here extends to $GL(2, \mathbb{C})$ for full polarimetric interferometry as presented in a companion paper, which generalizes and clarifies earlier work. Our findings are relevant to state of the art copolar and full polarimetric very long baseline interferometry measurements to determine features very near the event horizons of black holes at the centers of M87, Centaurus A, and the Milky Way.

I. INTRODUCTION

Interferometry, which began with Young's double slit experiment, has blossomed over the years into a technique now widely used in physics, biology, and astronomy. Examples include the use of astronomical interferometry in the first determination of the double-lobed morphology of Cygnus A [1], the first imaging of the event horizon of a black hole in the center of the M87 galaxy [2–8] and inner jet structures around the black hole in Centaurus A [9], determination of crystal structures ([10–14] and references therein), seismic imaging ([15] and references therein), and remote sensing using radar and sonar ([16,17] and references therein).

In each of these applications, accurate measurement of both the amplitude and phase of the coherence (represented as complex values) is critical for success. Often, accurate measurements of the complex-valued correlations is made difficult by corruptions introduced by the propagation medium, and nonideal behavior of the instrument as well as of the measurement process. For example, in radio interferometry, the electromagnetic wave fronts are corrupted by ionospheric and tropospheric turbulences at low and high frequencies, respectively, as well as by corrupting factors in the array element responses. In optical interferometry, the atmospheric turbulence and imperfections in the telescope's surface geometry tend to destroy the phase information.

These element-based effects can be calibrated if there is a standard object of known morphology available [18,19]. An alternative strategy is self-calibration [20,21], which is a self-consistent scheme which uses iteration and feedback to

[*]Nithyanandan.Thyagarajan@csiro.au; https://tnithyanandan.wordpress.com
[†]rajaram.nityananda@gmail.com
[‡]sam@rri.res.in

043019-1

successively refine the image starting with an initial model. But this is not always possible at the desired level of accuracy when the signal-to-noise ratio or the *a priori* knowledge of the calibrator object is inadequate. Therefore, there is considerable interest in calibration-independent quantities that are unaffected by these element-specific factors, and are hence true observable properties of the object's structure.

Such interferometric invariants are typically constructed using the product of pairwise correlations measured on a closed loop of array elements, and we call them *closure invariants*. In copolar interferometry (correlations between measurements of the same polarization at all the array elements) at radio and optical wavelengths, they are widely known as closure phases and amplitudes ([18,19,22] and references therein). The analog of closure invariants include the triplet and quartet invariants in x-ray crystallography ([10–14] and references therein), and Bargmann invariants in quantum mechanics [23]. A detailed geometric insight into closure phase and its connections to various disciplines is provided in [24].

In astronomical interferometry, closure quantities can be traced to [25,26]. Since then, they have been invaluable tools for interferometry at optical [22,27–30] and radio frequencies ([20,31–33] and references therein). This is particularly true of high spatial resolution applications using very long baseline interferometry (VLBI) at radio frequencies. A famous example is the recent imaging of the event horizon of the supermassive black hole at the center of the M87 galaxy [2–8]. Recently, new approaches of probing the structure formation in the intergalactic medium through application of closure phases on faint spectral line emission from the early Universe during the cosmic reionization epoch ($z \gtrsim 6$) are also being explored [34–36].

In the context of copolar correlations in astronomical interferometry, a detailed mathematical approach to determining the number of generic closure invariants (including closure phases and closure amplitudes), and implications for the resulting signal-to-noise ratios was presented by [37]. Recently, first steps in the extension of closure invariants (called "closure traces") to correlations of full polarimetric antenna measurements were taken in [38], which relied on including antenna autocorrelations to derive an independent set of closure traces involving four correlations.

In this and an accompanying paper [39] (hereafter papers I and II, respectively), we establish a formalism using a combination of group-theoretic and linear algebraic approaches that advances the previous work [37,38]. In contrast to [38], we do not rely on the use of autocorrelations (which in radio astronomy tend to be susceptible to significant noise biases, besides instabilities caused by radio frequency interference and instrumental systematics), but we can incorporate them if reliably measured, and derive a complete and independent set of closure invariants using

gauge theory. In this paper, we employ the $GL(1, \mathbb{C})$ gauge group[1] and its associated gauge freedom to treat and derive the complete and independent set of closure invariants in the copolar case. The closure phases and closure amplitudes familiar in radio interferometry emerge naturally from our formalism which treats triangular loops as fundamental units, and thus unifies the prescription for obtaining closure phases and closure amplitudes. In paper II, we build on the foundations presented in this paper and provide a formalism using the $GL(2, \mathbb{C})$ gauge group and its associated gauge freedoms for deducing the complete and independent set of closure invariants in the general case of full polarimetric interferometry. In both papers, we also confirm our findings with a parallel and independent viewpoint using conventional linear algebra along with numerical simulations. We also generalize the analysis to provide a prescription for extracting all the independent closure invariants in an N-element interferometer array. These closure invariants represent true observables about the system under observation, which in the case of interferometric imaging in astronomy corresponds to the true physical properties of the target object's morphology.

This paper is structured as follows. Section II lays out the copolar interferometry context within which we seek a full set of independent invariants. In Sec. III, we present the expected number of independent closure invariants through a dimension count analysis. Section IV develops a formalism that provides multiple but equivalent methods to produce a complete and independent set of closure invariants, which can be readily identified with closure phases and closure amplitudes familiar in radio interferometry. These invariants are listed explicitly in the case of three, four, and N array elements. Section V summarizes this work. In Appendix A, we describe a scheme for reliably measuring autocorrelations as the coincidence limit of cross-correlations if an extra short-spaced pair of elements is available. In Appendix B, we place our work in relation to the widely used "self-calibration" technique, which implicitly uses closure quantities. Appendix C describes the numerical scheme used to confirm our analytic results, and Appendix D brings out how our approach clarifies some aspects of the choice of invariants as it relates to the effects of noise and to object model parameters. In Appendix E, we highlight the explicit connection between the closure phase and Pancharatnam phase [40,41] in optics. Motivated by the radio astronomy application, we extend the Pancharatnam rule to include amplitudes as well.

II. COPOLAR INTERFEROMETRY

Consider an interferometer array with N elements labeled by the indices $a, b = 0, \ldots, N - 1$. Each element, a, in the

---

[1]In general, $GL(n, \mathbb{C})$ refers to the group of nonsingular $n \times n$ matrices under multiplication, so $GL(1, \mathbb{C})$ is the group of nonzero complex numbers under multiplication.

array measures the amplitude and phase of the same polarization (copolar, or scalar) state of electric field, e_a (represented by a complex number), incident on it, which is stochastic. The true correlation of the stochastic fields between pairs of array elements is, by definition, obtained by cross multiplication and averaging, $S_{ab} := \langle e_a e_b^\dagger \rangle$, where $\dagger$ denotes the conjugate transpose operator, which reduces to complex conjugation for $GL(1, \mathbb{C})$ matrices in the copolar interferometric case dealt in this paper.

The complex electric fields, e_a, may be subject to arbitrary gains G_a at each element, where G_a are nonzero complex numbers representing both amplitude and phase distortion of the measured signal. In ideal conditions, all the gains G_a would be unity and the measured signal at each baseline would accurately reflect the true signal correlation, S_{ab}, of emission from the object. In fact, the measured correlation, C_{ab}, which is a corrupted form of the true correlation, can be written as

$$C_{ab} = G_a \langle e_a e_b^\dagger \rangle G_b^\dagger = G_a S_{ab} G_b^\dagger \qquad (1)$$

because of local gain distortions at each element. Generally, the cross-correlations are complex-valued with nonzero amplitudes. From the definition, cross-correlations satisfy $C_{ab} = C_{ba}^\dagger$. The auto-correlations, $A_{aa} := C_{aa}$, are real and positive. Our objective is to construct quantities which are immune to the gain distortions and actually reflect true properties of the object's morphology, rather than local conditions at the individual elements of the interferometer.

We relate this problem to the "gauge" theories of fundamental forces like electromagnetism. The local gains G_a are regarded as gauge transformations and our objective is to isolate a maximal set of gauge invariant quantities which we will call closure invariants. This term encompasses both closure amplitudes and closure phases. Our treatment places them on an equal footing by working with a $GL(1, \mathbb{C})$ gauge group.

