
Supplementary information

A close quasar pair in a disk–disk galaxy merger at $z = 2.17$

In the format provided by the
authors and unedited

A Close Quasar Pair in a Disk-Disk Galaxy Merger at $z = 2.17$

Yu-Ching Chen^{1,2}, Xin Liu^{1,2}, Adi Foord³, Yue Shen^{1,2}, Masamune Oguri^{4,5}, Nianyi Chen⁶, Tiziana Di Matteo^{6,7,8}, Miguel Holgado⁶, Hsiang-Chih Hwang⁹, Nadia Zakamska¹⁰

¹*Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

²*National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 605 East Springfield Avenue, Champaign, IL 61820, USA*

³*Kavli Institute of Particle Astrophysics and Cosmology, Stanford University, Stanford, CA 94305, USA*

⁴*Center for Frontier Science, Chiba University, Chiba 263-8522, Japan*

⁵*Department of Physics, Graduate School of Science, Chiba University, Chiba 263-8522, Japan*

⁶*McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

⁷*NSF AI Planning Institute for Physics of the Future, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

⁸*OzGrav-Melbourne, Australian Research Council Centre of Excellence for Gravitational*

Wave Discovery, Melbourne, VIC 3122, Australia

*⁹School of Natural Sciences, Institute for Advanced Study, Princeton, 1 Einstein Drive,
NJ 08540, USA*

*¹⁰Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD
21218, USA*

Supplementary Information

Target Selection SDSS J0749+2255 is one of the dual quasar candidates selected from spectroscopically confirmed quasars in the SDSS DR7 and DR14 quasar catalogs^{[5][32][33]} using the new astrometric technique dubbed “Varstrometry”^{[6][7]}. Varstrometry utilizes the centroid jitters due to the asynchronous variability of an unresolved dual quasar. Dozens of high-redshift kpc-scale dual quasar candidates were selected by applying Varstrometry to Gaia DR2 data and follow-up *HST* optical dual-band imaging^{[7][34]}. SDSS J0749+2255 is a high-priority target for detailed follow-up observations given the high redshift of $z = 2.17$ and the small separation of $0''.46$. Supplementary Table 1 lists its basic photometric and spectroscopic properties. Previous high angular-resolution VLBA observations showed two compact radio cores, confirming both sources as quasars^[7]. However, the alternative scenarios of a lensed quasar or a pre-merger projected quasar pair at larger physical separations could not be completely ruled out based on the previous *HST* dual-band optical and VLBA single-band radio images.

Comparison with Known Dual Quasar Candidates Supplementary Figure 1 shows SDSS J0749+2255 in comparison with all the previously known $1 < z \lesssim 3$ galactic-scale (sub-10 kpc) candidate dual quasars from the literature. Four of the candidates^{[22][24]} are nearly identical quasars (NIQs; shown in grey) – quasars with nearly identical opti-

cal spectra, which are generally thought to be lensed quasars lacking a definitive detection of foreground lenses due to observational limitations^[24]. Three have more noticeable spectral differences^[20,21,25] (shown in magenta), but without the host galaxy mergers unambiguously observed, they are still debatable as either pre-merger projected quasar pairs at larger physical separations or gravitational lenses. The dual quasar candidate^[7] (shown in blue) exhibits differences in the broad emission line strengths^[35] but is still controversial as a quasar lens or a pre-merger projected quasar pair at larger physical separations due to the lack of high-resolution deep NIR imaging. Finally, a $z = 2.4$ quadruply lensed quasar (shown in green) has been proposed to contain a lensed kpc-scale dual quasar^[27], although the interpretation is subject to lens modeling uncertainties. Importantly, lensing can enhance angular resolution and astrometry to identify such rare systems of small-scale dual quasars that are also strongly lensed^[36], as reported^[37,38] in the dual AGN candidate in the lensed AGN MG B2016+112 at $z = 3.273$ with a separation of hundreds of parsec.

HST/WFC3 NIR Imaging High-resolution NIR imaging of SDSS J0749+2255 was conducted on 2 February 2022 UT (Program GO 16892; PI: X. Liu) with the *HST* Wide Field Camera 3 (WFC3) in the F160W filter in a single orbit. The F160W filter operates in the wide *H* band with a pivot wavelength of 1536.9 nm and a width of 268.3 nm^[39]. The total net exposure time was 2055 s. The IR detector has a pixel size of $0''.13$, which undersamples the PSF with a full width at half maximum (FWHM) of $0''.15$. We dithered

the observations to properly sample the PSFs while rejecting cosmic rays and bad pixels. We reduced the *HST* data following standard procedures⁴⁰. Geometric distortion and pixel area effects were corrected. We combined dithered frames while rejecting cosmic rays and hot pixels. The final image has a pixel scale of $0''.06$ for the *H*-band image to be Nyquist-sampled. The photometric zeropoint is 25.936 (28.177) in the AB (ST) system, which was derived based on observations of four white dwarf standard stars³⁹. Figure 1 shows the *HST* F160W image of SDSS J0749+2255.

Keck AO-Assisted NIR Imaging Keck AO-assisted NIR imaging was conducted on 21 December 2021 UT (Program N072; PI: Y. Shen) with the near-infrared camera NIRC2 in the K_p filter. The K_p filter has a central wavelength of 2.124 μm and a bandpass width of 0.351 μm . The observations were conducted in the Laser Guide Star (LGS) mode. The artificial LGS was used as a reference object to correct for atmospheric distortion. We used the star N85C000573 (located $52''.62$ from SDSS J0749+2255) as the reference tip-tilt star. The images were dithered using a $4'' \times 4''$ 9-position box. The total net exposure time was 1350 s. The individual images were reduced following standard procedures and retrieved through the Keck Observatory Archive. Individual exposures are then co-added together after correcting for the dither offsets. The final image has a pixel scale of $0''.01$. The zeropoint of the Keck K_p -band image is 24.80 (Vega), which was derived based on measurements taken on AO engineering stars during laser checkouts. Supplementary

Figure 2 shows the Keck image of SDSS J0749+2255.

Host Galaxy Morphology and Structural Decomposition We study host galaxy morphology and perform PSF-host structure decomposition for the *HST* F160W (H-band) and Keck AO K_p images. We adopt GALFIT³¹ to model the 2D surface brightness profiles for the galaxy merger. A field star is adopted to build the PSF model for the *HST* image. For the Keck image, a standard star, S852-C, which was observed during the same night closest to SDSS J0749+2255, is adopted to build the PSF model. Since the tip-tilt reference star is much fainter than the standard star, the strehl ratio of SDSS J0749+2255’s Keck image is much lower. Therefore, the standard star is convolved with a wider Gaussian to match the PSF size of SDSS J0749+2255’s observation (with a FWHM of $0''.05$) to correct for this mismatch.

The baseline model adopts two PSFs (for the two unresolved quasars), two Sérsic models (for the two extended host galaxies), and a constant background to fit the image. The Sérsic model is given by

$$\Sigma(r) = \Sigma_e \exp \left[-\kappa \left(\left(\frac{r}{r_e} \right)^{1/n} - 1 \right) \right], \quad (1)$$

where $\Sigma(r)$ is the pixel surface brightness at radial distance r , Σ_e is the pixel surface brightness at the effective radius r_e , and κ is a parameter related to the Sérsic index n . $n = 1$ for an exponential profile, whereas $n = 4$ for a de Vaucouleurs profile. Bulge-

dominated galaxies have high n values ($n > 2$), while disk-dominated galaxies have n close to unity. We try different models until reaching the minimum reduced χ^2 (i.e., χ^2_ν) for the smallest number of parameters. In all models considered, the centers of the Sérsic profiles were left as free parameters, but the best-fit Sérsic profile centers in all the 2 PSF + 2 Sérsic models are consistent with the best-fit PSF centers (Figure 1 and Supplementary Figure 2), suggesting that the best-fit models are reasonable. The only exception was the 2 PSF + 2 Sérsic model for the Keck image, where the center of the Sérsic profile for the NE host was fixed at the NE nucleus PSF model center to help break model degeneracy given the relative shallowness of the AO imaging.

Figure 1 shows the best-fit GALFIT model for the *HST* F160W image. The extended host galaxies are detected after the PSFs are subtracted (i.e., Panel c “Data-PSFs”). After we further subtract the extended hosts (i.e., Panel d “Data-PSFs-Sersics”), low surface-brightness tidal features are revealed in the residual image, providing unambiguous evidence for an ongoing galaxy merger. We quantify the detection significance of the faint tidal features by using manual apertures that roughly trace the boundaries of the four tidal clumps indicated in Panel d of Figure 1. The extracted total fluxes in the aperture are all detected at $> 10\sigma$ significance. Supplementary Table 7 summarizes the detection properties of the tidal features. For the brightest tidal clump T1 (SW of the pair), the detection significance is $\sim 20\sigma$, and the total AB magnitude is 25.47 ± 0.07 , which is much fainter

than the host galaxies. The surface brightness depth of the HST imaging is ~ 24.85 AB mag/arcsec², which is consistent with the expected depth given the integration time.

Supplementary Table 5 lists the best-fit parameters of the GALFIT models for the *HST* F160W image. In addition to the baseline model (with 2 PSFs + 2 Sérsic), we also consider a model for the lensing scenario (Model 2; with 2 PSFs for the quasars plus 1 Sérsic for the foreground lens galaxy) and a more complicated model to attempt the bulge-disk decomposition (Model 3; with 2 PSFs for the quasars plus 4 Sérsic for the two hosts, where the Sérsic indices are fixed at $n = 1$ for the disk and $n = 4$ for the bulge for each host to help break model degeneracy). The baseline model is statistically preferred over the lensing model based on the χ^2_ν value comparison.

While the more complicated model (Model 3) yields a slightly smaller χ^2_ν than the baseline model, the improvement is too small to warrant the larger number of parameters. For the *HST* image, the best-fit Sérsic index in the baseline model for the host of the SW nucleus is $n = 5.2$, suggesting a bulge-dominated host (but see caveats below) with an effective radius of 1.4 kpc, whereas that of the host of the NE nucleus is $n = 0.9$, suggesting a disk-dominated host with an effective radius of 4.3 kpc.

Supplementary Figure 2 shows the best-fit GALFIT model for the Keck AO K_p im-

age. While the sensitivity of the Keck image is much lower than that of the *HST* image, the Keck image offers a higher angular resolution of $\sim 0''.05$ with much better PSF sampling ($0''.01 \text{ pixel}^{-1}$), which is useful for better disentangling the quasars and the hosts, characterizing the host structures, and revealing any foreground lens galaxy in the lensing scenario. After we subtract the two unresolved PSFs, the extended host galaxies of the two quasars are also detected in the Keck AO image. Importantly, most of the extended emission is centered on each of the quasar point sources, suggesting that it is associated with the host galaxies.

Supplementary Table 6 lists the best-fit parameters of GALFIT models for the Keck AO K_p image. The lensing scenario (i.e., with 2 PSFs + 1 Sérsic) is statistically ruled out based on the χ^2_ν value comparison. A more complicated model is not warranted to attempt the bulge-disk decomposition for the Keck image, given its much lower sensitivity, which is not enough to detect any additional extended disks. The tidal features seen in the *HST* image (in the GALFIT model residual) are too faint to detect in the Keck AO image.

For the Keck AO image, the best-fit Sérsic index is $n = 0.2$ for the host of the SW nucleus (with an effective radius of 1.1 kpc) and $n = 0.6$ for the host of the NE nucleus (with an effective radius of 1.8 kpc), suggesting disk-dominated galaxies for both hosts. The Sérsic index estimate for the NE nucleus host is broadly consistent between the *HST*

and Keck images, although the effective radius estimate is larger for the *HST* image than that for the Keck AO image, likely due to the differences in the image resolution (i.e., the more compact disk component is better resolved with Keck AO) and sensitivity (i.e., the *HST* image is deep enough to reveal any additional extended disk component). On the other hand, the SW nucleus host's morphological type is more controversial. The discrepancy may be caused by the different image resolutions and PSF samplings, i.e., the smaller SW host is more challenging to robustly resolve with *HST* imaging given its lower angular resolution and poorer PSF sampling. We conclude that both hosts are likely to be disk-dominated galaxies based on the Keck-based measurements, given the higher angular resolution (PSF FWHM of $0''.05$ for Keck compared to $0''.15$ for *HST*) and finer PSF sampling ($0''.01 \text{ pixel}^{-1}$ for Keck compared to $0''.06 \text{ pixel}^{-1}$ for *HST*) for more robustly disentangling the quasars and their host galaxies and measuring the host galaxy structures, even though the *HST* PSF is more stable.

Galaxy merger simulations have shown that galaxy structure plays a dominant role in regulating nuclear activity in major mergers: inflows in bulgeless galaxies occur earlier in galaxy mergers whereas stellar bulges stabilize galaxies against bar modes and inflows until the galaxies merge⁴¹. The disk dominance of the hosts in SDSS J0749+2255 may be critical for efficient SMBH fueling to produce two luminous quasars in the early-stage merger without the stabilization effect of stellar bulges which are often present in low-

redshift low-luminosity dual AGNs⁴⁸.

Strong Lensing Tests After subtracting the two quasar PSFs, we also detect significant extended emission around the two quasars in both HST and Keck AO images. At redshift $z \sim 2$, the host galaxies of luminous quasars are detectable in high-resolution near-IR imaging from HST or ground-based AO⁴². Therefore it is reasonable to attribute the extended emission to quasar host galaxies. Importantly, the Keck AO imaging has much better spatial resolution than HST to show that most of the extended emission is at the locations of the two quasars, rather than in between as resulting from a putative foreground lens galaxy. Even in the HST image (Panel c in Figure 1), the centroid of the extended emission lies closer to the brighter SW nucleus. Thus the locations of the extended emission contradict a simple lens model where the image closer to the lens is expected to be fainter. Assuming most of the detected extended emission is from the host galaxy emission, the residual map in Panel d of Figure 1 shows no clear presence of a central putative lens galaxy. But the required lens galaxy with a minimum total brightness $H \lesssim 22$ (AB magnitude) would be strongly detected (see below). In addition, if the two quasars were lensed images, their host galaxy images would show noticeable distortions, which are not observed (i.e., the 2 Sersic profiles in Model 1 are well fit). The average surface brightness for the two hosts within the half-light radius is also markedly different (Supplementary Tables 5 and 6), contradictory to the strong lensing expectation.

Alternatively, we could interpret the extended emission as mostly from a foreground lens galaxy, with limited contribution from the host galaxy of the background, strongly lensed quasar. In this case the lens galaxy would be bright, with $H \sim 20.7$ (AB magnitude). Given the image separation and source redshift, the lensing galaxy should be brighter than $H \sim 22$ AB mag irrespective of its redshift as predicted from the Faber-Jackson relation⁴³. A lens galaxy as bright as $H \sim 20.7$ (AB) would be possible if the lens galaxy were at redshift $z \sim 0.3$ or $z \sim 1.8$ (the two solutions from the lens model). This way we could reconcile the lack of distorted morphology of the lensed host galaxy and the difference in the surface brightness of the lensed images. However, the locations of the extended emission, and the expectation of detecting significant host emission in $z \sim 2$ quasars are at odds with this alternative scenario. For example, in this scenario, the estimated total stellar mass of the quasar host would be much lower, resulting in exceedingly massive black holes with respect to the host stellar mass.

