Supplemental Material
Cosmic Bell Test using Random Measurement Settings from High-Redshift Quasars

Dominik Rauch,[1, 2, *] Johannes Handsteiner,[1, 2] Armin Hochrainer,[1, 2] Jason Gallicchio,[3] Andrew
S. Friedman,[4] Calvin Leung,[1, 2, 3, 5] Bo Liu,[6] Lukas Bulla,[1, 2] Sebastian Ecker,[1, 2] Fabian
Steinlechner,[1, 2] Rupert Ursin,[1, 2] Beili Hu,[3] David Leon,[4] Chris Benn,[7] Adriano Ghedina,[8] Massimo
Cecconi,[8] Alan H. Guth,[5] David I. Kaiser,[5, †] Thomas Scheidl,[1, 2] and Anton Zeilinger[1, 2, ‡]

[1]*Institute for Quantum Optics and Quantum Information (IQOQI),*
Austrian Acadamy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria
[2]*Vienna Center for Quantum Science & Technology (VCQ), Faculty of*
Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria
[3]*Department of Physics, Harvey Mudd College, Claremont, California 91711, USA*
[4]*Center for Astrophysics and Space Sciences, University of California, San Diego, La Jolla, California 92093, USA*
[5]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[6]*School of Computer, NUDT, 410073 Changsha, China*
[7]*Isaac Newton Group, Apartado 321, 38700 Santa Cruz de La Palma, Spain*
[8]*Fundación Galileo Galilei—INAF, 38712 Breña Baja, Spain*

Causal Alignment

As described in Ref. [1], the time-dependent locations of astronomical sources on the sky relative to our ground-based experimental site complicates the enforcement of the space-like separation conditions needed to address both the locality and freedom-of-choice loopholes. For example, the photon from quasar emission event $\mathcal{Q}_A$ must be received by Alice's cosmic-photon receiving telescope (Rx-CP) before that photon's causal wavefront reaches either the Rx-CP or the entangled-photon receiving telescope (Rx-EP) on Bob's side, and vice versa.

In this section we first present our main result for the causal-alignment windows, $\tau_{\text{valid}}^k(t)$ (for sides $k, \ell = \{\mathcal{A}, \mathcal{B}\}$), within which settings chosen by astronomical photons remain valid, and then derive the various terms in our expression. As shown in Figure 3 of the Main paper, we parameterize $\tau_{\text{valid}}^k(t)$ as

$$\tau_{\text{valid}}^k(t) = \bar{\tau}_{\text{geom}}^k(t) - \tau_{\text{set}}^k - \tau_{\text{buffer}}^k, \quad (1)$$

where $\tau_{\text{geom}}^k(t)$ arises from the geometrical arrangement of the quasars relative to the locations of relevant instrumentation on Earth, and $\bar{\tau}_{\text{geom}}^k(t) \equiv \min_t[\tau_{\text{geom}}^k(t)]$ is the minimum value of $\tau_{\text{geom}}^k(t)$ during an observing window. The term τ_{set}^k indicates the time required to electronically output a bit and implement the detector setting, while τ_{buffer}^k accommodates total delays due to atmosphere, telescope optics, and detector response. Negative validity times for either k would indicate an instantaneous configuration that was out of "causal alignment," in which at least one setting would be invalid for the purposes of closing the locality loophole.

For $\tau_{\text{geom}}^k(t)$, we find

$$\tau_{\text{geom}}^k(t) = \frac{1}{c}\hat{\mathbf{n}}_k(t) \cdot (\mathbf{r}_k - \mathbf{m}_\ell) + \frac{n}{c}\left[|\mathbf{m}_k - \mathbf{s}| - |\mathbf{m}_\ell - \mathbf{s}|\right]$$
$$- \frac{\gamma_k}{c}|\mathbf{r}_k - \mathbf{m}_k|, \quad (2)$$

where $\mathbf{r}_k$ and $\mathbf{m}_k$ are the spatial 3-vectors for the locations of the cosmic receiving telescopes (Rx-CP) and entangled-particle detectors (Rx-EP) for side k, respectively; $\mathbf{s}$ is the spatial 3-vector for the location of the source of entangled particles; and c is the speed of light in vacuum. The time-dependent unit vector $\hat{\mathbf{n}}_k(t)$ points toward the quasar used to set detector settings on side k, and is computed using astronomical ephemeris calculations. Additionally, n is the index of refraction of air and γ_k, which acts like an index of refraction, parameterizes the group velocity delay through fiber optics and/or electrical cables connecting the telescope and entangled photon detector on side k.

We work with a space-time metric signature $(+, -, -, -)$, so that space-time events represented by four-vectors $\mathcal{A}^\mu$ and $\mathcal{B}^\mu$ will be space-like separated if $(\mathcal{A}^\mu - \mathcal{B}^\mu)^2 < 0$. We represent spatial and temporal intervals of cosmological magnitude—such as the interval between emission of light from a distant quasar and its detection on Earth today—in terms of a spatially flat Friedmann-Lemaître-Robertson-Walker (FLRW) line element because on length-scales greater than $\mathcal{O}(100)$ Mpc, our universe has been measured to be homogeneous [2], isotropic [3], and spatially flat [4] to high accuracy. We have

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu$$
$$= c^2 dt^2 - R_0^2 a^2(t)\left[d\chi^2 + \chi^2 d\Omega_{(2)}^2(\theta, \phi)\right], \quad (3)$$

where t is cosmic time, equal to the proper time recorded by a freely falling observer at the origin of the spatial coordinate system, χ is a (dimensionless) comoving distance, and $d\Omega_{(2)}^2(\theta, \phi) = d\theta^2 + \sin^2\theta d\phi^2$ is the line-element for a unit 2-sphere. In this section we ignore possible complications from inhomogeneities, such as gravitational-lensing effects.

In Eq. (3), $a(t)$ is the (dimensionless) cosmic scale factor and the constant $R_0 \equiv c/H_0$ has dimensions of length.

We use $H_0 = 67.74 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ as the present value of the Hubble parameter [4], corresponding to a Hubble time $t_H = H_0^{-1} = 14.43 \, \mathrm{Gyr}$ and hence $R_0 = 14.43 \, \mathrm{Glyr}$. Physical distances at a given cosmic time, $r(t)$, are related to comoving distances by $r(t) = R_0 a(t) \chi$. The observed redshift, z, for astronomical objects arising from cosmic expansion is given by

$$1 + z = \frac{a(t_0)}{a(t_e)}, \tag{4}$$

where t_0 is the present time and t_e is the time of emission. We set $a(t_0) = 1$ and take $t = 0$ to be the time of the hot big bang (following any primordial phase of inflation, if inflation occurred). We further assign the origin of the spatial coordinates to be the center of the Earth. Errors introduced by treating the rotating Earth as an inertial frame are less than one part in 10^6, and are easily accommodated within τ_{buffer}.

Quasar emission event $\mathcal{Q}_k$ occurred a long time ago, in a galaxy far, far away. Hence it is convenient to introduce (dimensionless) conformal time, $d\eta \equiv H_0 dt / a(t)$, to take into account the cosmic expansion between the emission and detection of the cosmic photons. Then Eq. (3) becomes

$$ds^2 = R_0^2 a^2(\eta) \left[d\eta^2 - d\chi^2 - \chi^2 d\Omega_{(2)}^2 \right], \tag{5}$$

and (radial) null geodesics correspond to $d\eta = d\chi$. In these coordinates, the 4-vector corresponding to the receipt of a quasar photon at detector k on Earth may be written $\mathcal{R}_k^\mu = (\eta_{r_k}, \boldsymbol{\chi}_{r_k})$, and the 4-vector for the emission of the photon at the quasar is

$$\mathcal{Q}_k^\mu = \left(\eta_{q_k}, \boldsymbol{\chi}_{r_k} + (\eta_{r_k} - \eta_{q_k}) \, \hat{\mathbf{n}}_k \right), \tag{6}$$

where $\boldsymbol{\chi}_{r_k}$ is the (comoving) spatial location of the quasar photon detector Rx-CP on side k, and $\hat{\mathbf{n}}_k(t)$ is the unit spatial 3-vector pointing from the center of the Earth toward the quasar. The entangled pair is emitted from the source at $\mathcal{S}^\mu = (\eta_s, \boldsymbol{\chi}_s)$. See Figure 1.

The entangled-pair emission event should be space-like separated from the arrival of the quasar photons, which requires $(\mathcal{Q}_k - \mathcal{S})^2 < 0$. Assuming $(\eta_{r_k} - \eta_s), |\boldsymbol{\chi}_{r_k} - \boldsymbol{\chi}_s| \ll (\eta_{r_k} - \eta_{q_k})$, this implies

$$-\hat{\mathbf{n}}_k \cdot (\boldsymbol{\chi}_{r_k} - \boldsymbol{\chi}_s) < \eta_{r_k} - \eta_s. \tag{7}$$

(We further note that in the limit $(\eta_{r_k} - \eta_{q_k}) \gg |\boldsymbol{\chi}_{r_k}|$, the spherical waves emitted from the quasar arrive at the Earth as plane waves to a very good approximation.) The quantities in Eq. (7) all refer to Earthbound events, and hence we may transform back to coordinates more convenient for describing a given experimental trial. Upon recalling that $a(t_0) = 1$, we have $\boldsymbol{\chi}_{r_k} = R_0^{-1} \mathbf{r}_k$, where $\mathbf{r}_k$ is the present (physical) spatial location of the quasar-photon detector Rx-CP as reckoned from the center of
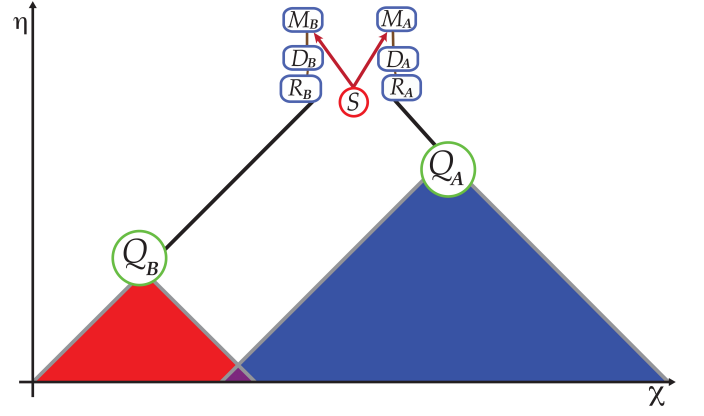

FIG. 1. Schematic space-time diagram (not to scale) of our cosmic Bell test, in (dimensionless) conformal time η versus comoving distance χ. In these coordinates, null geodesics appear as $45°$ diagonals. On each side $k \in \{A, B\}$, light from quasar emission event $\mathcal{Q}_k$ is received on Earth at event $\mathcal{R}_k$ and used to determine the detector setting at event $\mathcal{D}_k$. Meanwhile, spacelike-separated from events $\mathcal{Q}_k$, $\mathcal{R}_k$, and $\mathcal{D}_k$, the source $\mathcal{S}$ emits a pair of entangled particles, which are measured at events $\mathcal{M}_k$.

the Earth, and likewise $\boldsymbol{\chi}_s = R_0^{-1} \mathbf{s}$. We also have

$$\eta_{r_k} - \eta_s = H_0 \int_{t_s}^{t_{r_k}} \frac{dt'}{a(t')}. \tag{8}$$

The cosmic scale factor varies imperceptibly during the course of the experiment, so we may expand $a(t_0 + \delta t) = a(t_0) + \delta t / t_H + \mathcal{O}[(\delta t / t_H)^2]$, given that $t_H^{-1} = H_0 = \dot{a}(t_0)/a(t_0)$, and overdots denote derivatives with respect to cosmic time, t. During a given experimental trial, $\delta t = t_{r_k} - t_s$ is typically a fraction of a second, so we have $(\eta_{r_k} - \eta_s) = H_0(t_{t_k} - t_s) + \mathcal{O}(10^{-17})$. Then Eq. (7) becomes

$$-\frac{\hat{\mathbf{n}}_k}{c} \cdot (\mathbf{r}_k - \mathbf{s}) < t_{r_k} - t_s. \tag{9}$$

Each measurement on an entangled particle should be completed before the causal wavefront from the quasar emission on the other side arrives, which requires $(\mathcal{Q}_k - \mathcal{M}_\ell)^2 < 0$. Representing $\mathcal{M}_k^\mu = (\eta_{m_k}, \boldsymbol{\chi}_{m_k})$, using Eq. (6) for $\mathcal{Q}_k^\mu$, and proceeding as above to relate $(\eta_{m_k}, \boldsymbol{\chi}_{m_k})$ to $(t_{m_k}, \mathbf{r}_{m_k})$, we find

$$(t_{r_k} - t_{m_\ell}) + \frac{\hat{\mathbf{n}}_k}{c} \cdot (\mathbf{r}_k - \mathbf{m}_\ell) > 0. \tag{10}$$

Meanwhile, each quasar photon must be received, processed, and converted to a stable detector setting before the entangled photon from the source arrives. To calculate $\tau_{\mathrm{geom}}^k(t)$, we consider only the spatial arrangement of the various events, since we separately accommodate additional delays (from telescope optics, electronics, cables, and the like) with the factors τ_{set}^k and τ_{buffer}^k. For