III. COUNTING ARGUMENTS

Let us first do a dimensional count to see how many closure invariants we would expect to find. The number of measured cross-correlations is $N(N-1)/2$, this being the number of baselines or element pairs, i.e., the number of nonrepeating combinations among N elements taken two at a time. Since each cross-correlation is a complex number we have $N(N-1)$ real numbers. If n_A autocorrelations (described by n_A real numbers) are also measured, we would have to add n_A to the above count. Without loss of generality, we can assume that n_A is either zero or one.[2] In radio astronomy, autocorrelation measurements are

unreliable because they are dominated by nonastronomical systematics. Our formalism for closure invariants works with or without autocorrelations.

The unknown element-based gains are N complex numbers, but note that there may be sets of gains G_a which do not affect the correlations. That is, they satisfy $G_a C_{ab} G_b^\dagger = C_{ab}$ for all a, b. This yields $G_a G_b^\dagger = 1$ and so $G_a G_a^\dagger G_b G_b^\dagger = 1$. If the correlations themselves are unchanged, then so will the triple products of correlations. Thus, using $G_a G_b^\dagger G_b G_c^\dagger G_c G_a^\dagger = 1$ (assuming $N \geq 3$, and distinct a, b, c) gives us $G_c G_c^\dagger = 1$ for all c. This means $G_a = e^{i\theta}$ for all a, which is simply an overall phase factor that cancels out in Eq. (1) and does not affect the correlations. Thus, counting out this single phase parameter from the $2N$ real parameters in the gains gives us a count of effectively $2N - 1$ unknown real numbers in the gains.

Assuming none of the measured correlations is redundant (they are all independent of each other) and all of them are used, the number of closure invariants (that are independent of any choice of the unknown gains) is the difference

$$N(N-1) + n_A - (2N - 1) = N^2 - 3N + 1 + n_A. \qquad (2)$$

Note that this equals the number of closure phases, $(N-1)(N-2)/2$, plus the number of closure amplitudes, $N(N-3)/2 + n_A$ [18,37]. For example, this equals 1 (one closure phase) or 5 (three closure phases and two closure amplitudes) when $n_A = 0$ for $N = 3$ or $N = 4$, respectively, and 2 (one closure phase and one closure amplitude) or 6 (three closure phases and three closure amplitudes) when $n_A = 1$ for $N = 3$ and $N = 4$, respectively. As N becomes large, the structural information about the object that can be extracted using closure phases and closure amplitudes asymptotically approaches that which can be extracted using correlations.

IV. FORMALISM FOR CLOSURE INVARIANTS

From a gauge theory perspective, we can regard each of the N elements $a = 0, \ldots, N-1$ as vertices in a graph. The gains that represent element-based corruption factors are local variables and multiplication by the gain factor results in a gauge transformation at each vertex. Each baseline is an edge ab or link carrying "connection" variables that are bilocal since they are defined on the link connecting two vertices. These are the cross-correlations. In the language of gauge theory, each triangle, $\Delta_{(abc)} \equiv (a, b, c)$ is called an "elementary plaquette" and defines a "Wilson loop." The Wilson loop of any closed circuit can be determined from those of the elementary triangular plaquettes. Our objective is to list a complete and independent set of elementary Wilson loops. These are the closure invariants of radio astronomy.

In a $U(1)$ gauge theory like electromagnetism, the gauge group would be unitary and complex conjugation in Eq. (1)

would be the same as inversion. One can then form independent closure invariants from each elementary triangular plaquette abc by simply multiplying correlations around the triangle, as in $\mathcal{B}_{abc} = C_{ab}C_{bc}C_{ca}$. ($\mathcal{B}_{abc}$ is called the "bispectrum" in interferometry). However, our gauge group $\mathrm{GL}(1, \mathbb{C})$ is *not* unitary. Using $\mathcal{G} = \{G_a\}$, $a = 0, \ldots, N-1$ to denote a general gauge transformation at every vertex, a, on the graph, the bispectrum and other higher order products ("polyspectra") of a set of C_{ab} around a closed loop, Γ, transform as

$$\mathcal{P}_\Gamma \overset{\mathcal{G}}{\mapsto} |G_a|^2 |G_b|^2 \ldots \mathcal{P}_\Gamma. \qquad (3)$$

The phase of $\mathcal{P}_\Gamma$ is invariant under gauge transformations $\mathcal{G} = \{G_a\}$, $a = 0, \ldots, N-1$ and this is familiarly known as the closure phase. However, because the gauge group is not unitary, the amplitude of $\mathcal{P}_\Gamma$ is not invariant as it still depends on the gain amplitudes, $|G_a|$.

Similarly, one can define closure amplitudes for every even-edged loop. For example, given the loop $\Gamma = abcda$, the quantity

$$\mathcal{Q}_\Gamma = C_{ab}C_{bc}^{-1}C_{cd}C_{da}^{-1} \qquad (4)$$

transforms as

$$\mathcal{Q}_\Gamma \overset{\mathcal{G}}{\mapsto} G_a G_a^{\dagger-1} G_b G_b^{\dagger-1} G_c G_c^{\dagger-1} G_d G_d^{\dagger-1} \mathcal{Q}_\Gamma. \qquad (5)$$

It follows, since $|G_a G_a^{\dagger-1}| = 1$ that $|\mathcal{Q}_\Gamma|$, the modulus of $\mathcal{Q}_\Gamma$, is a closure invariant:

$$|\mathcal{Q}_\Gamma| \overset{\mathcal{G}}{\mapsto} |\mathcal{Q}_\Gamma|. \qquad (6)$$

However, the phase of $\mathcal{Q}$ is *not* an invariant because it still depends on the phases of G_a at the loop vertices.

The virtues of $\mathcal{P}_\Gamma$ and $\mathcal{Q}_\Gamma$ can be combined in a single quantity $\mathcal{C}_\Gamma$ by defining a hat operator, $\widehat{Z} = (Z^\dagger)^{-1}$, on nonzero complex numbers. We introduce the term *covariants* as the set of even number of products of correlations around a closed loop, Γ, with even numbered terms "hatted" starting with the second. For example, a covariant on a 4-vertex loop can be written as $\mathcal{C}_\Gamma = C_{ab}\widehat{C}_{bc}C_{cd}\widehat{C}_{da}$. Hence, covariants transform as

$$\mathcal{C}_\Gamma \overset{\mathcal{G}}{\mapsto} G_0 \mathcal{C}_\Gamma G_0^{-1}. \qquad (7)$$

Because the gauge group of G_a, $\mathrm{GL}(1, \mathbb{C})$, is Abelian,

$$\mathcal{C}_\Gamma \overset{\mathcal{G}}{\mapsto} \mathcal{C}_\Gamma. \qquad (8)$$

Thus, the covariants are invariant[3] under the gauge transformation effected by the element-based corruption, and form the closure invariants we are seeking in copolar interferometry. Some of the notation above is deliberately

kept general so that it transfers easily to the more difficult non-Abelian case discussed in paper II.

A. Triangular plaquettes

The need to have an even number of correlations in the product around a closed loop appears to impose the requirement of elementary quadrilaterals rather than triangles. However, it is unclear beforehand which even-vertexed loops will provide us with a complete set of independent invariants. We show below that such a set can be obtained by considering variables on elementary triangular plaquettes that are independent, rather than quadrilateral plaquettes. The latter can be entirely derived from the former.

The independent triangles in an N-element interferometer can be obtained by fixing a base vertex (for example, 0) and choosing all the triangles that contain this vertex [18]. There are $N_\Delta = (N-1)(N-2)/2$ such independent triangles. We construct elementary triangular plaquette variables pinned at vertex 0 that includes just three elements $(0, a, b)$: $\mathcal{A}_{ab} = C_{0a}\widehat{C}_{ab}C_{b0}$. This quantity is neither a bispectrum nor a covariant, but acts as a building block that can be used to construct all the closure invariants. These triangular variables are clearly independent and complete as all closed loops can be decomposed into triangles and each distinct triangle is independent. Around the closed loop pinned at vertex 0 (Γ_0), they undergo gauge transformations with G_0 as

$$\mathcal{A}_{ab} \overset{\mathcal{G}}{\mapsto} |G_0|^2 \mathcal{A}_{ab}. \qquad (9)$$

We term such quantities as *advariants* in this paper. A special advariant is $\mathcal{A}_0 = A_{00}$, the autocorrelation. The notation clarifies that this autocorrelation advariant is a local variable based at 0. Note that advariants have an advantage that their gauge freedom is restricted to just a scaling by a single $|G_0|^2$ in contrast to polyspectra (including bispectrum) whose gauge transformations depend on $|G_a|$ of all the vertices in the loop. Thus, $|G_0|^2$ is the only unknown quantity in the advariants that has to be eliminated to arrive at invariant quantities. However, the phase of a triangular advariant is equal to the phase of the corresponding bispectrum, and thus gives the closure phase [25], which is an invariant.