To more quantitatively test the strong lensing hypothesis where most of the extended emission comes from the foreground lens galaxy, we perform strong lensing mass modeling tests using the glafic software⁴⁴ for both the *HST* and the Keck images where the putative lens galaxy's position and mass can both be constrained with GALFIT. In this scenario, what we observe in the *HST* image would be two quasar images (two PSFs) plus the lens galaxy (extended emission in the PSF-subtracted residuals). We then assume that

“Model 2” listed in Supplementary Table 5 (from the GALFIT analysis of the *HST* image) tells us the quasar image positions as well as the position of the lens galaxy in the lensing scenario. We have seven constraints from the observation, i.e., the positions of the two quasar images (4), the flux ratio (1), and the position of the lens galaxy (2). The strong lensing mass modeling is illustrated in Supplementary Figure 3.

Our baseline model assumes a simple singular isothermal ellipsoid (SIE) lens. The SIE model contains seven parameters, i.e., the mass strength (1), the lens position (2), the ellipticity and its position angle (2), and the source position (2). Therefore, the adoption of an SIE model is justified given the number of constraints.

We assume positional errors of $0''.01$ for PSFs and $0''.03$ for the lensing galaxy. These positional uncertainties were inflated by a factor of ~ 5 from the GALFIT uncertainties, which are statistical only and likely underestimated. Adopting a positional error of $0''.05$ instead for the lensing galaxy does not affect the conclusion but makes the fitting more unstable. Assuming an SIE lens model, we were able to obtain a mass model that fit the positions and flux ratio of the double optical nuclei, so the lensing hypothesis cannot be rejected based on the Model 2 best-fit positions for the *HST* image (Supplementary Table 5) alone. However, Model 1 is statistically favored over Model 2 based on GALFIT results, which means that the dual quasar scenario is preferred by the data over the strong

lensed quasar scenario, even though the quantitative strong lensing mass modeling test cannot rule out lensing (where the extended emission is assumed to be the foreground lens). In addition, the residual map of Model 2 (the strong lensing hypothesis) shows excess fluxes at the locations of the quasars, indicating host galaxy emission that is not properly accounted for by the model.

A more definitive test of the strong lensing hypothesis comes from the Keck AO imaging, given its higher resolution and better localized extended emission. We performed the same lensing mass modeling for the Keck image using glafic, assuming positional errors of $0''.01$ for PSFs and $0''.03$ for the lensing galaxy. Assuming a simple SIE lens model, we were unable to obtain a mass model that could fit the positions and flux ratio of the double optical nuclei. With a more complicated SIE plus shear model, we got a reasonably good fit, but it would require very large external shear (with a best-fit value of ~ 0.6) and therefore quite unrealistic. Thus our mass modeling test suggests that the Model 2 best-fit positions for the Keck AO image (Supplementary Table 6) are inconsistent with the lensing hypothesis.

Supplementary Table 8 lists all the parameters (mean values and 1σ uncertainties) obtained using glafic for the SIE or the SIE+shear models from the Markov Chain Monte Carlo (MCMC) analysis for both the *HST* and Keck images. Since the number of degree of

freedom for the SIE model is zero, if the model is reasonable, we expect to have $\chi^2_{\min} \sim 0$ for the best-fit model. If this is not the case, it means that either the assumption on the lens mass model is incorrect, or it is not a strong gravitational lens system. For some of the parameters, the posterior distribution is not necessarily Gaussian, so the mean can be different from the best-fit value (e.g., in the Keck SIE+shear model, the best-fit value of external shear is ~ 0.6 but the mean is 0.48 ± 0.12).

HST/STIS Optical Slit Spectroscopy Spatially resolved optical slit spectroscopy was conducted for SDSS J0749+2255 on 16 February 2021 UT (Program GO 16210; PI: X. Liu) with the *HST*/Space Telescope Imaging Spectrograph (STIS) using the G750L grating. A slit width of $0''.2$ was used. The slit direction was oriented at P.A. = 236.4° (E of N) to simultaneously cover the two nuclei. The total net exposure time was 2136 s. The data were reduced using the *stis.cti* package, which removes trails and other artifacts caused by Charge Transfer Inefficiency effects⁴⁵. The exposures were flat-fielded and combined to reject cosmic rays. The final 1-dimensional (1D) spectra for each nucleus were extracted using an aperture size of 7 pixels, corresponding to $0''.35$. Supplementary Figure 4 shows the *HST*/STIS 2-dimensional (2D) spectrum and demonstrates the extraction of the 1D spectra for each nucleus. The two nuclei are well separated by HST/STIS. Figure 4 shows the final STIS spectra for the double nucleus.

Gemini/GMOS Optical Slit Spectroscopy Gemini spatially resolved optical slit spectroscopy was conducted on April 20 2022 UT (Program GN-2022A-Q-139; PI: X. Liu) with the Multi-Object Spectrographs (GMOS) on Gemini-North. The R150 grating and a slit width of $0''.5$ were used, yielding a spectral resolution of $R \sim 631$. The slit direction was oriented at P.A.= 236.1° (E of N) to simultaneously cover the two nuclei. The total net exposure time was 1876 s. The average seeing during the observations was $\sim 0''.49$ in the optical wavelengths. The spectra cover the wavelength range of 400-1200 nm. The data were reduced using the Gemini PyRAF/IRAF pipeline following standard procedures. The reduction steps included bias subtraction, flat-field correction, cosmic-ray rejection, wavelength calibration, and flux calibration. The CuAr spectra with the same instrumental configurations were used for the wavelength calibration. EG131 was observed as the standard star for spectrophotometric calibration.

Supplementary Figure 5 shows the Gemini/GMOS 2D spectrum and demonstrates the source decomposition of the 1D spectra for each nucleus. Based on the 1D spatial profile of the wavelength-averaged spectrum, the two nuclei are well modeled by two Gaussian components whose separation is consistent with that observed in the *HST* optical images. We have decomposed the two nuclei using two Gaussians using a fixed separation of $0''.46$ across all wavelengths to help with the decomposition. The FWHM of the Gaussian components is a free parameter during the fitting process to account for the

variable seeing as a function of wavelength. Figure 4 shows the final decomposed GMOS spectra for the double nucleus.

Gemini/GNIRS NIR Slit Spectroscopy Gemini spatially resolved NIR slit spectroscopy was conducted on 22 December 2020 UT (Program GN-2020B-FT-113; PI: X. Liu) with the Near Infra-Red Spectrograph (GNIRS) on Gemini-North. We used the cross-dispersion (XD) 32/I mode with the blue camera and a slit width of $0''.45$, yielding a spectral resolution of $R \sim 1000$. The total net exposure time was 2400 s. The average seeing during the observations was $\sim 0''.34$ in the NIR wavelengths. The detector controller artifacts in the raw images were removed using the `cleanir.py` script. Then, the data were reduced with the Gemini IRAF pipeline following standard procedures. The reduction steps included flat-field correction, S-distortion correction, wavelength calibration, telluric correction, and flux calibration. The pinhole images were used to correct the S-distortion in the images and then the Argon spectra were used for the wavelength calibration. HIP26000 was observed as the standard star for telluric correction and spectrophotometric calibration.

Supplementary Figure 6 shows the Gemini/GNIRS 2D spectrum and demonstrates the source decomposition of the 1D spectra for each nucleus. Based on the 1D spatial profile of the wavelength-averaged spectrum, the two nuclei are marginally resolved. The 1D profile is consistent with two Gaussian components with separation of $0''.46$, although

a third faint extended component is seen near the SW nucleus, which is likely due to the extended host galaxy emission in the merger. We adopted the local minimum between two Gaussian as the central aperture boundaries. The final 1D spectra for each nucleus were extracted using a boxcar aperture with a size of 21 pixels (corresponding to $1''.05$) for the SW nucleus and 7 pixels (corresponding to $0''.35$) for the NE nucleus. The aperture for the NE nucleus does not fully cover the tail of the emission (>91 pixels) because it reaches the edge of the slit. Figure 4 shows the final decomposed GNIRS NIR spectra for the double nucleus.

Joint Spectral Modeling of the HST/STIS and Gemini/GNIRS Data We jointly fit the optical and NIR spectra of each nucleus to estimate the systemic redshift and measure emission-line properties for each quasar. We use the PyQSOFit fitting code^[7] following the procedures as described in detail in^[46]. Here we briefly summarize the fitting procedures and model assumptions. Each combined spectrum has been de-reddened for Galactic extinction using the Milky Way reddening law^[47] and $E(B - V)$ derived from the dust map^[48]. The spectrum is modeled as a linear combination of a pseudo-continuum (consisting of a power-law plus polynomial quasar continuum and Fe II emission), broad, and narrow emission lines. A Balmer continuum component has also been considered in the models but made negligible contribution in the best-fit results. First, to estimate the systemic redshift, we perform a global fitting to the joint optical-NIR spectrum. Given the moderate S/N of

the spectra, the velocities of all the narrow and broad emission lines have been tied to be the same to help break model degeneracy. The best fit systemic redshift is 2.1679 ± 0.0003 for the SW nucleus and 2.1665 ± 0.0010 for the NE nucleus which we take as our baseline estimates. Alternatively, allowing for a nonzero velocity offset between the narrow and broad emission lines yields a best fit systemic redshift of 2.1686 ± 0.0007 for the SW nucleus and 2.1653 ± 0.0021 for the NE nucleus based on the narrow emission lines, which are consistent with the baseline estimates within uncertainties. The redshift uncertainties are statistical uncertainties only as quasar emission lines tend to have different velocity shifts from the systemic velocity that depend on the line species⁴⁹. Nevertheless, these statistical uncertainties are still useful to assess the redshift difference for the strongly lensed quasar scenario.

Second, to make emission-line measurements and derive physical quasar properties, we perform local fittings for each emission-line complexes. Given the moderate S/N of the existing spectra, each narrow and broad emission line component is modeled by a single Gaussian, where an upper limit of $\text{FWHM} = 1200 \text{ km s}^{-1}$ has been imposed on the narrow lines. The velocities of all the narrow emission lines have been fixed to be the same as the best-fit systemic redshift of each quasar, whereas non-zero velocity offsets are allowed for each broad emission lines (i.e., with respect to the systemic redshift) to better account for possible velocity offsets in different broad lines due to ionization stratification, accretion

disk winds, and/or IGM absorption. The widths of all the narrow emission lines are tied to be the same to help break model degeneracy, but the widths of the broad emission lines are allowed to vary among different lines. The flux ratios of the [O III] and [N II] doublets are fixed to be their theoretical values to help break model degeneracy and reduce ambiguities in decomposing the emission line complexes. Supplementary Figure 7 shows the best-fit spectral models for both the global and the local fits. Supplementary Table 2 lists the emission-line measurements from the local best-fit models.

Chandra ACIS-S X-ray 0.5–8 keV Imaging Spectroscopy, X-ray Data Reduction, and Data Analysis Using BAYMAX We observed SDSS J0749+2255 with the ACIS-S on board the *Chandra* X-ray Observatory on 23 Dec 2021 UT (Program 23700377, ObsID=25710 and 26245; PI: X. Liu). SDSS J0749+2255 was observed on axis on the S3 chip with a total exposure time of 30 ks. We follow a standard data reduction procedure for each observation using *Chandra* Interactive Analysis of Observations software (CIAO) v4.13⁵⁰. We correct for astrometry by cross-matching *Chandra*-detected point-like sources with available optical catalogs. We require a minimum of 3 matches between the X-ray observations and optical catalogs, and we required each matched pair to be separated by less than 2". For each *Chandra* observation, we find background flaring to be negligible, with no time interval containing a background rate 3σ above the mean level. We restrict our analysis to photons with energies between 0.5–8 keV. We analyze the pho-

tons that are contained within a $20'' \times 20''$ box centered on the nominal X-ray coordinates of SDSS J0749+2255. Supplementary Figure 8 shows the binned *Chandra* data, with the soft X-rays (0.5–2 keV) shown in red on the left (total of 114 counts) and the hard X-rays (2–8 keV) shown in purple on the right (total of 126 counts).

We analyze the *Chandra* observations for the presence of two X-ray point sources using `python` tool BAYMAX (Bayesian AnalYsis of Multiple AGN in X-rays)^{[51][53]}. BAYMAX allows for a quantitative analysis of whether a source in a given *Chandra* observation is more likely composed of one or two point sources. The main component of BAYMAX is the calculation of the Bayes factor $\mathcal{B}$, which represents the ratio of the plausibility of observed data, given each model. BAYMAX takes calibrated *Chandra* events and compares them to the expected distribution of counts for a single point source (M_1) versus a dual point source (M_2) model.

When calculating the Bayes Factor, BAYMAX uses `python` nested sampling package `nestle`, where both probability and prior densities need to be defined. The probability densities of M_1 and M_2 are estimated by comparing the *Chandra* sky coordinates (x, y) and energies (E) of each detected X-ray event to simulations based on single and multiple point source models^[51]. The prior vectors for M_1 and M_2 are $\theta_1 = \{\mu_1, f_{\text{bkg}}\}$ and $\theta_2 = \{\mu_1, \mu_2, f, f_{\text{bkg}}\}$, respectively. Here, the sky-x, sky-y coordinates of each point source

(encompassed by μ) are uniform distributions across the entire $20'' \times 20''$ field of view. The parameter f_{bkg} represents the number of counts associated with a spatially uniform background component versus a point-source component; the prior density is represented in log-space with log-normal distribution with a mean of -2 and a large standard deviation ($\sigma=0.5$) to allow flexible movement in prior space. Lastly, the parameter f represents the count ratio between the secondary and primary point-source; the prior density is represented in log-space with a log-normal distribution set between -4 and 0 . For sources with more than one *Chandra* observations, as in the case of SDSS J0749+2255, the code also fits for astrometric shifts between each observation, although the point spread function model used in the probability densities are unique to each observation.

These are the standard “non-informative” priors used by BAYMAX when analyzing X-ray point sources^{[52](#)[53](#)}. Running BAYMAX on the *Chandra* observations of SDSS J0749+2255 we find $\log \mathcal{B} = 3.3 \pm 0.7$. The error bars are returned by `nestle` and represent the statistical sampling error of the nested sampling procedure. Previous analyses have found this error to be consistent with the $\sim 68\%$ confidence interval when running BAYMAX 100 times on the same source^{[52](#)}.