$\tau^k_{\mathrm{geom}}(t)$, we may therefore parameterize

$$t_{d_k} - t_{r_k} = \frac{\gamma_k}{c}|\mathbf{r}_k - \mathbf{m}_k|, \qquad (11)$$

where t_{d_k} is the time when the setting for the entangled-photon detector on side k is set. Eq. (11) takes into account the fact that the quasar-photon reception and the detector-setting event can occur at different spatial locations. For measurements on the future light cone of the entangled-particle emission, we can use the particles' travel time from the source to the detectors to write

$$t_{m_k} = t_s + \frac{n}{c}|\mathbf{m}_k - \mathbf{s}|. \qquad (12)$$

For the setting to be valid, meanwhile, it must be set before the measurement, $t_{d_k} < t_{m_k}$. Then we may rearrange Eqs. (11) and (12) to write

$$t_{r_k} - t_s < \frac{n}{c}|\mathbf{m}_k - \mathbf{s}| - \frac{\gamma_k}{c}|\mathbf{r}_k - \mathbf{m}_k|. \qquad (13)$$

Similarly substituting Eq. (12) for t_{m_ℓ} into Eq. (10) we have

$$\frac{n}{c}|\mathbf{m}_\ell - \mathbf{s}| - \frac{\hat{\mathbf{n}}_k}{c} \cdot (\mathbf{r}_k - \mathbf{m}_\ell) < t_{r_k} - t_s. \qquad (14)$$

We want the most conservative limit on $\tau^k_{\mathrm{geom}}(t)$. Subtracting the lefthand side of Eq. (9) from the lefthand side of Eq. (14) yields $c^{-1}(n + \cos\theta)|\mathbf{m}_\ell - \mathbf{s}| \geq 0$, since the index of refraction satisfies $n \geq 1$, and $|\cos\theta| \leq 1$, where θ is the angle between $\hat{\mathbf{n}}_k$ and $(\mathbf{m}_\ell - \mathbf{s})$. Therefore the lefthand side of Eq. (14) provides the tighter lower bound on $(t_{r_k} - t_s)$. The validity time is determined by the difference between the upper and lower bounds on $(t_{r_k} - t_s)$: subtracting the lefthand side of Eq. (14) from the righthand side of Eq. (13) yields our expression for $\tau^k_{\mathrm{geom}}(t)$ in Eq. (2).

For our experiment, we may set $\mathbf{m}_k \simeq \mathbf{r}_k$ for each side, and accommodate measured delays for signal propagation and processing within the factors τ^k_{set}. Then we may compute values for $\tau^k_{\mathrm{geom}}(t)$ using the coordinates for the various experimental stations shown in Table I. We used $\tau_{\mathrm{buffer}} = 150$ ns on each side, as well as $\tau^{\mathcal{A}}_{\mathrm{set}} = 325$ ns and $\tau^{\mathcal{B}}_{\mathrm{set}} = 430$ ns. Incorporating these values for τ^k_{set} and τ_{buffer} as in Eq. (1) yielded the final values of τ^k_{valid} that we used, shown in Table II.

Component	Lat.$^{\circ}$	Lon.$^{\circ}$	Elev. [m]	diam$_k$ [m]
Rx-CP $\mathcal{A}$	28.75410°	-17.88915°	2375	3.58
Source $\mathcal{S}$	28.757189°	-17.884961°	2385	
Rx-CP $\mathcal{B}$	28.760636°	-17.8816861°	2352	4.20

TABLE I. Latitude, Longitude, and Elevation for the cosmic-photon receiving telescopes (Rx-CP) for Alice ($\mathcal{A}$) and Bob ($\mathcal{B}$), and for the Source (S). Also shown are the aperture diameters (diam$_k$) for the telescopes.

Excluded Spacetime Regions

By using light from distant quasars to determine detector settings, we may constrain the space-time region within which any putative local-realist mechanism could have engineered the observed correlations among measurements on the entangled particles. In this section we consider the space-time regions excluded from such local-realist scenarios.

Following the discussion in Ref. [5], we may relate the measured redshift for a given astronomical object to the conformal time at which the light we receive on Earth was emitted by the object, η_q. We take $\eta = 0$ to correspond to the time of the hot big bang. We may also compute the lookback time to the emission event (in cosmic time), t_{lb}, reckoned from the present, t_0.

We parameterize the Friedmann equation governing the evolution of $a(t)$ in terms of the function

$$E(a) \equiv \frac{H(a)}{H_0} = \sqrt{\Omega_\Lambda + \Omega_k a^{-2} + \Omega_M a^{-3} + \Omega_R a^{-4}}, \qquad (15)$$

where $H(a)$ is the Hubble parameter for a given scale factor $a = a(t)$, and we again use the best-fit value $H_0 = 67.74\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ [4]. The $\Omega_i \equiv \rho_i/\rho_c$ are the present-day ratios of the energy densities of dark energy (ρ_Λ), cold matter (ρ_M), and radiation (ρ_R) to the critical density, $\rho_c = 3H_0^2/(8\pi G)$, where G is Newton's gravitational constant. (The quantity ρ_M includes contributions from both baryonic matter and cold dark matter.) We also define the total fractional density of dark energy, cold matter, and radiation ($\Omega_T \equiv \Omega_\Lambda + \Omega_M + \Omega_R$), and the fractional density associated with spatial curvature ($\Omega_k \equiv 1 - \Omega_T$). We assume that ρ_Λ arises from a genuine cosmological constant with equation of state $w = p/\rho = -1$, and hence $\Omega_\Lambda a^{-3(1+w)} = \Omega_\Lambda$, which is consistent with observations [4]. We adopt the best-fit cosmological parameters from Ref. [4],

$$\vec{\Omega} = (\Omega_\Lambda, \Omega_M, \Omega_R) = (0.6911, 0.3089, 9.16\times10^{-5}), \qquad (16)$$

consistent with $|\Omega_k| < \mathcal{O}(10^{-3})$. Here $\Omega_R = \Omega_M/(1 + z_{\mathrm{eq}})$, with the redshift for matter-radiation equality given by $z_{\mathrm{eq}} = 3371$ [4].

Redshifts for the three quasars we observed are listed in Table II. For QSO J0831+5245, which was observed for both quasar pairs, we use a reported host galaxy redshift of $z = 3.9114 \pm 0.0003$ from Ref. [6]. For QSO B0350-073, we use the reported redshift of $z = 0.9635389 \pm 0.00011$ from the Sloan Digital Sky Survey (SDSS) Quasar Catalog Fourteenth Data Release [7]. For QSO B0422+004, reported redshifts include $z = 0.31$ [8] and $z = 0.268$ [9]. Neither reported redshift included an uncertainty, so we conservatively adopt the smaller value, $z = 0.268$. Given the small redshift uncertainties for QSO J0831+5245 and QSO B0350-073, and that no redshift uncertainties were

Pair	Side	Simbad ID	RA°	DEC°	$\alpha°$	$az_k°$	$alt_k°$	z	η_k	t_{lb}^k (Gyr)	F_{excl}	τ_{valid}^k [μs]
1	$\mathcal{A}$	QSO B0350−073	58.127300	−7.183976	83.81	233	38	0.964	2.46	7.78	0.960	2.34
	$\mathcal{B}$	QSO J0831+5245	127.923750	52.754860		35	57	3.911	1.56	12.21		0.90
2	$\mathcal{A}$	QSO B0422+004	66.195175	0.601758	72.84	246	38	0.268	2.95	3.22	0.635	2.20
	$\mathcal{B}$	QSO J0831+5245	127.923750	52.754860		21	64	3.911	1.56	12.21		0.53

TABLE II. Quasars whose light was used to determine detector settings for Alice ($\mathcal{A}$) and Bob ($\mathcal{B}$) for pairs 1 and 2. For each quasar, we list its QSO ID number from the Simbad database, celestial coordinates, angular separations (α) of each pair, azimuth (clockwise from due North) and altitude above horizon at the start of each observing run, and redshift (z). We also list the (dimensionless) conformal time of quasar emission (η_k), the lookback times (in Gyr) to each quasar emission event (t_{lb}^k), as well as the fraction of the physical 4-volume of the past light cone of our experiment, extending back to the big bang, from which any local-realist mechanism that might account for the measured violations of the Bell-CHSH inequality is excluded (F_{excl}). Finally, we list the validity time τ_{valid}^k from Eq. (1), which gives the minimum time that detector settings are valid for side $k = \{\mathcal{A}, \mathcal{B}\}$ during each experimental run, taking into account various delays and safety margins.

reported for QSO B0422+004, we assume that all redshift uncertainties are negligible.

The conformal time for the emission event from a distant quasar at redshift z may be written [5]

$$\eta_k(z) = \int_0^{1/(1+z)} \frac{da}{a^2 E(a)}, \qquad (17)$$

upon using $a_e = 1/(1+z)$ from Eq. (4) (and recalling our convention that $a(t_0) = 1$). Using $d\eta = H_0 dt/a(t)$, we may similarly compute the (cosmic time) lookback time to the emission event from today as

$$t_{lb}^k(z) = \int_{1/(1+z)}^1 \frac{da}{aH(a)}, \qquad (18)$$

again using $a_e = 1/(1+z)$. In Table II we list the quasars used in pairs 1 and 2 for $k = \mathcal{A}, \mathcal{B}$, their measured redshifts, z, and the corresponding values of $\eta_k(z)$ and $t_{lb}^k(z)$ for each quasar. The present age of the universe corresponds to $\eta_0 = \eta(z=0) = 3.20$, and the lookback time to the hot big bang is $t_{lb}(\infty) = 13.80$ Gyr.

Neglecting (for the moment) any possible effects from inhomogeneities along the lines of sight between the quasar emission events and our receipt of the cosmic photons on Earth, we assume that any local-realist mechanism that could have engineered the observed violations of the Bell-CHSH inequality must have acted within the past lightcone of either quasar emission event. Only within those spacetime regions could the local-realist mechanism have altered or previewed the bit that we would later receive on Earth, and shared that information (at or below the speed of light) with other elements of our experimental apparatus, such as the source of entangled particles or the detectors on the other side of our experiment [1, 10]. For pair 1, any such local-realist mechanism is constrained to have acted no more recently than $t_{lb} = 7.78$ Gyr ago, while for pair 2 the constraint is $t_{lb} = 3.22$ Gyr ago.

We may further characterize the space-time region within which any local-realist mechanism could have acted in order to produce the observed violations of the

Bell-CHSH inequality. That region consists of the union of the past lightcones from the quasar emission events utilized for a given experimental run, $V_Q^{(4)}$, which we may compare with the space-time 4-volume of the past lightcone of the experiment itself, $V_{exp}^{(4)}$. To calculate $V_Q^{(4)}$, we must consider the 4-volume of the past light cone from each emission event and subtract the 4-volume of those light cones' intersection.

We calculate the 4-volume contained within the past light cone of a quasar emission event by integrating the invariant volume element $dV = \sqrt{-g}\, d^4x$ over the region bounded by past-directed null geodesics extending from the quasar emission event, where $g = \det[g_{\mu\nu}(x)]$ is the determinant of the space-time metric. As we saw above, null geodesics take the form $d\eta = d\chi$ in the coordinates of Eq. (5). Taking the spatial origin to lie along the worldline of quasar k, the 4-volume of the past light cone from emission event $\mathcal{Q}_k$ may be evaluated as

$$V^{(4)}(\eta_k) = R_0^4 \int \Theta(\eta_k - \eta - \chi)\, a^4(\eta) \chi^2 \sin\theta d\eta d\chi d\theta d\phi$$
$$= \frac{4\pi R_0^4}{3} \int_0^{\eta_k} d\eta\, a^4(\eta)(\eta_k - \eta)^3, \qquad (19)$$

where $\Theta(x)$ is the Heaviside step function. In a FLRW universe, the most recent (conformal) time at which the past light cones from emission events $\mathcal{Q}_A$ and $\mathcal{Q}_B$ overlap is given by [5]

$$\eta_{AB} = \frac{1}{2}(\eta_A + \eta_B - \chi_L). \qquad (20)$$

Here χ_L is the comoving spatial distance between the worldlines of quasars A and B, which, in a spatially flat universe, is given by

$$\chi_L = \sqrt{\chi_A^2 + \chi_B^2 - 2\chi_A\chi_B \cos\alpha}, \qquad (21)$$

where α is the angle between the quasars as seen from Earth, and [5]

$$\chi_k(z) = \int_{1/(1+z)}^1 \frac{da}{a^2 E(a)}. \qquad (22)$$

Along any $\eta = $ constant surface, with $0 \leq \eta \leq \eta_{AB}$, the past light cones from emission events $\mathcal{Q}_A$ and $\mathcal{Q}_B$ appear as three-dimensional spheres that partially overlap. In Euclidean space, the (spatial) three-volume of the intersection region of two spheres of radii r_1 and r_2, with distance between their centers d, is given by [11]

$$V_I^{(3)} = \frac{\pi}{12d} (r_1 + r_2 - d)^2 \left[d^2 + 2d(r_1 + r_2) - 3(r_1 - r_2)^2 \right]. \tag{23}$$

Upon substituting $d \to \chi_L$, $r_1 \to \eta_A - \eta$, and $r_2 \to \eta_B - \eta$, making use of Eq. (20) for η_{AB}, and performing some straightforward algebra, we find the space-time 4-volume of the intersection region of the past light cones from emission events $\mathcal{Q}_A$ and $\mathcal{Q}_B$ to be

$$
\begin{aligned}
V_I^{(4)}&(\eta_A, \eta_B, \alpha) \\
&= 4\pi R_0^4 \int_0^{\eta_{AB}} d\eta \, a^4(\eta) \left[\frac{1}{3} (\eta_{AB} - \eta)^3 \right. \\
&\left. + (\eta_{AB} - \eta)^2 \left(\frac{\chi_L^2 - (\eta_A - \eta_B)^2}{4\chi_L} \right) \right]. \tag{24}
\end{aligned}
$$

The union of the past light cones from emission events $\mathcal{Q}_A$ and $\mathcal{Q}_B$ therefore has the 4-volume

$$V_Q^{(4)}(\eta_A, \eta_B, \alpha) = V^{(4)}(\eta_A) + V^{(4)}(\eta_B) - V_I^{(4)}(\eta_A, \eta_B, \alpha), \tag{25}$$

while the 4-volume of the past light cone of our experiment is given by $V_{\rm exp}^{(4)} = V^{(4)}(\eta_0)$.

