DETECTION OF LENSING SUBSTRUCTURE USING ALMA OBSERVATIONS OF THE DUSTY GALAXY SDP.81

Item Type	Article
Authors	Hezaveh, Y.; Dalal, Neal; Marrone, Daniel P.; Mao, Yao-Yuan; Morningstar, Warren; Wen, Di; Blandford, Roger D.; Carlstrom, John E.; Fassnacht, Christopher D.; Holder, Gilbert P.; Kemball, Athol; Marshall, Philip J.; Murray, Norman; Levasseur, Laurence Perreault; Vieira, J. D.; Wechsler, Risa H.
Citation	DETECTION OF LENSING SUBSTRUCTURE USING ALMA OBSERVATIONS OF THE DUSTY GALAXY SDP.81 2016, 823 (1):37 The Astrophysical Journal
DOI	10.3847/0004-637X/823/1/37
Publisher	IOP PUBLISHING LTD
Journal	The Astrophysical Journal
Rights	© 2016. The American Astronomical Society. All rights reserved.
Download date	24/08/2022 09:30:39
Item License	http://rightsstatements.org/vocab/InC/1.0/
Version	Final published version
Link to Item	http://hdl.handle.net/10150/621261

DETECTION OF LENSING SUBSTRUCTURE USING ALMA OBSERVATIONS OF THE DUSTY GALAXY SDP.81

YASHAR D. HEZAVEH[1,12], NEAL DALAL[2,3,4,5], DANIEL P. MARRONE[6], YAO-YUAN MAO[1,7], WARREN MORNINGSTAR[1], DI WEN[2], ROGER D. BLANDFORD[1,7], JOHN E. CARLSTROM[8], CHRISTOPHER D. FASSNACHT[9], GILBERT P. HOLDER[10], ATHOL KEMBALL[2], PHILIP J. MARSHALL[7], NORMAN MURRAY[11,13], LAURENCE PERREAULT LEVASSEUR[1], JOAQUIN D. VIEIRA[2], AND RISA H. WECHSLER[1,7]

[1] Kavli Institute for Particle Astrophysics and Cosmology and Department of Physics, Stanford University, 452 Lomita Mall, Stanford, CA 94305-4085, USA
[2] Astronomy Department, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana IL 61801, USA
[3] School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, NJ 08540, USA
[4] Kavli Institute for the Physics and Mathematics of the universe, TODIAS, The University of Tokyo, Chiba, 277-8583, Japan
[5] Department of Chemistry and Physics, University of Kwa-Zulu Natal, University Road, Westville, KZN, South Africa
[6] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA
[7] Kavli Institute for Particle Astrophysics and Cosmology and Department of Particle Physics and Astrophysics; SLAC National Accelerator Laboratory, Menlo Park, CA 94305, USA
[8] Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[9] Department of Physics, University of California, One Shields Avenue, Davis, CA 95616, USA
[10] Department of Physics, McGill University, 3600 Rue University, Montreal, Quebec H3A 2T8, Canada
[11] CITA, University of Toronto, 60 St. George St., Toronto ON M5S 3H8, Canada
Received 2016 January 19; accepted 2016 March 7; published 2016 May 19

ABSTRACT

We study the abundance of substructure in the matter density near galaxies using ALMA Science Verification observations of the strong lensing system SDP.81. We present a method to measure the abundance of subhalos around galaxies using interferometric observations of gravitational lenses. Using simulated ALMA observations we explore the effects of various systematics, including antenna phase errors and source priors, and show how such errors may be measured or marginalized. We apply our formalism to ALMA observations of SDP.81. We find evidence for the presence of a $M = 10^{8.96 \pm 0.12} M_\odot$ subhalo near one of the images, with a significance of 6.9σ in a joint fit to data from bands 6 and 7; the effect of the subhalo is also detected in both bands individually. We also derive constraints on the abundance of dark matter (DM) subhalos down to $M \sim 2 \times 10^7 M_\odot$, pushing down to the mass regime of the smallest detected satellites in the Local Group, where there are significant discrepancies between the observed population of luminous galaxies and predicted DM subhalos. We find hints of additional substructure, warranting further study using the full SDP.81 data set (including, for example, the spectroscopic imaging of the lensed carbon monoxide emission). We compare the results of this search to the predictions of ΛCDM halos, and find that given current uncertainties in the host halo properties of SDP.81, our measurements of substructure are consistent with theoretical expectations. Observations of larger samples of gravitational lenses with ALMA should be able to improve the constraints on the abundance of galactic substructure.

Key words: dark matter – gravitational lensing: strong

1. INTRODUCTION

Over the past few decades, multiple probes of the large-scale structure of the universe have established a clear picture of our cosmology that is described well by a simple model, the ΛCDM model (e.g., Dodelson 2003). This model predicts that structures including galaxies arise through a process of hierarchical structure formation, which is driven by the gravitational instability of cold dark matter (CDM) density perturbations seeded by quantum fluctuations generated during an early epoch of exponential expansion called inflation. On large scales ($\gtrsim$ megaparsecs) and across all epochs, the predictions of this model agree remarkably well with a wide variety of observations, including cosmic microwave background anisotropies (Planck Collaboration et al. 2015), galaxy clustering (Anderson et al. 2014), and weak gravitational lensing (Heymans et al. 2012). On smaller, subgalactic scales where structure has become highly nonlinear (e.g., $\lesssim 10$ kpc) at low redshift, the agreement between CDM predictions and observations of galactic structure has been less clear.

One of the most famous examples of CDM's small-scale difficulties is the missing satellite problem (e.g., Kravtsov 2010, and references therein). Numerical simulations of the formation of galactic dark matter (DM) halos robustly predict the presence of a large population of bound subhalos orbiting within all virialized objects such as the Milky Way. These subhalos span a broad spectrum of masses from objects like the Magellanic Clouds and extending down to the resolution limits of the simulations $M \sim 10^3 M_\odot$ (Diemand et al. 2008; Stadel et al. 2009; Navarro et al. 2010). Observations of dwarf galaxies in the Local Group find an almost factor of 10 deficit of satellites compared to ΛCDM predictions, at masses corresponding to $V_{\rm circ} \sim 10$ km s^{-1} (Kravtsov 2010).

Two main classes of solutions have been proposed to resolve this discrepancy. The first solution involves modifying the microphysics of DM particles in a manner that would suppress the abundance of low-mass substructure while remaining consistent with the existing bounds on structure at $k \lesssim 10h$ Mpc^{-1} from the Lyman α forest (Seljak et al. 2006; Marković & Viel 2014). Examples of proposed modifications to CDM include warm dark matter (WDM), in which the thermal streaming motions of DM particles wipe out small-scale structure (Bode et al. 2001; Abazajian 2006), or self-

---

[12] Hubble Fellow.
[13] Canada Research Chair in Astrophysics.

interacting DM, in which DM particles can have non-gravitational interactions (Spergel & Steinhardt 2000). In principle, fluid descriptions of DM could also suppress satellites (Khoury 2015).

Alternatively, the solution to the missing satellite problem may lie within baryonic astrophysics. Numerous processes have been suggested for suppressing the efficiency of star formation within low-mass halos and subhalos, such as photoevaporation during reionization or feedback from massive stars, supernovae, or black holes (see Kravtsov 2010, for a review). In this scenario, a large number of DM subhalos should exist around most galaxies; however, most of them would be devoid of stars and therefore difficult to detect using conventional means.

Gravitational lensing has long been suggested as a method of distinguishing between these possibilities (various properties of tidal streams and their disruption have also been suggested as a method to detect subhalos in the Milky Way, e.g., Siegal-Gaskins & Valluri 2008; Carlberg 2009; Erkal & Belokurov 2015a, 2015b). Strong gravitational lensing is sensitive to galactic substructure even if that substructure is completely devoid of baryons (e.g., Mao & Schneider 1998; Metcalf & Madau 2001; Dalal & Kochanek 2002; Moustakas & Metcalf 2003; Kochanek & Dalal 2004; Koopmans 2005; Keeton et al. 2006; Vegetti & Koopmans 2009; Vegetti et al. 2012; Nierenberg et al. 2014; Cyr-Racine et al. 2015; Xu et al. 2016). Dalal & Kochanek (2002) analyzed a sample of strongly lensed quasars and found that the lenses must have a large abundance of substructure. They found that the fraction of projected mass within ~10 kpc of lens galaxies in the form of substructure was $0.3\% < f_{\mathrm{sub}} < 3.5\%$, at 90% confidence. Vegetti et al. (2014) analyzed a sample of strongly lensed galaxies observed using the *Hubble Space Telescope* (*HST*) and Keck AO data and found $0.2\% < f_{\mathrm{sub}} < 2.8\%$, at 68% confidence, with two detections reported in Vegetti et al. (2010, 2012). These measurements of projected substructure are in broad agreement with expectations for the massive elliptical galaxies that are typical of strong lenses (see Section 6), but the large uncertainties do not yet allow for a discriminating test of DM models.

Recently, Hezaveh et al. (2013a) proposed that ALMA observations of the large population of strongly lensed dusty galaxies discovered in sub-mm surveys like the South Pole Telescope (Hezaveh et al. 2013b; Vieira et al. 2013), the Atacama Cosmology Telescope (Marsden et al. 2014) or *Herschel* (Negrello et al. 2010; Bussmann et al. 2013; Wardlow et al. 2013) could significantly improve constraints on DM substructure. Toward that end, in this paper we present a method for detecting DM subhalos using interferometric data, showing a few examples of the performance of the pipeline on mock data. Finally, we apply our formalism to the recent ALMA Science Verification observations of the strongly lensed dusty galaxy SDP.81 (ALMA Partnership et al. 2015b).

The outline of the paper is as follows. In Section 2 we present a formalism for modeling extended lensed sources using interferometric observations. In Section 3 we present a perturbative method for searching for subhalos using interferometric observations. Section 4 describes our implementation of this method and presents the results of various tests and analysis of mock data used to test the pipeline. In Section 5 we analyze ALMA observations of SDP.81 using this pipeline, quantifying the abundance of subhalos in the lensing galaxy. In

Table 1
Best-fit Lens Parameters with 68% Uncertainties

Parameter	Definition	Value
α	radial slope	1.06 ± 0.03
$\log_{10}(M_{10\,\mathrm{kpc}})$	mass within 10 kpc ($M_\odot$)	11.60 ± 0.006
ϵ_x	ellipticity x	0.371 ± 0.019
ϵ_y	ellipticity y	-0.046 ± 0.008
x_{lens}	lens x ($''$)	0.481 ± 0.006
y_{lens}	lens y ($''$)	0.154 ± 0.005
γ_1	external shear	0.0004 ± 0.006
γ_2	external shear	0.0017 ± 0.006
A_3	$m = 3$ multipole	$[5.90 \pm 6.26] \times 10^{-3}$
B_3	$m = 3$ multipole	$[25.44 \pm 6.00] \times 10^{-3}$
A_4	$m = 4$ multipole	$[12.53 \pm 10.10] \times 10^{-3}$
B_4	$m = 4$ multipole	$[6.52 \pm 11.20] \times 10^{-3}$
$\log_{10}(M_{\mathrm{sub}})$	subhalo mass ($M_\odot$)	8.96 ± 0.12
x_{sub}	subhalo position x ($''$)	-0.694 ± 0.025
y_{sub}	subhalo position y ($''$)	-0.749 ± 0.044

Note. Best-fit parameter values from a joint fit to bands 6 and 7 data. Positions are in arcseconds relative to the ALMA phase center.

Section 6 we compare the constraints obtained in the previous section with the predictions of the ΛCDM model, and finally discuss the results in Section 7. For all of the modeling presented in this paper, we assume a flat ΛCDM expansion history with mean matter density $\Omega_M = 0.267$ and current expansion rate $H_0 = 71\ \mathrm{km\ Mpc^{-1}\ s^{-1}}$.

2. FRAMEWORK: FITTING VISIBILITIES WITH A PIXELATED SOURCE

Data from interferometric observations consist of a large number of complex visibilities. We fit these data using a model for the distribution of matter in the lensing galaxy, the background source emission, and certain aspects of the measurements such as time-varying antenna phase errors. To describe the source emission, we use a pixelated source map containing many pixels for each observed channel. We can think of the source map pixels as parameters in our lens model, along with parameters describing the lens mass distribution and other nuisance parameters like antenna phase errors. Below, we will use the notation η to denote lens mass parameters (see Table 1), S to denote source pixel parameters, ϕ to denote other parameters like phase errors, and p to denote the full set of parameters, i.e., $p = \{\eta, S, \phi\}$.

The general framework of fitting strongly lensed images with pixelated sources is described in detail in Warren & Dye (2003) and Suyu et al. (2006). In general, given data vector D, model predictions $M(p)$, which depend on lensing parameters η (e.g., lensing mass), and source parameters S (i.e., pixel values), we can write the posterior probability distribution (PDF) for the parameters as $\mathcal{L} \propto e^{-\mathcal{E}/2}$, where $\mathcal{E}$ is defined as

$$\begin{aligned} \mathcal{E}(p) &= (D - M(p))^{\mathsf{T}} \mathbf{C}_{\mathbf{N}}^{-1} (D - M(p)) + p^{\mathsf{T}} \mathbf{C}_{\mathbf{p}}^{-1} p \\ &= (D - M)^{\mathsf{T}} \mathbf{C}_{\mathbf{N}}^{-1} (D - M) + S^{\mathsf{T}} \mathbf{C}_{\mathbf{s}}^{-1} S + \phi^{\mathsf{T}} \mathbf{C}_{\phi}^{-1} \phi. \end{aligned} \quad (1)$$

In this expression, the first term on the right-hand side corresponds to the usual goodness-of-fit χ^2, and the second term corresponds to a prior on the model parameters p. $\mathbf{C}_{\mathbf{N}}$ is the noise covariance matrix and $\mathbf{C}_{\mathbf{p}}$ is the prior covariance matrix describing our assumed multivariate Gaussian prior

PDF for source parameters. We assume that the covariance matrix is block diagonal, $\mathbf{C_p} = \mathrm{diag}(\mathbf{C_\eta}, \mathbf{C_s}, \mathbf{C_\phi})$ i.e., we assume no prior covariance between lens parameters, source parameters, and phase parameters. In the second line of Equation (1) and for the rest of this paper, we set $\mathbf{C_\eta}^{-1} = 0$, i.e., we assume no prior on the lens model parameters.

Of particular importance is the prior on source pixel parameters S. We denote that block within $\mathbf{C_p}$ as $\mathbf{C_s}$, the source prior covariance matrix, which gives a term $S^T\mathbf{C_s}^{-1}S$ in Equation (1). This matrix is often written as $\mathbf{C_s} = \lambda\hat{\mathbf{C}}_\mathbf{s}$, where λ is a scaling parameter (Suyu et al. 2006). We describe our procedure for determining the strength of the source prior, λ, in Section 2.1. In Equation (1) the parameter vector S is defined without loss of generality so that the Gaussian prior is centered at $S = 0$. More generally, if we define S so that the prior is centered at some S_c, then we replace $S \to S - S_c$ in the second to last term in Equation (1).

We note that the number of source parameters S is quite large, and as noted in previous works (e.g., Suyu et al. 2006) the source prior will act to regularize the reconstruction of the source parameters and avoid overfitting the data. This Gaussian prior, described by covariance matrix $\mathbf{C_s}$, is discussed below in Section 2.1. For visibility data, the noise covariance $\mathbf{C_N}$ is diagonal and its amplitude can be determined from the data using a method described in detail in Section 4.