In addition to calculating a Bayes factor, BAYMAX estimates optimal values for parameters such as the separation (r) and count ratio (f) using `python` MCMC package

PYMC3. We find best-fit values of $r = 1''.54^{+0''.39}_{-0''.36}$ and $\log f = -1.43^{+0.30}_{-0.30}$, at the 95% confidence interval. The overestimation of r as compared to the resolved separation of the two optical cores (along with the large error-bars) are likely a result of the secondary X-ray point source having, on average, only 6 counts. The low number of counts associated with a putative secondary X-ray point source (~ 6) affects the accuracy of estimations on r . However, the estimated position angle between the two X-ray point sources (350^{+37}_{-24} deg.; 95% confidence range) is consistent with the two resolved cores in the *HST* imaging (330 deg.). Deeper *Chandra* observations can help better discern whether the X-ray emission is spatially consistent with the *HST* imaging.

Sampling from the posterior distributions, the code will probabilistically assign counts to different model components (which, for M_2 is the primary point source, the secondary point source, and a background component), from which a spectra can be created and analyzed. This allows for the individual spectral fitting of each X-ray point source in a putative AGN pair. Iterating this process results in a set of 100 spectral realizations for each model component. We fit and model each spectral realization, for each X-ray point source, using XSPEC⁵⁴. This allows us to quantify the average values of the spectral shape and X-ray luminosity, across all spectral fits. With the fitted X-ray spectra, we estimate the hard X-ray (2–8 keV) flux ratio of the two nuclei (NE/SW) to be 45^{+41}_{-33} .

Supplementary Figure 9 shows the spectral fits to the primary (left) and secondary (right) X-ray point source detected by BAYMAX. We find both sets of spectral realizations are best-fit with a simple absorbed power law (XSPEC model components: $\text{phabs} \times \text{zphabs} \times \text{zpow}$). Given the low number of counts for both components (average of ~ 191.5 and ~ 6 counts for the primary and secondary, respectively), we fix the power law index, Γ , to a value of 1.8 for all spectral fits. Supplementary Table 3 lists the median values and the confidence interval of the best-fit spectral parameter distributions.

X-ray-to-optical Luminosity Ratio as Characterized by the Spectral Slope α_{OX} The slope of a hypothetical power law from 2500 Å to 2 keV is usually defined by $\alpha_{\text{OX}} = 0.3838 \log(l_{2500 \text{ Å}}/l_{2 \text{ keV}})$, where $l_{2500 \text{ Å}}$ is the rest-frame monochromatic optical luminosity (in units of $\text{erg s}^{-1} \text{ Hz}^{-1}$) at 2500 Å measured from the dereddened spectrum for each nucleus and $l_{2 \text{ keV}}$ is the rest-frame 2 keV luminosity density (in units of $\text{erg s}^{-1} \text{ Hz}^{-1}$) ⁵⁵. The X-ray-to-optical spectral indices α_{OX} are calculated as 1.46 ± 0.01 for the NE nucleus and 2.32 ± 0.13 for the SW nucleus. As an intrinsic probe of the accretion processes, α_{OX} characterizes the relation between the optical/UV thermal blackbody radiation from the accretion disk and the X-ray emission from the hot corona. α_{OX} of optically selected quasars has been observed to strongly correlate with optical luminosity with a possible residual evolution with redshift ⁵⁵. Compared with the typical optically selected quasars at similar luminosities ⁵⁵, α_{OX} of the NE nucleus is consistent the general population. On the

other hand, α_{OX} of the SW nucleus (which also has a higher estimated Eddington ratio) is much larger, which may be evidence for the SMBH experiencing some unusual accretion state induced by the galaxy merger. Alternatively, the large α_{OX} of the SW nucleus could be evidence of high circumnuclear absorption as expected in galaxy mergers, although the absorption material would have to be at the smallest scales that only affect the X-ray emitting region but do not obscure the optical emitting region given the Type 1 nature of the optical nucleus.

VLA Observations, Data Reduction, and Measurements Dual-band radio imaging was conducted on 11 and 12 December 2020 UT (Program 20B-242; PI: X. Liu) with the VLA in its most extended A configuration. C-band (6.0 GHz) and Ku-band (15.0 GHz) were adopted with an on-source time of 0.5 hours for each band. The images were calibrated and reduced through the standard VLA Calibration Pipeline version 5.6.2 in the Common Astronomy Software Applications package (CASA)⁵⁶. Sources 3C147 and 3C286 were used as the flux and bandpass calibrators. Source J0738+1742 was used as the gain calibrator as well as the phase reference source in the Ku-band. We fit 2D Gaussian components to each nucleus to obtain the flux densities. Two nuclei are well resolved in the Ku-band image and are consistent with two point sources. The two nuclei are marginally resolved in the C-band image, therefore, we fit the C-band image using two Gaussian components with a fixed shape, same as the synthesized beam, to obtain accurate flux estimations. Figure 3

shows the VLA A-config two-band radio images for SDSS J0749+2255. Supplementary Table 4 lists the radio measurements.

Estimation of Virial Black Hole Masses, Bolometric Luminosities, and Eddington Ratios We estimate SMBH properties for each nucleus using the combined GMOS+GNIRS spectra. These include SMBH masses, bolometric luminosities, and Eddington ratios. The GMOS spectrum is adopted in our baseline estimates because of the higher S/N than that of the STIS spectrum.

We estimate SMBH masses using the single-epoch virial mass estimator⁴⁶. The virial SMBH mass is estimated as

$$\log_{10} \left(\frac{M_{\text{BH}}}{M_{\odot}} \right) = a + b \log_{10} \left(\frac{\lambda L_{\lambda}}{10^{44} \text{ erg s}^{-1}} \right) + c \log_{10} \left(\frac{\text{FWHM}}{\text{km s}^{-1}} \right), \quad (2)$$

where L_{λ} is the monochromatic continuum luminosity at wavelength $\lambda = 5100 \text{ \AA}$, FWHM is the full width at half maximum of the $\text{H}\alpha$ line, and $a = 1.390$, $b = 0.555$, $c = 1.873$ are the coefficients from the calibration of⁴⁶. The $\text{H}\alpha$ -based virial BH masses are $\sim 10^{9.1 \pm 0.4} M_{\odot}$ for the NE nucleus and $\sim 10^{9.2 \pm 0.4} M_{\odot}$ for the SW nucleus, whereas the errors indicate total 1σ uncertainties including systematic errors. We have also estimated the virial BH masses using the broad Mg II lines as an independent double check. The Mg II -based BH masses are estimated as $\sim 10^{8.9 \pm 0.4} M_{\odot}$ for the NE nucleus and $\sim 10^{9.3 \pm 0.4} M_{\odot}$ for the SW nucleus, which are consistent with the $\text{H}\alpha$ -based BH masses within the systematic

uncertainty of ~ 0.4 dex⁵⁷.

We estimate the bolometric luminosity using the rest-frame optical flux density at 5100 Å and the relevant Bolometric Correction (BC)³. The bolometric luminosity is estimated as

$$L_{\text{bol}} = \lambda L_{\lambda} \times BC, \quad (3)$$

where L_{λ} is the monochromatic continuum luminosity at rest-frame wavelength $\lambda=5100$ Å and $BC = 10.33$ is the bolometric correction at 5100 Å³. The bolometric luminosities are estimated as $\sim (2.0 \pm 0.4) \times 10^{46}$ erg/s for the NE nucleus and $\sim (5.0 \pm 1.0) \times 10^{46}$ erg/s for the SW nucleus.

To infer the Eddington ratio, we calculate the Eddington luminosity from the BH mass as,

$$L_{\text{Edd}} \sim 1.26 \times 10^{38} \left(\frac{M_{\text{BH}}}{M_{\odot}} \right) \text{ ergs}^{-1}. \quad (4)$$

The Eddington ratios $L_{\text{bol}}/L_{\text{Edd}}$ are estimated as $\sim 0.12 \pm 0.03$ for the NE nucleus and $\sim 0.25 \pm 0.07$ for the SW nucleus.

In addition to the inferences based on the optical measurements, we also estimate the bolometric luminosities and Eddington ratios based on the X-ray measurements as an independent double check. We use the X-ray luminosity $L_{2-8 \text{ keV}}$ and apply the X-ray

bolometric correction for radio-loud quasars (Equation 14 in [58](#)). The estimated bolometric luminosities are $\sim 10^{46.59^{+0.03}_{-0.02}}$ erg/s for the NE nucleus and $\sim 10^{45.72^{+0.17}_{-0.30}}$ erg/s for the SW nucleus. The inferred Eddington ratios are $\sim 0.23^{+0.04}_{-0.03}$ for the NE nucleus and $\sim 0.03^{+0.01}_{-0.02}$ for the SW nucleus. The bolometric luminosities and Eddington ratios are in broad agreement with the optically based estimates considering substantial scatters in the X-ray bolometric correction [58](#).

X-ray-to-Radio Multi-wavelength Spectral Energy Distributions Supplementary Figure 10 shows the multi-wavelength spectral energy distributions (SEDs) for both nuclei in SDSS J0749+2255. The data include photometry in the radio, infrared, optical and X-ray wavelengths, as well as the optical and NIR spectra. For context, also shown are various optically-selected quasar SED templates from the literature [3](#)[59](#). Overall, the SEDs of both nuclei are consistent with the SED templates of typical optically selected (single) quasars. Both nuclei are classified as radio-loud quasars.

Estimation of Host-galaxy Stellar Masses We estimate the total stellar mass for the host galaxy of each nucleus using the *HST* F160W and Keck K_p -band PSF-subtracted host-only magnitudes. We first obtain the *HST* F160W magnitudes and Keck K_p -band magnitudes of the host galaxies by decomposing the 2D surface brightness profiles using GALFIT. We then run the stellar population synthesis code GALAXEV [60](#) to create composite stellar

population (CSP) models to fit the observed two-band host-only photometry. While only two-band data are available for the hosts, they allow a reasonable estimate of the mass-to-light ratio (M/L) to estimate the stellar mass. The BaSeL spectral library, a Chabrier initial mass function (IMF)^[61], and an exponentially decaying star formation history (SFH) are assumed to build the CSP SED. The evolution tracks for three different metallicities ($Z=0.02$, 0.008 , and 0.004)^{[62][63]} are adopted to model any CSP change due to different metallicities. Dust attenuation assuming a fixed optical depth is adopted. We create a grid of CSPs from assuming different stellar population parameters. The fitted free parameters include the total stellar mass, age, metallicity, the exponential decay star formation timescale, and the dust optical depth. The CSP models are fit to the PSF-subtracted *HST* F160W and Keck K_p -band magnitudes to estimate the best-fit stellar population parameters of the host galaxies.

Supplementary Figure 11 shows the best-fit CSP models and the PSF-subtracted host-only two-band photometry. The total stellar masses are estimated as $\sim 10^{11.46^{+0.02}_{-0.26}} M_{\odot}$ for the NE nucleus and $10^{11.50^{+0.04}_{-0.27}} M_{\odot}$ for the SW nucleus, whereas the errors include 1σ statistical uncertainties and the partial systematics from surface brightness profile fitting. This partial systematic uncertainty of total stellar masses is estimated to be ~ 0.2 dex from assuming different models during the surface brightness decomposition (e.g., 2 Sérsic versus 1 Sérsic for each host) and should be taken as a lower limit since it does not include

the additional systematics from adopting different assumptions about the IMF, the SFH, or the unknown dust configuration, which may result in an additional ~ 0.3 dex uncertainty at least^[64].

Black Hole Mass versus Stellar Mass Scaling Relations To put SDSS J0749+2255 into context, we study the relation between BH mass and host-galaxy total stellar mass by comparing SDSS J0749+2255 with observations of the host galaxies of single quasars and AGNs, as well as inactive galaxies in the literature. Supplementary Figure 12 shows the relation between BH mass and host total stellar mass for SDSS J0749+2255, in comparison with various literature samples. Included for comparison are the local samples of both inactive and active galaxies^[65-67] and samples of AGNs and quasars at $z > 0.2$ selected with various methods^{[42][68-71]}. Overall, the two nuclei of SDSS J0749+2255 are broadly consistent with both the local scaling relations for inactive galaxies^[65] and *HST*-imaged AGNs^[66] and the high-redshift quasar samples, considering the large systematic errors of our measurements. The hosts of SDSS J0749+2255 already follow the local scaling relations without prominent stellar bulges. This suggests that at least some SMBHs may have formed before their host stellar bulges, in contrast to naive expectations from the canonical SMBH-bulge co-evolution scenario where SMBHs and their host galaxy stellar bulges form coevally from merger-induced starbursts and quasar fueling^{[72][73]}.

Given the high redshift and high luminosities of SDSS J0749+2255, it is challenging to construct a large tailored sample of control quasars for a fairer comparison with similar selection methods, properties, and data qualities. The high angular resolution images in the rest-frame optical are expensive to obtain for even larger samples. Previous works have studied optical selected quasars and estimated the total stellar mass from decomposing *HST* NIR images^{[69][71]}. However, some of the studied quasars^[71] are at significantly lower redshifts ($0.2 < z < 0.6$) than that of SDSS J0749+2255. Others are focused on quasars with high SMBH masses ($10^9 - 10^{10} M_\odot$) only^[69]. While some work also studied optical selected quasars^[68], the stellar mass was estimated from fitting the broad-band SED rather than spatially decomposing the PSF and the host. Some works were based on X-ray selected AGNs in the COSMOS field^{[42][70]}, which are likely to trace a different SMBH population than optically selected quasars given that X-ray selected AGNs in small deep extragalactic fields tend to have lower bolometric luminosities than typical optically selected bright quasars from wide-area surveys.

Estimation of Merging Timescales Chandrasekhar dynamical friction timescale arguments are adopted to estimate characteristic timescales for the galaxy-galaxy merger and the inspiraling timescales of the two SMBHs. Assuming singular isothermal spheres for the host galaxy density profiles and circular orbits, the dynamical friction timescale for a satellite galaxy with velocity dispersion σ_S inspiraling from radius r in a host galaxy with

velocity dispersion σ_H is estimated as^[74]

$$t_{\text{fric}}^{\text{gal}} = \frac{2.7}{\ln \Lambda} \frac{r}{30 \text{ kpc}} \left(\frac{\sigma_H}{200 \text{ km s}^{-1}} \right)^2 \left(\frac{100 \text{ km s}^{-1}}{\sigma_S} \right)^3 \text{ Gyr}, \quad (5)$$

where Λ is the Coulomb logarithm with typical values of $3 \lesssim \ln \Lambda \lesssim 30$ ^[75]. We assume $\ln \Lambda \sim 2$, which is appropriate for equal mass mergers^[76]. Assuming the SW galaxy as the host (primary) and the NE one as the satellite (secondary), the two galaxies would merge in $t_{\text{fric}} \sim 7_{-5}^{+64}$ Myr, corrected for projection effects assuming random orientation (i.e., $r = r_p / \sin < I >$, where $r_p = 3.8$ kpc is the projected separation between the two galaxies and the median inclination angle is $< I > = 60^\circ$ for random orientations).