For an experimental run using a pair of quasars with redshifts z_A, z_B, and relative angle α, the space-time region that is *excluded* from playing any role in an explanation based on a local-realist mechanism is given by $V_{\rm exp}^{(4)} - V_Q^{(4)}$. As a fraction of $V_{\rm exp}^{(4)}$, this may be written

$$F_{\rm excl} = 1 - \left(\frac{V_Q^{(4)}(\eta_A, \eta_B, \alpha)}{V^{(4)}(\eta_0)} \right). \tag{26}$$

The relative angle (as seen from Earth) between the quasars in pair 1 is $\alpha = 83.81°$, and for the quasars of pair 2 is $\alpha = 72.84°$. Given the redshifts for each quasar listed in Table II, we find $F_{\rm excl} = 0.960$ for pair 1, and $F_{\rm excl} = 0.635$ for pair 2. See Figure 2 and Figure 3.

We may compare these values for $F_{\rm excl}$ with the corresponding values for our Vienna pilot test involving Milky Way stars [1]. We again use Eq. (4) to relate redshift to cosmic scale factor. Since the motions of Milky Way stars are dominated by local peculiar velocities independent of Hubble expansion, astronomers do not measure cosmic-expansion redshifts for Milky Way stars. Nonetheless, we may compute effective values of z with which to parameterize the various emission times. For the nearby sources used in Ref. [1], we may Taylor expand

$$a(t_e) = a(t_0) - H_0(t_0 - t_e) + \mathcal{O}([(t_0 - t_e)/t_H]^2), \tag{27}$$

where, as above, $H_0 = \dot{a}(t_0)/a(t_0)$ is the present value of the Hubble parameter, $t_H = H_0^{-1} = 14.43$ Gyr, and we

scale $a(t_0) = 1$. Neglecting uncertainties on the measured distances to the Milky Way stars that we used in our pilot test, we have $(t_0 - t_A) = 604$ years and $(t_0 - t_B) = 1930$ years for run 1, and $(t_0 - t_A) = 577$ years and $(t_0 - t_B) = 3624$ years for run 2. Using Eqs. (4) and (27), these correspond to effective redshifts $z_A = 4.19 \times 10^{-8}$ and $z_B = 1.32 \times 10^{-7}$ for run 1, and $z_A = 4.00 \times 10^{-8}$ and $z_B = 2.51 \times 10^{-7}$ for run 2. Given $\alpha_1 = 119°$ for run 1 and $\alpha_2 = 112°$ for run 2, the times when the past light cones from emission events A and B most recently overlapped were $t_{AB} = 2409$ years ago (run 1) and $t_{AB} = 4039$ years ago (run 2), corresponding to $z_{AB} = 1.67 \times 10^{-7}$ (run 1) and $z_{AB} = 2.80 \times 10^{-7}$ (run 2). Repeating the calculation as above, we then find $F_{\rm excl} = 1.38 \times 10^{-7}$ for run 1, and $F_{\rm excl} = 1.45 \times 10^{-7}$ for run 2. In other words, the Vienna pilot test excluded about one hundred-thousandth of one percent of the relevant space-time volume, compared to the exclusion of 96.0% (pair 1) and 63.5% (pair 2) achieved in the present experiment.

Effects of Inhomogeneities along the Line of Sight

Our discussion to this point has assumed that the quasar photons were emitted from point-like astronomical objects. In reality, quasars are large, messy objects; a given photon may be subject to complicated interactions involving optically thick atmospheres before escaping from the quasar. To address such scenarios, we consider the "effective emission time" to be the latest time η_k that any local interactions associated with the quasar could have altered the wavelength of the photon. Any corrections arising from strong electromagnetic fields, plasma effects, or related atmospheric phenomena near the quasar would affect a precise calculation of η_k by some small quantity (with cosmic-time values for the corrections $\Delta t \ll t_{\rm lb}^k$). Compared to our simple estimates of $\eta_k(z)$ and $t_{\rm lb}^k(z)$ based on the measured redshifts of the quasars, any such corrections would be indiscernible, given $t_{\rm lb}^k(z) \sim \mathcal{O}(1-10)$ Gyr for the quasars used in pairs 1 and 2.

Atmospheric or related interactions at the quasar could introduce delays between the arrival at Earth of the causal wavefront from the emission event of a given photon and the receipt of that photon on Earth. In principle, a local-realist mechanism could therefore exploit information about the wavelength of the incoming quasar photon prior to its detection, in order to engineer the observed violations of the Bell-CHSH inequality. However, any such advanced signal about the incoming quasar photon would need to be correlated with the detector-setting photon itself, and therefore the information carried by the advanced signal must also have originated within the past light cone of the quasar emission event $\mathcal{Q}_k$, with effective emission time $\eta_k(z)$.
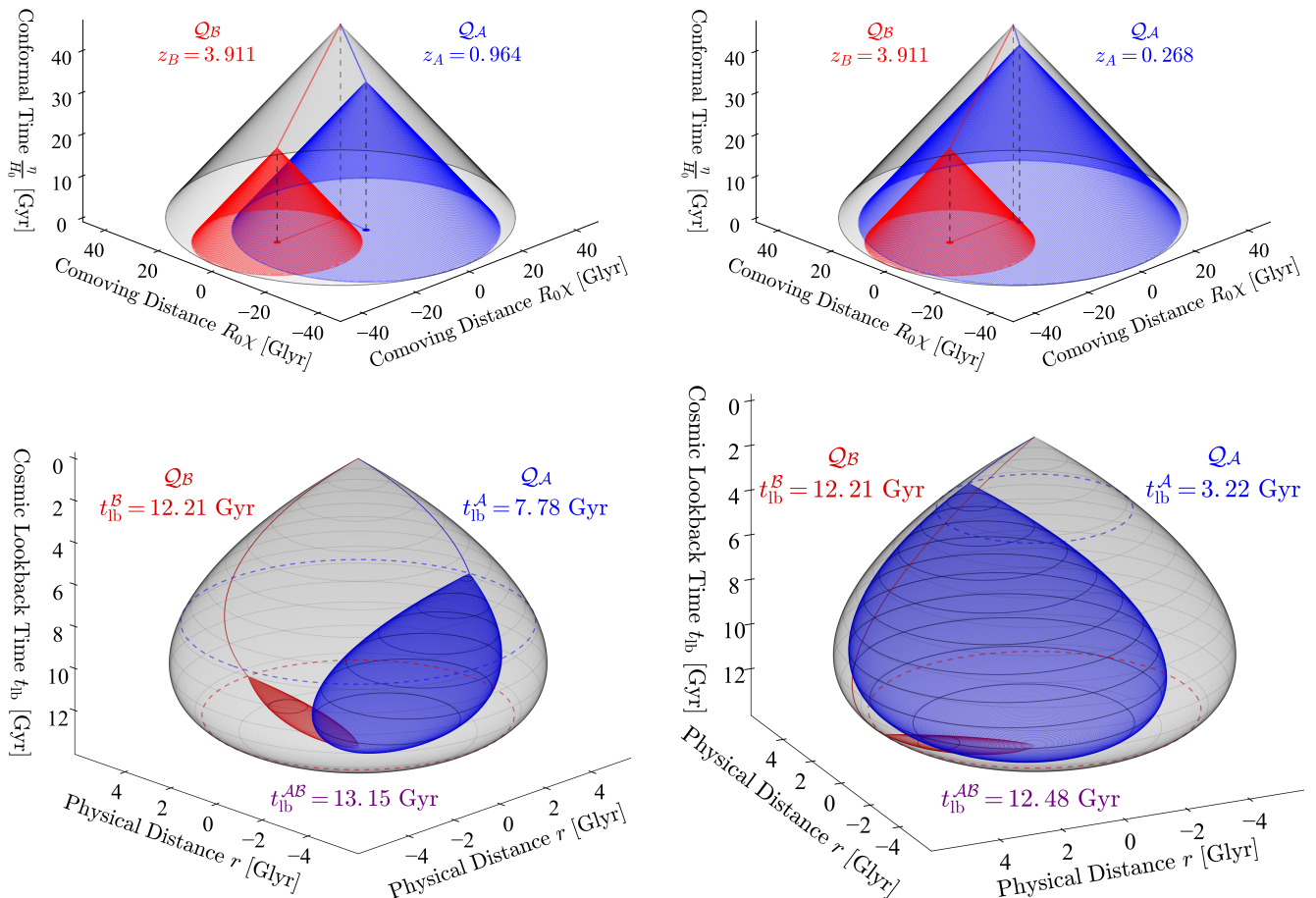
Similar considerations apply to the case in which pho-

FIG. 2. $(2+1)$D space-time diagrams (with one spatial dimension suppressed) for the quasars in pairs 1 (*left*) and 2 (*right*), showing the past light cones from quasar emission events $\mathcal{Q}_A$ (blue), $\mathcal{Q}_B$ (red), and the past light cone of our experiment (gray). Top row figures show (rescaled) conformal time η/H_0 and comoving distance $R_0\chi$; bottom row figures show cosmic lookback time $t_{\rm lb}$ and physical distance $r = R_0 a(t)\chi$. In conformal coordinates, the big bang singularity is mapped to the surface $\eta = 0$, much as the Earth's poles in a Mercator projection are depicted as lines of comparable length to the Earth's equator. The angle between the red and blue vectors in the $\eta = 0$ plane in the conformal diagrams (top row) represents the quasar pair's angular separation on the sky, given by $\alpha = 83.81°$ and $\alpha = 72.84°$ for pairs 1 and 2, respectively. The bottom-row plots show the lookback times for each emission event $(t_{\rm lb}^k)$ and the lookback time to when their past light cones most recently intersected $(t_{\rm lb}^{\mathcal{AB}})$. Note that the current age of the universe is $t_0 = 13.80$ Gyr. The teardrop shape of the past light cones when plotted in (t, r) coordinates arises from the varying expansion rate of the universe over time, given by the scale factor $a(t)$. The paths traveled by the quasar photons along the surface of the gray past light cone of the experiment are shown in red and blue. In each figure, a local-realist mechanism could have exploited information from the red and/or blue regions (and their overlap) to engineer the observed violations of the Bell-CHSH inequality. The gray regions outside of the red and blue regions are *excluded* from any such local-realist explanation. Assuming negligible uncertainties for the reported redshifts, these excluded regions amount to $F_{\rm excl} = 96.0\%$ (pair 1) and 63.5% (pair 2) of the total space-time 4-volume within the past light cone of our experiment, spanning all of cosmic history since the big bang.

tons from a given quasar are subject to strong grav-itational lensing en route from the quasar to Earth. For example, it is known that light from quasar QSO J0831+5245 (which we used to determine Bob's settings in pairs 1 and 2) is lensed by a large, intervening mass [12, 13], producing multiple images of the original quasar as seen from Earth. The multiple images arise from dif-ferent paths that quasar photons take between the lens and Earth. In the case of this particular quasar, it has been estimated that the distinct paths correspond to a

difference in arrival times at Earth of as much as 1 day [13].

Given the delay in arrival times, it is possible (in prin-ciple) that a local-realist mechanism could receive infor-mation from a short-path photon about the wavelength of a long-path photon before the latter arrives at Earth, and exploit that advanced signal to engineer the observed vi-olations of the Bell-CHSH inequality. Much like the case of atmospheric delays at the quasar itself, however, any information of value to the local-realist mechanism would
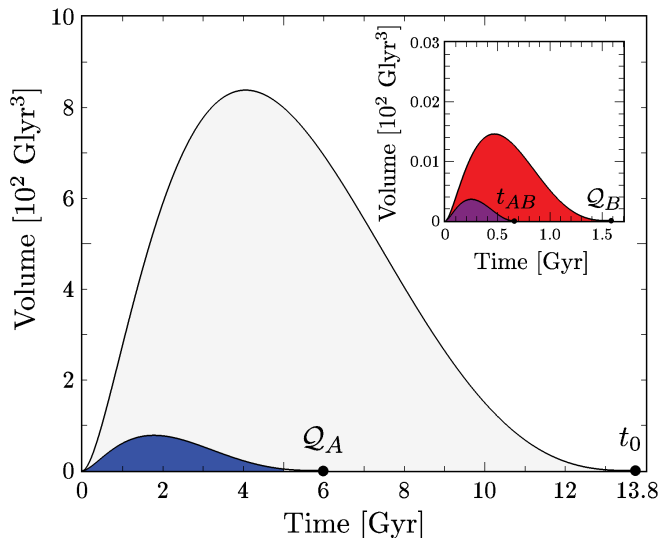
FIG. 3. Spatial volumes as a function of cosmic time t for the relevant past light cones for pair 1. The gray area shows the full past light cone of our experiment, and the blue area shows the past light cone of the emission event $\mathcal{Q}_A$ at $t_\mathcal{A} = 6.02$ Gyr ($t_{\mathrm{lb}}^\mathcal{A} = 7.78$ Gyr ago). The spacetime volume of this region is only 4.0% of the full past light cone. The insert shows emission event $\mathcal{Q}_B$, at $t_\mathcal{B} = 1.59$ Gyr ($t_{\mathrm{lb}}^\mathcal{B} = 12.21$ Gyr ago). The space-time volume of its past light cone (red) is only 0.020% of the full past light cone. The past light cones of $\mathcal{Q}_\mathcal{A}$ and $\mathcal{Q}_\mathcal{B}$ last intersected at $t_{\mathcal{AB}} = 0.65$ Gyr ($t_{\mathrm{lb}}^{\mathcal{AB}} = 13.15$ Gyr ago), and the space-time volume of their intersection (purple) is only 0.0023% of the total.

need to have originated within the past light cone of the emission event, with effective emission time, $\eta_k(z)$. Such scenarios are therefore constrained to the same space-time regions described above. See Figure 4.