For ALMA observations, the data vector D consists of the real and imaginary parts of complex visibilities. We can write our model visibilities as

$$M(\eta, S, \phi) = \mathbf{F}(\phi)\mathbf{BL}(\eta)S, \qquad (2)$$

where S is a vector of source pixel values, $\mathbf{L}$ is a lensing operator that maps the brightness of each source pixel to the image pixels (sky emission), $\mathbf{B}$ is a diagonal primary beam operator that multiplies the sky emission with the primary beam, and $\mathbf{F}$ is a dense Fourier operator whose ijth element is equal to $e^{-i(\mathbf{k}_i\cdot\mathbf{r}_j + \phi_{l_1} - \phi_{l_2})}$, corresponding to a visibility at uv-coordinate $\mathbf{k}_i$ from baseline l (composed of two antennas, labeled l_1 and l_2) and an image pixel located at $\mathbf{r}_j$. Note that rows of $\mathbf{F}$ with equal values of l have a common phase error (e.g., visibilities from the same baseline within the segmentation time of the antenna phase corruption parameters). To calculate $\mathbf{L}$ for a set of lens parameters η (e.g., mass and ellipticity) we solve the nonlinear lensing equation using a ray-tracing approach. Note that $\mathbf{L}$, $\mathbf{B}$, and $\mathbf{F}$ are all linear operators and that S is a subset of the model parameter vector $p = \{\eta, S, \phi\}$. Application of $\mathbf{FB}$ to a sky emission model produces M, the vector of the model visibilities.

We treat the source parameters S (the source pixel intensities) as nuisance parameters and marginalize over them. Our goal is then to compute the lensing parameter posterior described by Equation (1), marginalized over source parameters. Because the observables are linear in the source pixels and $\mathcal{E}$ is quadratic in the observables, then the likelihood is a Gaussian function of the source pixels. Since our assumed source prior is also Gaussian, the posterior is Gaussian. This allows us to analytically marginalize over the source nuisance parameters using Gaussian integrals to determine the posterior PDF for the remaining parameters.

The difference in marginalized log posterior between two models is then

$$\Delta\mathcal{E} = \Delta[B^T\mathbf{A}^{-1}B - \log(\det[\mathbf{A}])], \qquad (3)$$

where $\mathbf{A} = (\mathbf{FBL})^T\mathbf{C_N}^{-1}(\mathbf{FBL}) + \mathbf{C_s}^{-1}$ and $B = (\mathbf{LBF})^T\mathbf{C_N}^{-1}D$ (Suyu & Halkola 2010). The source reconstruction that maximizes the unmarginalized posterior (at a fixed lens model) is also analytic and is given by (Warren & Dye 2003; Suyu et al. 2006)

$$S = \mathbf{A}^{-1}B. \qquad (4)$$

We note that the above formalism is general to multi-frequency data. Vectors D and S can be concatenations of multiple data and reconstructed source vectors in different frequencies, while matrices $\mathbf{C_N}$ and $\mathbf{C_s}$ could be formed as block diagonal matrices including the noise and source prior in each channel. It is also possible to include regularization between different frequencies by allowing non-zero elements in $\mathbf{C_s}$ in off-block-diagonal elements.

2.1. Source Structural Prior

As Equation (1) makes explicit, the posterior PDF depends on our choice of the source prior, $\mathbf{C_s}$. Various forms of source priors are explored in the literature, including so-called gradient and curvature priors (Warren & Dye 2003; Suyu et al. 2006). More generally, we can construct a source prior covariance that is based on the expected statistical properties of the source emission. For example, if the covariance is stationary (i.e., depends only on the distance between two pixels) then we can describe the covariance matrix completely by an isotropic correlation function, or equivalently, by a diagonal power spectrum. The gradient and curvature priors, for example, correspond to power-law power spectra $P(k) \propto k^{-n}$ for appropriate n. We do not have to restrict ourselves to power-law power spectra, however. For instance, we could employ more physically motivated priors that are based on the expected morphology of the specific sources under investigation.

The dust and molecular line morphologies of early star-forming galaxies are expected to be well represented by a number of star-forming clumps embedded in a larger structure such as an exponential disk. One can use this structure to construct a source prior by calculating the power spectrum and covariance of such a clustered source model. Suppose that we have N_c clumps in our source galaxy whose distribution within the galaxy has profile $U_c(r)$. We normalize U_c to have unit integral, $\int U_c(r)d^2r = 1$. Its Fourier transform is $U_c(k)$. Clump i has luminosity L_i and profile $u_i(r)$, normalized to have unit integral, $\int u_i(r)d^2r = 1$. Then the power spectrum of the source emission is proportional to

$$P_{\mathrm{src}}(k) \propto \left[\sum_i^{N_c} L_i^2|u_i(k)|^2 + \sum_{i\neq j}^{N_c} L_iL_j|U_c(k)|^2 u_i^*(k)u_j(k)\right]. \qquad (5)$$

The Fourier transform of this power spectrum gives the correlation function of the source emission, and the source covariance $\mathbf{C_s}$ is determined from this correlation function. Note that the normalization of the power spectrum (and hence the normalization of $\mathbf{C_s}$) has not been specified. In principle we could normalize $\mathbf{C_s}$ using the observed intensity, but this has been magnified by an a priori unknown lensing magnification. Instead we normalize $\mathbf{C_s}$ by maximizing the marginalized likelihood (Bayesian evidence) for the fixed-parameter lens

model, as discussed in Suyu et al. (2006). We can scale an arbitrarily normalized source covariance matrix $\hat{\mathbf{C}}_\mathbf{s}$ (which in essence only defines the form of the prior) to get an appropriately scaled matrix $\mathbf{C_s} = \lambda\hat{\mathbf{C}}_\mathbf{s}$, where λ is a scaling parameter which can be determined by solving

$$N_s - \lambda\,\mathrm{Tr}([\mathbf{FL} + \hat{\mathbf{C}}_\mathbf{s}]^{-1}\hat{\mathbf{C}}_\mathbf{s}^{-1}) - \lambda S^\mathsf{T}\hat{\mathbf{C}}_\mathbf{s}S = 0, \qquad (6)$$

where N_s is the number of source pixels and S is determined using Equation (4). This equation can then be solved nonlinearly.

3. SEARCHING FOR SUBHALOS

Equation (1) tells us how well any lens model fits the data. That model can contain a smooth potential of the main lens and a number of subhalos. As we add subhalos to the mass model, the model parameter space grows considerably in dimensionality, and searching that high-dimensional parameter space can become computationally expensive. In this section we describe a simple approach toward expediting that search. Our method relies on first making a linear approximation to the effect of subhalos on the model predictions to identify promising points in parameter space where subhalos could significantly improve the fit. Once we have identified those promising parameter values using the linearized search we then conduct searches using the full nonlinear lens equations, starting at those points in parameter space.

As discussed above, lens model predictions are in general highly nonlinear functions of most of the lens parameters and are linear only in the source parameters. In this section we show that once a maximum posterior model is found using the procedure described above, we can linearize the model predictions in a small local neighborhood of these fiducial parameters even when substructure is added to the mass model. The key aspect of low-mass substructure that permits a linearized treatment, even approximately, is that the effects of substructure are perturbative when expanded in the deflection angle.

For the purpose of our discussion below it will be useful to separate the subhalo parameters $\boldsymbol{p}_\mathrm{sub}$ from the rest of the parameters $\boldsymbol{p}$ in our notation. The parameter set $\boldsymbol{p}$ therefore consists of macrolens density parameters, background source parameters, and other nuisance parameter (e.g., antenna phases), while we reserve the subhalo parameters as $\boldsymbol{p}_\mathrm{sub}$.

To infer the presence of subhalos we study the posterior PDF for the subhalo parameters, $\boldsymbol{p}_\mathrm{sub}$, marginalized over all the rest of the parameters $\boldsymbol{p}$, and we compare with the marginalized posterior PDFs in models with different numbers of subhalos. In particular, the ratio of the marginalized posterior of a subhalo model with mass $M_\mathrm{sub} > 0$ to that of a model without subhalos (or equivalently, with $M_\mathrm{sub} = 0$) provides a route to the marginal posterior for the presence of the given subhalo.

In our approach we first find the maximum posterior smooth model using the procedure described in Section 2. Next, we consider the effect of subhalos. Note that the effects of low-mass subhalos on the lensing deflection angles entering the matrix $\mathbf{L}$ are quite small in general. More precisely, low-mass substructure produces only a small perturbation to the deflection angle at every location. This suggests a linearized treatment of subhalos (Dalal & Kochanek 2002; Hezaveh et al. 2013a). For a given subhalo (which could be parameterized by its location and mass), we add the subhalo

deflections to the deflections arising from the best-fit smooth lensing potential model. We write the marginalized posterior, $\mathcal{L}_m$, for the subhalo of that mass and that location as

$$\mathcal{L}_m(\boldsymbol{p}_\mathrm{sub}) = \int \mathcal{L}(\boldsymbol{p}, \boldsymbol{p}_\mathrm{sub})d\boldsymbol{p} \propto \int e^{-\frac{1}{2}\mathcal{E}}d\boldsymbol{p} \qquad (7)$$

where $\mathcal{L}$ is the joint posterior PDF and $\mathcal{E}$ is given by Equation (1), which we rewrite as

$$\mathcal{E}(\boldsymbol{p}, \boldsymbol{p}_\mathrm{sub}) = [\boldsymbol{D} - \boldsymbol{M}(\boldsymbol{p}, \boldsymbol{p}_\mathrm{sub})]^\mathsf{T}\mathbf{C}_\mathbf{N}^{-1}[\boldsymbol{D} - \boldsymbol{M}(\boldsymbol{p}, \boldsymbol{p}_\mathrm{sub})]$$
$$+ \boldsymbol{p}^\mathsf{T}\mathbf{C}_\mathbf{p}^{-1}\boldsymbol{p}, \qquad (8)$$

to make the dependence on $\boldsymbol{p}_\mathrm{sub}$ explicit. Here, $\mathbf{C}_\mathbf{p}$ is the prior covariance matrix of all parameters $\boldsymbol{p}$, and $\mathbf{C}_\mathbf{p}$ contains the block $\mathbf{C_s}$ as described in the previous section. Note that as in Equation (1) our expression for $\mathcal{E}$ still includes a prior term so that it does not solely reflect the goodness-of-fit. The integral in Equation (7) is analytic when we can use a linearized treatment of parameters, i.e., when subhalos are perturbative (Hezaveh et al. 2014) (also known as the Laplace approximation). To leading order, we can write

$$\boldsymbol{M}(\boldsymbol{p} = \boldsymbol{p}_0 + \Delta\boldsymbol{p}) = \boldsymbol{M}(\boldsymbol{p}_0) + \frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}}\Delta\boldsymbol{p} \qquad (9)$$

for some fiducial parameter set $\boldsymbol{p}_0$, which we take to be the smooth model parameters from the maximum posterior model without substructure. Substituting this expression for $\boldsymbol{M}$ into $\mathcal{E}$ gives us a quadratic expression in the parameter shift $\Delta\boldsymbol{p}$,

$$\mathcal{E}(\boldsymbol{p}_0 + \Delta\boldsymbol{p}, \boldsymbol{p}_\mathrm{sub}) = \mathcal{E}_0 + 2W^\mathsf{T}\Delta\boldsymbol{p} + \Delta\boldsymbol{p}^\mathsf{T}\boldsymbol{R}^{-1}\Delta\boldsymbol{p}, \qquad (10)$$

where

$$\mathcal{E}_0 = \Delta\boldsymbol{D}^\mathsf{T}\mathbf{C}_\mathbf{N}^{-1}\Delta\boldsymbol{D} + \boldsymbol{p}_0^\mathsf{T}\mathbf{C}_\mathbf{p}^{-1}\boldsymbol{p}_0 \qquad (11)$$

$$\mathbf{R} = \left[\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}}^\mathsf{T}\mathbf{C}_\mathbf{N}^{-1}\frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}} + \mathbf{C}_\mathbf{p}^{-1}\right]^{-1} \qquad (12)$$

$$\boldsymbol{W} = \frac{\partial \boldsymbol{M}}{\partial \boldsymbol{p}}^\mathsf{T}\mathbf{C}_\mathbf{N}^{-1}\Delta\boldsymbol{D} - \mathbf{C}_\mathbf{p}^{-1}\boldsymbol{p}_0 \qquad (13)$$

and $\Delta\boldsymbol{D} = \boldsymbol{D} - \boldsymbol{M}(\boldsymbol{p}_0, \boldsymbol{p}_\mathrm{sub})$ consists of the residuals between the data and the predictions of the *unadjusted* smooth model plus the subhalo. Performing the integral in Equation (7) then gives the likelihood for the subhalo parameters $\boldsymbol{p}_\mathrm{sub}$, marginalized over all nuisance parameters $\boldsymbol{p}$,

$$\mathcal{L}_m(\boldsymbol{p}_\mathrm{sub}) \propto |\mathbf{R}|^{-1}e^{-\frac{1}{2}\mathcal{E}_0}\,e^{\frac{1}{2}W^\mathsf{T}\mathbf{R}^{-1}W} \equiv e^{-\frac{1}{2}\mathcal{E}_m(\boldsymbol{p}_\mathrm{sub})}, \qquad (14)$$

where we define an effective marginalized $\mathcal{E}_m(\boldsymbol{p}_\mathrm{sub})$ in terms of the log of the marginalized likelihood,

$$\mathcal{E}_m = \mathcal{E}_0 - W^\mathsf{T}\mathbf{R}W + \mathrm{const.} \qquad (15)$$

Again, $e^{-0.5\,\mathcal{E}_0}$ is proportional to the posterior if we assign uniform priors to the subhalo parameters and we add a subhalo to our best-fit smooth model without adjusting the parameters (i.e., for $\boldsymbol{p} = \boldsymbol{p}_0$): the $W^\mathsf{T}\mathbf{R}W$ term accounts for the marginalization over the smooth model parameters $\boldsymbol{p}$. Since $\mathbf{R}$ does not depend on subhalo parameters $\boldsymbol{p}_\mathrm{sub}$, we can write the difference of log posterior for two models as

$$\Delta\mathcal{E}(\boldsymbol{p}_\mathrm{sub}) \equiv \mathcal{E}_m(\boldsymbol{p}_\mathrm{sub}) - \mathcal{E}_{m0} \qquad (16)$$

where $\mathcal{E}_{m0}$ corresponds to the marginalized likelihood for the model with no subhalos.

Because we parameterize the source emission using pixel brightnesses on a fixed grid rather than using parameters like the locations of source clumps that can freely move, one subtlety that arises in the above expressions is that the matrix $\partial M/\partial p$ must involve terms corresponding to the deflection from the subhalo. Specifically, we evaluate the derivative $\partial M/\partial p$ in Equation (9) using a deflection field that includes deflection from the subhalo model under consideration. This correction introduces certain nonlinear terms which naively might appear to be small, but in practice we have found that neglecting these terms can lead to significant errors in estimating the ability of the smooth (macro) model to compensate for the effects of substructure, in particular for the most massive subhalos.

Using the procedure described above we can map out the marginalized posterior of models as a function of the subhalo parameters (e.g., location and mass). If we find any locations within the substructure parameter space that appear to significantly improve the fit, we then initiate an optimization of the model parameters using the full nonlinear lens equations starting from the most promising points found using the linearized search.

It is worth stressing that the linearized lens model is very accurate over most of the subhalo parameter space as long as subhalo effects are small. The only locations where the linearized model predictions noticeably differ from fully nonlinear model predictions are where subhalos produce significant changes to the posterior, i.e., where $|\Delta\mathcal{E}| \gtrsim 20$, which in general only occurs within a small fraction of the possible parameter space. Thanks to the high accuracy of the linearized approximation, we can use the maps of $\mathcal{E}_m$ not only to identify promising starting points for our nonlinear search but also to exclude subhalos over much of the possible parameter space where subhalos are ruled out. In other words, we can use the maps of $\mathcal{E}_m$ to derive constraints on the presence of subhalos and, in principle, derive constraints on the mean abundance of subhalos. To derive bounds on the abundance, we would similarly need to map out $\mathcal{E}_m$ for models that have two subhalos, three subhalos, and so on for all possible subhalo locations. However, because we are considering subhalos of low mass, their effects on the observables are perturbative and restricted to localized regions on the sky. We will argue below that this allows us to derive approximate constraints on the subhalo abundance from maps of $\mathcal{E}_m$ from models containing one subhalo.