After merging together with their individual hosts, the two SMBHs would inspiral to the center of the merged galaxy under dynamical friction with the stellar background. Assuming a singular isothermal sphere for the density distribution of the merged host galaxy, the dynamical friction timescale of a BH of mass M_{BH} on a circular orbit of radius r is estimated as

$$t_{\text{fric}}^{\text{BH}} = \frac{19}{\ln \Lambda} \left(\frac{r}{5 \text{ kpc}} \right)^2 \frac{\sigma_H}{200 \text{ km s}^{-1}} \frac{10^8 M_\odot}{M_{\text{BH}}} \text{ Gyr}, \quad (6)$$

where $\ln \Lambda \sim 6$ for typical values^[75]. We estimate $t_{\text{fric}} \sim 0.22_{-0.16}^{+0.50}$ Gyr for the SMBH in the NE nucleus reaching the center of the merger products and forming a gravitationally bound binary SMBH system with the SW nucleus, assuming a radius of $r = 4$ kpc at the start of the orbital decay.

The effects of friction by gas have been neglected in the estimation, which will accelerate the merger processes. On the other hand, the effects of tidal stripping of stars in the satellites have been neglected, which will delay the merger. The effect of BH accretion have also been neglected in the simple estimation, which may change the masses of the BHs and the resulting merger timescales. The actual merger involving the stellar and BH components with gas friction and gas accretion onto two SMBHs is likely much more complicated. Numerical simulations with model parameters tailored to those of SDSS J0749+2255 are needed to make more realistic predictions of the subsequent merger of the system^[77]. Regardless of the actual sequence of the mergers, the SMBHs will form a gravitationally interacting bound binary system.

Implications on the Abundance of $z \sim 2$ Kpc-Scale Dual Quasars Observational constraints on the occurrence rate of kpc-scale dual quasars at cosmic noon are still accumulating^[78,79]. A recent study using Gaia DR3 resolved pairs around luminous quasars at $z \sim 2$ reported a dual quasar fraction of $\sim 6 \times 10^{-4}$ over $\sim 3 - 30$ kpc pair separations^[79]. About one third of the sample have pair separations less than ~ 10 kpc. In this measurement both members of the pair are optically-unobscured quasars with similar physical properties as SDSS J0749+2255. Some of the double quasars in the sample^[79] are gravitationally lensed quasars, which require further confirmation in follow-up observations; therefore the quoted dual quasar fraction is an upper limit. Confirming SDSS J0749+2255 as a gen-

uine kpc-scale dual quasar is important in the sense that it implies at least some of the sub-arcsec double quasars at $z \sim 2$ are physical pairs, thus justifying dedicated follow-up observations to compile a statistical sample of such dual quasars at cosmic noon.

Comparison with Cosmological Hydrodynamic Simulations To compare SDSS J0749+2255 with theoretical predictions for similar systems, we utilize the very recent large-volume, high-resolution (gravitational softening of $1.5 \text{ kpc}/h$; dark matter mass resolution of $9.6 \times 10^6 M_\odot$) cosmological hydrodynamic simulation **ASTRID**^[80-82]. **ASTRID** includes sub-grid models for star formation, SMBH accretion, and the associated supernova and AGN feedback, recently updated with an SMBH seed population between $3 \times 10^4 M_\odot/h$ and $3 \times 10^5 M_\odot/h$ and a sub-grid dynamical friction model to follow the SMBH dynamics down to kpc scales. With a comoving volume of $(250 \text{ Mpc}/h)^3$, **ASTRID** is the largest galaxy formation simulation up to date that covers the epoch of the cosmic noon, and can provide a statistical sample of the rare quasar pair population down to galactic scales.

By searching through a redshift range of $z = 2.0\text{--}2.3$ in **ASTRID**, we have identified 8 quasar pairs with $L_{\text{bol}} > 10^{46} \text{ erg/s}$ for both quasars in the pair and with 3D spatial separations of $\Delta r < 30 \text{ kpc}$. Out of these 8 close quasar pairs, 6 only become both luminous ($L_{\text{bol}} > 10^{46} \text{ erg/s}$) after the host galaxy merger and when the pair separation is below $\sim 2 \text{ kpc}$. Only two pairs are in the early stages of the galaxy merger and hence have

separate host galaxies, similar to SDSS J0749+2255. For both pairs (PairA and PairB) we find that the quasars each contain SMBHs with masses of $M_{\text{BH},2} \sim [0.8 - 4] \times 10^9 M_{\odot}$, and are embedded in galaxies with $M_{\text{gal}} \sim \text{a few } 10^{11} M_{\odot}$.

Supplementary Figure 13 shows the $M_{\text{BH}} - M_{\text{gal}}$ relation for the eight bright quasar pairs. The colored lines mark PairA and PairB with apparently distinct galaxy bulges, although the kinematic bulge-disk classification adopted by simulations cannot be reliably used in interacting systems, whose morphology is typically disturbed and the disks are in the process of being disrupted in the merger process. PairB is the closest match with SDSS J0749+2255 in both the SMBH masses and the galaxy stellar masses.

In order for the pairs to be the closest match with SDSS J0749+2255 (i.e. $L_{\text{bol},2} > 10^{46} \text{ erg/s}$ and in separate galaxies), they are likely in an early phase of the galaxy merger, and in particular at their first few pericentric passages. This is a stage where both the quasar luminosities and the host galaxy star formation are enhanced from the merger-induced gas inflow. Supplementary Figure 14 shows the orbits of the two quasars in one of such pairs along with the host galaxy merger. This pair has a very eccentric orbit with an eccentricity of 0.94 before the merger, and significant host galaxy tidal disruption is seen, after which the secondary quasar becomes offset from the galaxy merger remnant. The shown pair (PairB) merges within $\sim 80 \text{ Myrs}$ after its properties are closest to those of

SDSS J0749+2255. PairA is at a later stage of the quasar pairing than pairB, and merges in the simulation within ~ 10 Myrs. Both merging timescales are broadly consistent with the estimate of the SDSS J0749+2255 dynamical friction time (Equations [5](#) and [6](#)).

Implications for Gravitational Waves After dynamical friction, SDSS J0749+2255 is expected to undergo hardening via loss-cone scattering of nearby stars, then GW-dominated evolution in the pulsar-timing array (PTA) band. Given that the virial SMBH mass estimates for the two quasars are comparable, dynamical friction should proceed without the binary stalling at $\sim 10\text{--}100$ pc^{[83](#)}. Regardless of the binary mass ratio, tidal interactions between the nuclear star clusters may accelerate the binarys orbital decay^{[84](#)}. Depending on how gas-rich the resulting nucleus is, the sub-pc SMBH binarys orbital evolution may be affected by circumbinary-disk accretion at sub-pc separations. Depending on how long this phase lasts and how massive the gas supply is, the binary may also accrete an appreciable amount of mass, which in turn decreases the GW coalescence timescale^{[85](#)}.

Various PTAs have observed a common-spectrum stochastic process in their data that may be consistent with a stochastic GW background^{[86-88](#)}. Hellings-Downs correlations^{[89](#)}, however, have not yet been observed, which are needed in order to determine that this signal is of GW origin. If the common-process signal that current PTAs are observing is assumed to be due to a stochastic GW background and if the cosmic SMBH binary popu-

lation at nanohertz GW frequencies is assumed to be the sole source of this background, then the amplitude of this signal may be used to constrain the cosmic merger rate of SMBH binaries. This can then be compared to the observed occurrence rate of dual AGN from current and future wide-area deep surveys. The growing number of observations of dual AGNs/quasars across cosmic time as well as PTAs improving their characterization of a potential stochastic GW background will continue to provide vital tests for astrophysical models of SMBH binary formation.

		F475W	F814W	F160W _{PSF}	$K_{p,PSF}$	F160W _{host}	$K_{p,host}$	$\log M_*$	$\log M_{BH}$	$\log M_{BH, vir}$
Nucleus	Redshift	(mag)	(mag)	(mag)	(mag)	(mag)	(mag)	($M_\odot$)	($M_\odot$)	($M_\odot$)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
SDSS J0749+2255 NE	2.1665 ± 0.0010	21.08	20.56	20.61	20.85	21.35	20.75	11.46	9.07	9.12
SDSS J0749+2255 SW	2.1679 ± 0.0003	19.48	19.19	19.34	19.62	21.05	20.51	11.50	9.12	9.20

Supplementary Table 1: Basic photometric and spectroscopic properties of SDSS J0749+2255. Col. 1: NE: northeast nucleus. SW: southwest nucleus; Col. 2: Spectroscopic redshift measured from the best-fit model of the joint optical-NIR spectral fitting (Supplementary Figure 7); Cols. 3 & 4: HST optical dual-band AB magnitude from [34](#); Cols. 5 & 6: HST and Keck NIR PSF AB magnitude; Cols. 7 & 8: HST and Keck NIR model AB magnitude for the extended host galaxies. Col. 9: Total host galaxy stellar mass estimate derived from $K_{p,host}$ and the F160W_{host} - $K_{p,host}$ color (Supplementary Figure 11). Col. 10: Black hole mass estimate inferred from host galaxy total stellar mass assuming the M_{BH} - M_* relation observed in local *HST*-imaged broad-line AGNs with virial BH mass estimates [66](#). Col. 11: Virial black hole mass estimate inferred from broad H α assuming the calibration of [46](#).

Emission-line Measurements	SDSS J0749+2255 NE	SDSS J0749+2255 SW
C III] Flux (10^{-17} erg s $^{-1}$ cm $^{-2}$) (1)	246 $\pm$ 13	736 $\pm$ 16
Mg II Flux (10^{-17} erg s $^{-1}$ cm $^{-2}$) (2)	110 $\pm$ 4	466 $\pm$ 8
H β Flux (10^{-17} erg s $^{-1}$ cm $^{-2}$) (3)	51 $\pm$ 14	233 $\pm$ 27
[O III] λ 5007 Flux (10^{-17} erg s $^{-1}$ cm $^{-2}$) (4)	8 $\pm$ 4	22 $\pm$ 9
Broad H α Flux (10^{-17} erg s $^{-1}$ cm $^{-2}$) . (5)	151 $\pm$ 10	309 $\pm$ 21
Narrow H α Flux (10^{-17} erg s $^{-1}$ cm $^{-2}$) (6)	8 $\pm$ 3	21 $\pm$ 10
C III] EW ($\text{\AA}$) (7)	29 $\pm$ 2	22 $\pm$ 1
Mg II EW ($\text{\AA}$) (8)	28 $\pm$ 1	32 $\pm$ 1
H β EW ($\text{\AA}$) (9)	45 $\pm$ 14	81 $\pm$ 13
[O III] λ 5007 EW ($\text{\AA}$) (10)	7 $\pm$ 4	8 $\pm$ 3
Broad H α EW ($\text{\AA}$) (11)	181 $\pm$ 16	150 $\pm$ 14
Narrow H α EW ($\text{\AA}$) (12)	9 $\pm$ 4	10 $\pm$ 5
C III] FWHM (km s $^{-1}$) (13)	8480 $\pm$ 1200	7900 $\pm$ 290
Mg II FWHM (km s $^{-1}$) (14)	3980 $\pm$ 160	4450 $\pm$ 100
H β FWHM (km s $^{-1}$) (15)	7600 $\pm$ 1800	7400 $\pm$ 1000
[O III] λ 5007 FWHM (km s $^{-1}$) (16)	910 $\pm$ 410	680 $\pm$ 200
Broad H α FWHM (km s $^{-1}$) (17)	5550 $\pm$ 400	4680 $\pm$ 430

Supplementary Table 2: Emission-line measurements. The measurements are from jointly analyzing the Gemini/GMOS and Gemini/GNIRS spectra. Lines 1–6: emission-line flux intensity. For Mg II, only the broad component is listed here. Lines 7–12: rest-frame equivalent width. Lines 13–17: Full width at half maximum. All errors quoted are 1- σ uncertainties estimated from Monte Carlo simulations.

Nucleus	Counts	flux _{0.5–2 keV}	flux _{2–8 keV}	HR	Γ_X	N_H	log $L_{0.5–2 \text{ keV}}$	log $L_{2–8 \text{ keV}}$
		(10^{-6} photons $\text{cm}^{-2} \text{s}^{-1}$)				(10^{22}cm^{-2})	(erg s^{-1})	(erg s^{-1})
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
SDSS J0749+2255 NE	191.5	$12.26^{+3.71}_{-1.06}$	$10.52^{+0.24}_{-0.48}$	$0.37^{+0.01}_{-0.02}$	1.8	$4.7^{+1.3}_{-2.6}$	$44.92^{+0.02}_{-0.03}$	$45.29^{+0.05}_{-0.03}$
SDSS J0749+2255 SW	6	$0.56^{+0.51}_{-0.41}$	$0.23^{+0.21}_{-0.17}$	$-0.34^{+0.48}_{-0.60}$	1.8	$<10^{-2}$	$43.17^{+0.32}_{-0.60}$	$43.62^{+0.32}_{-0.58}$

Supplementary Table 3: X-ray measurements from Chandra ACIS-S observations. Col. 2: Total net counts in 0.5–8 keV band. Cols. 3 & 4: Observed photon flux. Col. 5: Hardness ratio $\text{HR} \equiv (H - S)/(H + S)$ where H and S are the number of counts in the hard and soft X-ray bands, respectively. Col. 6: Photon index value (fixed), assuming a power-law model where $n(E) \propto E^{-\Gamma_X}$. Col. 7: Best-fit intrinsic column density assuming a power-law model. Cols. 8 & 9: Unabsorbed luminosity. All values and errors represent the median value and 99.7% confidence intervals from analyzing 100 spectral realizations of each X-ray point source.

	SDSS J0749+2255 NE	SDSS J0749+2255 SW
C-band (6.0 GHz)		
C (hh:mm:ss.ss+dd:mm:ss.s) (1)	07:49:22.99+22.55.12.0	07:49:22.96+22.55.11.7
S _{Peak} (mJy/beam) (2)	22.91±0.05	60.43±0.05
T _{Peak} (K) (3)	8420±20	22200±20
S _{Int} (mJy) (4)	22.94±0.05	60.51±0.05
log νL _{ν,6.0 GHz} (erg s ⁻¹) (5)	43.382±0.002	43.763±0.001
θ _{maj} ×θ _{min} (″×″) (6)	0′.44×0′.21	
PA (°) (7)	70.9	
Ku-band (15.0 GHz)		
C (hh:mm:ss.ss+dd:mm:ss.s) (8)	07:49:22.99+22.55.12.0	07:49:22.96+22.55.11.7
S _{Peak} (mJy/beam) (9)	15.64±0.08	44.66±0.08
T _{Peak} (K) (10)	7720±40	22050±40
S _{Int} (mJy) (11)	15.95±0.14	45.38±0.14
log νL _{ν,15.0 GHz} (erg s ⁻¹) (12)	43.622±0.004	44.036±0.002
θ _{maj} ×θ _{min} (″×″) (13)	0′.11×0′.10	
PA (°) (14)	−54.0	

Supplementary Table 4: Radio measurements. The measurements are based on 2D Gaussian fits from VLA A-config dual-band imaging. Col. 1: J2000 coordinates for the best-fit centers in 6 GHz. Col. 2: 6.0 GHz peak flux density. Col. 3: 6.0 GHz peak brightness temperature. Col. 4: 6.0 GHz integrated flux density from decomposing the two marginally resolved nuclei. Col. 5: Rest-frame 6.0 GHz luminosity density. Col. 6: Semi-major and semi-minor axis of the clean beam. Col. 7: Position angle of the clean beam. Col. 8-14: Similar to above, but for 15 GHz. As a comparison, the peak brightness temperatures from the archival VLBA X-band (8.4 GHz) image are 1.6×10^8 K for the SW nucleus and 5.2×10^7 K for the NE nucleus. All errors quoted represent 1σ statistical uncertainties.