B. Complete and independent set of invariants

1. Method 1

Consider all the N_Δ elementary triangular plaquettes pinned at vertex 0. Each of them gives us a complex number $\mathcal{A}_{ab}$. These complex numbers scale as in Eq. (9). The real

---

[3]Note that a four-edged covariant essentially resembles a *cross ratio*, which is a well-known invariant in projective geometry ([42,43] and references therein).

and imaginary parts of these complex numbers give us $M = 2N_\Delta$ real numbers, $\{\widetilde{I}_m\}$, $m = 1, \ldots, M$, all of which transform by the unknown scale, $|G_0|^2$, under gauge transformations. The *ratios* of these $\widetilde{I}_m$ with respect to any of them ($\widetilde{I}_0$, for example) or some symmetric combination ($[\sum_m \widetilde{I}_m^2]^{1/2}$, for example), will eliminate this unknown scale, and thus give us $M - 1$ real closure invariants, $\{\widetilde{I}_k/\widetilde{I}_0\}$, $k = 2, \ldots, M$. Their number is $N^2 - 3N + 1$, which is in exact agreement with the expectation in Eq. (2). This gives us a complete and independent set of closure invariants. Also, there is no need for autocorrelations in this construction. If we were to add A_{00} as a measured quantity, we would gain a single real invariant because $\mathcal{A}_0 = A_{00}$ being real and positive, would only contribute one more real number to $\{\widetilde{I}_m\}$. Note that the practical problem of systematic effects affecting autocorrelations can be mitigated by the use of cross-correlations between a pair of closely spaced elements in cases where the angular size of the object is much smaller than the angular resolution corresponding to the short spacing, and is described in Appendix A.

While this method of taking ratios completely solves the problem we sought to address, namely, the determination of a complete and independent set of copolar interferometric invariants, we provide below an alternative formalism that allows us to better visualize the closure phases and amplitudes, as they are familiarly known in radio interferometry, in terms of combinations of the invariants determined here.

2. Method 2

The elementary covariant quantity defined earlier as the closed loop product of an even number of correlations with alternate terms hatted starting with the second (a minimum of four terms) can be equivalently defined as the multiplication of an even number (minimum of two) advariants with every alternate term hatted starting with the second. For example, consider two triangles, $\Delta_1 \equiv \Delta_{(0,1,2)}$ (the base triangle) and $\Delta_\ell \equiv \Delta_{(0,a,b)}$ chosen from the set of N_Δ triangles, both pinned at vertex 0, and their reversed forms, $\nabla_1 \equiv \Delta_{(0,2,1)}$ and $\nabla_\ell \equiv \Delta_{(0,b,a)}$. The covariant, which is an invariant, can now be expressed by pairing the advariants on these triangles as

$$\mathcal{I}_{\Delta_1;\Delta_\ell} = \mathcal{C}_{\Delta_1;\Delta_\ell} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_\ell} \qquad (10)$$

$$\text{or,} \quad \mathcal{I}_{\Delta_1;0} = \mathcal{C}_{\Delta_1;0} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_0, \qquad (11)$$

which are in general complex valued. In the copolar case studied in this paper, since covariants are indeed invariants, they will be used interchangeably. However, the same is not true in the full polarimetric case presented in paper II.

The following properties of covariants are noted:

(i) $\mathcal{A}_{\Delta_1}$ scales as $|G_0|^2$, whereas $\widehat{\mathcal{A}}_{\Delta_\ell}$ and $\widehat{\mathcal{A}}_0$ scale as $|G_0|^{-2}$, thereby canceling the unknown scale factor

in Eqs. (10) and (11), and making the covariant an invariant.

(ii) A covariant formed with Δ_1 and Δ_ℓ has the same amplitude as that formed with Δ_1 and ∇_ℓ. That is, $|\mathcal{I}_{\Delta_1;\Delta_\ell}| = |\mathcal{I}_{\Delta_1;\nabla_\ell}|$, and thus do not provide independent amplitude information. Similarly, $|\mathcal{I}_{\Delta_\ell;\Delta_1}| = |\mathcal{I}_{\Delta_1;\Delta_\ell}|^{-1}$.

(iii) If ϕ_{Δ_1} and ϕ_{Δ_ℓ} are the phases of the advariants (closure phases), then the phases of $\mathcal{I}_{\Delta_1;\Delta_\ell}$ and $\mathcal{I}_{\Delta_1;\nabla_\ell}$ are $\phi_{\Delta_1} + \phi_{\Delta_\ell}$ and $\phi_{\Delta_1} - \phi_{\Delta_\ell}$, respectively.

(iv) Each invariant, $\mathcal{I}_{\Delta_1;\Delta_\ell}$, is complex valued and thus contains two real-valued invariants, except when $\Delta_\ell = \Delta_1$. In that case, $|\mathcal{I}_{\Delta_1;\Delta_1}| = 1$, and therefore, contains only one real invariant, namely, its phase which is twice the closure phase, $2\phi_{\Delta_1}$.

Using the above properties, a set of covariants (invariants) can be constructed by pairing all the independent triangular advariants, $\mathcal{A}_{\Delta_\ell}$ with the advariant of the base triangle, $\mathcal{A}_{\Delta_1}$, as $\mathcal{I}_{\Delta_1 \Delta_\ell} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_\ell}$, $\ell = 1, \ldots, N_\Delta$. Each of these complex invariants yields two real invariants for $\ell \neq 1$ and only one real invariant for $\ell = 1$ as noted above. Therefore, the total number of real invariants is $2N_\Delta - 1 = N^2 - 3N + 1$. The presence of an autocorrelation measurement will increase this count to $N^2 - 3N + 2$ due to one additional real-valued amplitude invariant in $\mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_0$. Therefore, this method provides $N^2 - 3N + 1 + n_A$ real invariants, which is a complete and independent set, and is consistent with the first method using ratios as well as with the dimension count analysis in Sec. III. We also verified these results numerically as detailed in Appendix C.

C. Closure phases and amplitudes

Using the second method, we can directly identify the closure phases and amplitudes familiar in radio interferometry with the closure invariants obtained here.

The independent closure phases are directly given by the phases of the N_Δ independent triangular advariants. But they can be derived from the covariants, $\mathcal{I}_{\Delta_1\Delta_\ell}$, as well, where Δ_ℓ is any triangle that shares vertex 0 with the reference triangle, Δ_1. The phase of $\mathcal{I}_{\Delta_1\Delta_1}$ is $2\phi_{\Delta_1}$, which yields ϕ_{Δ_1} to within an ambiguity of π. This ambiguity can be addressed by using the phase of the advariant, $\mathcal{A}_{\Delta_1}$, which gives ϕ_{Δ_1} directly without the aforementioned ambiguity. The phase of the rest of the covariants $\mathcal{I}_{\Delta_1\Delta_\ell}$ is $\phi_{\Delta_1} + \phi_{\Delta_\ell}$, which will directly yield ϕ_{Δ_ℓ} via plain substitution of the previously determined ϕ_{Δ_1}. This will yield N_Δ independent closure phases. The same construct also yields closure amplitudes.