4. IMPLEMENTATION, SIMULATIONS, AND TESTS

Perhaps the most challenging aspect of this analysis is the sheer volume of the visibility data provided by interferometers such as ALMA. For ALMA long-baseline campaign observations of SDP.81 (ALMA Partnership et al. 2015a), visibilities were typically recorded every 0.5 s for hundreds of baselines, each with more than 1000 spectral channels. This results in about 10^8 visibilities for a four-hour long observation. For a reasonable pixel size, the size of the resulting Fourier operator alone can exceed a few terabytes, and fitting many models to such matrices is beyond the capabilities of medium-size clusters. Fortunately, careful binning of the visibilities can make the problem tractable. Because of the non-zero antenna size, each visibility actually samples a range of spatial

frequencies around its nominal uv location and visibilities within an antenna diameter of each other are highly correlated. Binning visibilities within antenna-sized patches of the uv-plane can reduce the size of the data dramatically (to fewer than 10^6 visibilities per spectral channel) without significantly decreasing the information content of the data.

Even after binning this is still a computationally challenging task: the matrices are still too large to fit in the memory of single processors, but they can fit across distributed memory on large clusters. This requires the use of distributed linear-algebra libraries to store and manipulate these matrices. It is also important to note that unlike the blurring matrices used in analyzing CCD images of gravitational lenses, which are extremely sparse, the Fourier operator here is fully dense and does not permit the application of sparse libraries. For these reasons, we take advantage of the "*Elemental*" dense linear algebra library, which efficiently performs linear algebra operations over distributed cores (Poulson et al. 2013a, 2013b).

Our pipeline includes a "pre-processing" part, in which the visibilities are binned according to their position in the uv-plane and the noise covariance matrix $\mathbf{C_N}$ is determined. The result of this pre-processing is a vector of binned visibilities which can be modeled as described in Section 2.

The data sets provided by ALMA contain information about the standard deviation of noise for each visibility which are proportional to the system temperature T_{sys}. This information, however, is intended to be used for weighting of the visibilities for imaging purposes and is not adequately scaled to provide accurate root mean square (rms) noise in appropriate units (e.g., Jy). The noise variance cannot be taken directly from the visibilities because their variation contains a significant contribution from the sky signal, which must be removed. We achieve this by first grouping the visibilities into bins that probe the same signal and then taking differences between visibilities that null the sky within those groups, leaving only noise. We can then either find the best overall scaling for the provided system temperature, T_{sys}, to maximize the likelihood of the observed noise or, alternatively, we can simply assume that the variance in the subtracted visibilities is equal to twice the noise variance. We try both methods and find that they provide consistent results.

Explicitly, if we assume that the noise rms is proportional to T_{sys}, we can write $\mathbf{C_N} = A^2\mathbf{T^T T}$, where $\mathbf{T}$ is a diagonal matrix of system temperatures and A is an unknown scaling factor. Assuming Gaussian noise gives

$$\log(\mathcal{L}(A)) = -\frac{1}{2}\log(|A\mathbf{T}|) - \frac{1}{2}NATN + C, \qquad (17)$$

where $\mathcal{L}(A)$ is the likelihood of a certain value of A, N is the vector of noise (the residual after subtracting half of visibilities from the other half), and C is the sum of other terms that do not depend on A. By solving $\partial\log(\mathcal{L}(A))/\partial A = 0$ for A we get the most likely A,

$$A = \frac{n}{NTN}, \qquad (18)$$

where n is the number of visibilities used to estimate A (half of all visibilities, since one half is subtracted from the other to remove the signal). We build the noise covariance matrix $\mathbf{C_N}$ by scaling the system temperatures with this scaling. With appropriate variances in hand, we fit the visibilities with a

smooth lens model, and search for subhalos perturbatively using the procedure described in Section 3.

4.1. Mock Observations

We performed a large number of tests described below using mock data to ensure that the pipeline provided accurate results. We tested for errors in the noise estimation and for biases in the smooth lens model. These simulations also quantify our false positive detection rate and subhalo detection efficiency.

The mock observations for these tests were produced using a fully independent code to reduce the chances of duplicating errors. For each mock data set a complex background source composed of tens to hundreds of smaller clumps was lensed and the lensed images were used to simulate ALMA observations using the Common Astronomy Software Applications package. The data pre-processing steps described above (visibility binning and noise estimation) were applied to the mock data and the binned visibilities were modeled to find a best-fit smooth model.

The smooth density model is described by a singular elliptical power-law surface density profile of the form $\kappa(x, y) \propto [x^2 + y^2/(1 - \epsilon)^2]^{-\alpha/2}$ where α is the radial power-law index, and x and y are measured relative to the lens centroid x_{lens}, y_{lens} (Barkana 1998). The lens centroid and the subhalo positions are given relative to the ALMA phase center. To explore the lens parameter space we implemented the Markov chain Monte Carlo (MCMC) algorithm proposed by Goodman & Weare (2010). We follow the procedure described in Foreman-Mackey et al. (2013). Our implementation allows sampling the posterior in parallel over many cores, taking advantage of the MPI message-passing system. On the large computing resources that we used (e.g., Towns et al. 2014), this allowed much faster parameter explorations by evaluating multiple models simultaneously on different cores.

Figure 1 shows the one- and two-dimensional marginal posterior PDFs for all lens model parameters given a mock data set. The gray shaded contours show the recovered posterior when using a gradient covariance matrix as the source prior. The red contours show the posterior for the same data, but using a curvature prior on the source parameters. The black dashed crosses show the true parameters that were used to generate the mock data. As can be seen, in both cases we successfully manage to recover the parameters with no apparent biases.

However, in the course of our tests, we encountered two issues that could have resulted in biased parameters. The first resulted from using the power spectrum prior as defined in Equation (5). Since the power spectrum does not diverge at $k \rightarrow 0$, this prior tries to minimize the total emission of the unlensed source, i.e., it biases against low magnification. This bias toward high magnification favors shallow radial density slopes. To fix this issue, we removed the DC mode from this covariance matrix, which (similar to the gradient and curvature priors) renders it insensitive to the total sum of source pixel intensities. We can remove the zero mode of the covariance matrix by replacing

$$\bar{\mathbf{C}}_{\mathbf{s}}^{-1} = \mathbf{C}_{\mathbf{s}}^{-1} - \frac{\mathbf{C}_{\mathbf{s}}^{-1} \boldsymbol{a} \, \boldsymbol{a}^{\mathsf{T}} \, \mathbf{C}_{\mathbf{s}}^{-1}}{\boldsymbol{a}^{\mathsf{T}} \mathbf{C}_{\mathbf{s}}^{-1} \boldsymbol{a}}, \tag{19}$$

which follows from the Sherman–Morrison theorem[14] (Sherman & Morrison 1949). Here, $\boldsymbol{a}$ is a column vector containing only ones (i.e., the zero mode) and $\bar{\mathbf{C}}_{\mathbf{s}}$ is the modified covariance matrix with no DC sensitivity (i.e., $\bar{\mathbf{C}}_{\mathbf{s}} = \mathbf{C}_{\mathbf{s}} + \boldsymbol{a} \, \boldsymbol{a}^{\mathsf{T}}$). Removing the zero mode removes the bias toward high magnification and small source sizes.

Another potential source of bias was the choice of the area covered by source pixels. For a given lens model, if an image pixel is mapped outside the area covered by the source grid it will automatically be assigned a value of zero. A model that maps more image pixels to the defined source area may have more freedom in fitting the data compared with a model with fewer in-source-grid mappings. Therefore, under equal conditions, lens models that map all image pixels to some source pixel are preferred to models that do not. This can result in significant biases in lensing parameters (e.g., ellipticity and slope). To avoid this bias, the size of the source grid should be large enough to cover the lensing models during the fitting process (Suyu et al. 2006; Vegetti & Koopmans 2009).

4.2. Phase Corruption

Once a best-fit model is determined, we start the subhalo search by mapping $\Delta \mathcal{E}$ over the subhalo parameter space (position and mass). As mentioned earlier, in addition to lens and source parameters, we marginalize over other nuisance parameters in the linear subhalo finder, such as antenna phase errors. This is similar to the procedure described in Hezaveh et al. (2013b) with the addition of the possibility of imposing a prior on the antenna phase errors and allowing for time-segmented phase errors.

Figure 2 shows maps of $\Delta \mathcal{E}$ as a function of subhalo position for various cases of simulated data. For a simulation with no subhalo present (top left), the analysis correctly excludes the presence of a subhalo over the region of sensitivity. In the remaining panels the mocks contain a subhalo at the location of the blue circled cross. The simulation for the top right panel does not include phase errors in the measurements. The subhalo is correctly detected at $\sim 5\sigma$ with no false positives. The mocks for the two bottom panels include antenna phase errors. In the bottom left panel we have not corrected for these phase errors, resulting in numerous false positives. The bottom right panel shows the results for the same mock data when phase errors are marginalized over appropriately, resulting in the disappearance of the false detections. All four panels have the same noise realization.

This test illustrates one of the reasons why a careful analysis of the visibilities is essential to search for substructure using interferometric data. Given the challenges of analyzing large interferometric data sets, one might be tempted to simply analyze CLEAN images instead of visibilities. CLEAN images fix the value of the antenna phases and do not allow them to be marginalized over when comparing different models which, as illustrated here, could result in spurious detections. In addition,

---

[14] The Sherman–Morrison theorem states that for matrix $\mathbf{C}_{\mathbf{s}}$ and vectors $\boldsymbol{b}$ and $\boldsymbol{c}$,

$$(\mathbf{C}_{\mathbf{s}} + \boldsymbol{b} \, \boldsymbol{c}^{T})^{-1} = \mathbf{C}_{\mathbf{s}}^{-1} - \frac{\mathbf{C}_{\mathbf{s}}^{-1} \boldsymbol{b} \, \boldsymbol{c}^{T} \, \mathbf{C}_{\mathbf{s}}^{-1}}{1 + \boldsymbol{c}^{T} \, \mathbf{C}_{\mathbf{s}}^{-1} \boldsymbol{b}}.$$

This can be verified by inspection, by multiplying the above equation by $\mathbf{C}_{\mathbf{s}} + \boldsymbol{b} \, \boldsymbol{c}^{\mathsf{T}}$ on either the left or the right. We wish to adjust the covariance $\mathbf{C}_{\mathbf{s}}$ to give infinite variance to modes with uniform brightness across all pixels. In our notation this kind of zero mode is written as $f\boldsymbol{a}$, where $\boldsymbol{a}$ is a vector containing all ones in its components and f is a scalar. To give infinite variance to these modes, we set $\boldsymbol{b} = \boldsymbol{c} = f \, \boldsymbol{a}$ in the Sherman–Morrison formula and take the limit $f \rightarrow \infty$, which results in Equation (19).
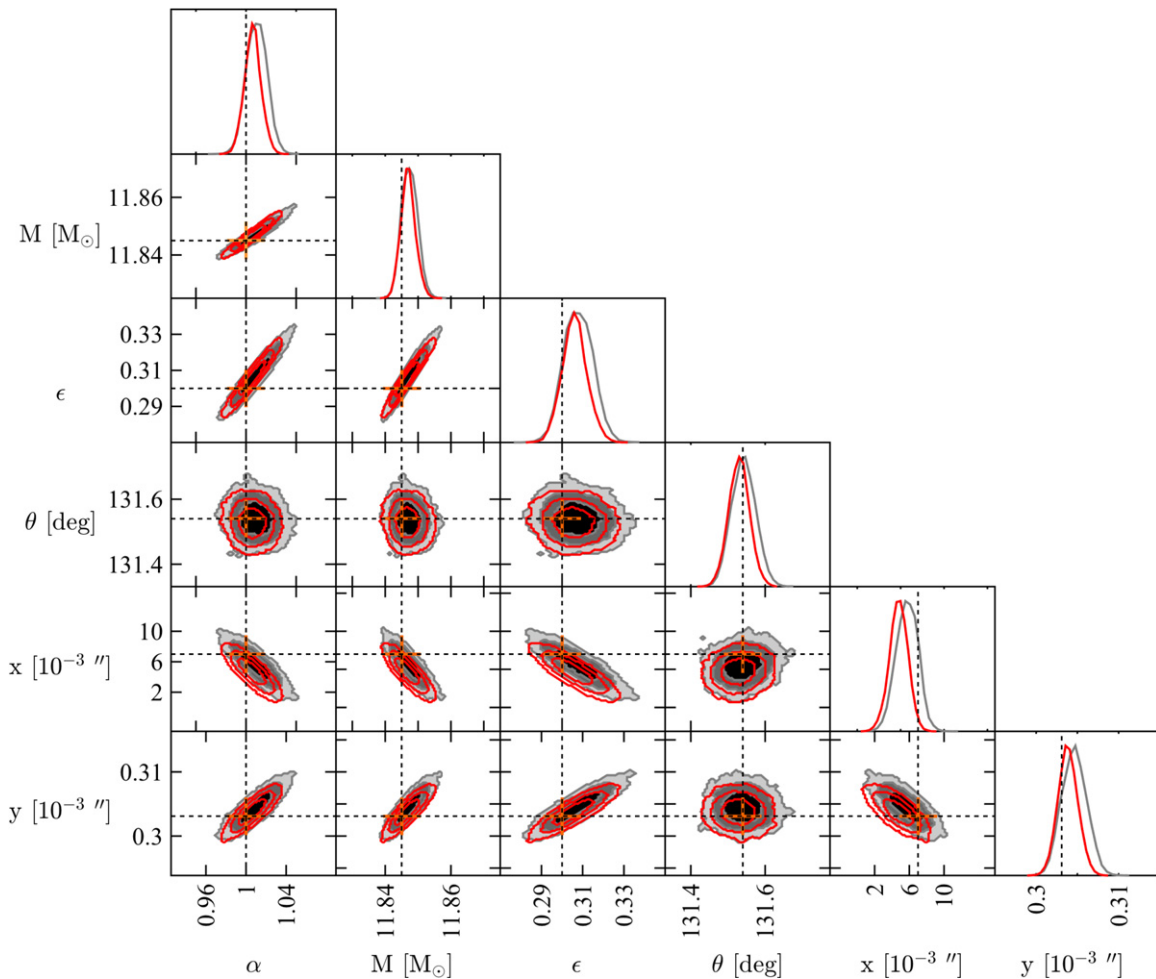
Figure 1. Posterior PDF for the lens model parameters for a mock data set of a 2 hr long observation of a 50 mJy source with an antenna configuration with maximum baselines of ∼2 km. The different colors correspond to different regularization schemes: the gray contours are for a gradient prior, and the red contours are for a curvature prior (Warren & Dye 2003). We see no evidence of systematic biases in the reconstructed parameters. The results shown here are typical of the mocks we simulated.

CLEAN images are produced through a nonlinear deconvolution procedure whose effects on the data and the correlated noise properties cannot be well quantified. It is worth stressing that the effects of low-mass substructure on the lensed images can be quite subtle, and so approximations and assumptions which may not lead to serious errors in other contexts could introduce significant errors for a substructure lensing analysis. This is why we adopt the approach of explicitly modeling the visibilities and marginalizing over the nuisance parameters.

Finally, we note that although in the two right panels of Figure 2 we only observe a single region with significant negative $\Delta\mathcal{E}$ (where the subhalo is detected), subhalos with larger masses can result in multiple islands. Figure 3 shows a mock generated using a simulation with a subhalo of mass $M_{sub} = 10^9 M_\odot$. The left panel shows the $\Delta\mathcal{E}$ map when searching for this subhalo. As can be seen, there are multiple islands where a subhalo can produce a better fit; however, the true position of the subhalo corresponds to the lowest $\Delta\mathcal{E}$. The right panel of Figure 3 shows the $\Delta\mathcal{E}$ map when the detected subhalo is added to the macro parameters and a search for an additional subhalo is performed, showing that all corresponding islands disappear.