One scenario in which gravitational lensing *could* affect our conclusions would arise if the wavelength of a quasar photon could somehow be measured without altering the photon's wavelength or trajectory. (No such measurement is possible according to quantum mechanics, but a local-realist mechanism, by design, is meant to be distinct from quantum mechanics.) If an "in-flight" measurement were possible, then a local-realist mechanism could potentially measure the wavelength of a quasar photon as it approaches the gravitational lens, and arrange for information about that photon's wavelength to arrive at Earth via a short path from the lens, rather than a long path. In such a scenario, the most recent time by which the local-realist mechanism would need to have acted would be bounded by the time the quasar photons reach the lens, which is more recent than the emission-time from the quasar. (We do not consider a scenario in which the local-realist mechanism could *alter* the wavelength of the quasar photon without changing its trajectory; any such mechanism would violate the conservation of energy and momentum.)
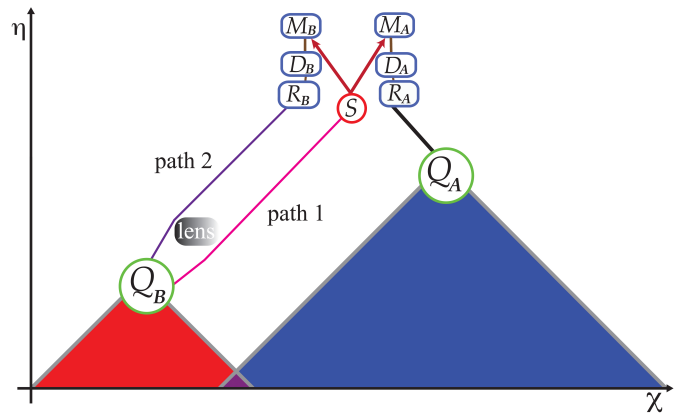

FIG. 4. A scenario in which light from quasar emission event $\mathcal{Q}_B$ is lensed by a large, intervening mass between the quasar and Earth. The lens can produce multiple images of the quasar as seen from Earth. In the scenario shown, photons following path 1 arrive at Earth before photons that follow path 2. However, to be valuable to a local-realist mechanism, the information that arrives along path 1 must be correlated with information that originates within the past light cone of the effective emission event $\mathcal{Q}_B$ (red or purple regions).

In the case of the lensed quasar in our study, the intervening lens has an estimated redshift $z_{\mathrm{lens}} \simeq 3$ [12]. This corresponds to a lookback time from Earth of $t_{\mathrm{lb}}(z_{\mathrm{lens}}) \simeq 11.6$ Gyr ago, compared to the quasar emission time $t_{\mathrm{lb}}^\mathcal{B} = 12.21$ Gyr. The time at which photons arrive at the lens is considerably earlier than the emission times from either of the quasars with which this quasar was paired ($t_{\mathrm{lb}}^\mathcal{A} = 7.78$ Gyr for pair 1 and $t_{\mathrm{lb}}^\mathcal{A} = 3.22$ Gyr for pair 2), so even such an in-flight measurement scenario would have no impact on our overall conclusions. Likewise, because the fraction of the relevant space-time 4-volume is dominated by the volume of the past light cone of the more recent emission event, re-calculating F_{excl} using $z_{\mathrm{lens}} \simeq 3$ rather than $z_B = 3.911$ yields virtually no change compared to the values computed above: $F_{\mathrm{excl,lens}} = 0.960$ for pair 1, and $F_{\mathrm{excl,lens}} = 0.635$ for pair 2.

Photons from distant quasars can be affected in other ways between emission and detection, beyond gravitational lensing. In particular, the intergalactic medium can affect the quasar spectra observed on Earth. Quasar sources typically have strong emission at the Lyman-α line, which, in the laboratory frame, corresponds to $\lambda_{\mathrm{emit}}^\alpha = 121.6$ nm (deep in the ultraviolet). However, en route, photons from high-redshift quasars encounter clouds of neutral hydrogen gas, which preferentially absorb photons at the Lyman-α wavelength — for quasar photons that have been redshifted during their travel to 121.6 nm by the time they encounter the gas cloud. Photons from very distant quasars can encounter multiple gas clouds en route to Earth, resulting in a dense "Lyman-α forest" of absorption lines at wavelengths shorter than

$\lambda_{\mathrm{obs}}^{\alpha} = (1+z)\lambda_{\mathrm{emit}}^{\alpha}$, where z is the redshift of the original quasar emission [14, 15].

In our experiment, Alice's receiving station observed quasars at redshifts $z_A = 0.964$ (pair 1) and $z_A = 0.268$ (pair 2). Hence the Lyman-α forest affected quasar photons that would have been received with wavelengths $\lambda_{\mathrm{obs}} < 238.8$ nm (pair 1) and $\lambda_{\mathrm{obs}} < 154.2$ nm (pair 2). But, as discussed further below, the detectors of our cosmic random number generators (CRNGs) were largely insensitive to $\lambda_{\mathrm{obs}} < 400$ nm. Hence selective absorption by the Lyman-α forest would have had no observable effect for either of the "quasar A" sources in our experiment. This also means that effects from transmission through the intergalactic medium could not have introduced correlations between the detected photons from quasars A and B, because any effects on the quasar-A photons would have fallen outside the sensitivity range of our detectors.

On the other hand, at Bob's receiving station we used a more distant quasar, with $z_B = 3.911$. The photons from this distant quasar certainly could have been affected by selective absorption within the Lyman-α forest, for observed wavelengths $\lambda < \lambda_{\mathrm{obs}}^{\alpha} = (1 + z_B)\lambda_{\mathrm{emit}}^{\alpha} = 597.2$ nm. Our CRNGs used dichroic filters to distinguish "red" from "blue" astronomical photons on either side of $\lambda_{\mathrm{filter}} = 630$ nm. Therefore the entire range of photons from quasar B that could have been affected by the Lyman-α forest falls within the "blue" channel. Akin to the gravitational-lensing scenario described above, one may imagine some local-realist "conspiracy" that used the Lyman-α forest to purposefully alter the spectra that would be received at Bob's station (by suppressing "blue" photons), and/or that could have "alerted" other elements of our experimental apparatus with a preview of the patterns in the upcoming sequence of "red" and "blue" detections at Bob's station.

However, given our detectors' sensitivity for $\lambda \geq 400$ nm, the *closest* gas cloud that could have affected the observable portion of the spectrum from quasar B would be at some redshift z_{cloud} such that $400\,\mathrm{nm} = (1 + z_{\mathrm{cloud}})\lambda_{\mathrm{emit}}^{\alpha}$, or $z_{\mathrm{cloud}} = 2.29$, corresponding to a lookback time at which the quasar photons encountered that cloud of $t_{\mathrm{lb}}^{\mathrm{cloud}} = 10.93$ Gyr ago. This falls considerably longer ago than the lookback times to the emission events from quasar A: $t_{\mathrm{lb}}^{A} = 7.78$ Gyr ago for pair 1, $t_{\mathrm{lb}}^{A} = 3.22$ Gyr ago for pair 2. Therefore, any "conspiracy" that might have occurred as recently as $t_{\mathrm{lb}}^{\mathrm{cloud}} = 10.93$ Gyr ago is consistent with our overall conclusion: we have constrained any such local-realist mechanisms to have occurred no more recently than t_{lb}^{A}. Likewise, we may calculate the excluded space-time volume fraction under the assumption that some "conspiracy" occurred at a gas cloud at $z_{\mathrm{cloud}} = 2.29$, by substituting $z_B \rightarrow z_{\mathrm{cloud}}$. Repeating the calculation as above, we then find $F_{\mathrm{excl}} = 0.958$ rather than 0.960 for pair 1, whereas for pair 2, we find $F_{\mathrm{excl}} = 0.635$, unchanged from our original calculation.
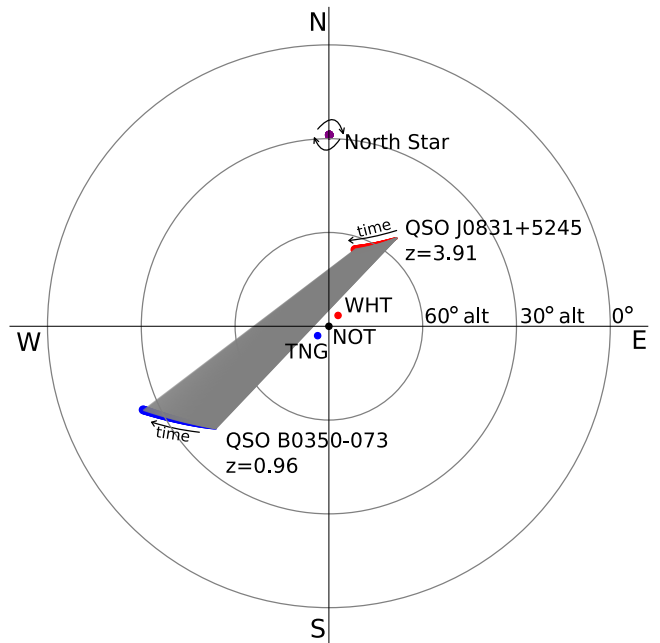

FIG. 5. Overhead view of the orientation of quasars in the sky with respect to the source at NOT and the receivers at TNG and WHT on 2018-01-11 between 00:00 and 01:30 UTC, ending with the pair 1 measurement.

Quasar Selection

In anticipation of our limited observing opportunities at the observatory, we searched for quasar pairs whose space-time arrangement would exclude the largest fraction of the 4-volume of the past light cone of our experiment, while maintaining a sufficiently high signal-to-noise ratio to yield a statistically significant result, given constraints on telescope time. (Following Refs. [1, 16, 17], we discuss requirements for signal-to-noise below.) We started with a database of the $\sim 62\,000$ quasars from the Simbad database that had a Sloan r'-band magnitude brighter than 19. We cross-referenced this to the SDSS DR14 database, taking the most conservative of their redshift estimates (or other redshifts from Simbad and the literature, where SDSS redshifts were not available). We considered only quasars that were the brightest for their distance within each $5°$ by $5°$ patch of sky. This left $\sim 4\,000$ quasars. For each pair of these, and for every minute of allotted telescope time, we simulated the number of quasar and skyglow photons that each telescope's detectors would record when pointed at the relevant instantaneous elevation angle $> 25°$ and observed through the associated airmass. Figure 5 shows the path on the sky that the quasars of pair 1 took during a period of 1.5 hours on one of our observing nights; our experiment with pair 1 was conducted near the end of that window.

To estimate the rate at which entangled photon coincidences would accumulate, we required that the rel-

evant signal-to-noise threshold be exceeded in each of the red/blue detector-setting ports for both quasars, and that detectors be triggered while both quasars were in causal alignment with respect to the experimental stations such that $\tau_{\text{valid}}^k(t) > 0$ for both sides $k = A, B$. This ensured that we only included entangled-photon coincidences while closing the locality loophole and ensuring the signal-to-noise was sufficiently high.

For a given experimental visibility, entangled-photon coincidence rate, and quasar-photon rate, we estimated the statistical significance we could achieve during the time window while all these conditions were met. Then we chose the highest-redshift pairs with the largest observable angular separations that our simulations predicted could yield significant results in the time allotted.

For the best of these candidates, we further required each object to have published spectra and verified redshifts (either SDSS DR14 or elsewhere in the literature). We manually vetted these to ensure they were legitimate quasar spectra and accurately estimated redshifts, and not, for example, misidentified stellar spectra of $r' \sim 13$-19 magnitude stars (which the SDSS algorithms sometimes misclassified as extragalactic objects with redshifts $z > 0.1$ [7]). These cases were non-negligible as contaminants to the subset of the SDSS DR14 database that our software preferentially selected, so the manual checks were required before choosing final target quasars to observe. For each of the four initially scheduled time slots on the telescopes, we performed this procedure to select ~10-20 vetted targets, yielding several possible pairs that would be optimal at the beginning of the scheduled observation window. The best pairs balanced the trade-off between the largest excluded cosmological 4-volume and the highest statistical significance that could be expected to accumulate within the relevant time window.

Although we were originally scheduled for observation windows over the course of four consecutive nights, due to bad weather and technical problems, we were only able to conduct experiments during the last of our scheduled sessions (early in the morning of 11 January 2018).