An inspection of $|\mathcal{I}_{\Delta_1\Delta_\ell}|$ shows that it is indeed a closure amplitude, $|C_{01}||C_{21}|^{-1}|C_{20}||C_{a0}|^{-1}|C_{ab}||C_{0b}|^{-1}$. In general, closure amplitudes formed from a pair of triangular advariants involves six such terms. However, when Δ_ℓ shares an edge with Δ_1 ($a = 2$, for example),

then the contributions, $|C_{20}|$ and $|C_{a0}|^{-1}$, associated with that edge cancel each other, and the closure amplitude reduces to the familiar form containing four terms, $|C_{01}||C_{21}|^{-1}|C_{2b}||C_{0b}|^{-1}$. Noting that $|\mathcal{I}_{\Delta_1\Delta_1}|$ does not contribute an amplitude (it always has unit amplitude), there are $N_\Delta - 1 = N(N-3)/2$ independent closure amplitudes (from the $N_\Delta - 1$ triangle pairs sharing vertex 0 with the base triangle Δ_1). The autocorrelation measurement, if present, adds one more independent closure amplitude through $|\mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_0| = |C_{01}||C_{21}|^{-1}|C_{20}|A_{00}^{-1}$. Therefore, there are $N_\Delta - 1 + n_A$ independent closure amplitudes in general. It may also be noted that if an autocorrelation can be reliably measured, we can obtain all the invariants just from triangles with the autocorrelation advariant as the base advariant using $\mathcal{I}_{0;\Delta_l} = \mathcal{A}_0\widehat{\mathcal{A}}_{\Delta_l}$, without requiring quadrilaterals or six-edged loops.

The set of closure amplitudes and phases contains a total of $2N_\Delta - 1 + n_A = N^2 - 3N + 1 + n_A$ real invariants as given in Eq. (2), which is complete and independent, and is in agreement with all the earlier analyses. Our approach using advariants and covariants (invariants) constructed only from these triangular plaquettes provides a unifying approach that directly identifies the traditional invariants known as closure phases and amplitudes in radio interferometry.

Then the following question arises: which form of closure invariants is preferable? In Appendix D, we consider this question from the viewpoint of correlations within the chosen set of invariants and their overall contribution to the determination of the likelihood of parameterized models. In brief, the choice of invariants is immaterial in the determination of the likelihood function of the model parameters as long as the set of invariants employed is complete. However, certain choices of coordinates in which the invariants are represented could lead to coordinate-induced correlations among the invariants, and also to situations when one or more of the invariants become ill-defined in those coordinates. A well-known example occurs when representing real and imaginary parts of random variables in polar form, where the amplitude and phase angle can become correlated in general and the phase angle becomes ill defined in low S/N regimes even if the real and imaginary parts are perfectly uncorrelated and well defined. For these reasons, even if the real and imaginary parts of the invariants in our approach may be correlated, we still prefer this representation over the polar form consisting of phases and amplitudes.

D. Illustration with examples

For illustration, we consider the familiar examples of three- and four-element arrays below.

1. Three-element array

Consider an array with $N = 3$ and one autocorrelation measured ($n_A = 1$). We have 1 and 6 real values in the auto-

and cross-correlations, respectively, totaling seven real measurements. We have $2N - 1 = 5$ unknown real-valued parameters in the three complex-valued element gains. From dimension counting used in Eq. (2), we expect to find two real-valued closure invariants. From these measurements, we can form a triangular advariant, $\mathcal{A}_{\Delta_1}$, corresponding to the triangle $\Delta_1 \equiv \Delta_{(0,1,2)}$, and an autocorrelation advariant, $\mathcal{A}_0$. The complex invariant

$$\mathcal{I}_{\Delta_1;0} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_0 = \frac{|C_{01}||C_{20}|}{|C_{12}|A_{00}} e^{i\phi_{\Delta_1}} \qquad (12)$$

gives both these real-valued invariants. The magnitude of this complex quantity is like a closure amplitude (with one autocorrelation) and the phase is the standard closure phase of three elements. These two quantities are evidently independent. $\mathcal{I}_{\Delta_1;\Delta_1} = e^{2i\phi_{\Delta_1}}$ will give no new invariants.

If autocorrelation is not measured ($n_A = 0$), then the number of real measurements reduces by 1 to 6. Hence, we now expect only one invariant from Eq (2). Using the only advariant available, $\mathcal{A}_{\Delta_1}$, we find that

$$\mathcal{I}_{\Delta_1;\Delta_1} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_1} = e^{2i\phi_{\Delta_1}} \qquad (13)$$

contains that expected invariant, namely, its phase, which is twice the standard closure phase associated with Δ_1, and carries an ambiguity of π. However, the closure phase, ϕ_{Δ_1}, can be unambiguously obtained from the phase of $\mathcal{A}_{\Delta_1}$ directly.

2. Four-element array

For an $N = 4$ array, we have $N_\Delta = 3$ independent triangles, $\Delta_1 \equiv \Delta_{(0,1,2)}$, $\Delta_2 \equiv \Delta_{(0,2,3)}$, $\Delta_3 \equiv \Delta_{(0,1,3)}$, and hence, three complex advariants, $\mathcal{A}_{\Delta_1}$, $\mathcal{A}_{\Delta_2}$, and $\mathcal{A}_{\Delta_3}$. If the autocorrelation $\mathcal{A}_0$ is measured, then we have an additional real advariant. The dimension count analysis predicts six invariants when the autocorrelation is measured and five when it is not.

When autocorrelation is not measured ($n_A = 0$), we combine the reference triangle Δ_1, first with itself, and then with the two other independent ones, Δ_2 and Δ_3.

$$\mathcal{I}_{\Delta_1;\Delta_1} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_1} = e^{2i\phi_{\Delta_1}}, \qquad (14)$$

$$\mathcal{I}_{\Delta_1;\Delta_2} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_2} = \frac{|C_{01}||C_{23}|}{|C_{21}||C_{03}|} e^{i(\phi_{\Delta_1}+\phi_{\Delta_2})}, \qquad (15)$$

$$\mathcal{I}_{\Delta_1;\Delta_3} = \mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_3} = \frac{|C_{20}||C_{31}|}{|C_{21}||C_{30}|} e^{i(\phi_{\Delta_1}+\phi_{\Delta_3})}. \qquad (16)$$

We see that there are five real-valued invariants in the equations above—three independent closure phases and two independent closure amplitudes familiar in radio

interferometry for a four-element copolar interferometer array.

When $n_A = 1$, we can write down the following three complex invariants—$\mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_0$, $\mathcal{A}_{\Delta_2}\widehat{\mathcal{A}}_0$, and $\mathcal{A}_{\Delta_3}\widehat{\mathcal{A}}_0$—which gives us six real invariants comprising of three closure phases and three closure amplitudes (two as above, and one more due to autocorrelation). The invariant phase in $\mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_{\Delta_1}$ is redundant due to the presence of $\mathcal{A}_{\Delta_1}\widehat{\mathcal{A}}_0$, and can therefore be ignored.

Thus, the closure phases and closure amplitudes emerge naturally in our approach which treats independent triangles as fundamental. We break this up into a single invariant associated with the reference triangle, Δ_1, and two for each of the rest. The autocorrelation measurement will provide one additional closure amplitude.

E. Quadrilateral closure amplitudes

We have so far emphasized the role of triangles in constructing advariants. The complete set of closure invariants we generate include four-sided loops (quadrilaterals) as well as six-sided loops. Traditionally, closure amplitudes have been discussed by astronomers using quadrilaterals. In order to connect with this discussion we ask the following: can we obtain a complete, independent set with only quadrilateral-based closure amplitudes? A clear methodology is provided in [44] which involves placing the N elements on a ring and considering two nonoverlapping pairs of elements where each pair is made of consecutive elements. Here, we provide an alternate method based on our formalism.