5. RESULTS FOR SDP.81

We use the recent ALMA Science Verification observations of the strongly lensed system SDP.81 (ALMA Partnership et al. 2015b) to illustrate the application of this method to real data. The system is comprised of a foreground galaxy ($z = 0.299$), lensing a background dusty, star-forming galaxy ($z = 3.042$). Figure 4 shows the lensing galaxy observed by the *HST* (grayscale) superimposed on the lensed arcs observed by ALMA. Our analysis shows that this data set is highly sensitive to the effects of low-mass subhalos. In the analysis presented here, we used only the calibrated continuum data from bands 6 and 7. In future work we will present our analysis of the full data set, including CO line data.

We estimate the noise variance for each visibility using the procedure described above and bin the visibilities in 12 m cells. However, we only bin visibilities that share the same baseline and are taken within a short (≲20 minutes) observing period. This allows us to assign a separate antenna phase error to each baseline at different time intervals (the phase errors could slowly change during the course of an observation). This results in ∼0.5 million binned visibilities, a factor of ∼20 fewer than the original ∼10 million. We note that the shorter (≲2 km) baselines in the binned visibilities have significantly higher
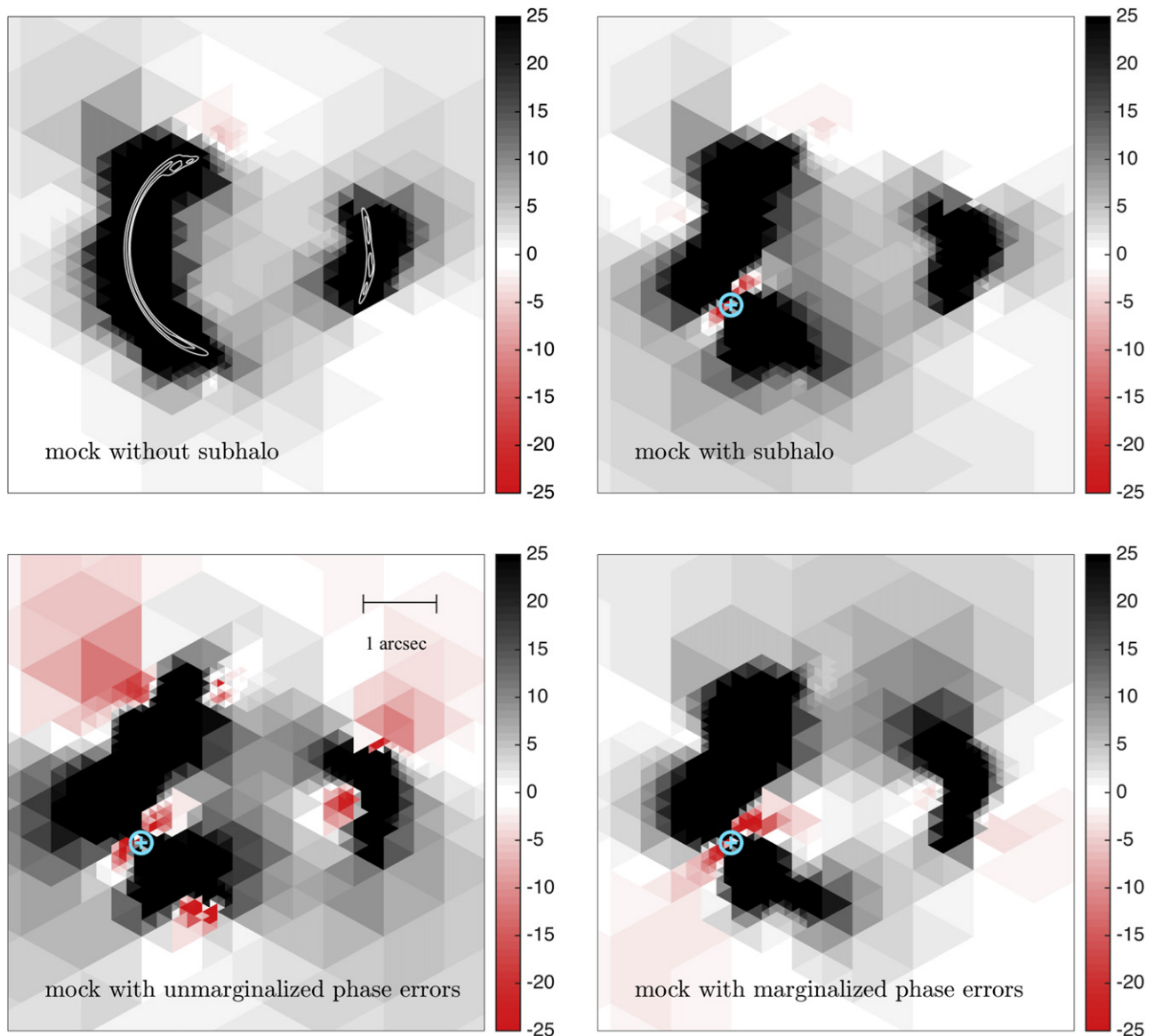
Figure 2. Maps of $\Delta\mathcal{E}$ (Equation (16)) as a function of location for a subhalo of mass $M = 4 \times 10^8 M_\odot$, using simulated data. Positive $\Delta\mathcal{E}$ corresponds to subhalos excluded at that location while negative $\Delta\mathcal{E}$ corresponds to subhalos improving the log posterior. The mock data for the top left panel are for a simulation without any substructure and our analysis excludes the presence of subhalos of this mass over the area of sensitivity. The simulated arcs have been overlaid with white contours. In the remaining panels, the mocks contain a subhalo at the location of the blue circled cross. The simulation for the top right panel does not include phase errors in the measurements. The lower left panel includes such errors but does not compensate for them and the lower right panel includes phase errors and marginalizes over them. All four panels have the same noise realization and a model subhalo mass of $M = 4 \times 10^8 M_\odot$.

signal-to-noise compared to the longer baselines because (a) for resolved sources, the signal diminishes on the longest baselines and (b) the visibilities are sampled more densely on short baselines, meaning that the short-baseline bins contain a larger number of visibilities. This allows us to speed up our search for the best-fitting parameters using the following approach. We first use only the subset of baselines shorter than 2 km in our initial MCMC analysis to localize the neighborhood of the best fit. This approach greatly expedites our MCMC optimization, as many evaluations of the likelihood are required to fully search the highly multidimensional parameter space of our smooth model. Once this initial localization has been achieved, we can use the full data set. We note, however, that the observations of SDP.81 are performed on various dates which span many weeks with significantly variable observational conditions, giving rise to a large range in noise properties. We

therefore decide to model binned visibilities with noise lower than 5.0 and 5.5 mJy in bands 6 and 7, respectively, which contain about 40% of the total number of unbinned visibilities.

In addition to the parameters described in Section 4.1, we allow for low-order angular multipoles in the main lens, of the form $\kappa_m(r, \phi) = [A_m \cos(m\phi) + B_m \sin(m\phi)](r/r_s)^{-\alpha}$ for $m = 3$, 4, where $r_s = 1''$. Note that the same radial slope α and centroid $(x_{\text{lens}}, y_{\text{lens}})$ are used for the multipoles and for the ellipsoidal piece. We also allow for external shear, parameterized by the usual components γ_1 and γ_2. To avoid degeneracy of the orientation angle when ellipticity is close to zero, we use fitting parameters ϵ_x and ϵ_y, defined so that $\epsilon = (\epsilon_x^2 + \epsilon_y^2)^{1/2}$ and the orientation angle is given by the arctangent of these components. Overall, therefore, our primary lens model contains 12 freely adjustable parameters.
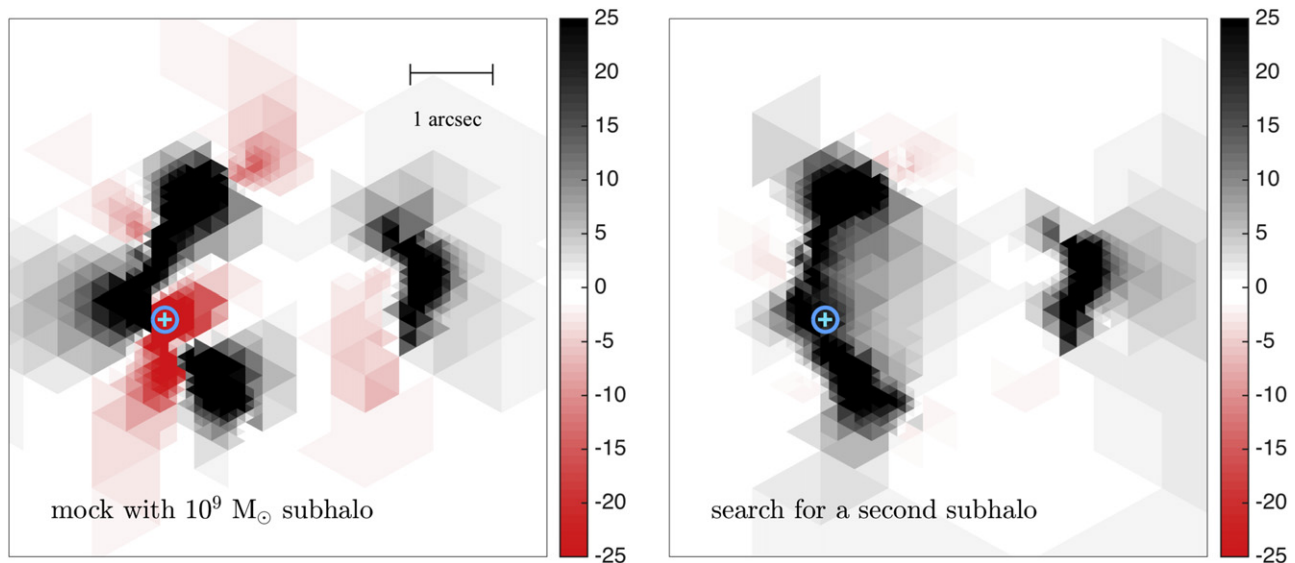
Figure 3. $\Delta\mathcal{E}$ maps for mock data including a massive ($M_{sub} = 10^9 M_\odot$) subhalo showing multiple islands where the addition of a subhalo could produce a better fit. We find that the island producing the lowest $\Delta\mathcal{E}$ corresponds to the true position of the subhalo (the blue cross). After adding the detected subhalo to the macro model and searching for a second subhalo the other islands disappear, confirming that they arise from the inability of the smooth model to fit perturbed observations at various locations on the arc. The position of the subhalo is marked with a circled cross. The mock data was not corrupted with phase errors, but phase errors have been marginalized.
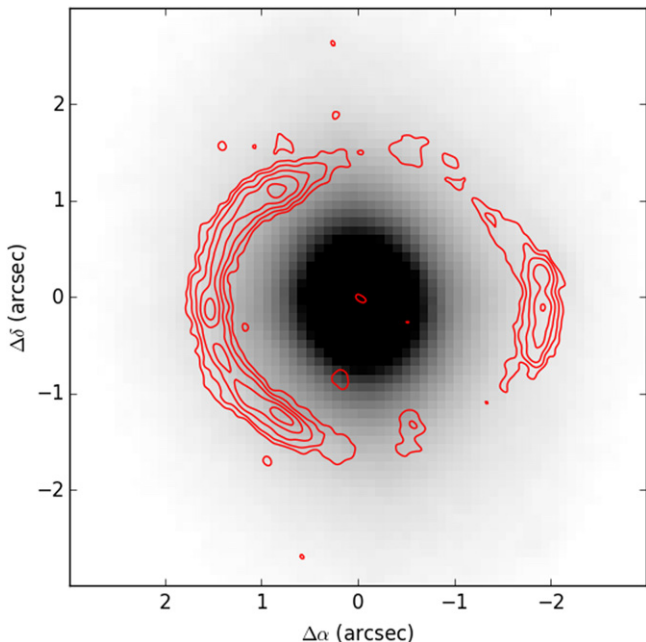
Figure 4. SDP.81 system. The grayscale shows *HST*/WFC3 F160W data, while the red contours show ALMA continuum emission in band 6.

5.1. Initial Subhalo Search

Once a smooth model is obtained, we use the best-fit parameters to perform a linearized search for subhalos. As we have mentioned above, these lens parameters, source parameters, antenna parameters, etc., all become nuisance parameters that we marginalize over for every different model when we search for subhalos. We follow Hezaveh et al. (2013a) and model the subhalo deflection field using a truncated isothermal surface density profile, also called a pseudo-Jaffe profile (Muñoz et al. 2001). This profile is characterized by a velocity

dispersion σ_v and truncation radius r_t, and the total mass of the subhalo is given by $M_{sub} = \pi\sigma_v^2 r_t/G$. To reduce the dimensionality of the subhalo parameter space, we assume that r_t is related to σ_v by $r_t = (\sigma_v/\sqrt{2}\,\sigma_G)r_E$, where σ_G is the velocity dispersion of the main lens, determined from its observed Einstein radius r_E. We search for subhalos over a range of subhalo masses, over a 8×8 arcsec area around the lens center. Figure 5 shows the results of our initial search. The figure plots $\Delta\mathcal{E}$, twice the difference in marginalized log posterior between a model with a subhalo compared to our smooth model, as a function of subhalo location for a subhalo mass $M = 10^{8.6}M_\odot$. As the figure indicates, there are several locations where adding a subhalo improves the posterior considerably, with the most significant having $\Delta\mathcal{E} = -22.2$.

As discussed above in Section 4, improper modeling of systematics and unknown errors can lead to spurious detections of substructure. We have attempted to mitigate these effects by marginalizing over many potential systematics, including time-varying antenna phase errors. Nevertheless, it is possible that the apparent detection of substructure indicated in Figure 5 could be due to an unknown interferometric data corruption such as visibility decorrelation or rapidly varying antenna phase errors. Given that such errors are temporally variable, an analysis of multiple data sets observed at different times can reveal if our analysis is affected by them. As a test of this, we analyzed bands 6 and 7 data separately, noting that they were obtained on different dates. Our analysis reveals a consistent pattern between the two bands (see Figure 5), giving us confidence that the level of unknown systematics from such effects is below our statistical uncertainties. Figure 6 illustrates the difference between our best-fitting model without substructure and the best-fitting model with substructure for bands 6 and 7. As expected, the subhalo's(s') effect is largely localized to its immediate vicinity and the counter-images of that region.