HST (F160W) GALFIT Parameters	SDSS J0749+2255 NE	SDSS J0749+2255 SW
Model 1 (2 PSFs + 2 Sérsics): $\chi^2_\nu = 1.331$		
m_{PSF} (AB) (1)	20.605±0.002	19.344±0.019
$m_{\text{Sérsic}}$ (AB) (2)	21.349± 0.034	21.053±0.095
R_e (kpc) (3)	4.29±0.07	1.44±0.28
n (4)	0.91±0.03	5.20±1.32
Model 2 (2 PSFs + 1 Sérsic): $\chi^2_\nu = 1.354$		
m_{PSF} (AB) (5)	20.613±0.003	19.260±0.001
C_{PSF} (hh:mm:ss.ss+dd:mm:ss.s) (6)	07:49:23.00+22:55:12.0	07:49:22.97+22:55:11.8
$m_{\text{Sérsic}}$ (AB) (7)	20.734±0.008	
$C_{\text{Sérsic}}$ (hh:mm:ss.ss+dd:mm:ss.s) (8)	07:49:22.98+22:55:11.9	
R_e (kpc) (9)	3.47±0.04	
n (10)	1.38±0.04	
Model 3 (2 PSFs + 4 Sérsics): $\chi^2_\nu = 1.327$		
m_{PSF} (AB) (11)	20.613±0.003	19.373±0.014
$m_{\text{Sérsic},1}$ (AB) (12)	21.476±0.038	21.257±0.057
$m_{\text{Sérsic},2}$ (AB) (13)	23.794±0.876	22.186±0.157
$R_{e,1}$ (kpc) (14)	3.89±0.09	1.94±0.29
$R_{e,2}$ (kpc) (15)	9.94±13.91	0.81±0.14
n_1 (16)	1	4
B/T (17)	0.11±0.08	0.70±0.05

Supplementary Table 5: HST image decomposition. GALFIT structural decomposition results for the *HST*/WFC3 F160W image. Lines 1 & 2: Best-fit component magnitude. Lines 3 & 4: Best-fit Sérsic component effective radius and index. Lines 5–10: Similar to above, but for a single Sérsic for the foreground lens. Lines 6 & 8: J2000 coordinates for the best-fit centers. Lines 11–17: Similar to above, but for two Sérsic models for each host to attempt bulge-disk decomposition. The Sérsic index is fixed at $n = 1$ ($n = 4$) for the disk (bulge) in each host to help break model degeneracy. Line 17: Bulge-to-total luminosity ratio. All error bars are 1σ .

Keck (K_p) GALFIT Parameters	SDSS J0749+2255 NE	SDSS J0749+2255 SW
Model 1 (2 PSFs + 2 Sérsics): $\chi^2_\nu = 1.416$		
m_{PSF} (AB) (1)	20.847 ± 0.006	19.620 ± 0.002
$m_{\text{Sérsic}}$ (AB) (2)	20.745 ± 0.017	20.514 ± 0.008
R_e (kpc) (3)	1.76 ± 0.04	1.07 ± 0.01
n (4)	0.56 ± 0.05	0.16 ± 0.02
Model 2 (2 PSFs + 1 Sérsic): $\chi^2_\nu = 1.521$		
m_{PSF} (AB) (5)	20.706 ± 0.004	19.573 ± 0.002
C_{PSF} (hh:mm:ss.ss+dd:mm:ss.s) (6)	07:49:23.00+22:55:12.0	07:49:22.97+22:55:11.8
$m_{\text{Sérsic}}$ (AB) (7)	19.925 ± 0.011	
$C_{\text{Sérsic}}$ (hh:mm:ss.ss+dd:mm:ss.s) (8)	07:49:22.97+22:55:11.7	
R_e (kpc) (9)	1.79 ± 0.04	
n (10)	1.405 ± 0.04	

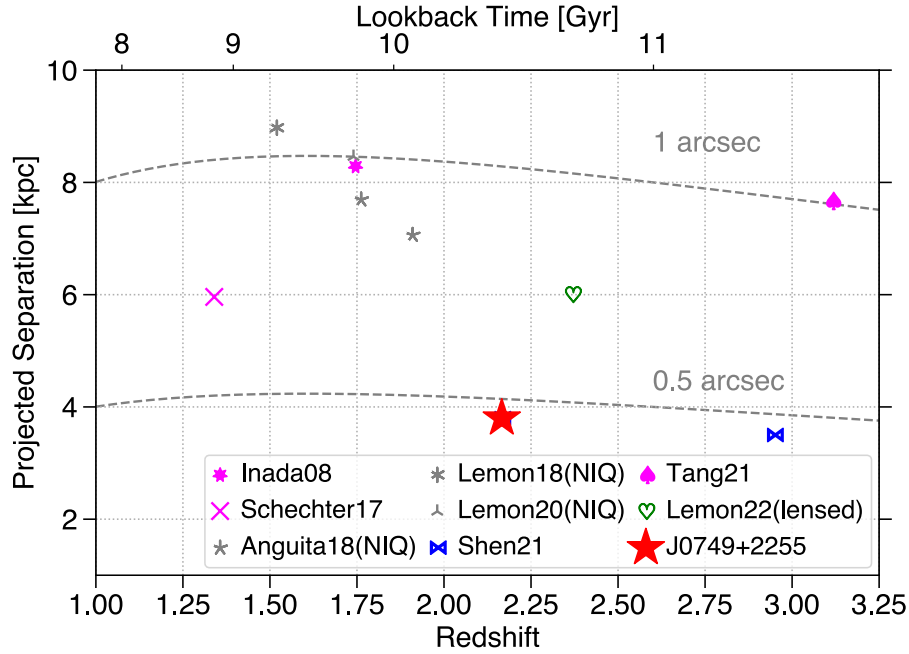
Supplementary Table 6: Keck image decomposition. GALFIT structural decomposition results for the Keck/NIRC2 AO K_p image. Lines 1 & 2: Best-fit component magnitude. Lines 3 & 4: Best-fit Sérsic component effective radius and index. Lines 5–10: Similar to above, but for a single Sérsic model for the foreground lens. Lines 6 & 8: J2000 coordinates for the best-fit centers. The absolute astrometry of the Keck image has been registered to that of the *HST* image according to the PSF centers. All error bars are 1σ .

Identifier	n_{pix}	mag/mag uncertainty	significance level	location
T1	47	25.47 ± 0.07	23	SW to the pair
T2	45	25.92 ± 0.07	15	S to the pair
T3	35	26.22 ± 0.06	13	N to the pair
T4	30	26.50 ± 0.06	11	E to the pair

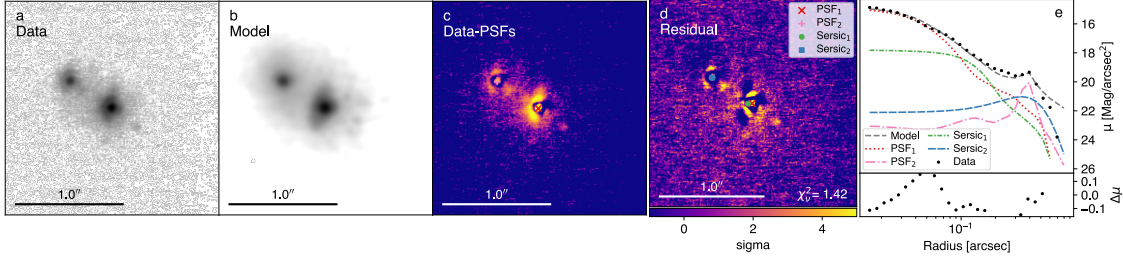
Supplementary Table 7: Tidal features. Identified tidal features from the HST residual map and their flux (AB magnitude) measurements. All error bars are 1σ .

Strong Lensing Mass Model Parameters	SIE (<i>HST</i>)	SIE (Keck)	SIE+shear (Keck)
$\chi^2_{\min}$ (1)	~ 0	39.9	3.2
θ_{Ein} (") (2)	0.24 ± 0.01	0.23 ± 0.01	0.13 ± 0.03
Δ_{RA} (") (3)	0.25 ± 0.03	0.22 ± 0.02	0.07 ± 0.02
Δ_{Dec} (") (4)	0.12 ± 0.03	0.10 ± 0.02	0.04 ± 0.01
e (5)	0.17 ± 0.07	0.06 ± 0.03	0.23 ± 0.14
PA_e ($^\circ$) (6)	37 ± 11	41 ± 8	56 ± 17
shear (7)	N/A	N/A	0.48 ± 0.12
PA_{shear} ($^\circ$) (8)	N/A	N/A	-35 ± 3

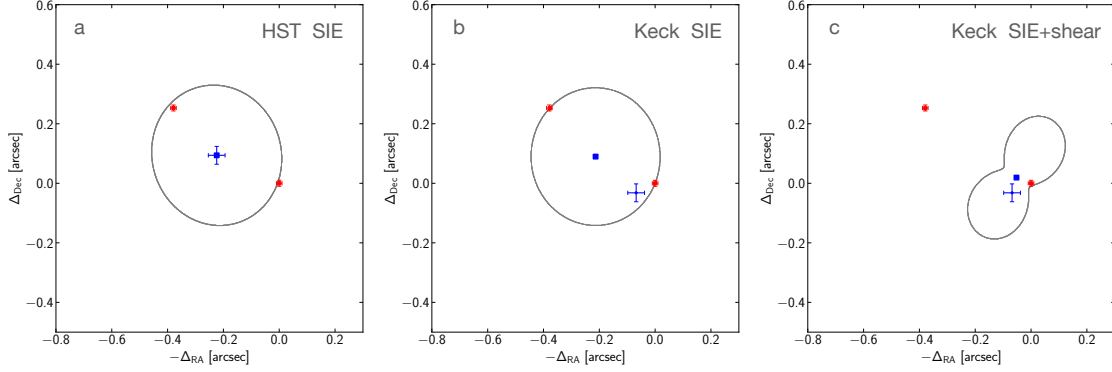
Supplementary Table 8: Lens modeling. Strong lensing mass modeling parameters (mean values and $1\text{-}\sigma$ uncertainties except for the minimum χ^2) obtained using glafic from the MCMC analysis for the *HST* (assuming the SIE model) and Keck (assuming the SIE or SIE+shear model) images. Line 1: Minimum χ^2 given by the best-fit model. Line 2: Einstein radius. Lines 3 & 4: Lens galaxy position with respect to the SW quasar image. Line 5: Ellipticity of the ellipsoid. $e = 0$ for a spherical case. Line 6: Position angle of the ellipsoid measured East of North. Line 7: External shear. Line 8: Position angle of the external shear measured East of North. All errors quoted are 1σ uncertainties estimated from the MCMC analysis.



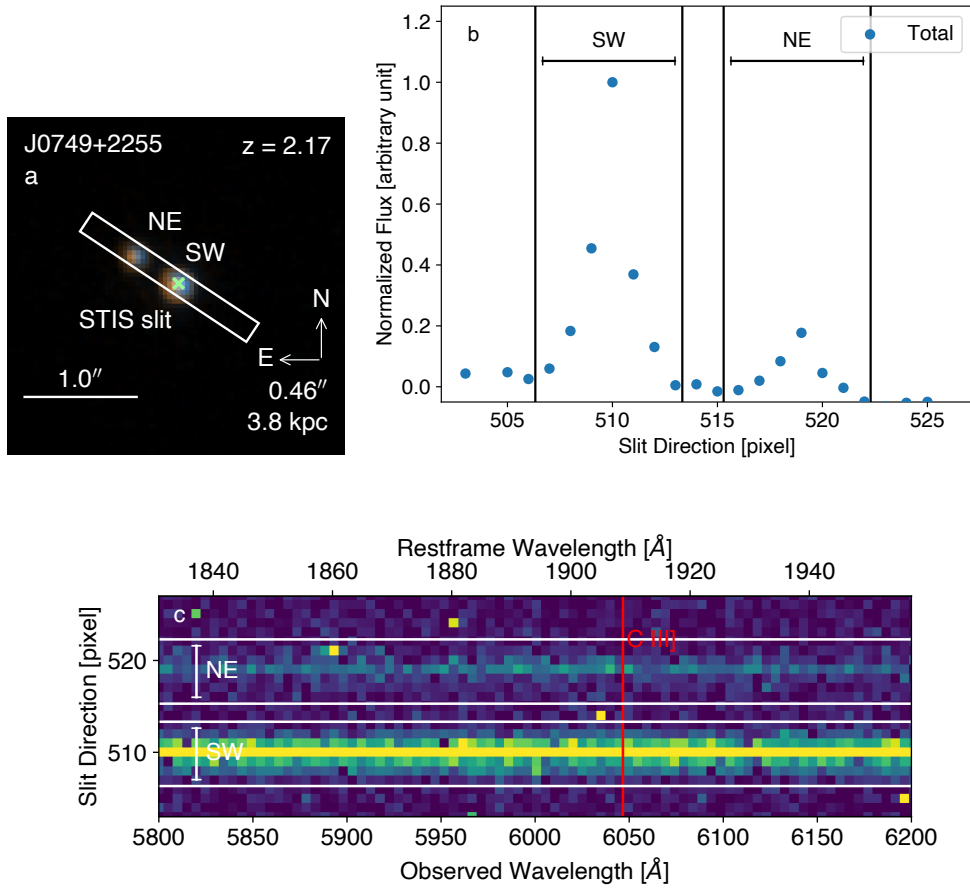
Supplementary Figure 1: Pair separation versus redshift. SDSS J0749+2255 in comparison to the previously known candidate dual quasars in the literature at $z > 1$ and with projected separations < 10 kpc. Nearly identical quasars (NIQs)^[22-24], shown with grey symbols, are marked with "(NIQ)" in the legend. They are generally thought to be lensed quasars lacking a definitive detection of foreground lenses due to observational limitations. Three other dual quasar candidates have more noticeable spectral differences^[20,21,25] (shown in magenta), but they do not show unambiguous host merger features, and are debatable as project pairs at much larger separations or lenses. The dual quasar candidate of^[7] (shown in blue) shows differences in the broad emission line strengths^[35] but is still controversial as a quasar lens due to the lack of high-resolution deep NIR imaging. The quadruply lensed quasar^[27] (shown in green) has been suggested to contain a lensed kpc-scale dual quasar, but the interpretation is subject to lense modeling. There are two more $z > 1$ candidate dual AGNs not shown given their much smaller separations: a strongly lensed binary AGN candidate at $z = 2.51$ in PKS 1830-211 with sub-parsec separation^[90] and a strongly lensed dual AGN candidate^[37,38] at $z = 3.273$ in MG B2016+112 with a separation of hundreds of parsec.