One way to ensure independence is to ensure that each loop contains a unique element pair not present in any other. To do this, we fix the element pair $(0,1)$ and consider the quadrilateral made with two more elements a and b. We arrange for $1 < a < b < N$ and choose the set (which we call $\mathbb{A}$) of closure amplitudes, $|C_{01}||C_{1a}|^{-1}|C_{ab}||C_{b0}|^{-1}$. So, each such choice of a pair (a,b) with $a < b$ gives us an independent closure amplitude. The number of closure amplitudes in the set $\mathbb{A}$ is therefore the number of such (a,b) pairs, which is $(N-2)(N-3)/2$. However, we know from basic counting [18] and linear algebra [37] that the total number of independent closure amplitudes is $N(N-3)/2$. We therefore need $(N-3)[N-(N-2)]/2 = N-3$ more independent closure amplitudes to complete the desired list. We supply these in a set $\mathbb{B}$, by observing that element 0 is only paired with b and hence never paired with element 2 in the set $\mathbb{A}$. We therefore add precisely $N-3$ more closure amplitudes by including the pair, $(0,2)$, as $|C_{01}||C_{1b}|^{-1}|C_{b2}||C_{20}|^{-1}$, which contains C_{02} in addition to our base choice of C_{01}. The set $\mathbb{B}$ is only indexed by b which can take $N-3$ values, and its members are clearly independent of each other, since each contains the element pair $(1, a)$ which is unique to it within this set $\mathbb{A}$. Every member of $\mathbb{B}$ is also independent of all those of set $\mathbb{A}$, since none of

those contains the pair $(0,2)$. This completes the explicit construction of $N(N-3)/2$ independent quadrilateral closure amplitudes. Although this construction has been described in a self-contained way, it was arrived at by considering quadrilateral closure amplitudes as the product of two advariants, sharing an edge.

The closure amplitudes from our approach and the ring-based approach in [44] are essentially equivalent. However, a consequence of the fundamental difference in the methodologies leads to one notable difference in the invariants constituting the complete and independent set when an $(N+1)$th element (indexed by $a = N$) is added to the array. In the approach of [44], the inclusion of a new member, $a = N$, in the ring will remove the earlier pairing of $(N-1, 0)$ because they are no longer adjacent and the corresponding second set of consecutive pairs will also get removed. In their place, two new pairs of consecutive indices, $(N-1, N)$ and $(N, 0)$, and the corresponding sets of second adjacent pairings will have to be included. In other words, some members of the original set of closure amplitudes will be removed and replaced with a larger but different set. A similar removal and replacement process also occurs when an array element has to be excluded. In our approach, the additional element yields $N-1$ triangles (and advariants) pinned at base vertex 0 containing this element that were not present before. When paired with the base triangle, we get $N-1$ new complex-valued invariants in addition to the existing set. That is, the members of the original set are undisturbed and $2N-2$ real-valued invariants are added, amounting to a total of $N^2 - N - 1$ real invariants for the $(N+1)$-element array. This difference may have practical consequences when certain array elements are flagged at specific times or frequencies due to poor quality of data. Our approach will be robust to such practical challenges.

Thus, we have provided multiple methods of explicitly listing the closure invariants in an N-element interferometer array. These methods do not yield the same set of invariants but carry exactly equivalent information. Our approach unifies the treatment of all copolar closure invariants, including the triangular closure phases and quadrilateral closure amplitudes familiar in radio interferometry as seen earlier.

V. SUMMARY

In this paper, we have explicitly enumerated from copolar interferometric measurements a complete and independent set of closure invariants that are immune to corruptions local to the elements of an interferometer array. The solution relies on mathematical ideas borrowed from gauge theories of particle physics and the geometric phase of quantum mechanics and optics. Specifically, the copolar interferometric invariants presented here are invariant under gauge transformations by the element-based corruption factors which belong to the Abelian gauge group, $GL(1, \mathbb{C})$.

Invariants such as closure phases and closure amplitudes have been in wide use in many disciplines for decades. However, they relied on using closed triangular and quadrilateral loops, respectively, and thus required different treatments. The solution presented here relies on using only independent triangular plaquettes, which form the simplest nontrivial loops and are well determined, as our basis. The quantity defined on these triangles, called advariants, can be combined with an advariant chosen on a reference triangle to produce a complete set of independent invariants. The main result of this work is that using this formalism we have unified the treatment of closure invariants, which have historically required separate methods. Thus, we have provided an altered and unified perspective which improves our global understanding of interferometric closure invariants from a symmetry viewpoint. The familiar closure phases and closure amplitudes emerge naturally from this approach.

Radio astronomers are aware that autocorrelation measurements are unreliable because they could be dominated by nonastrophysical systematic effects. Our formalism works even without autocorrelations but can naturally accommodate them if they are reliably measured. We have also outlined a methodology for using a cross-correlation on a short-spaced element pair in place of an autocorrelation to increase the number of invariants by 1, without introducing systematic errors associated with a direct single-element autocorrelation measurement.

Our approach clarifies certain aspects related to the choice of the form of closure invariants and their implications for their covariance properties and their impact on the likelihood of model parameters. While the methods presented here are specific to copolar interferometry, the concepts serve as a stepping stone and reveal their full power when applied to a discussion of invariants in full polarimetric interferometry in paper II. We expect that this solution will aid astronomers in processing data collected from VLBI measurements and other radio interferometry experiments.

There is also an unexpected spinoff from this application which is described in more detail in Appendix E. As explained there, there is a strong connection between closure phases in astronomy and the Pancharatnam phase in optics and quantum mechanics. The present application suggests how the Pancharatnam phase rule can be generalized to include amplitudes as well as phases. This rule goes beyond existing ideas in the physics literature and shows how an astronomical application can enrich the physics which is being applied.

APPENDIX A: AUTOCORRELATIONS FROM SHORT-SPACED ELEMENTS

We have seen that autocorrelations can be used to construct covariants, not just advariants, from a triangle of elements such as in Eq. (12). This option is usually unavailable because of systematic errors in measuring autocorrelations. However, when the angular size of the object (θ_{obj}) is small compared to the angular resolution determined by the element spacing, we can recover this advantage, by using a cross-correlation, $C_{00'}$, when the elements 0 and 0′ have a small separation ($D_{00'}$), such that $\theta_{obj} \ll \lambda/D_{00'}$, where λ is the wavelength of observation.

We assume that the gains of the these closely spaced elements are independent of each other, while their correlation in $C_{00'}$ is free from the problems that a single-element autocorrelation poses. One example would be an additional element, 0′, in close proximity to the base element 0. Another possibility is to use two physically close but independent subarrays (denoted by 0 and 0′) that are phased from a dense array that is being used as a single element in a "phased array" mode. An example is the Atacama Large Millimeter/submillimeter Array (ALMA) in the EHT observations of M87 and Centaurus A [3,9]. Although 0′ can be paired with all the other existing elements, the true correlations, $S_{a0'}$, will carry no new information that is not redundant[4] with S_{a0}, excepting $S_{00'}$, which is a good approximation to S_{00}.

Thus, we have two closely spaced elements, 0 and 0′, which do not resolve the object's features but have independent element gains, G_0 and $G_{0'}$. This means that $S_{00'} \approx S_{00} = S_{0'0'}$. We can form the covariant, $C_{00'}\widehat{C}_{0'a}C_{ab}\widehat{C}_{b0} = G_0 S_{00'}\widehat{S}_{0'a}S_{ab}\widehat{S}_{b0}G_0^{-1} \approx S_{00}\widehat{S}_{0a}S_{ab}\widehat{S}_{b0}$. Note that the G_0 terms are eliminated because our group is Abelian and we have used $S_{00'} \approx S_{00}$, which is the true autocorrelation. Thus, we have effectively included autocorrelations as the coincidence limit of cross-correlations, which will increase the number of real-valued independent invariants by 1 corresponding to the autocorrelation. The use of closely located elements was suggested in [38] as a diagnostic for non-element-dependent errors. Here, we are using them to provide one more invariant effectively involving an autocorrelation.

Since we have used advariants as the building blocks, we can also reformulate this in terms of the advariant $C_{00'}\widehat{C}_{0'a}C_{a0} = G_0 S_{00'}\widehat{S}_{0'a}S_{a0}G_0^{\dagger}$, where $a \notin \{0, 0'\}$. Under our assumption of closeness of elements 0 and 0′, this can be written as $\approx G_0 S_{00}\widehat{S}_{0a}S_{a0}G_0^{\dagger}$. Since $\widehat{S}_{0a}S_{a0} = 1$, we are therefore left with only $G_0 S_{00}G_0^{\dagger}$, which is exactly what an autocorrelation advariant pinned at 0 would have yielded.

---

[4]The redundancy, however, may be useful towards obtaining a calibration that is independent of knowledge of the sky brightness distribution, and forms the basis of redundant-calibration schemes in radio interferometry (for example, [45,46]).

This means the earlier discussion with $n_A = 1$ is applicable and we gain one invariant. This agrees with the preceding discussion based on covariants, as expected.