Based on the results of this initial linearized search, we then expanded our lens model to include a subhalo with three
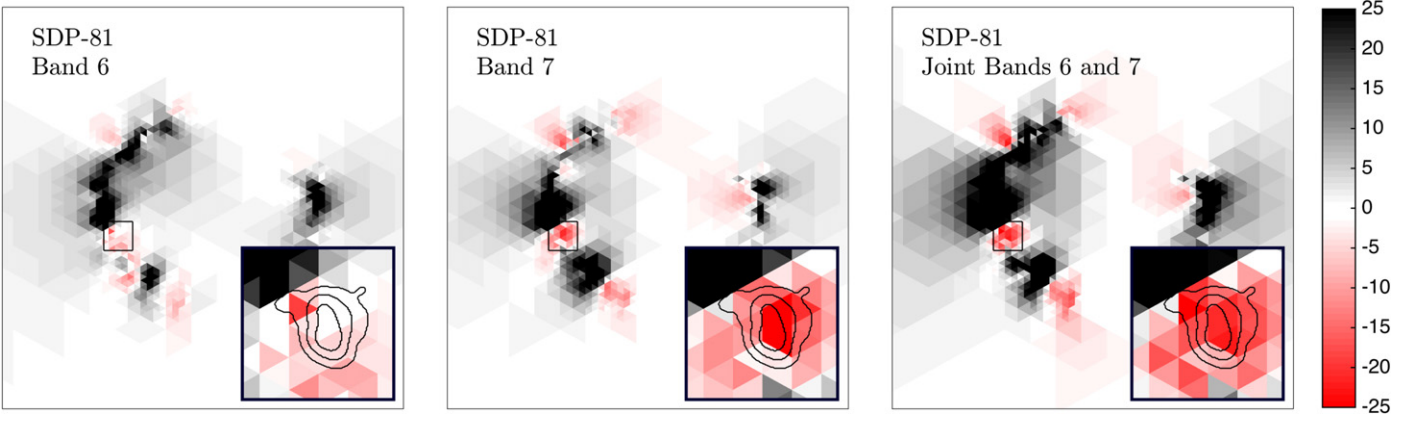
Figure 5. Initial subhalo search using ALMA Science Verification observations of SDP.81. Depicted are maps of linearized $\Delta\mathcal{E}$ from Equation (16) showing twice the difference in log marginalized posterior probability density between a smooth model without substructure and a model with a subhalo of mass $M = 10^{8.6}M_{\odot}$, as a function of location of that subhalo. The three panels correspond to analysis of Band 6 only (left), Band 7 only (middle), and joint Bands 6 and 7 (right). Based on the significant improvement to the fit provided by substructure (as indicated by the map), we subsequently added one subhalo to our lens model and re-optimized the model parameters (see Table 1). The contours in the insets show the 1-, 2-, and 3-σ confidence regions for the position of the subhalo from a nonlinear joint fit to the data.
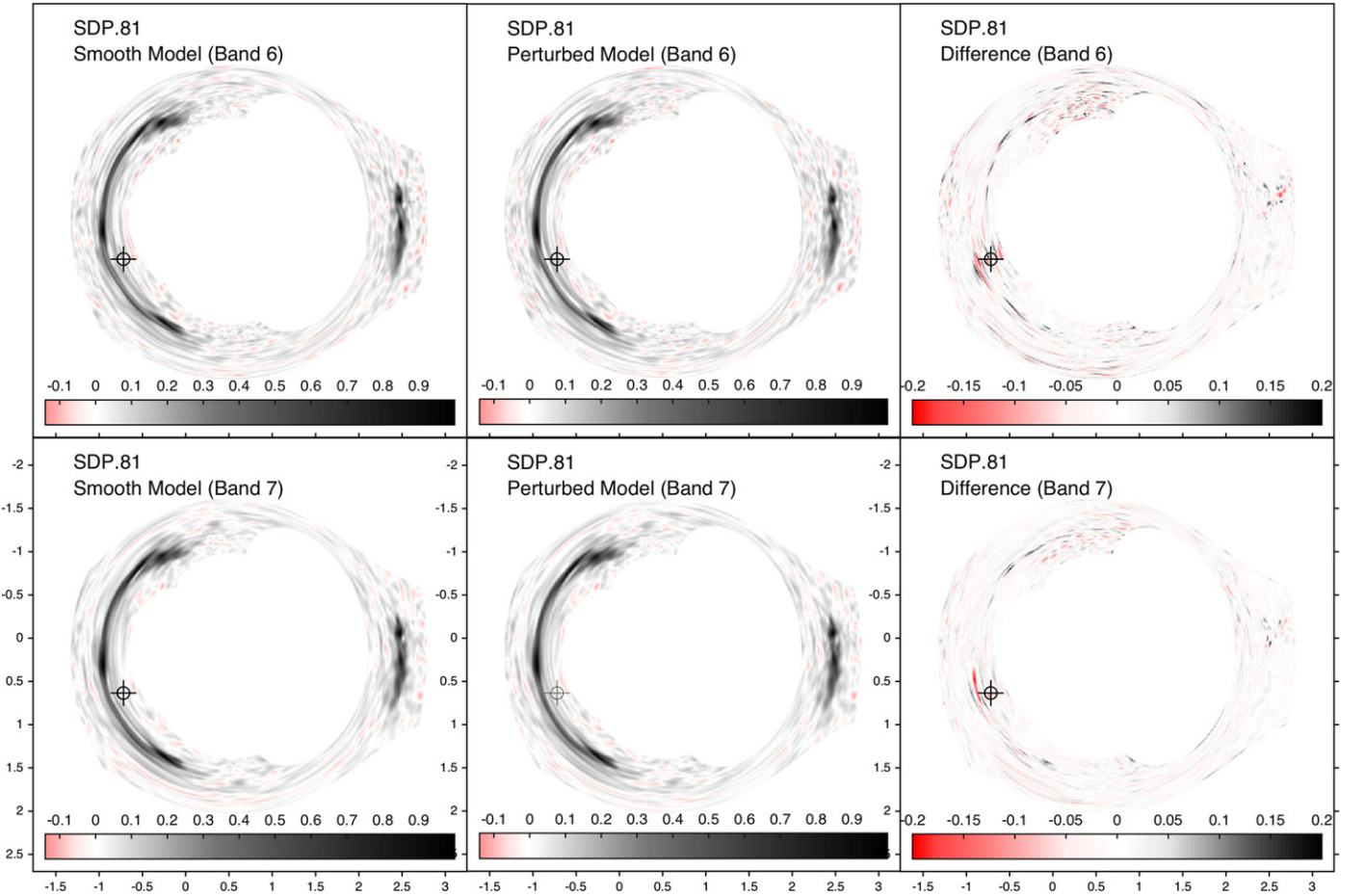
Figure 6. Top left: the sky emission model in band 6 for the best-fit smooth lens parameters for the SDP.81 data. Top middle: the same for the perturbed model. Top right: the difference between the two models. The bottom panels show the same for band 7. The bright feature in the difference plots is mainly caused by the astrometric anomaly of the arc. In each row, the images have been scaled to the peak flux of the smooth model.

adjustable parameters: mass M_{sub}, and 2D location x_{sub}. We then re-fit the joint data set, re-optimizing all the parameters fully nonlinearly. We find that a model with a subhalo of mass $M = 10^{8.96}M_{\odot}$ improves the marginalized log posterior fit by $\Delta\mathcal{E} = -47.3$ in the joint fit (note that the initial linear search

was performed at $M_{\mathrm{sub}} = 10^{8.6}M_{\odot}$). Based on this result, we conclude that the ALMA Science Verification observations of SDP.81 detect a subhalo in the projected mass distribution. Having found the best-fit parameters for the detected subhalo, we then sample the full parameter space (smooth lens and
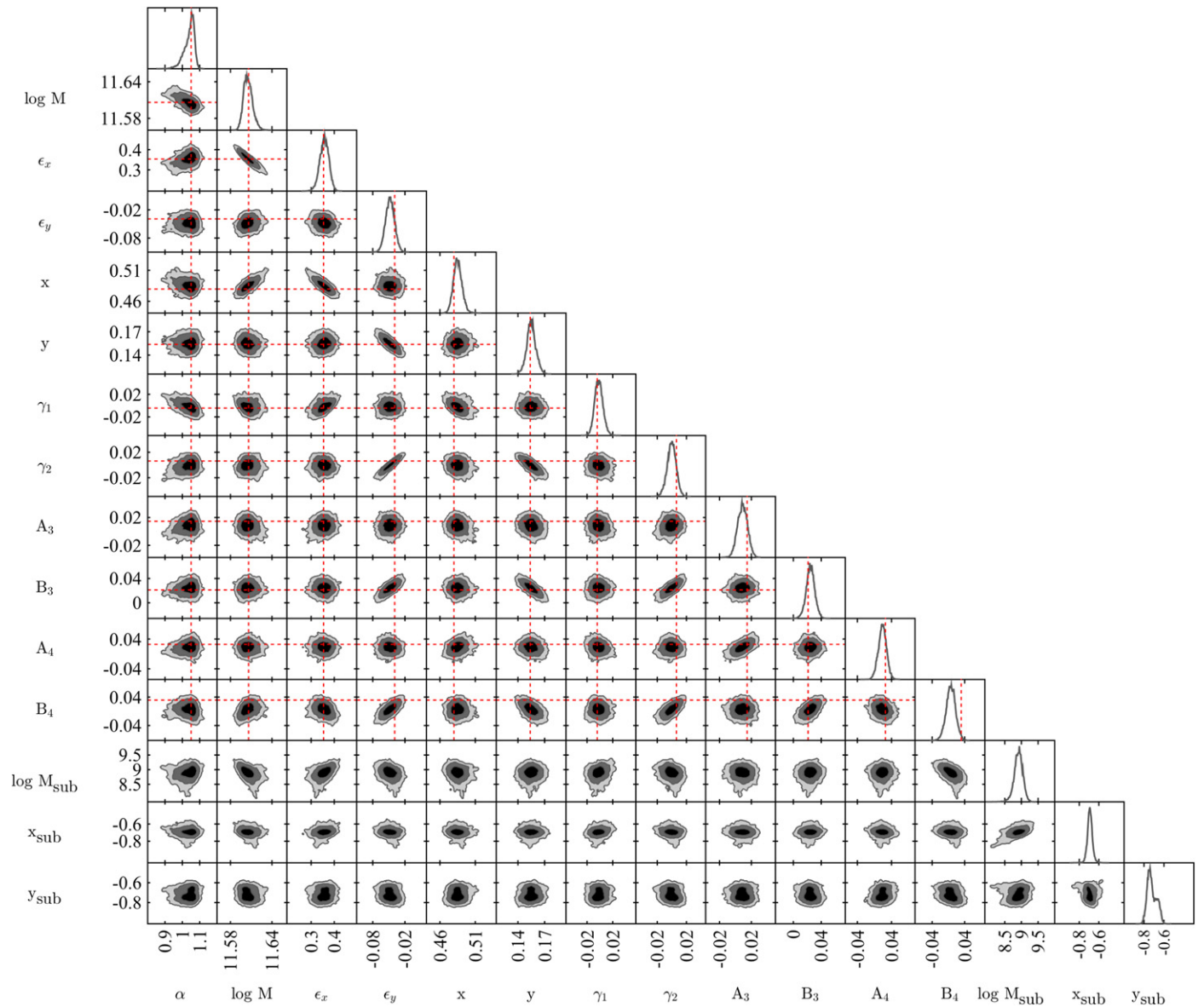
Figure 7. Contours show the 1-, 2-, and 3-σ (68.3, 95.5, and 99.7%) confidence regions for the parameters in our lens model for a joint analysis of bands 6 and 7 data. The red dashed lines show the maximum posterior smooth model parameters (in the absence of a subhalo in the model).

subhalo parameters) nonlinearly using our MCMC sampler. Figure 7 shows the error covariance of the reconstructed lens parameters for the joint fit to bands 6 and 7. We do not find evidence for significant degeneracies between the subhalo parameters and the parameters of the smooth lens model, including low-order multipoles in the gravitational potential. This confirms findings that such multipoles cannot mimic the effects of small-scale substructure for lenses with high-quality arcs (Kochanek & Dalal 2004).

The full set of best-fit lens model parameters are presented in Table 1. Many previous works have modeled the lens potential in SDP.81 using *HST* data (Dye et al. 2014), Submillimeter Array data (Bussmann et al. 2013), and ALMA data (Dye et al. 2015; Hatsukade et al. 2015; Rybak et al. 2015a, 2015b; Tamura et al. 2015; Wong et al. 2015). Our smooth model has a larger ellipticity compared with these models. We note, however, that our model has more degrees of freedom (e.g., angular multipoles) and phase errors, and that the degeneracy of some of these additional parameters with ellipticity may shift

its value. We do find that models with parameters given by these authors produce reasonable fits to the data. We also performed the linear subhalo search for these parameters, finding that they produce similar results and that the conclusion of the presence of the subhalo is robust against these variations. Figure 8 shows the reconstructed source using this model with pixel size of 10 mas in band 6 (top panel) and band 7 (bottom panel).

This model appears to be a good fit to the data when we fit the entire data set. The full data set, however, includes emission unrelated to SDP.81. The ALMA primary beam covers approximately $\sim 25''$, of which only the central few arcseconds are relevant for strong lens modeling. If we model the sky emission only over a 5×5 arcsec area centered on the lens, our model obtains $\chi^2 = 2 \times 10^5$ for 1.7×10^5 degrees of freedom, suggesting that not all of the signals in the data have been modeled. However, if we expand our source-plane image to cover the entire primary beam, additional flux is indicated away from the lensed galaxy and the χ^2 decreases to
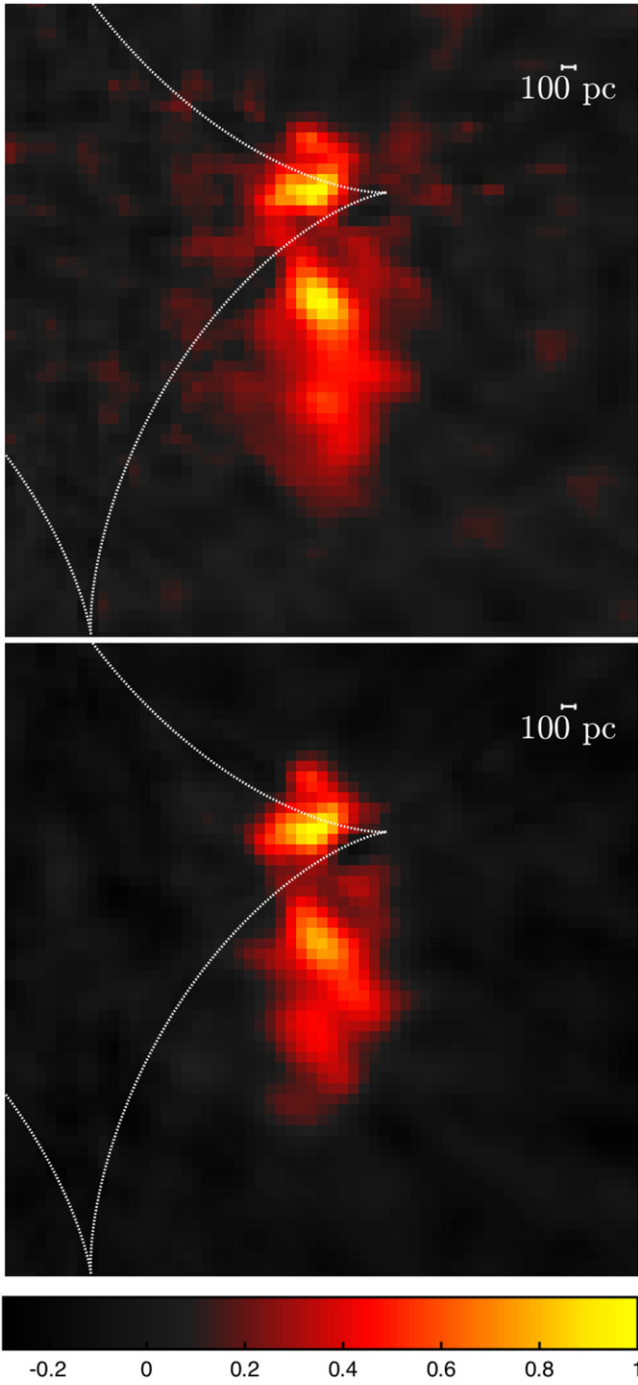
Figure 8. Reconstructed source continuum emission from Band 6 (top panel) and Band 7 (bottom panel) data on a 10 mas pixel grid. The white dashed curve shows the tangential caustic predicted by our best-fit smooth model.
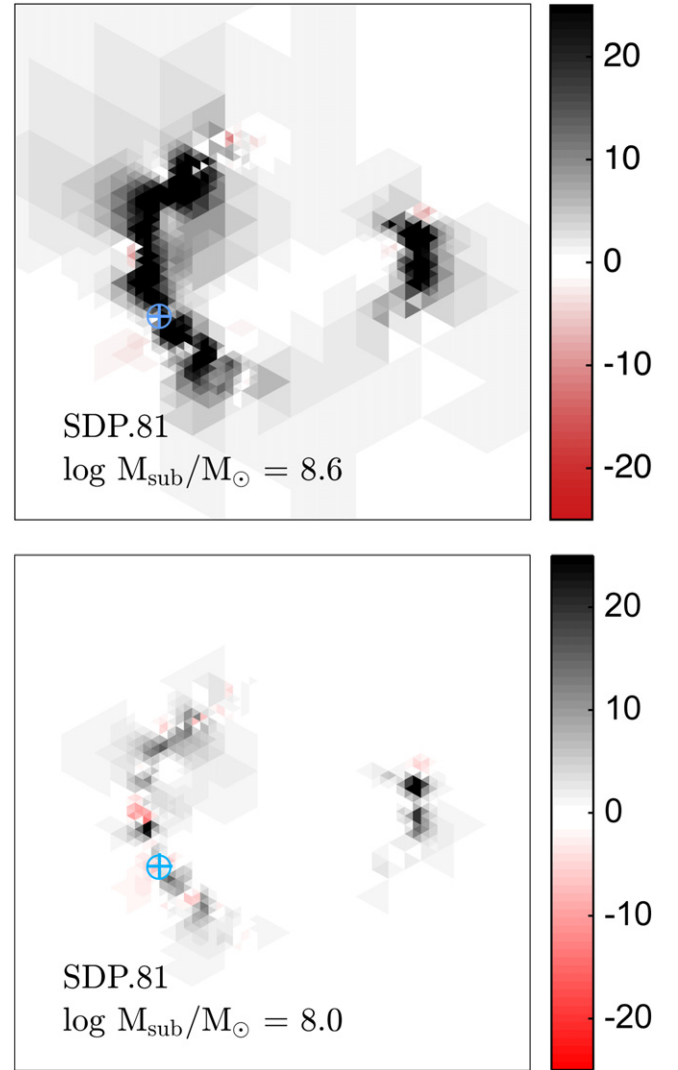