Experimental Details

We used Type-0 spontaneous parametric down conversion in a 30 mm periodically poled KTiOPO$_4$ (ppKTP) crystal placed in a Sagnac-interferometer loop to produce entangled photon pairs. The crystal was bi-directionally pumped by a grating-stabilized 405 nm laser to produce pairs of horizontally polarized down-converted photons at wavelengths of 773.6 nm and 850 nm with a spectral bandwidth of ≈ 2.5 nm FWHM. The entangled state was rotated near the maximally entangled Bell state $|\psi^-\rangle = \frac{1}{\sqrt{2}}(|HV\rangle - |VH\rangle)$ with fiber polarization controller paddles. The relative phase was adjusted by the polarization of the pump beam by optimizing the Bell

test visibility. We locally measured heralding efficiencies of 31% and 41%, which differed by about 1% before and after the experiment, confirming that these parameters remained stable over the duration of the measurement. For pair 1 (pair 2) the duty cycle of Alice's and Bob's measurements - i.e. the temporal sum of used valid setting intervals divided by the total measurement time per run - were 0.32 % (1.17 %) and 1.62 % (0.96 %), respectively, resulting in a duty-cycle for valid coincidence detections between Alice and Bob of 5.2×10^{-5} (1.1×10^{-4}).

Dichroic Selection of CRNGs

As emphasized in our previous work [1, 17], it is critical to select appropriate dichroic filters for generating astronomical random bits on the basis of their wavelength. The filters should be chosen to minimize the predictability of the random bits generated. This means minimizing cross-talk between the "red" and "blue" detection channels, and choosing the infrared cutoff of our red band to be opaque to skyglow, which at the Roque de los Muchachos observatory increases rapidly with increasing wavelength over the range $\sim 700 - 1000$ nm [18] due to transitions in the rovibrational states of OH radicals which are abundant in the upper atmosphere [19].

We employed a system of four dichroic elements which define our detection bands as shown in Figure 7. First, a longpass dichroic beamsplitter (BS) is used to reflect incoming light with $\lambda \lesssim 635$ nm and send it towards our "blue" detector. After the dichroic beamsplitter, we place additional shortpass (SP1: $\lambda < 620$ nm) and longpass filters (LP1: $\lambda > 637$ nm) in the blue and red output ports to reject photons near the transition wavelength, which may go either way at the dichroic beamsplitter. This way, the fraction of red photons detected in the blue arm and vice versa is negligible ($f_w < 2 \times 10^{-5}$). This represents a significant improvement over our previous experiment, in which the wrong-way fractions for each detector were $f_w \sim \mathcal{O}(10^{-2})$ [1]. Finally, after the longpass filter in the transmitted (red) output port, we define the long-wavelength cutoff of our red band with a shortpass filter (SP2) at ≈ 745 nm, a wavelength chosen to optimize the trade-off between rejection of infrared skyglow and transmission of quasar photons.

In order to make the selection of the dichroic elements BS, SP1, LP1, and SP2, we began with a hand-prepared list of available filtersets (BS, SP1, LP1) whose wavelengths were compatible as well as a list of available filters SP2. Then, for every combination of (BS, SP1, LP1) and SP2, we computed the signal-to-noise ratio in the red and blue channels up to an unknown overall constant which varies from quasar to quasar. We chose the final filter combination that yielded a strong signal-to-noise ratio for all three quasars observed. For every observation except

Path	% Transmit (red)	% Reflect(blue)
Through atmosphere	95-96	86-89
Through all optics	38-39	20-30
Detector quantum efficiency	75-76	30-44

TABLE III. The probability that a quasar photon in our transmit/reflect wavelength band is transmitted through the atmosphere, through the telescope/CRNG collection optics, and the probability of registering as an electronic pulse at the APD.

that of QSO B0422+004, our first-principles computation of the red-blue imbalance agreed with the measured count rates to within 11%. For the violently-variable BL Lac object QSO B0422+004, our observation was redder than predicted by our model by a factor of ≈ 1.6. This is consistent with recent photometric observations, which report a brightening of QSO B0422+004 in the V band by a factor of 2.5 between December 9, 2014, and December 29, 2015 [20]. Similar dramatic variations in brightness were also observed in the $1.2 - 2.2$ micron J, H, and K bands between February 2013 and January 2015 [21–23].

The optical efficiency of our CRNGs is shown in Table III. It varies from quasar to quasar due to their diverse spectral shape and different observing altitudes. For all computations, we employ a spectral model, formulated in Ref. [1], which takes as input a quasar spectrum (counts per second per wavelength), its observation altitude, and choices for the dichroic elements BS, SP1, LP1, and LP2. Quasar spectra are corrected for atmospheric extinction with tabulated values for atmospheric reddening at the Roque de los Muchachos [24]. Both quasar and skyglow spectra are weighted by the transmission curves of the telescopes' aluminum mirror coatings, our lenses' anti-reflection coatings, and our ID120 detectors' quantum efficiency curve. These constituent curves are plotted in Figure 6, and the weighted spectra are plotted in Figure 7.

Data Analysis

In this section we analyze the data from the two experimental runs, which were each conducted early in the morning of 11 January 2018. We make the assumptions of fair sampling and fair coincidences [27]. Thus, for testing local realism, all data can be postselected to coincidence events between Alice's and Bob's measurement stations. These coincidences were identified using a time window of 2.66 ns.

As in Ref. [1], we denote by N_{ij}^{AB} the number of coincidences in which Alice had outcome $A \in \{+, -\}$ under setting a_i (with $i = 1, 2$) and Bob had outcome $B \in \{+, -\}$ under setting b_j (with $j = 1, 2$). Then the number of all
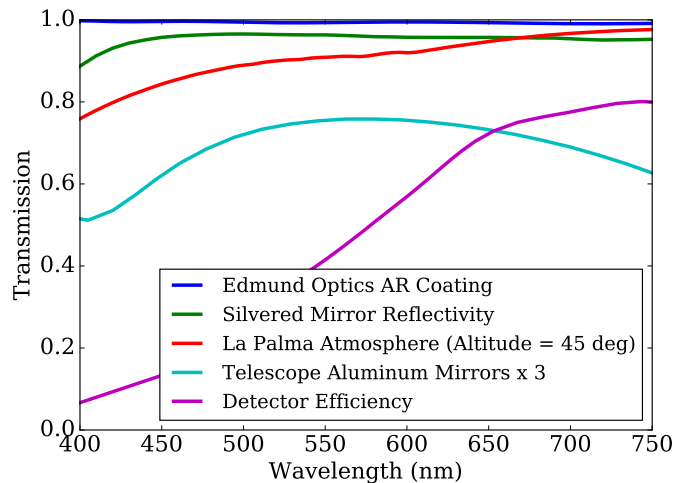

FIG. 6. Transmission spectra of optical elements in our Cosmic RNGs. Both the WHT and TNG have three aluminum mirrors, whose cumulative transmission is plotted here.
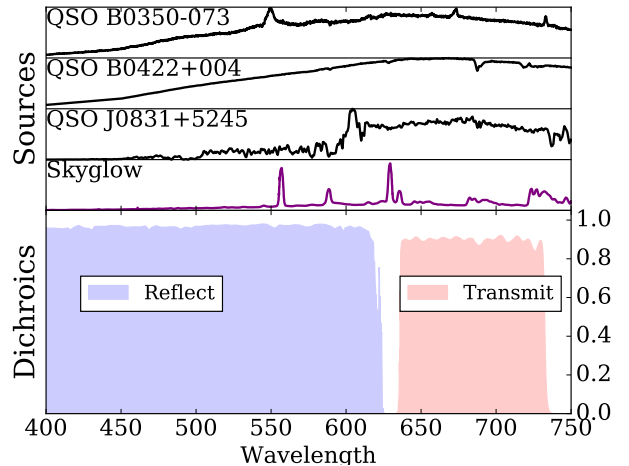

FIG. 7. Top: The number distribution of photons from each source, weighted by the transmission of optical elements and the detector efficiency. The quasar spectra are reddened by Rayleigh scattering. Bottom: Our system of four dichroic elements splits incoming photons cleanly into a reflect (blue) channel and a transmit (red) channel. The red channel is shortpassed to reject a large fraction of Meinel rovibrational emissions which grow rapidly at wavelengths longer than $\sim$ 700 nm. Quasar spectra were taken from: QSO J0831+5245 [25], QSO B0350-073 (SDSS DR14, [7]), and QSO B0422+004 [26]. The skyglow spectrum is from Ref. [18].

coincidences for settings combination $a_i b_j$ is given by

$$N_{ij} \equiv \sum_{A,B=+,-} N_{ij}^{AB}, \qquad (28)$$

and the total number of all recorded coincidences is

$$N \equiv \sum_{i,j=1,2} N_{ij}. \qquad (29)$$

A point estimate of the joint setting probabilities is given by

$$q_{ij} \equiv p(a_i b_j) = \frac{N_{ij}}{N}. \tag{30}$$

We may then test whether the probabilities q_{ij} can be factorized, that is, whether they can be written (approximately) as

$$p_{ij} \equiv p(a_i)p(b_j), \tag{31}$$

where

$$p(a_i) \equiv \frac{N_{i1} + N_{i2}}{N} = q_{i1} + q_{i2},$$
$$p(b_j) \equiv \frac{N_{1j} + N_{2j}}{N} = q_{1j} + q_{2j}. \tag{32}$$

We may also estimate the conditional probabilities for correlated outcomes in which both Alice and Bob observe the same result:

$$p(A = B|a_i b_j) = \frac{N_{ij}^{++} + N_{ij}^{--}}{N_{ij}}. \tag{33}$$

If we define the quantity

$$C \equiv -p(A = B|a_1 b_1) - p(A = B|a_1 b_2)$$
$$- p(A = B|a_2 b_1) + p(A = B|a_2 b_2), \tag{34}$$

then the Bell-CHSH inequality for local-realist models [28] takes the form $C \leq 0$ (if one neglects the "freedom-of-choice" loophole [16]). We may also construct the correlation functions

$$E_{ij} \equiv 2p(A = B|a_i b_j) - 1, \tag{35}$$

in terms of which we may construct the quantity

$$S \equiv |E_{11} + E_{12} + E_{21} - E_{22}|. \tag{36}$$

The Bell-CHSH inequality may then be written $S \leq 2$. The quantities C and S are related as $S = 2|C + 1|$.

The measured coincidence counts N_{ij}^{AB} for pair 1 are

Pair 1:

$ij \setminus AB$	++	+−	−+	−−	
11	145	956	1 229	291	
12	487	1 618	2 417	370	(37)
21	440	1 514	1 399	206	
22	3 229	418	593	2 321	

For pair 1, we have $N = 17,633$ total coincidence counts. For pair 2, we find

Pair 2:

$ij \setminus AB$	++	+−	−+	−−	
11	71	1 007	1 102	168	
12	654	1 394	1 975	494	(38)
21	315	771	702	186	
22	1 975	108	165	1 333	

For pair 2, we have $N = 12,420$ total coincidence counts.

Using the measured coincidence counts in Eqs. (38) and (37), we find the values for the Bell-CHSH parameters C and S of Eqs. (34) and (36) as shown in Table IV. There we also show the visibility fraction, defined as

$$\mathcal{V} \equiv \frac{S}{S_{\max}^{QM}}, \tag{39}$$

where $S_{\max}^{QM} = 2\sqrt{2}$, the Tsirelson bound [29], is the maximum value that the quantity S may attain according to quantum mechanics.

Pair	C	S	$\mathcal{V}$
1	0.3229	2.6457	0.935
2	0.3140	2.6281	0.929

TABLE IV. Values for the Bell-CHSH parameters and the visibility fraction for each experimental run.

Statistical Independence of Settings Choices

We first consider whether the joint settings frequencies q_{ij} for the data from pairs 1 and 2 may be factorized. For the individual settings probabilities $p(a_i)$ and $p(b_j)$ defined in Eq. (32), we find the values shown in Table VI. The joint frequencies q_{ij} and the inferred joint probabilities p_{ij} are shown in Table VII. Note that for each dataset, we find $\sum_{ij} q_{ij} = \sum_{ij} p_{ij} = 1.00$, as required.

We expect that for each dataset, $q_{ij} \simeq p_{ij}$. If we did not find $q_{ij} \simeq p_{ij}$, then (in principle) there could have been some common cause that established correlations among the various setting choices; the choices ij would not be independent. To test for any such violation of independence among the joint settings ij, we may conduct

Pair	Side	ID	n_r	s_r	n_b	s_b
1	$\mathcal{A}$	QSO B0350-073	288	2094	350	2774
1	$\mathcal{B}$	QSO J0831+5245	358	9320	408	5064
2	$\mathcal{A}$	QSO B0422+004	640	3970	684	3224
2	$\mathcal{B}$	QSO J0831+5245	389	10908	347	6213

TABLE V. Average signal and background count rates from the quasars measured during our experiment. All rates are reported in counts per second. While the fluctuations in signal rates varied significantly due to atmospheric seeing, fluctuations in the measured noise rates were consistent with $\sqrt{N}$ error over the course of a 5-minute noise measurement.

Pair	$p(a_1)$	$p(a_2)$	$p(b_1)$	$p(b_2)$
1	0.4261	0.5739	0.3505	0.6495
2	0.5527	0.4473	0.3480	0.6520

TABLE VI. Values for the single-side settings probabilities.