We can also explain this using the dimension counting principles. The presence of an additional $S_{00'} \approx S_{00}$, which is nearly real and positive, effectively increases the dimensionality of the real values in the true correlations by $n_A = 1$ to $N(N-1) + 1$. It would appear that we have introduced an unknown complex gain, $G_{0'}$, consisting of two real parameters in the process. However, because of the redundancy $S_{a0'} \approx S_{a0}$, $G_{0'}$ can be expressed as

$$G_{0'} \approx C_{0'a} C_{0a}^{-1} G_0, \tag{A1}$$

which is fully determined by G_0, and therefore not an independent degree of gauge freedom. Hence, the number of unknown real parameters in the gains is still $2N - 1$. Thus, the resulting number of real invariants increases by $n_A = 1$, to $N^2 - 3N + 2$, which confirms the alternate viewpoints presented above.

APPENDIX B: RELATION TO SELF-CALIBRATION

The parameters describing the element-dependent effects can be determined if there is a standard signal. In practice, measurements on a pointlike object are interspersed with those of the target object to get a preliminary calibration. A major improvement of this procedure is "self-calibration" [20,21,32]. The approximate calibration parameters are only used as an initial guess to produce an approximate image. The image is then refined by alternating steps of deconvolution, with adjustment of the instrumental parameters to best fit the current image at each stage. This converges, in favorable cases, to a much better image than was earlier possible. The word "better" reveals that criteria based on *a priori* information, such as positivity, smoothness, and compactness of the emission, play a role via the deconvolution step.

Our work is concerned with a related but distinct notion of using closure invariants, which directly characterize the source and are independent of any model or deconvolution scheme. Forming images using just these invariants alone has been explored, especially in VLBI, by "forward-modeling," that is, fitting a model to the measured closure invariants. This approach would be appropriate in cases where the data are not extensive enough to constrain a free-form fitting procedure like self-calibration. Invariants can be used to discriminate between different proposed models, purely in the domain of observations, without bringing in deconvolution with its attendant *a priori* assumptions. A fully converged self-calibration solution will, of course, automatically satisfy all invariants. However, invariants have a role even when self-calibration is not directly applicable.

APPENDIX C: NUMERICAL TEST FOR INDEPENDENCE OF INVARIANTS

Here, we numerically verify the independence of the invariants derived through various analytical methodologies described in this paper. We begin with the simplest case with three elements which gives three cross-correlations. We also include one autocorrelation. We generate multiple realizations of one random positive real number for the uncorrupted autocorrelation, S_{00}, and three random complex numbers for the uncorrupted cross-correlations, S_{01}, S_{12}, and S_{31}, and three more random complex numbers for the element-based gains (corruptions), G_0, G_1, and G_2. These can be used to construct the measured correlations, A_{00}, C_{01}, C_{12}, and C_{20} using Eq. (1), which are described by seven real values. From these, we construct the advariants, $\mathcal{A}_{\Delta_1}$ and $\mathcal{A}_0$ and compute the triangular invariant, $\mathcal{A}_{\Delta_1} \widehat{\mathcal{A}}_0$, whose real and imaginary parts are the two closure invariants.

Analytically, the test for independence of these invariants would be to look at the two real invariants as functions of the real and imaginary parts of the input true correlations. We construct the Jacobian matrix of partial derivatives relating first order changes in the output to those in the input. The rank of the Jacobian matrix gives the number of independent invariants. Numerically, we construct the elements of this Jacobian matrix by varying one input quantity x_i at a time by a small amount, and recording the output changes y_j as a column. In this case, there will be seven columns of length 2, so the Jacobian is a 2×7 matrix. The rank is checked numerically by carrying out a singular value decomposition (SVD) of the Jacobian and examining the list of nonzero singular values. Zero-valued singular values appear as very small values (compared to the nonzero singular values by many orders of magnitude) due to finite-precision computations. The number of zero singular values (also called the corank) gives the reduction in the number of independent invariants, compared to the total number being calculated. In geometric terms, we are finding the dimension of the surface onto which a given, general set of measured visibilities gets mapped when we compute a number of invariants of our own choosing. This dimension is in general less than the maximum possible rank, by the number of zeros.

The rank of a Jacobian can also be used to cross-check the dimension counting. While we have pointed out the most obvious redundancy in the gains, namely, an overall phase, one might want to verify that no others have been missed. In this case, we simulate random true correlations as the input, and apply randomly chosen gains to them to get the measured correlations as output. Now, we vary all the gain parameters in small increments and find the Jacobian via the partial derivatives of all the real-valued parameters in the measured correlation with respect to these changes in the $2N$ real gain parameters. Not surprisingly, in all cases the corank is 1, and the count of $2N - 1$ for the

number of independent variations of the gains which modify the measured correlations, is confirmed. This numerical scheme generalizes to full polarimetric measurements as well.

APPENDIX D: CHOICE OF INVARIANTS, NOISE, AND IMAGING

We have seen in Sec. IV that a complete and independent set of invariants is not necessarily unique and can take multiple forms. How do we choose the form for a complete and independent set of invariants from all possible sets? We approach this question from two considerations—one from that of likelihood on the inferred model parameter space, and another from anticipated noise properties of the invariants themselves.

Let us consider determining the likelihood of a parametrized set of models, e.g., a ring of emission with azimuthal asymmetries as in [7]. Here and below, we use a single symbol like $\mathcal{M}$ to denote an entire set of variables, in this case parameters describing the model. This model will lead to a predicted set of correlations from which a predicted set of invariants follows, $\mathcal{I}^p(\mathcal{M})$. The noise on the correlations, presumed known, can be propagated to a probability distribution for the invariants $\mathcal{I}$ around $\mathcal{I}^p(\mathcal{M})$, using the functional relation $\mathcal{I} = f(C)$. We denote this density by $P_{\mathcal{I}}(\mathcal{I})$. Evaluating this at the measured values of the invariants $\mathcal{I}^m$ gives the likelihood function on the model space, $L(\mathcal{M}|\mathcal{I}^m) = P_{\mathcal{I}}(\mathcal{I}^m)$.

Now consider working with two different, but complete, sets of invariants, $\mathcal{I}$ and $\widetilde{\mathcal{I}}$. There is a two-way functional relation between them. Because the probability densities for $\mathcal{I}$ and $\widetilde{\mathcal{I}}$ are related by $P_{\mathcal{I}}(\mathcal{I})\mathrm{d}\mathcal{I} = P_{\widetilde{\mathcal{I}}}(\widetilde{\mathcal{I}})\mathrm{d}\widetilde{\mathcal{I}}$, we get $L(\mathcal{M}|\mathcal{I}^m) = L(\mathcal{M}|\widetilde{\mathcal{I}}^m)|\mathbf{J}|$, where $|\mathbf{J}| := \det(\mathbf{J})$ is the determinant of the Jacobian matrix, $\mathbf{J} \equiv J_{pq} = \partial\widetilde{\mathcal{I}}_p/\partial\mathcal{I}_q$. $|\mathbf{J}|$ is evaluated at the *measured* values of the two sets of invariants, and does *not* depend on the model, $\mathcal{M}$. Hence, $|\mathbf{J}|$ appears as a simple proportionality factor in the space of $\mathcal{M}$. Therefore, the maximum likelihood solution does not depend on the choice of invariants.

In Bayesian approaches, including the maximum entropy methods, there is an additional factor, namely, a prior

depending solely on $\mathcal{M}$. It may be noted that the standard approach of self-calibration does not even use invariants explicitly, while satisfying them implicitly, so the question of a choice does not arise. It is therefore satisfying that the direct use of invariants in determining the maximum likelihood of the models is also independent of the chosen form of invariants.

While the maximum likelihood of the models does not depend on the chosen form of the invariants, there may however be other considerations of noise characteristics and interpretations that may favor one form of invariants over another. Here, we compare and contrast our approach to that of [44]. Their construction of $N(N-3)/2$ independent closure amplitudes, and the demonstration that all other such amplitudes can be constructed from these is elegant and complete, and has the convenience of interpretation in terms of amplitudes and phases.