Figure 9. Search for additional substructure. The top panel shows a map of linearized $\Delta\mathcal{E}$ for a second subhalo of mass $M = 10^{8.6}M_{\odot}$, following the inclusion of one subhalo of mass $M \approx 10^{9}M_{\odot}$ at the location of the blue circled cross, using a joint analysis of bands 6 and 7. After adding the subhalo to the main lens model, no additional subhalos of this mass are found. The bottom panel shows a similar analysis for a lower mass subhalo, showing a marginal improvement of $\mathcal{E}$ at another point near the first detection.

1.7×10^{5}. Because this emission originates from regions well separated from the lensed images (far beyond the correlation length of the dirty beam), it has no model covariance with the lens parameters and we therefore neglect it in the remainder of our analysis.

5.2. Search for Additional Substructure

The ALMA observations of SDP.81 allow us to search for additional substructure besides the subhalo detected in the previous subsection. Given our lens model (including one

subhalo of $M_{\rm sub} = 10^{8.96}M_{\odot}$), we next searched for additional substructure using the linearized treatment discussed in Section 3. We repeated our search for a second subhalo, by linearly expanding about a smooth model now containing a subhalo of mass $M_{\rm sub} = 10^{8.96}M_{\odot}$. As before, we marginalize over all parameters of the smooth model, including the mass and location of the detected subhalo discussed above.

The inclusion of the subhalo in our main lens model removes any improvement to the marginalized posterior from additional subhalos of mass $M_{\rm sub} \geqslant 10^{8.6}M_{\odot}$, as illustrated in top panel of Figure 9. Instead, additional subhalos of this mass are excluded from occurring near the observed arcs. For lower subhalo masses, however, we find that there are certain locations where the addition of a second subhalo can improve the marginalized posterior. The lower panel of Figure 9 illustrates this for $M = 10^{8}M_{\odot}$. The red regions in the map depict the locations where a subhalo of this mass can improve the marginalized posterior. Based on the improvement suggested in this figure,
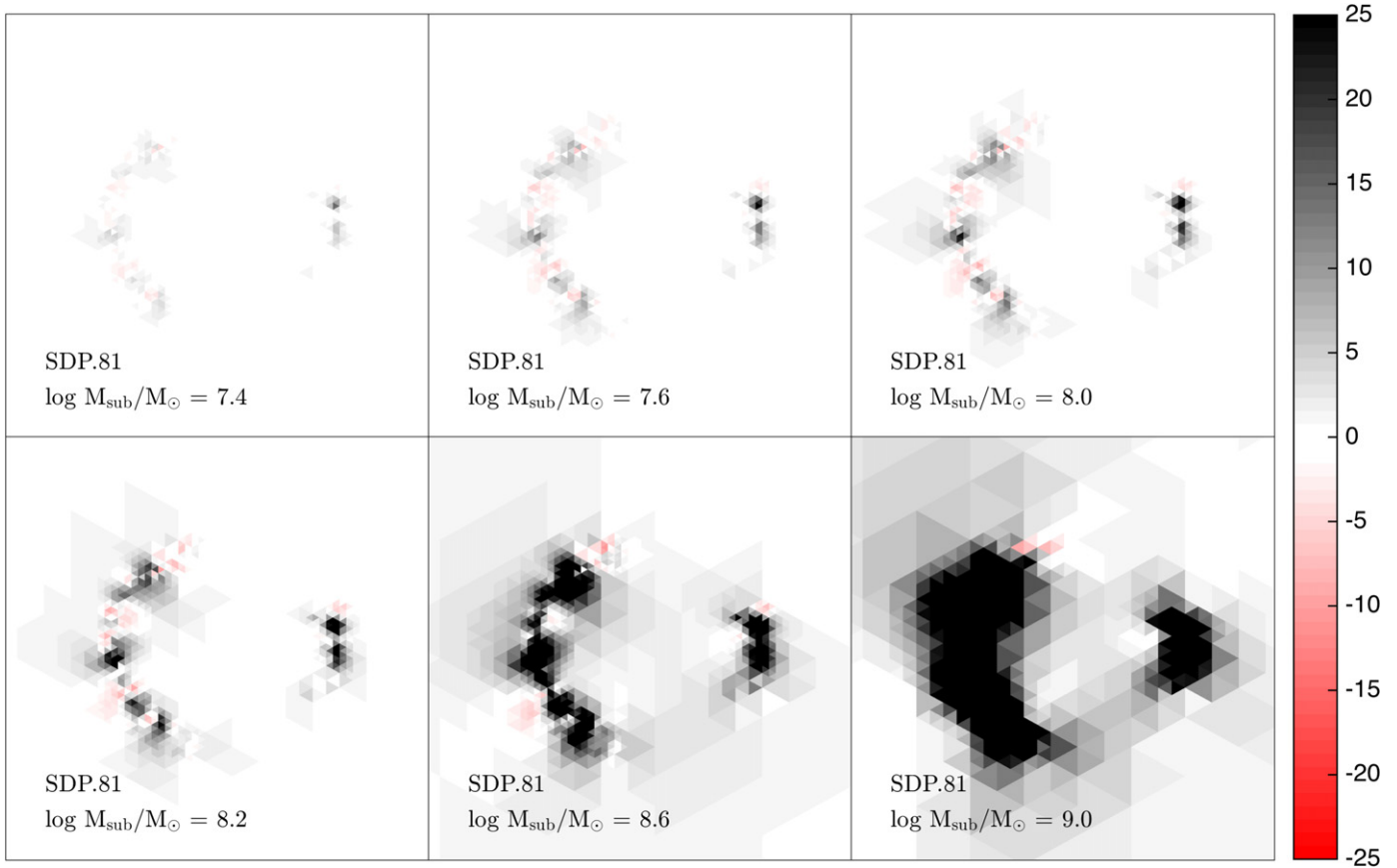
Figure 10. Maps of $\Delta\mathcal{E}$ for a 3rd subhalo. When two subhalos are included in our main lens model, no significant evidence for any additional substructure is found in the joint band 6 and 7 analysis.

we attempted adding a second subhalo to our main lens model. Our nonlinear fit found that a subhalo of mass $M \approx 10^8 M_\odot$ could improve the fit marginally, with $\Delta\mathcal{E} \approx -22$, compared with the 1-subhalo model.

We are hesitant to consider this a detection of a second object, however. First, the detection is marginal ($<5\sigma$). Also, the second object is spatially near the $M_{sub} = 10^{8.96} M_\odot$ subhalo. This suggests that the second object may possibly be an artifact of our modeling. We have assumed an extremely simple mass profile for the subhalos—a spherical tidally truncated pseudo-Jaffe model—which is unlikely to describe realistic subhalos in great detail. Small deviations in the actual mass profile compared to our model could show up as residuals that could be fitted by lower mass substructure. On the other hand, this object could be real. The close proximity to the first subhalo might arise simply because our area of sensitivity to detecting subhalos is relatively narrow. Given these ambiguities, we do not label the second object as a detection; however, we do include it in our main lens model. Inclusion of additional data, in particular the CO lines which we have not used in this analysis, may be able to determine conclusively whether a lower mass subhalo is present. To avoid biases arising from assuming a specific subhalo profile, it may be advantageous to reconstruct the substructure density field with more flexible models. For example, a pixelated substructure map (e.g., Vegetti & Koopmans 2009) would allow more freedom in the inferred substructure density field. Given the quality of the ALMA SDP.81 data set, such an approach may be warranted.

Following the inclusion of this second subhalo in our lens model, no significant evidence for additional substructure is found in the data set (see Figure 10).

5.3. Bounds on the Subhalo Mass Function

Our modeling of the mass distribution around SDP.81 tells us where subhalos appear to be present as well as where subhalos appear to be excluded, and by combining those two constraints we can derive bounds on the mean abundance $n(M)$ of DM subhalos in the vicinity of SDP.81. Specifically, we use our nonlinear mass model to tell us the number and masses of the detected subhalos as well as the area over which those subhalos are found, and we use our linearized $\Delta\mathcal{E}$ maps to determine the area on the sky where subhalos are excluded. Each piece (detections and exclusions) gives a likelihood for the mean number density of subhalos, $P[n(M)]$, and by multiplying the likelihoods derived from the detections and from exclusions we derive our full constraints on the mass function.

Calculation of the constraints from significantly detected subhalos is a straightforward application of Poisson statistics. The constraints from non-detections are slightly more complicated due to our spatially varying sensitivity. If we had a well-defined area over which subhalos were definitively excluded, then computing the Poisson statistics from that area would be straightforward. Instead, our sensitivity to substructure varies considerably over the area of interest. There are many pixels in our $\Delta\mathcal{E}$ maps that slightly disfavor the presence of subhalos, and collectively those weak constraints over many pixels can

combine to permit interesting bounds on the abundance. Below we describe the method we use to determine upper limits on the subhalo abundance from our linearized $\Delta\mathcal{E}$ maps.

We assume that the incidence of subhalos is a Poisson process with a mean projected number density $n(M)$ that is spatially uniform across the region over which we are sensitive to the effects of subhalos. In principle this allows us to use Bayes' theorem to derive the likelihood of abundance $n(M)$ given our observed data, $P(n|\text{data}) \propto P(\text{data}|n)$. To estimate $P(\text{data}|n)$, we can sum over all possible realizations of subhalo positions and masses, weighted by the Poisson probabilities for realizing each possible configuration. For clarity we will first derive the likelihood of density $n(M)$ for indistinguishable subhalos of a single mass M; we subsequently show how our expressions generalize for a spectrum of subhalo masses.

The likelihood to observe the measured data for abundance of subhalos in a narrow mass bin, $n = dn/dM \times dM$, is given by

$$
\begin{aligned}
P(\text{data}|n) = {} & P(\text{data}|0\ \text{subhalos})P(0\ \text{subhalos}|n) \\
& + P(\text{data}|1\ \text{subhalo})P(1\ \text{subhalo}|n) \\
& + P(\text{data}|2\ \text{subhalos})P(2\ \text{subhalos}|n) \\
& + \dots
\end{aligned}
\tag{20}
$$

Each term in this sum represents an integral over the relevant parameter space for 0 subhalos, 1 subhalo, and so on. The 0 subhalo term is simply the marginalized posterior of our smooth model. The 1 subhalo term is

$$
\begin{aligned}
& P(\text{data}|1\ \text{subhalo})P(1\ \text{subhalo}|n) \\
& = \int P(\text{data}|x, y)\frac{dP}{dA}(x, y|n)dxdy.
\end{aligned}
\tag{21}
$$

The two factors in the integrand on the right-hand side are given by the marginalized posterior and by Poisson statistics, respectively. The first factor, representing the marginalized posterior for a subhalo at position x, y, is $P(\text{data}|x, y) = P(\text{data}|0\ \text{subhalos}) \exp(-\Delta\mathcal{E}(x, y)/2)$. The second factor, $\frac{dP}{dA}(x, y|n)\,dxdy$ represents the Poisson probability to find 1 subhalo within area $dxdy$ at position x, y, and is independent of x, y because we assume Poisson statistics with uniform mean density. Indeed, the Poisson probability of finding 1 subhalo at location x, y, given mean number density n, is $P(1|n) = \exp(-\nu)\nu$, where the expected number $\nu = n\,dxdy$. The Poisson likelihood to observe 0 subhalos instead is $P(0|n) = \exp(-\nu)$, so that $P(1|n) = P(0|n)n\,dxdy$. Therefore, we see that

$$
\begin{aligned}
P(\text{data}|1)P(1|n) = {} & P(\text{data}|0)P(0|n) \\
& \times n \int \exp\left(-\frac{1}{2}\Delta\mathcal{E}(\boldsymbol{x})\right)d\boldsymbol{x},
\end{aligned}
\tag{22}
$$

where $\Delta\mathcal{E}(\boldsymbol{x})$ is given by Equation (16) and represents the improvement in log posterior between a model with a subhalo at location $\boldsymbol{x}$ and a model with no subhalos.

To compute the 1 subhalo, 2 subhalos, and higher terms in Equation (20), we need analogous $\Delta\mathcal{E}$ maps for all possible configurations and masses of two subhalos, three subhalos, and so on. This is an onerous calculation, and typically such integrals are performed using Monte Carlo. However, since our 1-subhalo maps do not reveal any significant detections (by construction), we can use the 1-subhalo maps to approximate

the higher-order terms. Specifically, we assume that each subhalo makes only a perturbative correction to the fit, i.e., $\Delta\mathcal{E} \ll \mathcal{E}_0$, so that the $\Delta\mathcal{E}$ from each subhalo adds linearly to the $\Delta\mathcal{E}$ from any other subhalos. For example, we assume

$$
\Delta\mathcal{E}(M_1, \boldsymbol{x}_1; M_2, \boldsymbol{x}_2) \approx \Delta\mathcal{E}(M_1, \boldsymbol{x}_1) + \Delta\mathcal{E}(M_2, \boldsymbol{x}_2).
\tag{23}
$$

This assumption of linear addition means that we can approximate the N-subhalo term in Equation (20) using an N-fold integral of our 1-subhalo $\mathcal{E}$ maps. More precisely, the assumption that each subhalo just adds linearly to $\mathcal{E}$ means that every pixel in our $\mathcal{E}$ map is assumed to be independent of every other pixel, which implies that we can construct the total likelihood for number density n by multiplying together the individual constraints on n from all the separate pixels. For a pixel of area dA with $\Delta\mathcal{E}$ from one subhalo, the constraint on n is

$$
\begin{aligned}
P(n) &\propto e^{-n\,dA}[1 + n\,dA\,e^{-\Delta\mathcal{E}/2} + \dots] \\
&= e^{-n\,dA}\sum_N [(n\,dA\,e^{-\Delta\mathcal{E}/2})^N/N!] \\
&= \exp[n\,dA(e^{-\Delta\mathcal{E}/2} - 1)].
\end{aligned}
\tag{24}
$$

Multiplying all the pixels, we obtain

$$
P(n) = P_0 \exp\left[n\,dA\sum_i (e^{-\Delta\mathcal{E}_i/2} - 1)\right].
\tag{25}
$$

We use Equation (25) to determine the constraints on the subhalo abundance from non-detections of subhalos in our $\Delta\mathcal{E}$ maps. The assumption of linearity underlying this equation clearly becomes invalid in the limit of large numbers of subhalos that in combination can produce $\Delta\mathcal{E} \sim \mathcal{E}_0$, or where subhalos overlap with each other. Because our 1-subhalo maps do not reveal any significant detections of subhalos, such configurations are probabilistically disfavored, and hence should not lead to significant errors in our approximation. In cases where subhalos do overlap, our tests indicate that our assumption of linear addition tends to underestimate the decrease in posterior probability density, suggesting that our bounds below are (slightly) conservative.