Pair	quantity	11	12	21	22
1	q_{ij}	0.1486	0.2774	0.2018	0.3721
	p_{ij}	0.1493	0.2767	0.2011	0.3728
2	q_{ij}	0.1890	0.3637	0.1589	0.2883
	p_{ij}	0.1923	0.3604	0.1556	0.2916

TABLE VII. Values for the joint settings frequencies q_{ij} and inferred joint probabilities p_{ij}.

a Pearson's χ^2 test by calculating the statistic

$$\chi^2 = N \sum_{ij} \frac{(q_{ij} - p_{ij})^2}{p_{ij}}, \qquad (40)$$

and then computing the p-value corresponding to this value of χ^2 [1]. For pair 1, we find $\chi^2 = 0.1504$, which implies that (under the assumption of independent setting choices) there is a purely statistical chance of $p = 0.698$ that the observed frequencies q_{ij} (or data that deviate even more from the inferred p_{ij}) would be obtained. For pair 2, we find $\chi^2 = 2.405$, corresponding to a statistical chance $p = 0.121$. Given that each of these p values is larger than typical thresholds (such as $p < 0.05$), we conclude that there is no strong statistical support to refute the hypothesis that the settings ij were independent of each other, for both pairs 1 and 2.

Testing for Violations of No-Signaling

Another statistical test to consider is whether our data are consistent with the assumption of no-signaling. To be consistent with the principle of no-signaling, the local outcome probabilities must not depend on the setting of the distant detector, which is to say [1]:

$$p(A = +|a_i b_j) = p(A = +|a_i b_{j'}),$$
$$p(B = +|a_i b_j) = p(B = +|a_{i'} b_j). \qquad (41)$$

The analogous expressions for the '−' outcomes follow from $p(A = -|a_i b_j) = 1 - p(A = +|a_i b_j)$. Point estimates for these conditional probabilities may be estimated from the measured coincidence counts. We define

$$N_{ij}^{A=+} \equiv N_{ij}^{++} + N_{ij}^{+-},$$
$$p(A = +|a_i b_j) \equiv \frac{N_{ij}^{A=+}}{N_{ij}^{A=+} + N_{ij}^{A=-}}, \qquad (42)$$

with analogous expressions for measurement outcomes $A = -$ and $B = \pm$. Point estimates for the various probabilities are shown in Table VIII.

We may assess whether these conditional probabilities show any statistical evidence of the violation of no-signaling by performing pooled two-proportion z-tests for each relevant pair, such as $p(A = +|a_1 b_1)$ and $p(A = +|a_1 b_2)$ [1], which yields the p-values listed in Table IX.

As can be seen in Table IX, nearly all of the single-sided outcomes from pairs 1 and 2 are consistent with the hypothesis of no-signaling. The only anomalously low p-value is for $p(A = +|a_2 b_j)$ for pair 2, which yields $p = 0.023 < 0.05$. However, for 8 independent tests, the probability that one shows a p-value at least as bad as 0.023 is $1 - (1 - 0.023)^8 = 0.170$, so the overall suite of tests is consistent with the hypothesis of no-signaling.

Predictability of Settings

In this section we consider imperfections in the experiment that can lead to an excess predictability [16] of the setting choices. Such excess predictability ϵ quantifies the fraction of runs in which one could predict a specific setting better than would be inferred from the overall bias of the setting choices, given all possible knowledge about the setting-generation process that could be available at the emission event for the entangled particles and thus at the distant measurement events. Loosely speaking, ϵ quantifies the fraction of runs in which the assumptions of locality and freedom-of-choice fail to hold.

In general, we consider a given trial "corrupt" if any one of three possibilities occurred at either Alice's or Bob's detector: it (1) involved a noise photon rather than a cosmic photon (where noise photons could arise from either detector dark counts or skyglow), or (2) involved a cosmic photon that was misdirected by a dichroic filter, or (3) involved a cosmic photon that was previewed by the local-realist model for the purpose of considering a dichroic-mirror error, but was passed over because the cosmic photon already had the desired color.

Given these considerations, we parameterize the excess predictabilities for each detector setting as [1]

$$\epsilon_{a_i} = \frac{1}{r_{a_i}} \left(n_{a_i} + f_{i' \to i}^{(A)} s^{(A)} \right),$$
$$\epsilon_{b_j} = \frac{1}{r_{b_j}} \left(n_{b_j} + f_{j' \to j}^{(B)} s^{(B)} \right), \qquad (43)$$

where the rate of detected photons at Alice's detector is given by

$$r_{a_i} = (1 - f_{i \to i'}^{(A)}) s_i^{(A)} + f_{i' \to i}^{(A)} s_{i'}^{(A)} + n_{a_i}, \qquad (44)$$

with a comparable expression for Bob's detector, r_{b_j}. Here n_{a_i} is the measured rate of noise photons (dark counts and skyglow), which may be quantified by pointing the receiving telescope marginally away from its quasar target; $f_{i \to i'}^{(A)}$ is the fraction of cosmic photons whose color (if correctly identified) would have led to setting choice a_i, but which were misdirected by the dichroic mirror toward the wrong port, yielding setting $a_{i'}$; and $s_i^{(A)}$ is the detected rate of cosmic photons which have a color that (when correctly identified) yield setting choice a_i. Because the fractions of misdirected cosmic photons

Pair	$p(A=+|11)$	$p(A=+|12)$	$p(A=+|21)$	$p(A=+|22)$	$p(B=+|11)$	$p(B=+|21)$	$p(B=+|12)$	$p(B=+|22)$
1	0.4201	0.4303	0.5490	0.5559	0.5242	0.5167	0.5936	0.5825
2	0.4591	0.4534	0.5502	0.5817	0.4996	0.5152	0.5820	0.5976

TABLE VIII. Point estimates for the conditional probabilities for measurement outcomes at Alice's and Bob's detectors, with $p(k=+|ij) \equiv p(k=+|a_ib_j)$.

Pair	p for $p(A=+|a_1b_j)$	p for $p(A=+|a_2b_j)$	p for $p(B=+|a_ib_1)$	p for $p(B=+|a_ib_2)$
1	0.395	0.503	0.562	0.234
2	0.653	0.023	0.308	0.156

TABLE IX. Probabilities p that the observed data (or worse) are obtained under the null hypothesis of no-signaling.

were so small (as noted above), with $f_w < 2 \times 10^{-5}$ for both red-to-blue and blue-to-red at Alice's and Bob's stations, we may neglect the effects from nonzero f_w. This simplifies our analysis compared to Ref. [1], in which the various wrong-way fractions were as large as $f_w = \mathcal{O}(10^{-2})$. For pairs 1 and 2, we find the signal rates and noise rate shown in Table V.

As in Refs. [1, 16], we assume that a local-realist model could exploit each "corrupt" trial so as to produce measurement outcomes that exceed the usual Bell-CHSH inequalities of $C \leq 0$ or $S \leq 2$. In particular, we make the conservative assumption that predictable setting events do not occur simultaneously at both detectors, so that the total fraction of corrupt joint settings is simply the sum of the corrupt settings on each side:

$$\epsilon_{ij} \equiv \epsilon_{a_i} + \epsilon_{b_j}. \tag{45}$$

If any value calculated as in Eq. (45) exceeds 1, then the corresponding ϵ_{ij} is set to 1. We further assume that the local-realist model can maximally exploit each corrupted trial, so that the maximum value of C that the local-realist model could attain by exploiting excess predictabilities of detector settings would be [16]

$$C \leq \epsilon, \tag{46}$$

where

$$\epsilon \equiv \max_{ij}\epsilon_{ij} = \max_i \epsilon_{a_i} + \max_j \epsilon_{b_j}. \tag{47}$$

Following Refs. [1, 16, 17], to ensure that a sufficient number of genuine quasar photons are detected compared to skyglow, dark current, and misdirected photons, the inequality of Eq. (46) places a constraint on the visibility fraction, such that we require

$$\epsilon < \mathcal{V}\sqrt{2} - 1 \tag{48}$$

for all times during the experiment, where the visibility fraction $\mathcal{V}$ is defined in Eq. (39). Eq. (48), in turn, sets a constraint on the signal-to-noise ratio on each of the four settings ports. Given the values of $\mathcal{V}$ shown in Table IV, the constraint of Eq. (48) becomes $\epsilon < 0.322$ (pair 1) and $\epsilon < 0.314$ (pair 2). As shown in Table XI,

all values for the excess predictabilities ϵ_{ij} easily satisfy this constraint, for both pairs of quasars.

Meanwhile, each of the corrupt fractions ϵ_{a_i} and ϵ_{b_j} has some statistical uncertainty, $\sigma_{\epsilon_{a_i}}$ and $\sigma_{\epsilon_{b_j}}$, solely due to fluctuating skynoise during the measurement run. Since the total number of runs is recorded, the only unknown is in the total number of runs that were conducted with noise photons in our CRNGs over our measurement period. This is estimated by measuring the average noise count rates before and after the measurement period at each telescope and using the higher of the two count rates to estimate the total number of corrupt runs. In each of our noise measurements we find that the estimated unbiased mean-square error in the count rates is consistent with Poisson noise.

We temporarily drop the labels for Alice and Bob, and assume that the rates r_i and n_i are independent (which follows from our assumption of fair sampling for all detected photons). Then we find

$$\sigma_{\epsilon_i}^2 = \sigma_{n_i}^2/\bar{r}_i^2 + \mathcal{O}(f_w). \tag{49}$$

If we further assume that Alice's and Bob's predictabilities are independent, then we find

$$\sigma_{\epsilon_{ij}} = \sqrt{\sigma_{\epsilon_{a_i}}^2 + \sigma_{\epsilon_{b_j}}^2}, \tag{50}$$

with an estimated uncertainty on ϵ, as defined in Eq. (47), of

$$\sigma_\epsilon = \sqrt{\sigma_{\max_i\epsilon_{a_i}}^2 + \sigma_{\max_j\epsilon_{b_j}}^2}. \tag{51}$$

Values of $\epsilon_{a_i} \pm \sigma_{\epsilon_{a_i}}$ and $\epsilon_{b_j} \pm \sigma_{\epsilon_{b_j}}$ for pairs 1 and 2 are shown in Table X, and values of the excess predictabilities for joint settings, $\epsilon_{ij} \pm \sigma_{\epsilon_{ij}}$, are shown in Table XI.

Since the exact number of runs is known and recorded, we consider the probabilities ϵ_{ij} to be known (to within some uncertainty $\sigma_{\epsilon_{ij}}$), but the actual number of corrupt trials to be subject to statistical fluctuations. In other words, the occurrence of a corruption in any trial is taken to be an independent random event, which has probability ϵ_{ij} that depends on the settings pair ij. We assume that for "uncorrupt" trials, the local-realist model has no

Pair	$\epsilon_{a_1} \pm \sigma_{\epsilon_{a_1}}$	$\epsilon_{a_2} \pm \sigma_{\epsilon_{a_2}}$	$\epsilon_{b_1} \pm \sigma_{\epsilon_{b_1}}$	$\epsilon_{b_2} \pm \sigma_{\epsilon_{b_2}}$
1	0.1441	0.1334	0.0653	0.0342
	$\pm 1.21 \times 10^{-3}$	$\pm 0.88 \times 10^{-3}$	$\pm 0.46 \times 10^{-3}$	$\pm 0.13 \times 10^{-3}$
2	0.1326	0.1679	0.0537	0.0342
	$\pm 0.46 \times 10^{-3}$	$\pm 0.54 \times 10^{-3}$	$\pm 0.93 \times 10^{-3}$	$\pm 0.26 \times 10^{-3}$

TABLE X. Values for the fractions of "corrupt" detector settings at each detector.

Pair	$\epsilon_{11} \pm \sigma_{\epsilon_{11}}$	$\epsilon_{12} \pm \sigma_{\epsilon_{12}}$	$\epsilon_{21} \pm \sigma_{\epsilon_{21}}$	$\epsilon_{22} \pm \sigma_{\epsilon_{22}}$
1	0.2095	0.1783	0.1987	0.1676
	$\pm 1.30 \times 10^{-3}$	$\pm 1.22 \times 10^{-3}$	$\pm 0.99 \times 10^{-3}$	$\pm 0.89 \times 10^{-3}$
2	0.1862	0.1667	0.2216	0.2021
	$\pm 1.04 \times 10^{-3}$	$\pm 0.53 \times 10^{-3}$	$\pm 1.07 \times 10^{-3}$	$\pm 0.60 \times 10^{-3}$

TABLE XI. Values for the excess predictabilities for various joint detector settings.

Pair/Side	Quasar	L	$\hat{I}(m)$
1/$\mathcal{A}$	QSO B0350-073	5668580	2.0×10^{-4}
1/$\mathcal{B}$	QSO J0831+5245	9010082	1.6×10^{-4}
2/$\mathcal{A}$	QSO B0422+004	6338028	5.1×10^{-5}
2/$\mathcal{B}$	QSO J0831+5245	13336320	2.7×10^{-4}

TABLE XII. We characterize the statistical predictability of our random bitstreams generated by each quasar by computing the mutual information of each bit with the $m = 17$ bits preceeding it. This measure corresponds to the probability of guessing the bit correctly given knowledge of the prior 17 bits. The number of lookback bits m for which our estimator is valid is set by L, the length of the bitstream.

information about what the settings pair will be beyond the joint settings probabilities q_{ij} [1, 16].