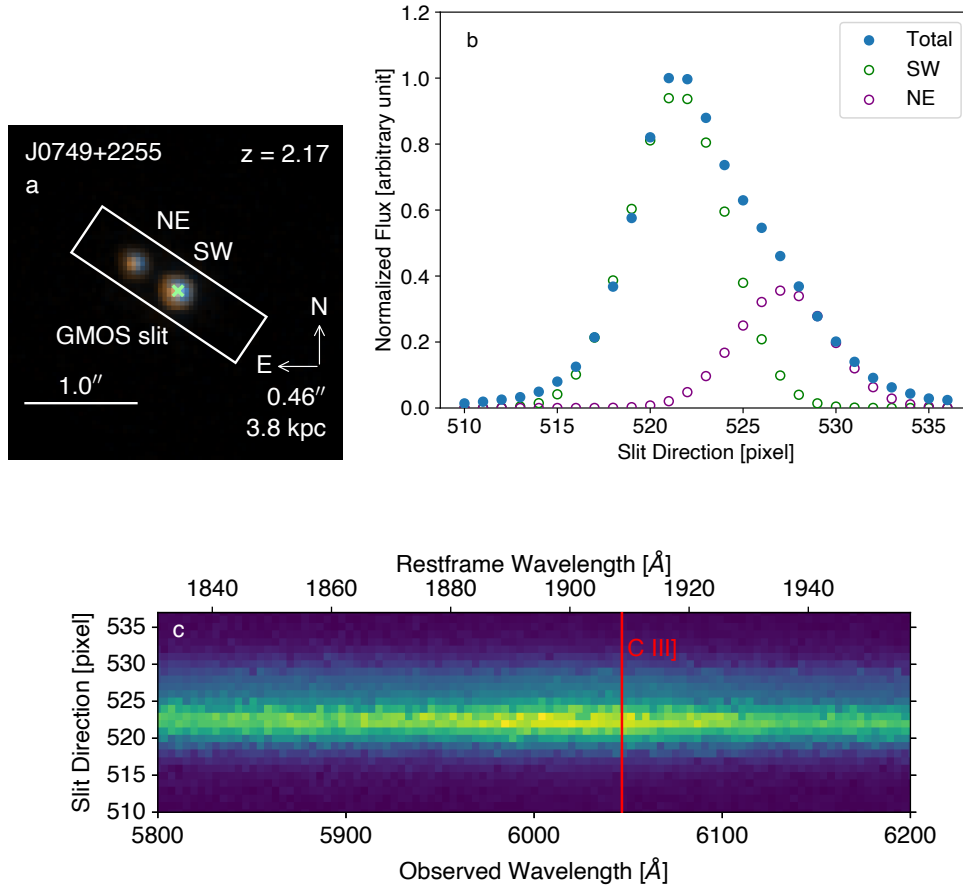
Supplementary Figure 2: Keck imaging. Similar to Figure 1, but for Keck AO-assisted K_p imaging of SDSS J0749+2255 and 2D structural decomposition results from GALFIT analysis. The best-fit GALFIT model contains two PSFs for the two unresolved quasars and two Sérsic models for the two host galaxies with the χ^2_ν value labeled on the residual image. After we subtract the two PSFs, the extended host galaxies are detected (i.e., Data-PSFs). The tidal features detected in the *HST* image (in the GALFIT model residuals) are too faint to detect in the shallower Keck image. The lensing hypothesis is disfavored based on both the χ^2_ν values and lensing mass model tests.



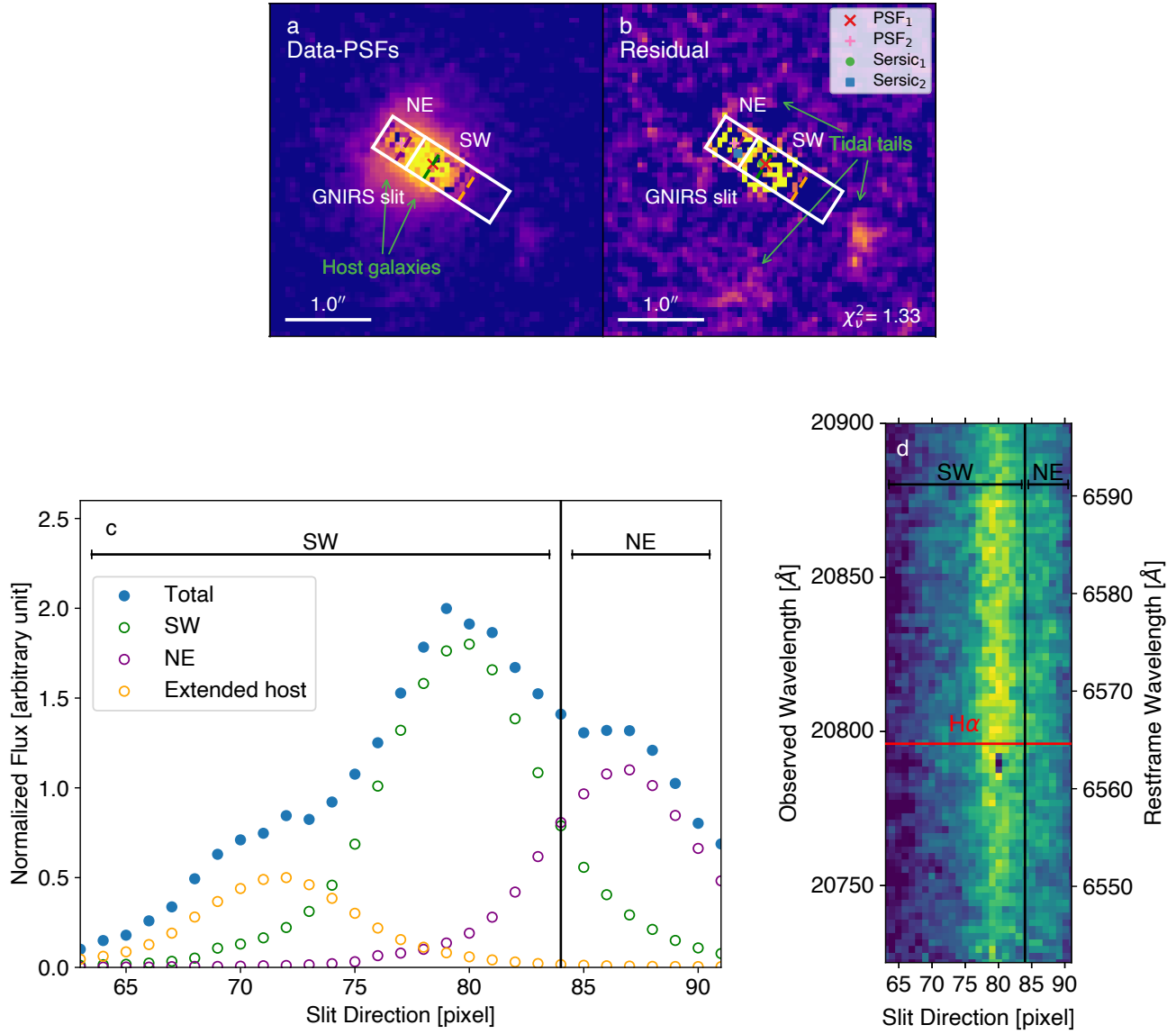
Supplementary Figure 3: Strong Lensing Mass Models. Strong lensing tests from glafic. Panel a and panel b show the strong lensing mass model results using a singular isothermal ellipsoid (SIE) lens based on the HST and Keck images, respectively. Panel c shows a more complicated SIE plus shear model from the Keck image. The solid lines are the critical curves. The filled circles with error bars (1σ) indicate the positions of quasars (red) and the putative lensing galaxy (blue) from HST and Keck images. The filled blue squares are the positions of the lensing galaxy predicted by the best-fitting mass models.



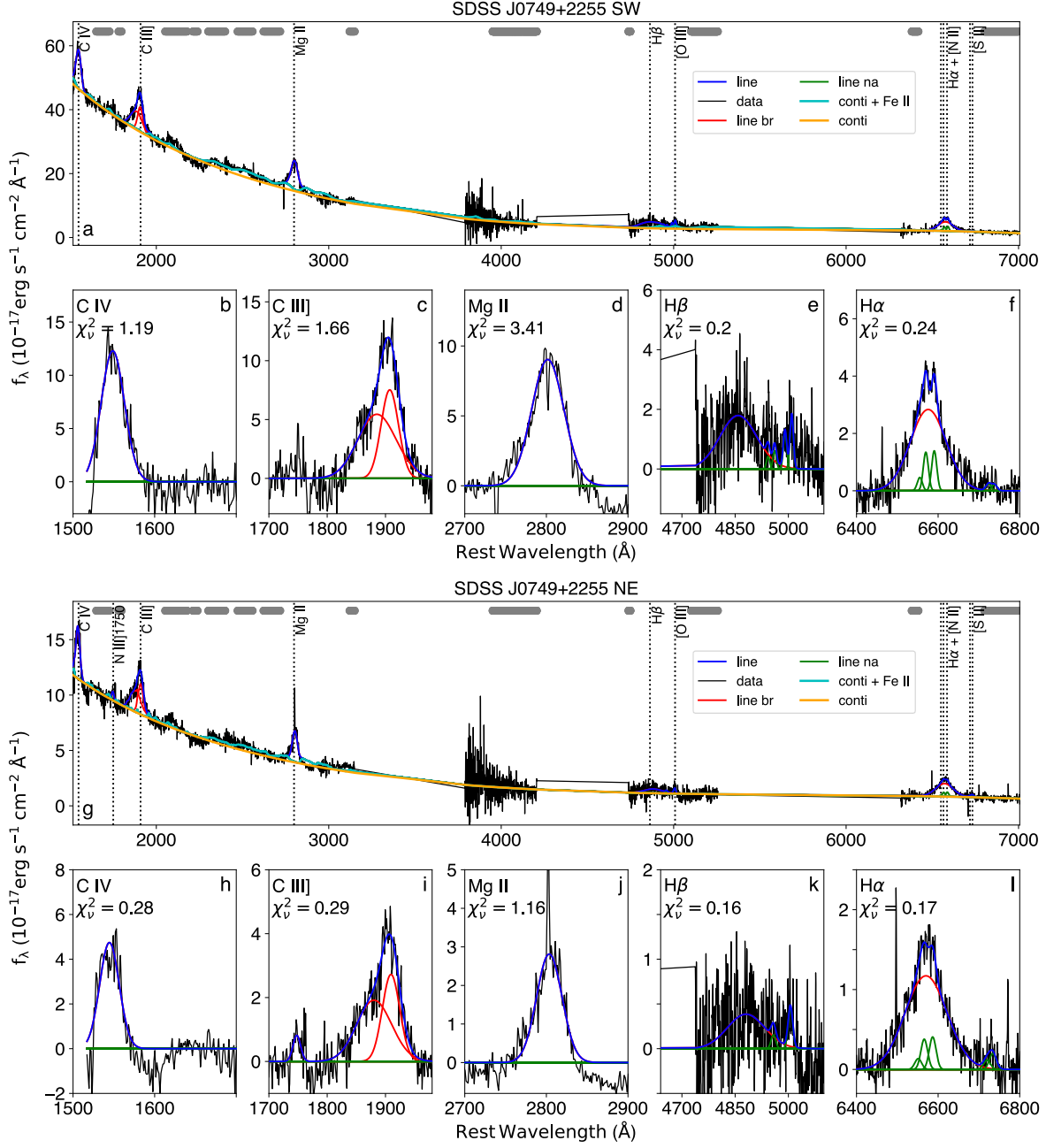
Supplementary Figure 4: HST/STIS spectrum. 2D HST/STIS spectrum for SDSS J0749+2255 and illustration of the slit position and 1D aperture extraction for each nucleus. Panel a: Illustration of the STIS slit position and orientation. The background shows the *HST* optical two-band (F475W and F814W) color image⁷. The width of the white rectangle denotes the slit width. Panel b: Projected 1D spatial profile of the wavelength-averaged STIS spectrum. The blue filled circles denote the total spectrum. The black vertical lines mark the extraction apertures for the two nuclei. Panel c: 2D STIS spectrum around the C III] line. The white horizontal lines mark the extraction apertures for the two nuclei.



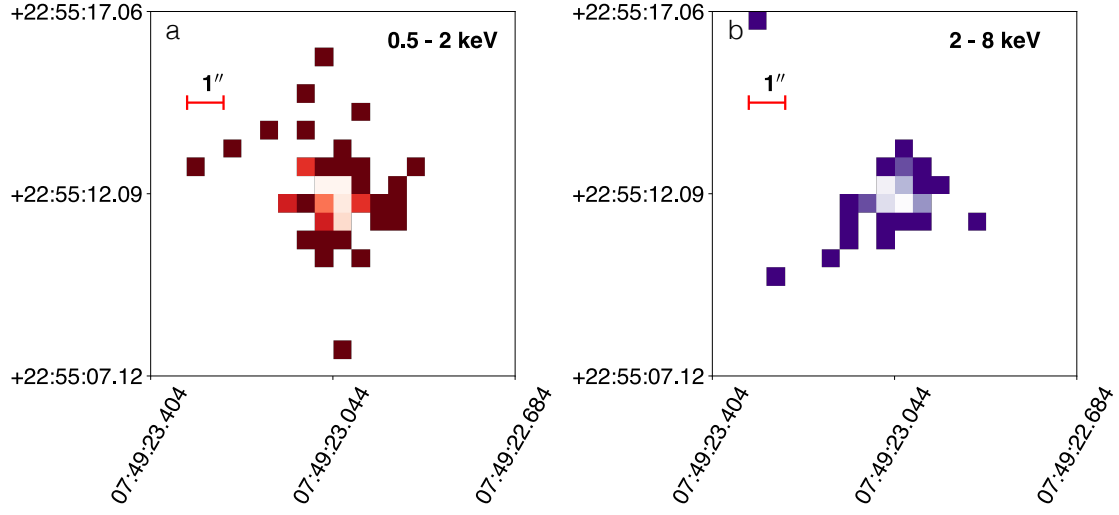
Supplementary Figure 5: Gemini/GMOS spectrum. 2D Gemini/GMOS spectrum for SDSS J0749+2255 and illustration of the slit position and source decomposition for each nucleus. Panel a: Illustration of the GMOS slit position and orientation. The background shows the *HST* optical two-band (F475W and F814W) color image⁷. The width of the white rectangle denotes the slit width. Panel b: Projected 1D spatial profile of the wavelength-averaged GMOS spectrum. The blue filled circles denote the total spectrum. The green (purple) open circles represent the decomposed spatial profile of the SW (NE) nucleus. Panel c: 2D GMOS spectrum around the C III] line.