So far, our discussion has focused on constraining the number density of identical subhalos in a narrow mass bin dM. It is straightforward to generalize Equation (25) to allow for distinguishable subhalos of different masses. Repeating the same argument used to derive Equation (25) when there are subhalos of varying masses, it is straightforward to see that the likelihood for a mass function dn/dM factorizes into a product of terms from each mass bin,

$$
\begin{aligned}
P\left(\frac{dn}{dM}\right) &= P_0 \prod_j \exp\left[n_j\,dA\sum_i (e^{-\Delta\mathcal{E}_{ji}/2} - 1)\right] \\
&= P_0 \exp\left[\sum_{i,j}\frac{dn}{dM}(M_j)dM\,dA(e^{-\Delta\mathcal{E}_{ji}/2} - 1)\right]
\end{aligned}
\tag{26}
$$

where i runs over angular pixels on the sky, j runs over subhalo mass bins, and we have written the number density of subhalos in bin j as $n_j = dn/dM(M_j)dM$.

To summarize, Equation (26) tells us the effective area over which subhalos are excluded. Our main lens model tells us the number of subhalos that were detected and the area over which they were found to occur. Combining those two measurements,
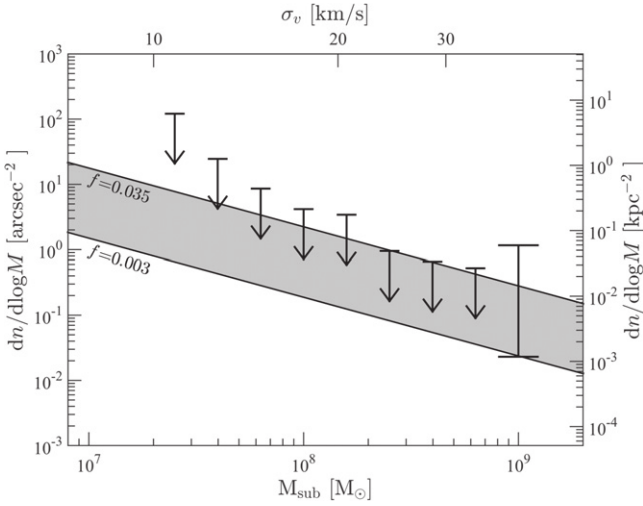
Figure 11. Error bars indicate the 95% confidence limits on the projected differential number density of subhalos around SDP.81 derived using the non-detection regions shown in Figure 10 and the detection of the $10^9 M_\odot$ subhalo. For comparison, the shaded band shows the 90% confidence region from Dalal & Kochanek (2002).
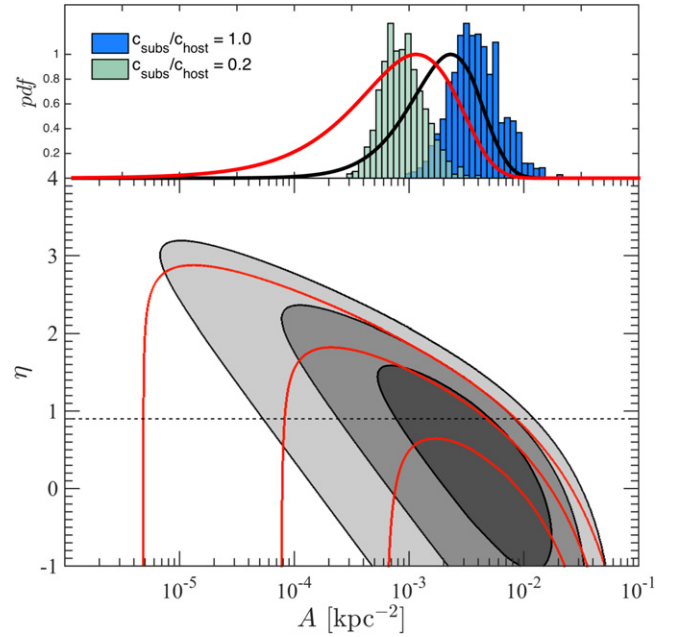

Figure 12. Limits on the normalization (A) and slope (η) of the mass function $dn/d \log M = A(M/M_{\mathrm{pivot}})^{-\eta}$ using the bounds in Figure 11. Here we use $M_{\mathrm{pivot}} = 10^9 M_\odot$. The gray contours show constraints derived using Equation (26), while the red contours show how the constraints change if we neglect the marginally detected subhalo with $M \approx 10^8 M_\odot$. The top panel shows the probability at $\eta = 0.9$. The red and black curves simply show a slice of the probability of the lower panel at $\eta = 0.9$. For comparison, the histograms show the distribution of A using assumptions based on ΛCDM simulations assuming two different values of $c_{\mathrm{subs}}/c_{\mathrm{host}}$, which are intended to be representative. These values assume $\eta = 0.9$ and a distribution of host halo masses and concentrations given by abundance matching. See Section 6 for details.

we derive Poisson constraints on the underlying subhalo abundance.

Figure 11 shows the resulting constraints on the differential subhalo mass function, $dn/d \log M (M_{\mathrm{sub}})$, derived from the maps of $\Delta \mathcal{E}$ shown in Figure 10. In mass bins where no subhalos were detected, the downward arrows indicate 95% upper limits. For the mass bin at $M_{\mathrm{sub}} = 10^9 M_\odot$ where we have a detected subhalo, the central 95% confidence region is 0.012 arcsec$^{-2} < n < 0.2$ arcsec^{-2}. If we instead define the confidence region in terms of levels of equal posterior encompassing 95% of the posterior, we obtain 0.003 arcsec$^{-2} < n < 0.1806$ arcsec^{-2}. The reason these two ranges are somewhat different is that the likelihood is asymmetric.

Combining the bounds from the different mass bins, we can derive constraints on the subhalo mass function using Equation (26). We describe the mass function using a simple parametrization, $dn/d \log M = A(M/M_{\mathrm{pivot}})^{-\eta}$, and show in Figure 12 the constraints on these parameters. In the next section we compare these constraints to the amount of substructure expected for lens galaxies like SDP.81 in ΛCDM cosmologies.

6. COMPARISON TO ΛCDM PREDICTIONS

In this section, we compare the constraints on the subhalo abundance in SDP.81 found above, with predictions from ΛCDM simulations, and also discuss the neighboring environment of this system. To predict the subhalo mass function down to the small masses probed while fully accounting for the halo-to-halo scatter, we follow the methodology presented in Mao et al. (2015), which captures the dominant source of the halo-to-halo scatter by considering both mass and concentration of host halos. The model is able to reproduce the subhalo abundance found in high-resolution zoom-in simulations (e.g., Xu et al. 2015) as well as larger statistical samples of halos.

We assume the cumulative subhalo mass function has the form of

$$\langle n(>M_{\mathrm{sub}}) \rangle = \left(\frac{M_{\mathrm{sub}}}{M_0} \right)^{-\eta} - \left(\frac{M_{\mathrm{host}}}{M_0} \right)^{-\eta}, \qquad (27)$$

where M_{host} is the host halo mass, M_0 and η are the normalization and the log–log slope, respectively, of the subhalo mass function. We then use ΛCDM simulations to calibrate the relation between the parameter M_0 and the mass and concentration of the host halo. To calibrate this relation, we use the same set of high-resolution zoom-in simulations described in Mao et al. (2015) with the addition of a very high-resolution cosmological box, (4096^3 particles in a 400 Mpc/h box, ds14_i) from the Dark Sky Simulations (S-killman et al. 2014).[15] This calibration is done by: (1) assuming a constant log–log slope (η); (2) then finding the best-fit M_0 for each host halo in the simulations; and (3) finally, finding the best-fit values of (α, β, γ) for all host halos in

$$M_0 = \alpha M_{\mathrm{host}}^\beta c_{\mathrm{host}}^\gamma. \qquad (28)$$

With this model we can then predict the subhalo mass function given the host halo mass and concentration and the log–log slope.

The subhalo abundance predicted in the procedure described above is for all subhalos within the virial radius of the host halo. To convert our prediction to the relevant quantity probed

---

[15] http://darksky.slac.stanford.edu

by strong lensing measurements, we need to assume a spatial distribution for the subhalos. Here we make three simplifying assumptions: (1) the subhalo spatial distribution is *independent* from the subhalo mass function (i.e., subhalos of different mass halos have the same spatial distribution); (2) the angular distribution of subhalos is isotropic (see, however, Nierenberg et al. 2011); and (3) the radial distribution of subhalos within their host halos follows an NFW profile with a characteristic concentration c_{subs}. In other words, we assume the subhalo abundance factorizes into a mass dependence and radial dependence, $n(M, r) = n(M) f(r)$, where the radial dependence $f(r)$ is an NFW profile of concentration c_{subs}.

To predict the projected abundance of substructure, our model requires a prescription for the concentration of the subhalo distribution, c_{subs}. In ΛCDM simulations, generally the radial distribution of subhalos is less centrally concentrated than the DM distribution of the host halo (i.e., $c_{subs}/c_{host} < 1$) (e.g., Nagai & Kravtsov 2005; Gao et al. 2012), and at small radii the subhalo distribution may become shallower than an NFW profile (e.g., Xu et al. 2015). Observational results for real galaxies are less clear: some are consistent with $c_{subs}/c_{host} \simeq 1$ (e.g., Guo et al. 2012; Yniguez et al. 2014), while others imply that galaxies are less concentrated (e.g., Hansen et al. 2005) than the total mass distribution in their hosts. Also note that our assumption of spherical symmetry might lead us to underestimate the average substructure abundance around lenses, since strong lenses are preferentially viewed along the major axis of their host halos (Hennawi et al. 2007; Rozo et al. 2007).

Given the uncertainty in predictions for c_{subs}, we treat it as a free parameter along with other parameters describing the lens halo: the host halo mass and concentration (M_{host}, c_{host}), and the log–log slope (η) of the subhalo mass function. Using these model ingredients, we can predict $dn/d\log M$ projected at the Einstein radius. The histograms in the top panel of Figure 12 show an example, the distribution of A, i.e., $dn/d\log M$ at $M = 10^9 M_\odot$ computed with this model. For this figure we assume the mass function slope is $\eta = 0.9$ and we show two possible values for the subhalo concentration, $c_{subs}/c_{host} = 0.2$ and 1.0, which should span the range of uncertainty described above. For the other two parameters we marginalize over possible values of the host halo mass and concentration using the following prior. We first assign galaxy luminosity to DM halos and subhalos with the abundance matching technique (e.g., Conroy et al. 2006; Reddick et al. 2013), and find that the joint distribution of mass and concentration corresponding to the luminosity of the lens galaxy, which is $M_r = -21.88 \pm 0.015$ using the SDSS DR10 magnitudes, k-corrected to $z = 0.1$ using a red galaxy template. For the abundance matching procedure, we assume the AGES luminosity function (Kochanek et al. 2012) and use the maximal circular velocity (v_{max}) of the halo at its peak value along its trajectory as the matching proxy and also apply a constant scatter of 0.2 dex on the luminosity. This gives us the allowed spread of halo mass and concentration typical for galaxies of luminosity similar to the lensing galaxy in SDP.81. The typical halo mass we found with this procedure is roughly $8.7 \times 10^{12} M_\odot h^{-1}$ with an approximate uncertainty of a factor of 3.5.

Comparing the lines and the histograms in the top panel of Figure 12, we can see that substructure limits from SDP.81 are currently consistent with theoretical predictions. The result does hint at a lower normalization of $dn/d\log M$, or equivalently, a smaller value of c_{subs}/c_{host}. However, the four parameters used here are highly degenerate because all of them affect the normalization of the projected central density in similar fashions. Figure 13 illustrates this degeneracy by varying only one of the four parameters in the model at a time and shows how each of these four parameters affects the predicted projected density over reasonable ranges in parameters space. Given this degeneracy, at this stage it is difficult to jointly constrain these parameters from the observed limits. The comparison shown in Figure 13 also demonstrates, however, that the bounds from this Science Verification data set are already in an interesting regime that can rule out some portion of the parameter space. We also note that our estimates for substructure around strong lensing galaxies appear to be consistent with independent estimates from other high-resolution simulations (Fiacconi et al. 2016).

The environment of the host halo can also affect the subhalo population. Fortuitously, SDP.81 is in the SDSS footprint, allowing us to examine its immediate environment. This system appears to be in close proximity to a massive galaxy cluster in the SDSS DR8 RedMaPPer cluster catalog (Rykoff et al. 2014). Here we use RedMaPPer v6.3 (Rykoff et al 2016, in preparation). In the RedMaPPer catalog, this galaxy is a member (with 87% probability) of a more massive system, with richness $\lambda = 32$, where λ corresponds to the number of red sequence galaxies brighter than $0.4 L_\star$. Assuming the mass–richness relationship of Rykoff et al. (2014), this corresponds to a mass of $\sim 2.6 \times 10^{14} M_\odot$. According to the cluster catalog, SDP.81 is a member galaxy of this system, 640 kpc h^{-1} away from the most likely central galaxy (in projection).

The proximity of SDP.81 to a galaxy cluster has the potential to cloud our predictions for its abundance of substructure. In principle, small subhalos unbound to SDP.81 but within the cluster could project into the strong lensing region. For the current configuration, we have calculated the expected number of subhalos from such a cluster at the measured distance (using the same model described in the previous section) and find that these subhalos are subdominant to those expected from the halo of SDP.81. However, this underscores the importance of determining reliable estimates of lens halo masses and nearby environments before deriving any bounds on cosmological models from observed lensing systems.

7. DISCUSSION

In this paper, we present a method to analyze interferometric measurements of strong gravitational lenses to constrain the abundance of DM substructure. We apply this method to the ALMA Science Verification observations of the lens system SDP.81 and we report the detection of a subhalo with mass $M_{sub} \approx 10^9 M_\odot$. We also find hints for additional substructure at lower mass, but defer a detailed analysis of additional substructure for future work. We compare our measurements of substructure abundance with previous measurements and to theoretical predictions from CDM, and find that our results are consistent with both.