In our approach based on gauge theory, the use of triangles as the generators of all other loops and their associated invariants is more natural. These emerge as complex quantities, the magnitudes being generalized closure amplitudes (since some have six baselines) and the phases are sums of the standard closure phases on two triangles. We have chosen to work with the real and imaginary parts rather than the amplitudes and phases of our set of invariants. Our primary motivation is described below.

Following [18], for simplicity, consider an ideal case of bivariate Gaussian distribution of two uncorrelated variables, X and Y, denoting the real and imaginary parts of a complex number, respectively. Their joint probability distribution, without loss of generality, is taken to be centered on $(X_0, 0)$ with a variance of σ^2, and is given by

$$P_{X,Y}(X, Y) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(X - X_0)^2 + Y^2}{2\sigma^2}\right]. \quad \text{(D1)}$$

Here, $P_{X,Y}(X, Y) = P_X(X)P_Y(Y)$. By change of variables, $X = A\cos\theta$ and $Y = A\sin\theta$, the joint probability distribution in polar form becomes

$$P_{A,\theta}(A, \theta) = \frac{A'}{2\pi\sigma^2}\exp\left[-\frac{(A - X_0)^2 - 2X_0^2\cos\theta + 4(A - X_0)X_0\sin^2\frac{\theta}{2}}{2\sigma^2}\right], \quad \text{(D2)}$$

where $A' = A\exp[-X_0^2/\sigma^2]$. The first and the second terms inside the exponential in Eq. (D2) depend only on $A - X_0$ and θ, respectively. However, the third term depends on both, which makes A and θ correlated in general. Only when $S/N \gg 1$ ($|A - X_0| \ll |X_0|$), the apparent correlation term may be neglected and the joint distribution becomes separable into amplitude and phase

terms, $P_{A,\theta}(A, \theta) = P_A(A)P_\theta(\theta)$, where $P_\theta(\theta)$ reduces to a von Mises distribution, that can be further approximated as a Gaussian distribution in θ [18,44]. However, when $S/N \lesssim 1$, approximating that A and θ are uncorrelated is not only invalid, but the phase is also not well defined. So, the joint distribution is preferably represented in real and imaginary coordinates (X and Y) rather than their polar

form (A and θ). Thus, it is evident that purely on account of the choice of the coordinate system on which the joint probability distribution is represented, it can not only induce covariance among the random variables but also cause one or more variables to be poorly defined. Therefore, in such cases, a different coordinate system is clearly preferred.

The closure invariants, regardless of the form they are represented in, are generally higher order functions of the correlations, and are thus expected to be neither Gaussian distributed nor have an uncorrelated behavior. When expressed as amplitudes and phases, they could suffer from induced covariance and singularities in low S/N regimes. Thus, owing to such scenarios, we consider the real and imaginary parts of the covariants as "better" variables and prefer them over their amplitude and phase representation, even though they carry the same physical information.

APPENDIX E: CONNECTION TO PANCHARATNAM PHASES

The electric field time series, $e_a(t)$, received at element, a, as a function of time can be viewed as a vector in Hilbert space [47,48],

$$|e_a\rangle = \{e_a(t)|0 \le t \le T\}, \qquad (E1)$$

where T is the integration time. The correlation function, $\langle e_a e_b^\dagger \rangle$, provides an inner product between these vectors,

$$\langle e_b|e_a\rangle = C_{ab} = \frac{1}{T}\int_0^T e_b^*(t)e_a(t)\mathrm{d}t. \qquad (E2)$$

Closure phases can now be understood as a Pancharatnam phase [40,41] (see [33] for an earlier radio astronomy application). Pancharatnam's work in polarization optics gives us a rule (known in mathematics as a "connection") for comparing phases between vectors based at a and b. We fix $|e_a\rangle$ at element a and define $|e_b\rangle$ to be *in phase* with $|e_a\rangle$ if the correlation $\langle e_b|e_a\rangle$ is real and positive. The physical motivation is that the intensity of the superposed beam is a maximum under this condition. This gives a rule for

transporting a phase from element a to b, that is, modifying the phase of b so that it is "in agreement" with a. Iterating this rule and going in a closed loop from element a to b to c and back to a, we find on returning to a that the cyclic application of the rule gives a nontrivial phase change with respect to the original phase of $|e_a\rangle$. In words, $|e_b\rangle$ can be "in phase" with $|e_a\rangle$, and $|e_c\rangle$ with $|e_b\rangle$, but then $|e_c\rangle$ in general is not "in phase" with $|e_a\rangle$. The Pancharatnam phase is

$$\arg\langle e_b|e_a\rangle\langle e_c|e_b\rangle\langle e_a|e_c\rangle = \arg C_{ab}C_{bc}C_{ca}, \quad (E3)$$

which astronomers will recognize as the closure phase [25]. Closure phases, thus, emerge as the curvature of the Pancharatnam rule for comparing phases. This is a discrete version of the curvature familiar from parallel transport along a closed curve on a sphere.

A natural question arises at this point: is there a transport rule for the *amplitude* as well as the phase? The idea is to modify the vector $|e_b\rangle$ representing the signal at b so that it is "in agreement" with $|e_a\rangle$ in both phase and amplitude. This is achieved as follows. The vector $|e_b\rangle' = f_b|e_b\rangle$ is proportional to $|e_b\rangle$, but is now rescaled by a complex factor f_b, i.e., in amplitude and phase so that $\langle e_a|e_b'\rangle = f_b\langle e_a|e_b\rangle = 1$ (note that in the case of phases this was only made real and positive, but the magnitude was left undetermined). This gives $f_b = 1/\langle e_a|e_b\rangle$. At the next stage, the vector $|e_c\rangle$ is rescaled to $f_c|e_c\rangle$ so that its inner product with $f_b|e_b\rangle$ is 1, i.e., $f_b^* f_c\langle e_b|e_c\rangle = 1$. This gives $f_c = \langle e_a|e_b\rangle^*/\langle e_b|e_c\rangle$. The pattern is now clear. After an even number of steps, returning to a, we obtain a vector $f_a|e_a\rangle$ which can be compared in amplitude and phase to the original $|e_a\rangle$. The rescaling factor, f_a, comes out to be nothing but our four-element complex closure invariant, made up of correlations with alternate terms hatted. This transport rule, applicable to amplitudes and phases, can be regarded as a spinoff from radio astronomy to possible application in other areas. The restriction to an even number of steps has appeared before in relativity in a discussion of the analogue of Fermi transport for null curves [49].

[1] R. C. Jennison and M. K. Das Gupta, Fine structure of the extra-terrestrial radio source Cygnus I, Nature (London) **172,** 996 (1953).

[2] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. I. The shadow of the supermassive black Hole, Astrophys. J. Lett. **875,** L1 (2019).

[3] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. II. Array and instrumentation, Astrophys. J. Lett. **875,** L2 (2019).

[4] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon

Telescope results. III. Data processing and calibration, Astrophys. J. Lett. **875,** L3 (2019).

[5] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. IV. Imaging the central supermassive black hole, Astrophys. J. Lett. **875,** L4 (2019).

[6] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. V. Physical origin of the asymmetric ring, Astrophys. J. Lett. **875,** L5 (2019).

[7] K. Akiyama, A. Alberdi, W. Alef, K. Asada, R. Azulay, A.-K. Baczko, D. Ball, M. Baloković, J. Barrett *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. VI. The shadow and mass of the central black hole, Astrophys. J. Lett. **875,** L6 (2019).

[8] K. Akiyama, J. C. Algaba, A. Alberdi, W. Alef, R. Anantua, K. Asada, R. Azulay, A.-K. Baczko, D. Ball *et al.* (Event Horizon Telescope Collaboration), First M87 Event Horizon Telescope results. VII. Polarization of the ring, Astrophys. J. Lett. **910,** L12 (2021).

[9] M. Janssen *et al.* (The EHT Collaboration), Event Horizon Telescope observations of the jet launching and collimation in Centaurus A, Nat. Astron. **5,** 1017 (2021).

[10] H. Hauptman, The direct methods of x-ray crystallography, Science **233,** 178 (1986).

[11] H. Hauptman, Direct methods and anomalous dispersion (Nobel lecture), Angew. Chem., Int. Ed. Engl. **25,** 603 (1986).

[12] H. A. Hauptman, The phase problem of x-ray crystallography, Rep. Prog. Phys. **54,** 1427 (1991).