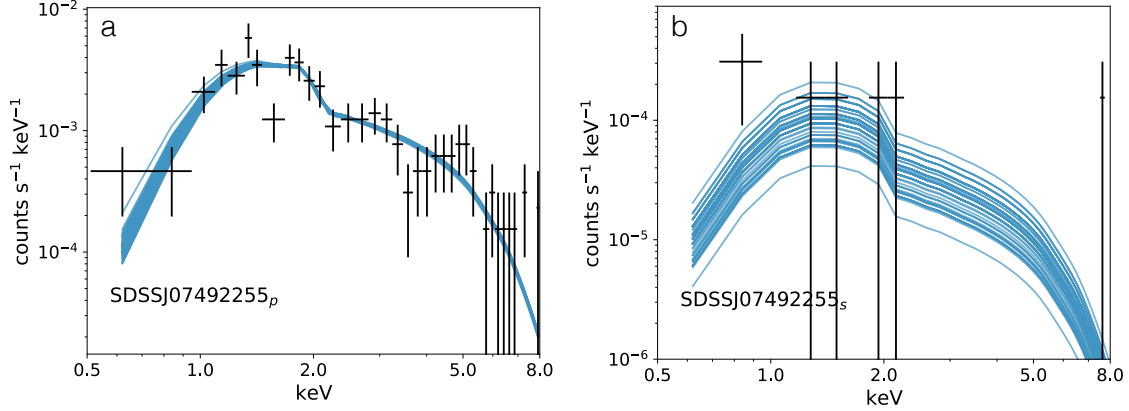
Supplementary Figure 6: Gemini/GNIRS spectrum. 2D Gemini/GNIRS spectrum for SDSS J0749+2255 and illustration of the slit position and source decomposition for each nucleus. Panels a & b: The GNIRS slit position and orientation. The background shows the *HST* F160W PSF-subtracted (Panel a) and residual (Panel b) images. The white rectangles marked with “NE” and “SW” show the extraction apertures of both nuclei. The three dashed lines mark the peak positions of three components seen in the projected 1D profile. Panel c: Projected 1D spatial profile of the wavelength-average GNIRS spectrum. The blue filled circles show the total spectrum. The open circles represent the SW nucleus (shown in green), the NE nucleus (purple), and the extended host (orange). The top horizontal black lines mark the extraction apertures for both nuclei. Panel d: 2D GNIRS spectrum around the H α line. The horizontal black lines denote the extraction apertures for both nuclei.



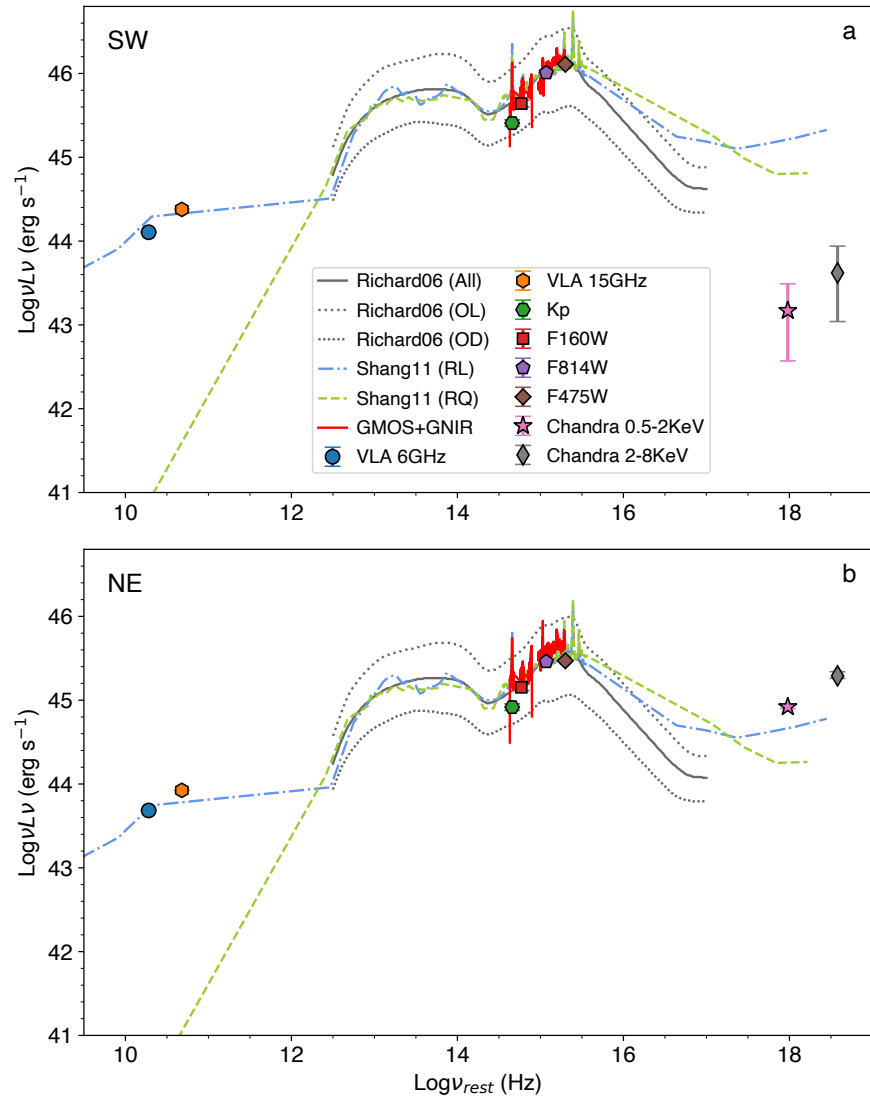
Supplementary Figure 7: Spectral fits. Global and local spectral fitting analysis for the joint optical-NIR spectrum of the two nuclei in SDSS J0749+2255. Panels a–f (Panels g–l) show results for the SW (NE) nuclei. Panels a & g: Global fit of the pseudo-continuum (in cyan), which consists of a power-law plus polynomial (in orange), and a fit with the Fe II template. The horizontal grey bars denote the wavelength windows adopted in the pseudo-continuum fit. Panels b–f, h–l: Local fit for individual emission lines (with the χ^2_ν values marked) including C IV through H α using the pseudo-continuum-subtracted spectra. Red curves show models for the broad emission lines, whereas the green and blue curves denote models for the narrow and total emission lines, respectively.



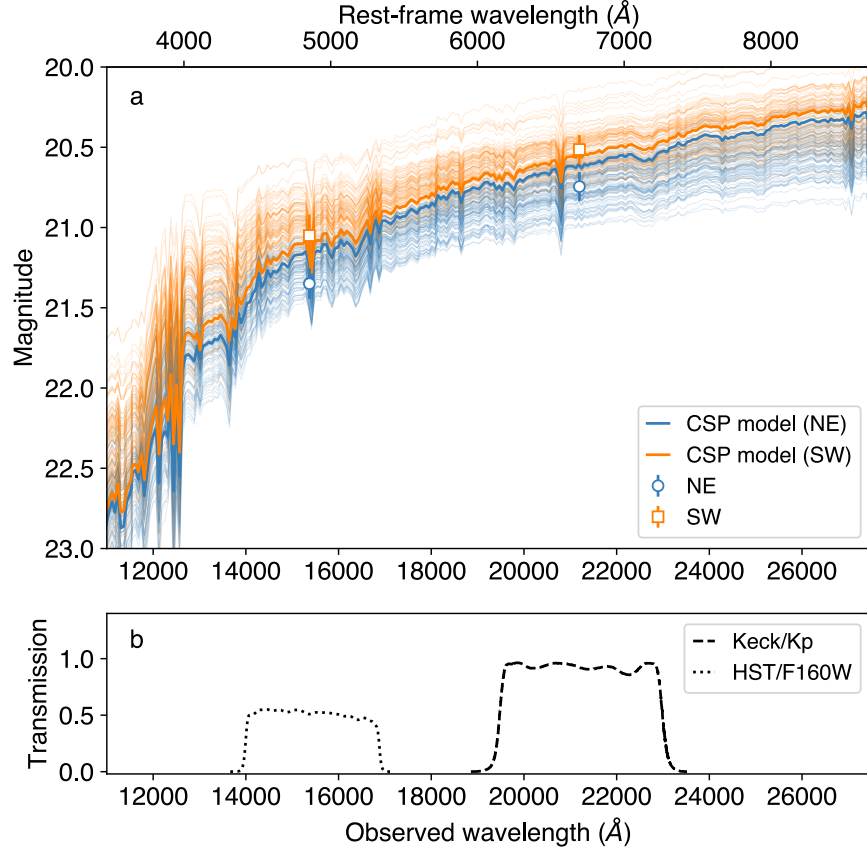
Supplementary Figure 8: X-ray analysis. Binned *Chandra* data, centered on the nominal X-ray coordinates of SDSS J0749+2255. Panel a shows soft photons (0.5–2 keV) (total of 114 counts) and Panel b shows hard photons (2–8 keV) (total of 126 counts). In general, we find that the hardness ratio of the primary point source is consistent with an AGN ($HR = 0.37^{+0.01}_{-0.02}$), while the secondary point source has a softer spectrum (albeit with larger error bars; $HR = -0.34^{+0.48}_{-0.60}$). Error bars here are 3σ .



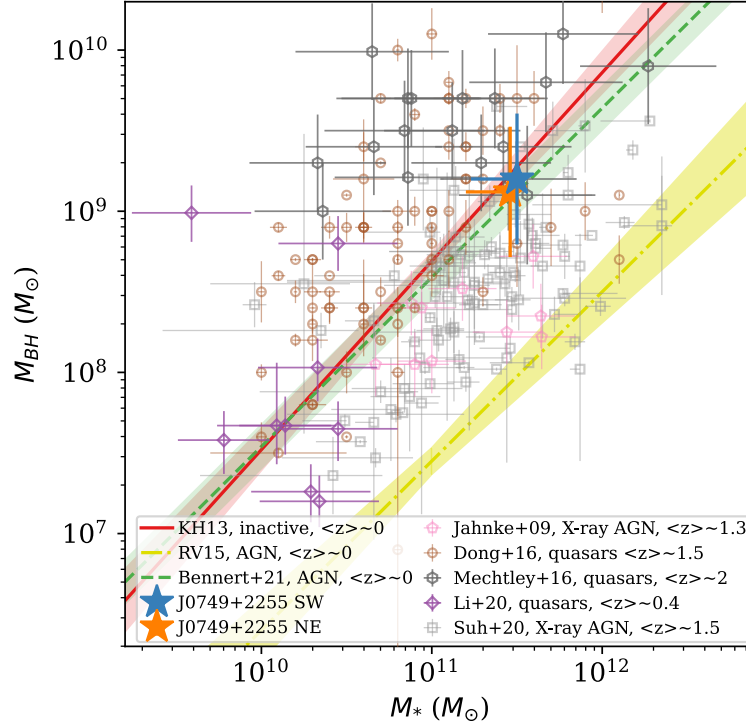
Supplementary Figure 9: X-ray spectral fits. Spectral fits to the primary (Panel a) and secondary (Panel b) X-ray point source detected by *BAYMAX*. We fit and model 100 spectral realizations for each X-ray point source (blue) using *XSPEC* to quantify the average values of the spectral shape and X-ray luminosity. We show one of these spectral realizations in black, for each X-ray point source. We find both sets of spectral realizations are best-fit with a simple absorbed power law (*XSPEC* model components: $\text{phabs} \times \text{zphabs} \times \text{zpow}$). Given the low number of counts for both components (average of ~ 191.5 and ~ 6 counts for the primary and secondary, respectively), we fix the power law index, Γ , to a value of 1.8 for all spectral fits. We list the median values and the confidence interval of the best-fit spectral parameter distributions in Supplementary Table 3. All error bars for the data points in this figure are 1σ .



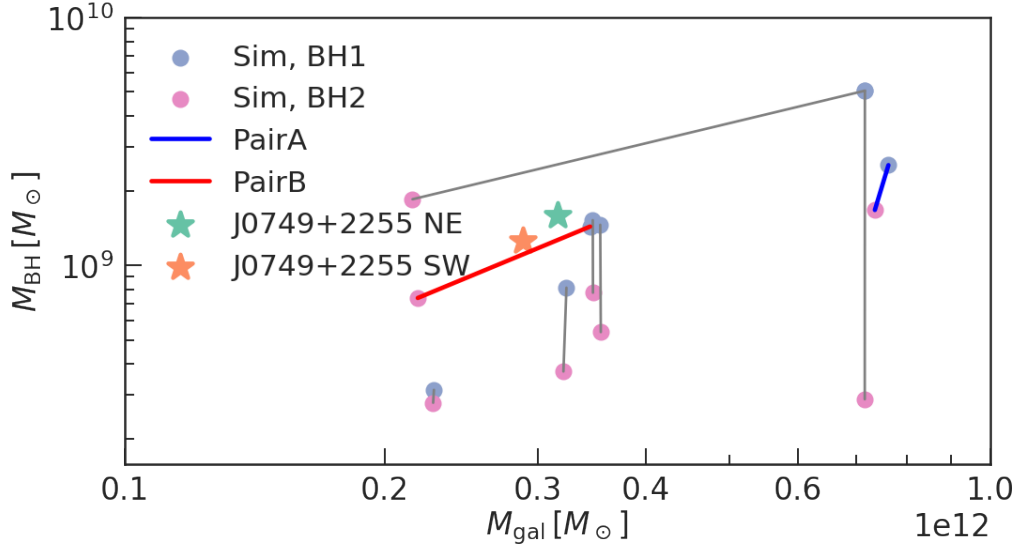
Supplementary Figure 10: Spectral energy distributions. Panel a (Panel b): Multi-wavelength Spectral Energy Distributions (SEDs) of the SW (NE) nuclei in SDSS J0749+2255. For comparison, the SEDs of typical optically selected SDSS quasars³ are shown in grey, with the mean values shown in solid (“All”), and the optically luminous (“OL”) and dim (“OD”) sub-populations shown in dotted curves. Also shown are the SEDs of optically bright, non-blazar quasars from⁵⁹, including both radio-loud (“RL”; shown in dash-dotted curves) and radio-quiet (“RQ”; shown in dashed curves) populations. All comparison SEDs are normalized to match the luminosities of SDSS J0749+2255 in *HST* F814W. All error bars are 1σ .