Although the subhalo analysis presented in this paper only used the continuum data, the method is directly applicable to observations of molecular lines. For example, in this work our joint analysis of bands 6 and 7 treats the two bands as two distinct frequency channels. The analysis of molecular lines, therefore, only differs in the larger number of frequency
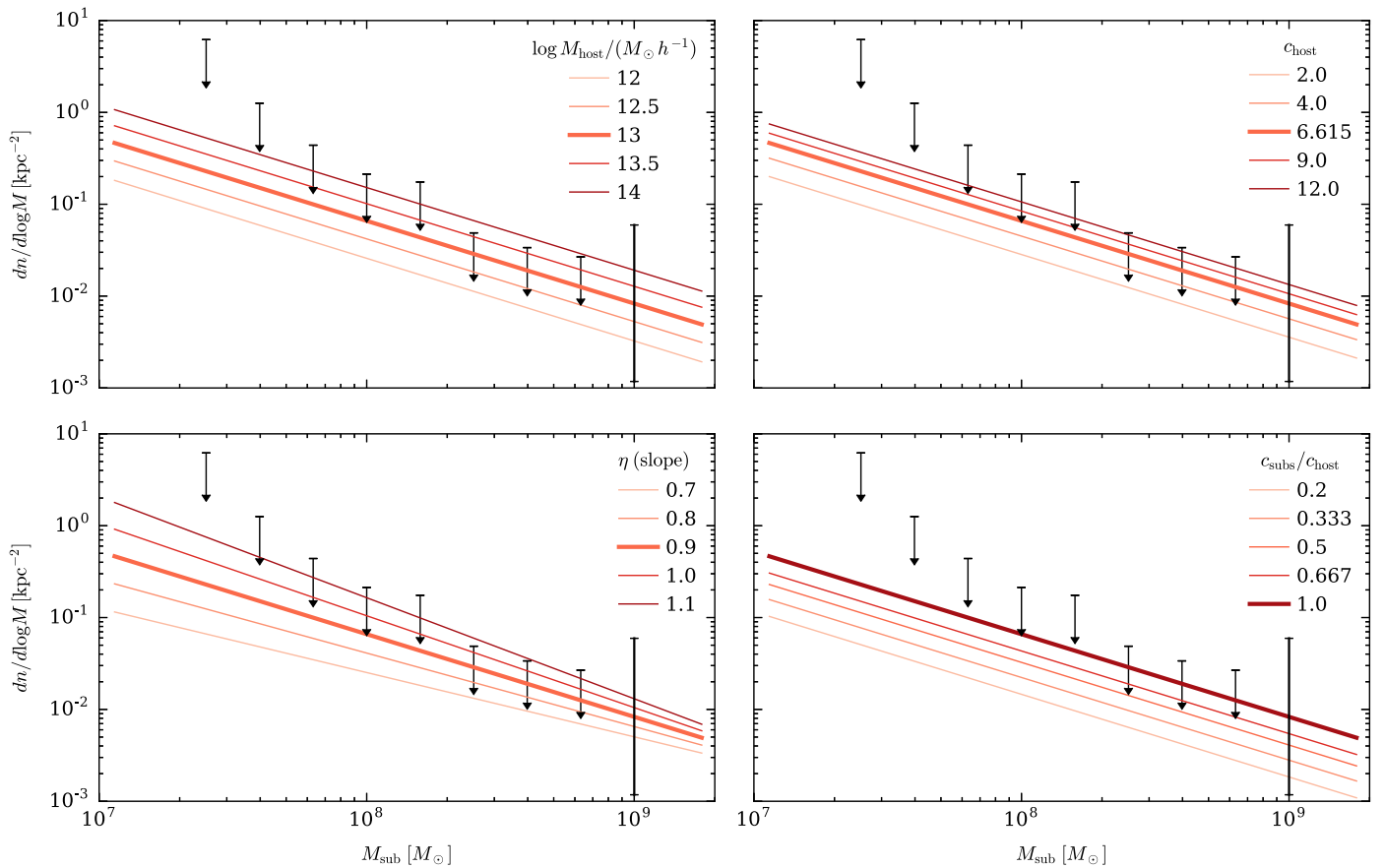
Figure 13. Comparison between the 95% confidence limits (the black squares) on the projected density of substructures, $dn/d\log M$, and the predictions (the lines) for various properties of the host halo and its substructure distribution. The thick line in each panel assumes $M_{\mathrm{host}} = 10^{13} M_{\odot}\,h^{-1}$, a median halo concentration for that mass ($c_{\mathrm{host}} = 6.615$), a typical log–log slope ($\eta = 0.9$) for the subhalo mass function, and that the radial distribution of subhalos matches that of dark matter ($c_{\mathrm{subs}}/c_{\mathrm{host}} = 1$). Each panel varies only *one* of the four assumptions listed above. The varying assumption is shown on the upper right corner of each panel.

channels used. The computational cost of such an analysis, however, would be higher. Given the strong hints of lower mass substructure that we find in the continuum data, an analysis of the line data would appear to be quite worthwhile.

The simulations and the analysis of independently generated mock data show that our pipeline can successfully quantify the lensing effects of $\gtrsim 10^7\ M_{\odot}$ subhalos in an interesting regime where CDM and some of the most popular WDM models have significantly different predictions (Li et al. 2015). This framework is designed to be able to marginalize over many nuisance parameters to avoid false detections. We studied the effects of complex source morphologies, source priors, visibility binning, antenna phase errors, pixel sizes, and grid width, finding that the current pipeline is a reliable tool to quantify the lensing effects of these subhalos, with systematic errors that are below our current statistical uncertainties. As we push to lower the subhalo masses and lower significance detection or non-detection regimes, however, we will need to quantify more subtle systematic effects. In particular, more detailed studies of weaker interferometric data corruption effects should be carefully studied. The effects of decorrelation, visibility smearing due to frequency averaging, choice of parameterization for antenna phase error corrections, and antenna amplitude errors are among these effects.

In this work we have used a singular elliptical power-law surface density profile for the macrolens. In SDP.81, as in many (but not all) strong lenses, the lensed images of

background sources span a narrow range in radius from the center of the lens and hence are only sensitive to the local amplitude and radial slope of the density as opposed to the full radial dependence. For this reason, this system can safely be modeled as a power law even though on larger scales or smaller scales there might be deviations from single power-law behavior. For example, the addition of a constant mass sheet breaks power-law behavior, but it would not necessarily be detectable from this lensing data, and if modeled would be degenerate with the radial slope of the macro model (e.g., Schneider & Sluse 2013). This mass-sheet degeneracy has been shown not to affect the inference of the presence of density perturbations (Dobler & Keeton 2006), however the inferred masses of the subhalos could be affected by this degeneracy by a fractional uncertainty of order κ_{sheet}. In standard CDM cosmologies external mass sheets on arcsecond scales are expected to contribute at few percent levels (Dalal et al. 2003), below the statistical uncertainties on our derived mass. Similarly, as noted above, SDP.81 appears to occur relatively near a massive galaxy cluster which should generate a smooth convergence and shear at the location of the lensed images, but once again our estimates for this external convergence from the cluster fall well below the fractional uncertainty in the derived subhalo mass.

As Figures 10 and 11 illustrate, the SDP.81 data set loses sensitivity for detecting individual subhalos at masses $\lesssim 10^7 M_{\odot}$. Large numbers of subhalos at low masses are

expected in all CDM cosmologies and the deflections from this population of objects combine to form an effectively stochastic field. Even though the objects generating this random field cannot be individually detected, the collective effects of the population may be significantly detected. For example, Hezaveh et al. (2014) showed that the power spectrum of density fluctuations of low-mass substructure may be accurately measured using observations of lenses like SDP.81, allowing us to probe the subhalo mass function below our nominal detection limits for individual subhalos.

Recently, Inoue et al. (2016) published an independent substructure lensing analysis of SDP.81, and it is interesting to compare their conclusions to ours. Inoue et al. analyze CLEAN images from continuum band 7 as well as CO 8-7 line data from band 6, and report evidence for substructure at a similar location to our reported detection. However, the properties of the substructure reported by Inoue et al. are significantly different than what we find. In particular, it appears that their results require the presence of a compact, highly underdense region next to one of the arcs. Underdensities are rarely compact in standard cosmologies, so this result appears puzzling. Our analysis does not confirm this finding: we find that compact regions of overdensity explain the SDP.81 data far better than adding compact regions of underdensity (or negative density). One possible origin for the discrepancy, as noted in Section 4, is that analyses of CLEANed images are subject to systematic biases arising from phase errors, which can lead to a host of spurious artifacts in substructure reconstructions (see Figure 2). Given the subtlety of the lensing effects of low-mass subhalos, we recommend that substructure analyses operate on the visibilities and thereby fully extract the information encoded in the interferometric measurements.

In summary, the Science Verification observations of SDP.81 demonstrate the potential of ALMA for probing DM structures. Our joint analysis of the bands 6 and 7 data detects a $M = 10^{8.96 \pm 0.12} M_\odot$ subhalo with a significance of 6.9σ and produces substructure bounds that are consistent with previous lensing measurements of other systems (Dalal & Kochanek 2002; Vegetti et al. 2014) and also consistent with theoretical expectations as described in Section 6. However, the constraints from this one single lens are already interesting, near the abundances expected for halos of this mass. More importantly, the analysis shows that the ALMA data are sensitive to low-mass substructure in a regime that can constrain the properties of DM models. Larger samples of similar lenses have the potential to put tight constraints on the mass function of DM substructures. Fortunately, large samples of similar lenses are already known from existing submillimeter surveys, suggesting that future ALMA observations have the potential to significantly advance our understanding of the abundance of DM substructure.

REFERENCES

Abazajian, K. 2006, PhRvD, 73, 063513
ALMA Partnership, Fomalont, E. B., Vlahakis, C., et al. 2015a, ApJL, 808, L1
ALMA Partnership, Vlahakis, C., Hunter, T. R., et al. 2015b, ApJL, 808, L4
Anderson, L., Aubourg, É., Bailey, S., et al. 2014, MNRAS, 441, 24
Barkana, R. 1998, ApJ, 502, 531
Bode, P., Ostriker, J. P., & Turok, N. 2001, ApJ, 556, 93
Bussmann, R. S., Pérez-Fournon, I., Amber, S., et al. 2013, ApJ, 779, 25
Carlberg, R. G. 2009, ApJL, 705, L223
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Cyr-Racine, F.-Y., Moustakas, L. A., Keeton, C. R., Sigurdson, K., & Gilman, D. A. 2015, arXiv:1506.01724
Dalal, N., Holz, D. E., Chen, X., & Frieman, J. A. 2003, ApJL, 585, L11
Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25
Diemand, J., Kuhlen, M., Madau, P., et al. 2008, Natur, 454, 735
Dobler, G., & Keeton, C. R. 2006, MNRAS, 365, 1243
Dodelson, S. 2003, Modern Cosmology (New York: Academic)
Dye, S., Furlanetto, C., Swinbank, A. M., et al. 2015, MNRAS, 452, 2258
Dye, S., Negrello, M., Hopwood, R., et al. 2014, MNRAS, 440, 2013
Erkal, D., & Belokurov, V. 2015a, MNRAS, 450, 1136
Erkal, D., & Belokurov, V. 2015b, MNRAS, 454, 3542
Fiacconi, D., Madau, P., Potter, D., & Stadel, J. 2016, ApJ, in press (arXiv:1602.03526)
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gao, L., Navarro, J. F., Frenk, C. S., et al. 2012, MNRAS, 425, 2169
Goodman, J., & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65
Guo, Q., Cole, S., Eke, V., & Frenk, C. 2012, MNRAS, 427, 428

Hansen, S. M., McKay, T. A., Wechsler, R. H., et al. 2005, ApJ, 633, 122
Hatsukade, B., Tamura, Y., Iono, D., et al. 2015, PASJ, 67, 93
Hennawi, J. F., Dalal, N., Bode, P., & Ostriker, J. P. 2007, ApJ, 654, 714
Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Hezaveh, Y., Dalal, N., Holder, G., et al. 2014, ApJ, submitted (arXiv:1403.2720)
Hezaveh, Y., Dalal, N., Holder, G., Kuhlen, M., & Marrone, D. 2013a, ApJ, 767, 9
Hezaveh, Y. D., Marrone, D. P., Fassnacht, C. D., et al. 2013b, ApJ, 767, 132
Inoue, K. T., Minezaki, T., Matsushita, S., & Chiba, M. 2016, MNRAS, 457, 2936
Keeton, C. R., Burles, S., Schechter, P. L., & Wambsganss, J. 2006, ApJ, 639, 1
Khoury, J. 2015, PhRvD, 91, 024022
Kochanek, C. S., & Dalal, N. 2004, ApJ, 610, 69
Kochanek, C. S., Eisenstein, D. J., Cool, R. J., et al. 2012, ApJS, 200, 8
Koopmans, L. V. E. 2005, MNRAS, 363, 1136
Kravtsov, A. 2010, AdAst, 2010, 8
Li, R., Frenk, C. S., Cole, S., et al. 2015, MNRAS, in press (arXiv:1512.06507)
Mao, S., & Schneider, P. 1998, MNRAS, 295, 587
Mao, Y.-Y., Williamson, M., & Wechsler, R. H. 2015, ApJ, 810, 21
Markovič, K., & Viel, M. 2014, PASA, 31, 6
Marsden, D., Gralla, M., Marriage, T. A., et al. 2014, MNRAS, 439, 1556
Metcalf, R. B., & Madau, P. 2001, ApJ, 563, 9
Moustakas, L. A., & Metcalf, R. B. 2003, MNRAS, 339, 607
Muñoz, J. A., Kochanek, C. S., & Keeton, C. R. 2001, ApJ, 558, 657
Nagai, D., & Kravtsov, A. V. 2005, ApJ, 618, 557
Navarro, J. F., Ludlow, A., Springel, V., et al. 2010, MNRAS, 402, 21
Negrello, M., Hopwood, R., De Zotti, G., et al. 2010, Sci, 330, 800
Nierenberg, A. M., Auger, M. W., Treu, T., Marshall, P. J., & Fassnacht, C. D. 2011, ApJ, 731, 44
Nierenberg, A. M., Treu, T., Wright, S. A., Fassnacht, C. D., & Auger, M. W. 2014, MNRAS, 442, 2434
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, arXiv:1502.01589
Poulson, J., Engquist, B., Li, S., & Ying, L. 2013a, SIAM Journal on Scientific Computing, 35, C194

Poulson, J., Marker, B., van de Geijn, R. A., Hammond, J. R., & Romero, N. A. 2013b, ACM Trans. Math. Softw., 39, 13
Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, ApJ, 771, 30
Rozo, E., Chen, J., & Zentner, A. R. 2007, arXiv:0710.1683
Rybak, M., McKean, J. P., Vegetti, S., Andreani, P., & White, S. D. M. 2015a, MNRAS, in press (arXiv:1503.02025)
Rybak, M., Vegetti, S., McKean, J. P., Andreani, P., & White, S. D. M. 2015b, MNRAS, 453, L26
Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104
Schneider, P., & Sluse, D. 2013, A&A, 559, A37
Seljak, U., Makarov, A., McDonald, P., & Trac, H. 2006, PhRvL, 97, 191303
Sherman, J., & Morrison, W. 1949, Ann. Math. Statist., 20, 620
Siegal-Gaskins, J. M., & Valluri, M. 2008, ApJ, 681, 40
Skillman, S. W., Warren, M. S., Turk, M. J., et al. 2014, arXiv:1407.2600
Spergel, D. N., & Steinhardt, P. J. 2000, PhRvL, 84, 3760
Stadel, J., Potter, D., Moore, B., et al. 2009, MNRAS, 398, L21
Suyu, S. H., & Halkola, A. 2010, A&A, 524, A94
Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, MNRAS, 371, 983
Tamura, Y., Oguri, M., Iono, D., et al. 2015, arXiv:1503.07605
Towns, J., Cockerill, T., Dahan, M., et al. 2014, CSE, 16, 62
Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 392, 945
Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017
Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969
Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Natur, 481, 341
Vieira, J. D., Marrone, D. P., Chapman, S. C., et al. 2013, Natur, 495, 344
Wardlow, J. L., Cooray, A., De Bernardis, F., et al. 2013, ApJ, 762, 59
Warren, S. J., & Dye, S. 2003, ApJ, 590, 673
Wong, K. C., Suyu, S. H., & Matsushita, S. 2015, ApJ, 811, 115
Xu, D., Sluse, D., Gao, L., et al. 2015, MNRAS, 447, 3189
Xu, D., Sluse, D., Schneider, P., et al. 2016, MNRAS, 456, 739
Yniguez, B., Garrison-Kimmel, S., Boylan-Kolchin, M., & Bullock, J. S. 2014, MNRAS, 439, 73