In general, the total rates (r_{a_i}, r_{b_j}) and noise rates (n_{a_i}, n_{b_j}) can vary during the course of an observing period. For example, when observing a given quasar over a substantial period of time, the receiving telescope will collect its light through varying amounts of airmass, as the quasar rises above or moves toward the horizon, thereby affecting the noise rate. In practice, however, our observing periods for both pairs 1 and 2 were brief enough (≤ 17 min) that the measured values for r_{a_i}, r_{b_j} and n_{a_i}, n_{b_j} did not change substantially; the largest variation among all four detector settings yielded a difference $\Delta \epsilon_{a_1} = 8.6\%$ in the excess predictability. Therefore we adopt the conservative approach of assuming constant values of each ϵ_{a_i} and ϵ_{b_j} during a given experimental run, and use the largest values for each detector setting. Such an approach will (modestly) underestimate the statistical significance of our results.

Statistical Predictability of Random Bits

Throughout our analysis, we assume that the bits within the sequence gathered from a given quasar are independent of each other. That is, we assume that there is no sequence of bits within the bitstream that contains any information about any future bit. If this independence did not hold, then (in principle) a local-realist mechanism would be able to exploit any excess predictability (beyond the natural bias) to engineer a measured violation of Bell's inequality. Although complete independence among the bits within each bitstream can never be rigorously proven, we can calculate bounds for the available information. As we assume fair-sampling for the cosmic photons, we included all detection events, instead of postselecting for those that satisfied the requirements of $\tau_{\text{valid}}^k(t)$ to yield a valid setting. For the calculation of the mutual information $\hat{I}(m)$ we adopt the

approach developed in Ref. [17]: We calculate the mutual information between every bit and every sequence of the $m = 17$ preceding bits, correcting the biased estimator for the finite size of our bitstream. The value m has been chosen such that it is strictly larger than $\lfloor \log_2(L) - 7 \rfloor$ for each measurement file, where L is the total number of detection events [30, 31]. The results are presented in Table XII. The calculated values of $\hat{I}(m)$ are $2 - 3$ orders of magnitude smaller than the values of the excess predictabilities ϵ_{a_i} and ϵ_{b_j} that we already incorporate in our data analysis. These small values of $\hat{I}(m)$ therefore make no quantitative impact on our analysis or conclusions.

Statistical Significance of Bell Violation

As discussed in Ref. [1], there exist several different approaches to estimating the statistical significance for Bell experiments. The result of any such statistical analysis is a p-value, that is, a bound for the probability that the null hypothesis—in our case, local realism with excess predictability of the detector settings ϵ_{ij}, biased detector-setting frequencies q_{ij}, fair sampling, fair coincidences, and any other additional assumptions—could have produced the experimentally observed data by a random, statistical variation.

Until recently, it was typical in the literature to make several simplifying assumptions when calculating the p-value for a Bell test, such as that each trial was independent and identically distributed, that the local-realist mechanism could not make any use of "memory" of the settings and outcomes of previous trials, and that excess predictabilities ϵ_{ij} could be neglected. Under those assumptions, one typically applied Poisson statistics for single coincidence counts. As emphasized in Ref. [1], however, such an approach assumes that the measured coincidence counts N_{ij}^{AB} are equal to their expected values, only to contradict that assumption by calculating the probability that the N_{ij}^{AB} could have values differing by several standard deviations from their expected values. More recent, sophisticated analyses do not make such assumptions, and represent significant improvements over

previous methods [16, 32–35]. However, even these newer approaches are not optimal for our particular experimental arrangement. For example, they are not optimized for experiments with unequal (biased) settings probabilities, and/or they yield non-tight upper bounds for p that can dramatically underestimate the statistical significance.

For these reasons, in Ref. [1] we presented an *ab initio* calculation of the p-value tailored more specifically to experiments like ours. We follow the same steps here, and refer readers to the more detailed description of the calculation in Ref. [1]. First we construct the quantity

$$W \equiv \sum_{ij} \frac{N_{ij}^{\text{win}}}{q_{ij}(1 - \epsilon_{ij})}, \qquad (52)$$

with $N_{ij}^{\text{win}} \equiv [N_{11}^{A \neq B}, N_{12}^{A \neq B}, N_{21}^{A \neq B}, N_{22}^{A = B}]$. If we estimate the conditional probabilities $p(A = B|a_i b_j)$ as in Eq. (33), then $W = (3 + C)N$, with C defined in Eq. (34). As described in Ref. [1], to avoid the ambiguity that would arise by the need to assume a specific prior probability distribution for the various detector settings, we take the actual number N_{ij} of the occurrences of each settings pair $a_i b_j$ as given. The relevant ensemble with respect to which we calculate probabilities is the ensemble of all possible *orders* in which the settings choices could have occurred. The p-value we calculate is then the fraction of orderings for which the local-realist mechanism, using its best strategy, could obtain a value of W greater than or equal to the value obtained in the experiment.

In the absence of noise or errors, the local-realist mechanism could specify which outcomes (A, B) will arise for each of the possible settings (a_i, b_j). The best plans will win for three of the four possible settings pairs, but will lose for one of the possible settings pairs. Therefore a plan may be fully specified by identifying which settings pair will be the loser for a given trial. In the presence of noise and errors, for each time the settings pair is $a_i b_j$, there is a probability ϵ_{ij} that the trial is corrupt. If the trial is corrupt, it automatically registers as a win. If it is not corrupt, then it has a probability P_{ij}^{win} of registering as a win, where we take $P_{ij}^{\text{win}} = p(A = B|a_i b_j)$ for $(ij) = (22)$, and $p(A \neq B|a_i b_j)$ for the other cases. These considerations motivate the form of W in Eq. (52) [1].

The ensemble average of W takes the form [1]

$$\langle W \rangle = N(3 + \bar{\epsilon}), \qquad (53)$$

where we have defined

$$\bar{\epsilon} \equiv \sum_{ij} \frac{\epsilon_{ij}}{1 - \epsilon_{ij}}. \qquad (54)$$

To calculate the statistical significance, we also must calculate $\sigma_W^2 \equiv \langle W^2 \rangle - \langle W \rangle^2$. As in Ref. [1], we calculate σ_W^{opt}, subject to the condition that the local-realist mechanism may choose the fractions f_{ij} so as to optimize its strategy, where the f_{ij} are defined as the fraction of trials for which the local-realist mechanism chooses settings pair (ij) to be the losing pair. In Ref. [1] we found

$$f_{ij}^{\text{opt}} = \frac{1}{2} - q_{ij} + \frac{N - 1}{2N} \left[\frac{\epsilon_{ij}}{1 - \epsilon_{ij}} - \bar{\epsilon} q_{ij} \right]. \qquad (55)$$

Each value of f_{ij} must be non-negative. (If any value is negative, we use a second Lagrange multiplier and recalculate the f_{ij}^{opt} [1].) For pairs 1 and 2, we find all values $f_{ij}^{\text{opt}} > 0$ when calculated as in Eq. (55), so we may use our expression for σ_W^{opt} from Ref. [1], namely

$$\begin{aligned}
\left(\sigma_W^{\text{opt}} \right)^2 = {} & \frac{N^2}{4(N - 1)} \left[\left(\sum_{ij} \frac{1}{q_{ij}} \right) - 4 \right] - N\bar{\epsilon} \\
& + \frac{N}{4} \sum_{ij} \frac{\epsilon_{ij}}{q_{ij}(1 - \epsilon_{ij})} \\
& - \frac{1}{4}(N - 1)\bar{\epsilon}^2 + \frac{1}{4} \sum_{ij} \frac{(N - \epsilon_{ij})\epsilon_{ij}}{q_{ij}(1 - \epsilon_{ij})^2}.
\end{aligned} \qquad (56)$$

We may then calculate the number of standard deviations $\bar{\nu}$ by which the correlations among the measured coincidence counts exceed those that the local-realist mechanism could have engineered by exploiting excess predictabilities:

$$\bar{\nu} = \frac{W - \langle W \rangle}{\sigma_W^{\text{opt}}}. \qquad (57)$$

The quantities W, $\langle W \rangle$, and σ_W^{opt} each depend on ϵ_{a_i} and ϵ_{b_j} as well as on the coincidence counts N_{ij}^{AB}, which we take as given. Whereas Eq. (57) takes into account the excess predictabilities, ϵ_{ij}, however, it does not incorporate the uncertainties on the excess predictabilities, $\sigma_{\epsilon_{a_i}}$ and $\sigma_{\epsilon_{b_j}}$. For the next step, we therefore propagate uncertainties on ϵ_{a_i} and ϵ_{b_j} to compute the uncertainty on $\bar{\nu}$, which we denote Δ_ν. The uncertainty Δ_ν takes the form [1]

$$\begin{aligned}
\Delta_\nu^2 = {} & \sum_{a_i} \left(\frac{\sigma_{\epsilon_{a_i}}}{\sigma_W^{\text{opt}}} \right)^2 \left[\sum_j \frac{\mathcal{E}_{ij}}{q_{ij}(1 - \epsilon_{ij})^2} \right]^2 \\
& + \sum_{b_j} \left(\frac{\sigma_{\epsilon_{b_j}}}{\sigma_W^{\text{opt}}} \right)^2 \left[\sum_i \frac{\mathcal{E}_{ij}}{q_{ij}(1 - \epsilon_{ij})^2} \right]^2, \quad (58)
\end{aligned}$$

where

$$\mathcal{E}_{ij} \equiv N_{ij}^{\text{win}} - N q_{ij} - \left(\frac{\bar{\nu} N}{2\sigma_W^{\text{opt}}} \right) f_{ij}^{\text{opt}}. \qquad (59)$$

The naive estimate of the number of standard deviations, $\bar{\nu}$ in Eq. (57), implicitly assumed $\sigma_{\epsilon_{a_i}}, \sigma_{\epsilon_{b_j}} = 0$, and therefore $\Delta_\nu = 0$. If we allow for an uncertainty in ν

equal to n times the 1-σ uncertainty in $\bar{\nu}$, then we should calculate the p-value using

$$\nu_n \equiv \bar{\nu} - n\Delta_\nu. \qquad (60)$$

If we choose n so that $n = \nu_n$, then we find

$$\nu_n = \frac{\bar{\nu}}{1 + \Delta_\nu}. \qquad (61)$$

Assuming a Gaussian distribution for large-sample experiments, we conclude that the conditional probability that the local-realist mechanism could achieve a value of W as large as the observed value W_{obs}, assuming that the true value of $\nu \geq \nu_n$, is given by

$$p_{\text{cond}} = \frac{1}{2}\text{erfc}(\nu_n/\sqrt{2}). \qquad (62)$$

Moreover, if we assume Gaussian statistics for the uncertainty in ν, then there is an equal probability that the true value of ν is less than ν_n, in which case we must conservatively assume that W might exceed W_{obs}. Therefore the p-value corresponding to the total probability that $W \geq W_{\text{obs}}$ is bounded by

$$p_{\text{no-m}} = 2p_{\text{cond}}. \qquad (63)$$

Again assuming Gaussian statistics, we may then calculate

$$\nu_{\text{no-m}} = \sqrt{2}\,\text{erfc}^{-1}(2p_{\text{no-m}}). \qquad (64)$$

The subscript "no-m" stands for "no-memory," and indicates that these quantities have been calculated without taking into account possible memory effects, which the local-realist mechanism might have been able to exploit. Using the coincidence counts in Eqs. (37) and (38) and the values of various quantities in Tables VII, X, and XI, we find the numerical values relevant to the calculation of $p_{\text{no-m}}$ and $\nu_{\text{no-m}}$ as shown in Table XIII.

The expressions for $\langle W \rangle$ and σ_W^{opt} in Eqs. (53) and (57) neglect any advantages that the local-realist model could gain by exploiting memory of previous trials. In the last step of our analysis, we incorporate possible effects from such memory, again following closely the discussion in Ref. [1]. We consider the quantity

$$\tilde{W} \equiv W - (3 + \bar{\epsilon})N, \qquad (65)$$

with W given in Eq. (52). From Eq. (53), we see that $\langle \tilde{W} \rangle = 0$ when averaged over all N trials. Hence the local-realist mechanism cannot change $\langle \tilde{W} \rangle$, but presumably it *could* affect the standard deviation of $\tilde{W}$. If we denote by $\tilde{W}_0$ the value of $\tilde{W}$ obtained in the experiment after all N trials, then the p-value we seek is the probability that the local-realist mechanism could have achieved $\tilde{W} \geq \tilde{W}_0$ by chance.

We again assume Gaussian statistics for experiments with sufficiently large numbers of trials, $N \gg 1$. Then we

expect that as long as $\tilde{W}_n \leq \tilde{W}_0$, the best strategy for the local-realist mechanism is to maximize $\sigma_{\tilde{W}}$, where $\tilde{W}_n$ is the value of $\tilde{W}$ after $n < N$ trials. In this way, the local-realist mechanism would require the smallest number of standard deviations to reach its goal. When and if $\tilde{W}_n$ exceeds $\tilde{W}_0$, on the other hand, then the best strategy for the local-realist mechanism is to minimize $\sigma_{\tilde{W}}$, so as to minimize the probability that $\tilde{W}$ might backslide to $\tilde{W} < \tilde{W}_0$. The quantity we aim to calculate is therefore p, which is bounded by

$$p \leq \frac{p_{\text{no-m}}}{1 - B}, \qquad (66)$$

where $p_{\text{no-m}}$ is given in Eq. (63). The quantity B is defined as

$$B \equiv \max_n\{p_{\text{left,max}}(n)\}, \qquad (67)$$

where $p_{\text{left,max}}(n)$ is the maximum probability that the quantity $\tilde{W}$ moves to the left after $n < N$ trials. As in Ref. [1], we compute B numerically using the values of q_{ij} and ϵ_{ij} for pairs 1 and 2. As we had found in Ref. [1], the maximum value of $p_{\text{left,max}}(n)$ occurs for $n = 1$. (See Fig. 8.) For pair 1, we find $B = 0.6001$, which yields $p \leq 7.41 \times 10^{-21}$, corresponding to $\nu = 9.29$. For pair 2, we find $B = 0.5937$, which yields $p \leq 7.03 \times 10^{-13}$, corresponding to $\nu = 7.08$.