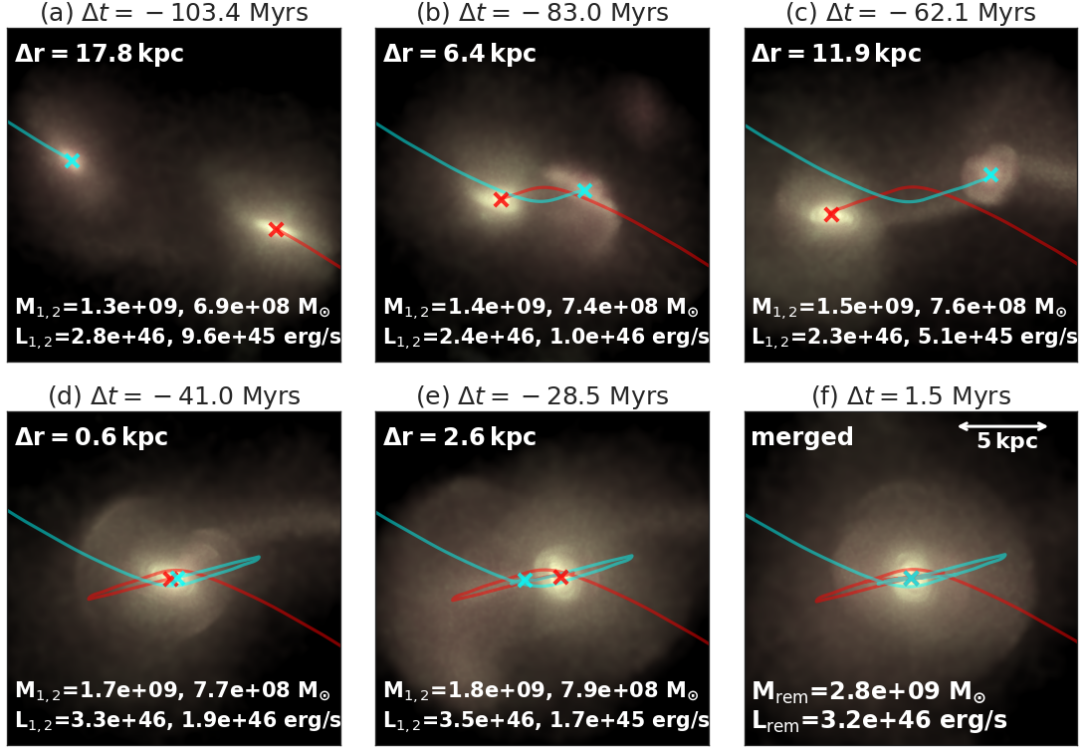
Supplementary Figure 11: Stellar population modeling. Panel a: *HST* F160W and Keck AO K_p PSF-subtracted host-only photometry and the best-fit composite stellar population (CSP) models for the NE (in cyan) and SW (in orange) nuclei of SDSS J0749+2255. The thick curves represent the best-fit CSP models whereas the thin ones are 100 random draws from 500k MCMC realizations (with 1k walkers and 1k steps with 500 burn-ins) for illustration. Panel b: Transmission curves for the *HST* F160W (dotted) and Keck K_p (dashed) filters, which roughly sample the rest-frame g and R bands for SDSS J0749+2255. All error bars are 1σ .



Supplementary Figure 12: Black hole mass versus host-galaxy total stellar mass. The two nuclei in SDSS J0749+2255 are plotted as blue (SW) and orange (NE) stars with the error bars representing 1σ uncertainties. Three local scaling relations are shown with a red solid line^[65], a green dashed line^[66], and a yellow dash-dotted line^[67]. The shaded regions denote the 1σ scatters for these relations. Optical selected quasars at $z > 0.2$ with HST IR images are shown as dark gray hexagons^[69] and purple diamonds^[71]. X-ray selected AGN are shown as light grey squares^[70] and light pink pentagons^[42]. Quasars whose stellar masses are estimated from broad-band SED fitting (i.e., rather than spatially decomposing the PSF and the host) are shown as light brown circles^[68]. All error bars are 1σ .



Supplementary Figure 13: Simulated black hole mass versus host stellar mass relation. $M_{\text{BH}} - M_{\text{gal}}$ relation for the eight simulated bright quasar pairs in *Astrid*. The colored lines mark the two pairs (PairA and PairB) with apparently distinct galaxy bulges, although the kinematic bulge-disk classification adopted by simulations cannot be reliably used in interacting systems whose morphology is typically disturbed and the disks are in the process of being disrupted in the merger process. PairB is the closest match with SDSS J0749+2255 in both the SMBH masses and the host galaxy stellar masses.



Supplementary Figure 14: Simulated dual quasars. A simulated quasar pair (shown in red and cyan crosses) at $z \sim 2$ closely matching the properties of SDSS J0749+2255, shown before and after the host galaxy merger. Δt refers to the time difference between the snapshot and the SMBH merger. Panel b is similar to SDSS J0749+2255 at the time of observation, with slightly lower SMBH and galaxy masses. The two SMBHs merge in the simulation 83 Myrs after the “observation”, and the merger remnant has $M_{\text{BH}} \sim 3 \times 10^9 M_{\odot}$.

32. Shen, Y. *et al.* A Catalog of Quasar Properties from Sloan Digital Sky Survey Data Release 7. *Astrophys. J. Suppl. Ser.* **194**, 45 (2011).
33. Pâris, I. *et al.* The Sloan Digital Sky Survey Quasar Catalog: Fourteenth data release. *Astron. Astrophys.* **613**, A51 (2018).
34. Chen, Y.-C. *et al.* Varstrometry for Off-nucleus and Dual Subkiloparsec AGN (VODKA): Hubble Space Telescope Discovers Double Quasars. *Astrophys. J.* **925**, 162 (2022).
35. Mannucci, F. *et al.* Unveiling the population of dual and lensed active galactic nuclei at sub-arcsec separations. *Nat. Astron.* **6**, 1185–1192 (2022).
36. Barnacka, A. Gravitational lenses as high-resolution telescopes. *Phys. Rep.* **778**, 1–46 (2018).
37. Spingola, C., McKean, J. P., Massari, D. & Koopmans, L. V. E. Proper motion in lensed radio jets at redshift 3: A possible dual super-massive black hole system in the early Universe. *Astron. Astrophys.* **630**, A108 (2019).
38. Schwartz, D., Spingola, C. & Barnacka, A. Resolving Complex Inner X-Ray Structure of the Gravitationally Lensed AGN MG B2016+112. *Astrophys. J.* **917**, 26 (2021).

39. Dressel, L., 2022. Wide Field Camera 3 Instrument Handbook, Version 14.0 (Baltimore: STScI).
40. Sahu, K.C., et al., 2021, WFC3 Data Handbook, Version 5.0, (Baltimore: STScI).
41. Mihos, J. C. & Hernquist, L. Gasdynamics and Starbursts in Major Mergers. *Astrophys. J.* **464**, 641–663 (1996).
42. Jahnke, K. *et al.* Massive Galaxies in COSMOS: Evolution of Black Hole Versus Bulge Mass but not Versus Total Stellar Mass Over the Last 9 Gyr? *Astrophys. J. Lett.* **706**, L215–L220 (2009).
43. Rusin, D. *et al.* The Evolution of a Mass-selected Sample of Early-Type Field Galaxies. *Astrophys. J.* **587**, 143–159 (2003).
44. Oguri, M. The Mass Distribution of SDSS J1004+4112 Revisited. *Publ. Astron. Soc. Jpn.* **62**, 1017–1024 (2010).
45. Anderson, J. & Bedin, L. R. An Empirical Pixel-Based Correction for Imperfect CTE. I. HST’s Advanced Camera for Surveys. *Publ. Astron. Soc. Jpn.* **122**, 1035–1064 (2010).
46. Shen, Y. & Liu, X. Comparing Single-epoch Virial Black Hole Mass Estimators for Luminous Quasars. *Astrophys. J.* **753**, 125 (2012).

47. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).
48. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
49. Shen, Y. *et al.* The Sloan Digital Sky Survey Reverberation Mapping Project: Velocity Shifts of Quasar Emission Lines. *Astrophys. J.* **831**, 7 (2016).
50. Fruscione, A. *et al.* CIAO: Chandra’s data analysis system. In Silva, D. R. & Doxsey, R. E. (eds.) *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 6270 (2006).
51. Foord, A. *et al.* A Bayesian Analysis of SDSS J0914+0853, a Low-mass Dual AGN Candidate. *Astrophys. J.* **877**, 17 (2019).
52. Foord, A. *et al.* A Second Look at 12 Candidate Dual AGNs Using BAYMAX. *Astrophys. J.* **892**, 29 (2020).
53. Foord, A., Gültekin, K., Runnoe, J. C. & Koss, M. J. AGN Triality of Triple Mergers: Detection of Faint X-Ray Point Sources. *Astrophys. J.* **907**, 71 (2021).

54. Arnaud, K. A. XSPEC: The First Ten Years. In Jacoby, G. H. & Barnes, J. (eds.) *Astronomical Data Analysis Software and Systems V*, vol. 101, 17 (1996).
55. Green, P. J. *et al.* A Full Year's Chandra Exposure on Sloan Digital Sky Survey Quasars from the Chandra Multiwavelength Project. *Astrophys. J.* **690**, 644–669 (2009).
56. McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. CASA Architecture and Applications. In Shaw, R. A., Hill, F. & Bell, D. J. (eds.) *Astronomical Data Analysis Software and Systems XVI*, vol. 376, 127 (2007).
57. Shen, Y. The mass of quasars. *Bull. Astron. Soc. India* **41**, 61–115 (2013).
58. Runnoe, J. C., Brotherton, M. S. & Shang, Z. Updating quasar bolometric luminosity corrections. *Mon. Not. R. Astron. Soc.* **422**, 478–493 (2012).
59. Shang, Z. *et al.* The Next Generation Atlas of Quasar Spectral Energy Distributions from Radio to X-Rays. *Astrophys. J. Suppl. Ser.* **196**, 2 (2011).
60. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
61. Chabrier, G. Galactic Stellar and Substellar Initial Mass Function. *Publ. Astron. Soc. Jpn.* **115**, 763–795 (2003).

62. Fagotto, F., Bressan, A., Bertelli, G. & Chiosi, C. Evolutionary sequences of stellar models with new radiative opacities. IV. $Z=0.004$ and $Z=0.008$. *Astron. Astrophys.* **105**, 29–38 (1994).
63. Bressan, A., Fagotto, F., Bertelli, G. & Chiosi, C. Evolutionary Sequences of Stellar Models with New Radiative Opacities. II. $Z = 0.02$. *Astron. Astrophys. Suppl. Ser.* **100**, 647 (1993).
64. Conroy, C. Modeling the Panchromatic Spectral Energy Distributions of Galaxies. *Annu. Rev. Astro. Astrophys.* **51**, 393–455 (2013).
65. Kormendy, J. & Ho, L. C. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. *Annu. Rev. Astro. Astrophys.* **51**, 511–653 (2013).
66. Bennert, V. N. *et al.* A Local Baseline of the Black Hole Mass Scaling Relations for Active Galaxies. IV. Correlations Between M_{BH} and Host Galaxy σ , Stellar Mass, and Luminosity. *Astrophys. J.* **921**, 36 (2021).
67. Reines, A. E. & Volonteri, M. Relations between Central Black Hole Mass and Total Galaxy Stellar Mass in the Local Universe. *Astrophys. J.* **813**, 82 (2015).
68. Dong, X. Y. & Wu, X.-B. Herschel Observed Stripe 82 Quasars and Their Host Galaxies: Connections between AGN Activity and host Galaxy Star Formation. *Astrophys. J.* **824**, 70 (2016).

69. Mechtley, M. *et al.* Do the Most Massive Black Holes at $z = 2$ Grow via Major Mergers? *Astrophys. J. Lett.* **830**, 156 (2016).
70. Suh, H. *et al.* No Significant Evolution of Relations between Black Hole Mass and Galaxy Total Stellar Mass Up to $z \sim 2.5$. *Astrophys. J.* **889**, 32 (2020).
71. Li, J. I. H. *et al.* The Sloan Digital Sky Survey Reverberation Mapping Project: The M_{BH} -Host Relations at $0.2 \lesssim z \lesssim 0.6$ from Reverberation Mapping and Hubble Space Telescope Imaging. *Astrophys. J.* **906**, 103 (2021).
72. Di Matteo, T., Springel, V. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* **433**, 604–607 (2005).
73. Di Matteo, T., Colberg, J., Springel, V., Hernquist, L. & Sijacki, D. Direct Cosmological Simulations of the Growth of Black Holes and Galaxies. *Astrophys. J.* **676**, 33–53 (2008).
74. Chandrasekhar, S. Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. *Astrophys. J.* **97**, 255–264 (1943).
75. Binney, J. & Tremaine, S. *Galactic dynamics* (1987).

76. Dubinski, J., Mihos, J. C. & Hernquist, L. Constraining Dark Halo Potentials with Tidal Tails. *Astrophys. J.* **526**, 607–622 (1999).
77. Renaud, F., Appleton, P. N. & Xu, C. K. N-body Simulation of the Stephan’s Quintet. *Astrophys. J.* **724**, 80–91 (2010).
78. Silverman, J. D. *et al.* Dual Supermassive Black Holes at Close Separation Revealed by the Hyper Suprime-Cam Subaru Strategic Program. *Astrophys. J.* **899**, 154 (2020).
79. Shen, Y. *et al.* Statistics of Galactic-Scale Quasar Pairs at Cosmic Noon. *arXiv e-prints* arXiv:2208.04979 (2022).
80. Bird, S. *et al.* The ASTRID simulation: galaxy formation and reionization. *Mon. Not. R. Astron. Soc.* **512**, 3703–3716 (2022).
81. Ni, Y. *et al.* The ASTRID simulation: the evolution of supermassive black holes. *Mon. Not. R. Astron. Soc.* **513**, 670–692 (2022).
82. Chen, N. *et al.* Massive black hole mergers with orbital information: predictions from the ASTRID simulation. *Mon. Not. R. Astron. Soc.* (2022).
83. Dosopoulou, F. & Antonini, F. Dynamical Friction and the Evolution of Supermassive Black Hole Binaries: The Final Hundred-parsec Problem. *Astrophys. J.* **840**, 31 (2017).

84. Ogiya, G., Hahn, O., Mingarelli, C. M. F. & Volonteri, M. Accelerated orbital decay of supermassive black hole binaries in merging nuclear star clusters. *Mon. Not. R. Astron. Soc.* **493**, 3676–3689 (2020).
85. Siwek, M. S., Kelley, L. Z. & Hernquist, L. The effect of differential accretion on the gravitational wave background and the present-day MBH binary population. *Mon. Not. R. Astron. Soc.* **498**, 537–547 (2020).
86. Arzoumanian, Z. *et al.* The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. *Astrophys. J. Lett.* **905**, L34 (2020).
87. Chen, S. *et al.* Common-red-signal analysis with 24-yr high-precision timing of the European Pulsar Timing Array: inferences in the stochastic gravitational-wave background search. *Mon. Not. R. Astron. Soc.* **508**, 4970–4993 (2021).
88. Goncharov, B. *et al.* On the Evidence for a Common-spectrum Process in the Search for the Nanohertz Gravitational-wave Background with the Parkes Pulsar Timing Array. *Astrophys. J. Lett.* **917**, L19 (2021).
89. Hellings, R. W. & Downs, G. S. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *Astrophys. J. Lett.* **265**, L39–L42 (1983).
90. Nair, S., Jin, C. & Garrett, M. A. Helical jet in the gravitationally lensed blazar PKS1830-211. *Mon. Not. R. Astron. Soc.* **362**, 1157–1166 (2005).