[13] C. Giacovazzo, F. Capitelli, C. Cuocci, and M. Ianigro, Direct methods and applications to electron crystallography, in *Microscopy, Spectroscopy, Holography and Crystallography with Electrons*, Advances in Imaging and Electron Physics Vol. 123, edited by P. W. Hawkes, P. G. Merli, G. Calestani, and M. Vittori-Antisari (Elsevier, New York, 2002), pp. 291–310, 10.1016/S1076-5670(02)80067-2.

[14] C. Giacovazzo, *Phasing in Crystallography: A Modern Perspective*, International Union of Crystallography Texts on Crystallography (Oxford University Press, Oxford, 2014), 10.1093/oso/9780199686995.001.0001.

[15] R. Snieder and E. Larose, Extracting earth's elastic wave response from noise measurements, Annu. Rev. Earth Planet Sci. **41,** 183 (2013).

[16] H. Callow, *Signal Processing for Synthetic Aperture Sonar Image Enhancement: A Thesis Presented for the Degree of Doctor of Philosophy in Electrical and Electronic Engineering at the University of Canterbury, Christchurch, New Zealand* (University of Canterbury, 2003), 10.26021/1573.

[17] P. Rosen, S. Hensley, I. Joughin, F. Li, S. Madsen, E. Rodriguez, and R. Goldstein, Synthetic aperture radar interferometry, Proc. IEEE **88,** 333 (2000).

[18] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy*, 3rd Ed. (Springer, Cham, 2017), 10.1007/978-3-319-44431-4.

[19] *Synthesis Imaging in Radio Astronomy II*, edited by G. B. Taylor, C. L. Carilli, and R. A. Perley, Astronomical Society

of the Pacific Conference Series Vol. 180 (Astronomical Society of the Pacific, San Francisco, Calif., 1999), https://www.aspbooks.org/a/volumes/table_of_contents/?book_id=292.

[20] T. J. Cornwell and P. N. Wilkinson, A new method for making maps with unstable radio interferometers, Mon. Not. R. Astron. Soc. **196,** 1067 (1981).

[21] R. D. Ekers, The almost serendipitous discovery of self-calibration, in *Serendipitous Discoveries in Radio Astronomy* (NRAO, Green Bank, WV, 1984), p. 154.

[22] J. D. Monnier, Phases in interferometry, New Astron. Rev. **51,** 604 (2007).

[23] V. Bargmann, Note on Wigner's theorem on symmetry operations, J. Math. Phys. (N.Y.) **5,** 862 (1964).

[24] N. Thyagarajan and C. L. Carilli, Invariants in interferometry: Geometric insight into closure phases, arXiv:2012.05254.

[25] R. C. Jennison, A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent, Mon. Not. R. Astron. Soc. **118,** 276 (1958).

[26] R. Q. Twiss, A. W. L. Carter, and A. G. Little, Brightness distribution over some strong radio sources at 1427 Mc/s, The Observatory **80,** 153 (1960), https://ui.adsabs.harvard.edu/abs/1960Obs....80..153T/abstract.

[27] J. D. Monnier, Astrophysics with closure phases, in *EAS Publications Series*, EAS Publications Series Vol. 6, edited by G. Perrin and F. Malbet (2003), p. 213, 10.1051/eas:2003019.

[28] J. D. Monnier, Optical interferometry in astronomy, Rep. Prog. Phys. **66,** 789 (2003).

[29] J. D. Monnier, J.-P. Berger, R. Millan-Gabet, W. A. Traub, F. P. Schloerb, E. Pedretti, M. Benisty, N. P. Carleton, P. Haguenauer, P. Kern, P. Labeye, M. G. Lacasse, F. Malbet, K. Perraut, M. Pearlman, and M. Zhao, Few skewed disks found in first closure-phase survey of Herbig Ae/Be stars, Astrophys. J. **647,** 444 (2006).

[30] J. D. Monnier, M. Zhao, E. Pedretti, N. Thureau, M. Ireland, P. Muirhead, J. P. Berger, R. Millan-Gabet, G. Van Belle, T. ten Brummelaar, H. McAlister, S. Ridgway, N. Turner, L. Sturmann, J. Sturmann, and D. Berger, Imaging the surface of Altair, Science **317,** 342 (2007).

[31] F. R. Schwab, Adaptive calibration of radio interferometer data, in *1980 Intl Optical Computing Conf I*, Vol. 0231, edited by W. T. Rhodes, International Society for Optics and Photonics (SPIE, Bellingham, WA, 1980), pp. 18–25, 10.1117/12.958828.

[32] T. J. Pearson and A. C. S. Readhead, Image formation by self-calibration in radio astronomy, Annu. Rev. Astron. Astrophys. **22,** 97 (1984).

[33] S. Bhatnagar and R. Nityananda, Solving for closure errors due to polarization leakage in radio interferometry of unpolarized sources, Astron. Astrophys. **375,** 344 (2001).

[34] N. Thyagarajan, C. L. Carilli, and B. Nikolic, Detecting Cosmic Reionization Using the Bispectrum Phase, Phys. Rev. Lett. **120,** 251301 (2018).

[35] N. Thyagarajan and C. L. Carilli, Detection of cosmic structures using the bispectrum phase. I. Mathematical foundations, Phys. Rev. D **102,** 022001 (2020).

[36] N. Thyagarajan *et al.*, Detection of cosmic structures using the bispectrum phase. II. First results from application to

cosmic reionization using the hydrogen epoch of reionization array, Phys. Rev. D **102**, 022002 (2020).

[37] A. Lannes, Phase and amplitude calibration in aperture synthesis. Algebraic structures, Inverse Probl. **7**, 261 (1991).

[38] A. E. Broderick and D. W. Pesce, Closure traces: Novel calibration-insensitive quantities for radio astronomy, Astrophys. J. **904**, 126 (2020).

[39] J. Samuel, R. Nityananda, and N. Thyagarajan, companion Letter, Invariants in Polarimetric Interferometry: A Non-Abelian Gauge Theory, Phys. Rev. Lett. **128**, 091101 (2022).

[40] S. Pancharatnam, Generalized theory of interference, and its applications. Part I. Coherent pencils, Proc. Indian Acad. Sci. A **44**, 247 (1956).

[41] S. Pancharatnam, Collected works of S. Pancharatnam, Phys. Bull. **27**, 265 (1976).

[42] F. Klein, Vergleichende Betrachtungen über neuere geometrische Forschungen (Verlag von Andreas Deichert, Erlangen, 1872) published later in Math. Ann. **43**, 63 (1893). English translation by M. W. Haskell, A comparative review of recent researches in geometry, Bull. N. Y. Math. Soc. **2**, 215 (1892–1893).

[43] R. Penrose and W. Rindler, *Spinors and Space-Time. Vol. 1: Two-Spinor Calculus and Relativistic Fields* (Cambridge University Press, 1984), 10.1017/CBO9780511564048.

[44] L. Blackburn, D. W. Pesce, M. D. Johnson, M. Wielgus, A. A. Chael, P. Christian, and S. S. Doeleman, Closure statistics in interferometric data, Astrophys. J. **894**, 31 (2020).

[45] M. H. Wieringa, An investigation of the telescope based calibration methods 'redundancy' and 'self-cal', Exp. Astron. **2**, 203 (1992).

[46] A. Liu, M. Tegmark, S. Morrison, A. Lutomirski, and M. Zaldarriaga, Precision calibration of radio interferometers using redundant baselines, Mon. Not. R. Astron. Soc. **408**, 1029 (2010).

[47] M. M. Komesaroff and I. Lerchet, Extending the fourier transform—the positivity constraint, in *Image Formation from Coherence Functions in Astronomy*, edited by C. Van Schooneveld (Springer Netherlands, Dordrecht, 1979), pp. 241–247, 10.1007/978-94-009-9449-2_27.

[48] M. M. Komesaroff, R. Narayan, and R. Nityananda, The maximum entropy method of image restoration properties and limitations, Astron. Astrophys. **93**, 269 (1981), https://ui.adsabs.harvard.edu/abs/1981A%26A....93..269K/abstract.

[49] J. Samuel and R. Nityananda, Transport along null curves, J. Phys. A Math. Gen. **33**, 2895 (2000).