---

* dominik.rauch@oeaw.ac.at
† dikaiser@mit.edu
‡ anton.zeilinger@univie.ac.at
[1] J. Handsteiner, A. S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H. T. Nguyen, I. Sanders, F. Steinlechner, R. Ursin, S. Wengerowsky, A. H. Guth, D. I. Kaiser, T. Scheidl, and A. Zeilinger, "Cosmic Bell Test: Measurement Settings from Milky Way Stars," Phys. Rev. Lett. **118**, 060401 (2017), arXiv:1611.06985 [quant-ph].

[2] P. Sarkar, J. Yadav, B. Pandey, and S. Bharadwaj, "The scale of homogeneity of the galaxy distribution in SDSS DR6," M. Not. Roy. Astron. Soc. **399**, L128–L131 (2009), arXiv:0906.3431 [astro-ph.CO].

[3] C. Marinoni, J. Bel, and A. Buzzi, "The scale of cosmic isotropy," J. Cosmol. Astropart. Phys **10**, 036 (2012), arXiv:1205.3309 [astro-ph.CO].

[4] P. A. R. Ade and Planck Collaboration, "Planck 2015 results. XIII. Cosmological parameters," Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].

[5] A. S. Friedman, D. I. Kaiser, and J. Gallicchio, "The shared causal pasts and futures of cosmological events," Phys. Rev. D **88**, 044038 (2013), arXiv:1305.3943 [astro-ph.CO].

[6] D. Downes, R. Neri, T. Wiklind, D. J. Wilner, and P. A. Shaver, "Detection of CO (4-3), CO (9-8), and Dust Emission in the Broad Absorption Line Quasar APM 08279+5255 at a Redshift of 3.9," Astrophys. J. Lett. **513**, L1–L4 (1999), arXiv:astro-ph/9810111.

Pair	W	$\langle W \rangle$	σ_W^{opt}	$\bar{\nu}$	Δ_ν	ν_n	p_{cond}	$p_{\text{no-m}}$	$\nu_{\text{no-m}}$	B	p	ν
1	$72\,224.1$	$69\,319.1$	290.222	10.01	0.0576	9.46	1.48×10^{-21}	2.96×10^{-21}	9.39	0.6001	7.41×10^{-21}	9.29
2	$51\,110.3$	$49\,268.0$	242.745	7.59	0.0395	7.30	1.43×10^{-13}	2.86×10^{-13}	7.21	0.5937	7.03×10^{-13}	7.08

TABLE XIII. Quantities relevant to calculating the statistical significance.
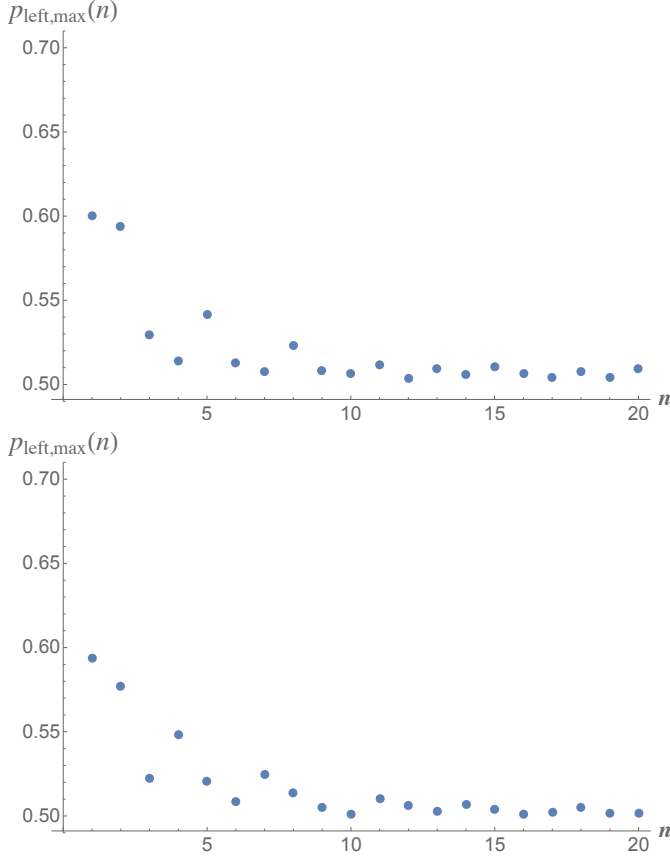
FIG. 8. The quantity $p_{\text{left,max}}(n)$ is the maximum probability that $\tilde{W}$ moves to the left in n trials. Shown here is $p_{\text{left,max}}(n)$ for pair 1 (top) and pair 2 (bottom).

[7] I. Pâris, P. Petitjean, É. Aubourg, A. D. Myers, A. Streblyanska, B. W. Lyke, S. F. Anderson, É. Armengaud, J. Bautista, M. R. Blanton, M. Blomqvist, J. Brinkmann, J. R. Brownstein, W. N. Brandt, É. Burtin, K. Dawson, S. de la Torre, A. Georgakakis, H. Gil-Marín, P. J. Green, P. B. Hall, J.-P. Kneib, S. M. LaMassa, J.-M. Le Goff, C. MacLeod, V. Mariappan, I. D. McGreer, A. Merloni, P. Noterdaeme, N. Palanque-Delabrouille, W. J. Percival, A. J. Ross, G. Rossi, D. P. Schneider, H.-J. Seo, R. Tojeiro, B. A. Weaver, A.-M. Weijmans, C. Yèche, P. Zarrouk, and G.-B. Zhao, "The Sloan Digital Sky Survey Quasar Catalog: Fourteenth data release," Astron. Astrophys. **613**, A51 (2018), arXiv:1712.05029 [astro-ph.GA].

[8] M. G. Mingaliev, Y. V. Sotnikova, R. Y. Udovitskiy, T. V. Mufakharov, E. Nieppola, and A. K. Erkenov, "RATAN-600 multi-frequency data for the BL Lacertae objects," Astron. Astrophys. **572**, A59 (2014), arXiv:1410.2835 [astro-ph.GA].

[9] M. S. Shaw, R. W. Romani, G. Cotter, S. E. Healey, P. F. Michelson, A. C. S. Readhead, J. L. Richards, W. Max-Moerbeck, O. G. King, and W. J. Potter, "Spectroscopy of the Largest Ever γ-Ray-selected BL Lac Sample," Astrophys. J. **764**, 135 (2013), arXiv:1301.0323 [astro-ph.HE].

[10] J. Gallicchio, A. S. Friedman, and D. I. Kaiser, "Testing Bell's Inequality with Cosmic Photons: Closing the Setting-Independence Loophole," Phys. Rev. Lett. **112**, 110405 (2014), arXiv:1310.3288 [quant-ph].

[11] W. F. Kern and J. R. Bland, *Solid Mensuration, With Proofs* (Wiley, 1938).

[12] E. Egami, G. Neugebauer, B. T. Soifer, K. Matthews, M. Ressler, E. E. Becklin, T. W. Murphy, Jr., and D. A. Dale, "APM 08279+5255: Keck Near- and Mid-Infrared High-Resolution Imaging," Astrophys. J. **535**, 561–574 (2000), arXiv:astro-ph/0001200.

[13] S. Oya, Y. Minowa, H. Terada, M. Watanabe, M. Hattori, Y. Hayano, Y. Saito, M. Ito, T.-S. Pyo, H. Takami, and M. Iye, "Spatially Resolved Near-Infrared Imaging of a Gravitationally Lensed Quasar, APM 08279+5255, with Adaptive Optics on the Subaru Telescope," Publ. Astron. Soc. Japan **65**, 9 (2013).

[14] M. Rauch, "The Lyman Alpha Forest in the Spectra of QSOs," Ann. Rev. Astron. Astrophys. **36**, 267–316 (1998), arXiv:astro-ph/9806286.

[15] P. McDonald, U. Seljak, S. Burles, D. J. Schlegel, D. H. Weinberg, R. Cen, D. Shih, J. Schaye, D. P. Schneider, N. A. Bahcall, J. W. Briggs, J. Brinkmann, R. J. Brunner, M. Fukugita, J. E. Gunn, Ž. Ivezić, S. Kent, R. H. Lupton, and D. E. Vanden Berk, "The Lyα Forest Power Spectrum from the Sloan Digital Sky Survey," Astrophys. J. Suppl. **163**, 80–109 (2006), arXiv:astro-ph/0405013.

[16] J. Kofler, M. Giustina, J.-Å. Larsson, and M. W. Mitchell, "Requirements for a Loophole-Free Photonic Bell Test using Imperfect Setting Generators," Phys. Rev. A **93**, 032115 (2016), arXiv:1411.4787 [quant-ph].

[17] C. Leung, A. Brown, H. Nguyen, A. S. Friedman, D. I. Kaiser, and J. Gallicchio, "Astronomical random numbers for quantum foundations experiments," Phys. Rev. A **97**, 042120 (2018), arXiv:1706.02276 [quant-ph].

[18] C. R. Benn and S. L. Ellison, "Brightness of the night sky over La Palma," New Astron. Rev. **42**, 503 (1998), arXiv:astro-ph/9909153.

[19] I. A. B. Meinel, "OH Emission Bands in the Spectrum of the Night Sky." Astrophys. J. **111**, 555 (1950).

[20] K. Z. Stanek, T. W.-S. Holoien, C. S. Kochanek, A. B. Davis, G. Simonian, U. Basu, N. Goss, J. F. Beacom, B. J. Shappee, J. L. Prieto, D. Bersier, S. Dong, P. R. Wozniak, J. Brimacombe, D. Szczygiel, and G. Pojmanski, "ASAS-SN Photometry of QSO BZB J0424+0036," The Astronomer's Telegram **6866** (2015).

[21] L. Carrasco, G. Escobedo, A. Porras, E. Recillas, V. Chabushyan, A. Carraminana, and D. Mayya, "NIR brightening of the blazar BZB J0424+0036," The Astronomer's Telegram **5712** (2014).

[22] L. Carrasco, A. Porras, E. Recillas, J. Leon-Tavares,

V. Chavushyan, and A. Carraminana, "An ongoing NIR Flare of the QSO BZB J0424+0036," The Astronomer's Telegram **6865** (2015).

[23] L. Carrasco, A. Porras, E. Recillas, J. Leon-Tavares, V. Chavushyan, and A. Carraminana, "NIR brightening of the QSO HB89 0422+004," The Astronomer's Telegram **6971** (2015).

[24] D. L. King, "Atmospheric Extinction at the Roque de los Muchachos Observatory, La Palma," RGO/La Palma technical note **31** (1985).

[25] F. G. Saturni, D. Trevese, F. Vagnetti, M. Perna, and M. Dadina, "A multi-epoch spectroscopic study of the BAL quasar APM 08279+5255. II. Emission- and absorption-line variability time lags," Astron. Astrophys. **587**, A43 (2016), arXiv:1512.03195 [astro-ph.GA].

[26] B. Sbarufatti, A. Treves, R. Falomo, J. Heidt, J. Kotilainen, and R. Scarpa, "ESO Very Large Telescope Optical Spectroscopy of BL Lacertae Objects. II. New Redshifts, Featureless Objects, and Classification Assessments," Astron. J. **132**, 1–19 (2006), arXiv:astro-ph/0601506.

[27] J.-Å. Larsson, M. Giustina, J. Kofler, B. Wittmann, R. Ursin, and S. Ramelow, "Bell-inequality violation with entangled photons, free of the coincidence-time loophole," Phys. Rev. A **90**, 032107 (2014), arXiv:1309.0712 [quant-ph].

[28] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," Phys. Rev. Lett. **23**, 880–884 (1969).

[29] B. S. Cirelson, "Quantum generalizations of Bell's inequality," Lett. Math. Phys. **4**, 93–100 (1980).

[30] A. Treves and S. Panzeri, "The upward bias in measures of information derived from limited data samples," Neural Computation **7**, 399–407 (1995).

[31] L. E. Bassham III, A. L. Rukhin, J. Soto, J. R. Nechvatal, M. E. Smid, E. B. Barker, S. D. Leigh, M. Levenson, M. Vangel, D. L. Banks, *et al.*, "Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications," Special Publication (NIST SP) **800-22 Rev 1a** (2010).

[32] R. D. Gill, "Time, Finite Statistics, and Bell's Fifth Position," in *Proceedings of Foundations of Probability and Physics - 2, Math. Modelling in Phys., Engin., and Cogn. Sc.*, Vol. 5 (Växjö University Press, 2003) p. 179, arXiv:quant-ph/0301059.

[33] R. D. Gill, "Statistics, causality and Bell's theorem," Statist. Sci. **29**, 512–528 (2014), arXiv:1207.5103 [stat.AP].

[34] P. Bierhorst, "A robust mathematical model for a loophole-free Clauser-Horne experiment," J. Phys. A **48**, 195302 (2015), arXiv:1312.2999 [quant-ph].

[35] D. Elkouss and S. Wehner, "(Nearly) optimal P values for all Bell inequalities," NPJ Quantum Information **2**, 16026 (2016), arXiv:1510.07233 [quant-ph].