# DISSECTING THE GRAVITATIONAL LENS B1608+656. II. PRECISION MEASUREMENTS OF THE HUBBLE CONSTANT, SPATIAL CURVATURE, AND THE DARK ENERGY EQUATION OF STATE[*]

S. H. SUYU[1], P. J. MARSHALL[2,3], M. W. AUGER[3,4], S. HILBERT[1,5], R. D. BLANDFORD[2], L. V. E. KOOPMANS[6], C. D. FASSNACHT[4], AND T. TREU[3,7]

*Draft version October 15, 2009*

## ABSTRACT

Strong gravitational lens systems with measured time delays between the multiple images provide a method for measuring the "time-delay distance" to the lens, and thus the Hubble constant. We present a Bayesian analysis of the strong gravitational lens system B1608+656, incorporating (i) new, deep *Hubble Space Telescope* (*HST*) observations, (ii) a new velocity dispersion measurement of $260 \pm 15 \, \mathrm{km \, s^{-1}}$ for the primary lens galaxy, and (iii) an updated study of the lens' environment. Our analysis of the *HST* images takes into account the extended source surface brightness, and the dust extinction and optical emission by the interacting lens galaxies. When modeling the stellar dynamics of the primary lens galaxy, the lensing effect, and the environment of the lens, we explicitly include the total mass distribution profile logarithmic slope $\gamma'$ and the external convergence $\kappa_{\mathrm{ext}}$; we marginalize over these parameters, assigning well-motivated priors for them, and so turn the major systematic errors into statistical ones. The *HST* images provide one such prior, constraining the lens mass density profile logarithmic slope to be $\gamma' = 2.08 \pm 0.03$; a combination of numerical simulations and photometric observations of the B1608+656 field provides an estimate of the prior for $\kappa_{\mathrm{ext}}$: $0.10^{+0.08}_{-0.05}$. This latter distribution dominates the final uncertainty on $H_0$. Fixing the cosmological parameters at $\Omega_{\mathrm{m}} = 0.3$, $\Omega_\Lambda = 0.7$, and $w = -1$ in order to compare with previous work on this system, we find $H_0 = 70.6^{+3.1}_{-3.1} \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$. The new data provide an increase in precision of more than a factor of two, even including the marginalization over $\kappa_{\mathrm{ext}}$. Relaxing the prior probability density function for the cosmological parameters to that derived from the WMAP 5-year data set, we find that the B1608+656 data set breaks the degeneracy between $\Omega_{\mathrm{m}}$ and $\Omega_\Lambda$ at $w = -1$ and constrains the curvature parameter to be $-0.031 < \Omega_{\mathrm{k}} < 0.009$ (95% CL), a level of precision comparable to that afforded by the current Type Ia SNe sample. Asserting a flat spatial geometry, we find that, in combination with WMAP, $H_0 = 69.7^{+4.9}_{-5.0} \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ and $w = -0.94^{+0.17}_{-0.19}$ (68% CL), suggesting that the observations of B1608+656 constrain $w$ as tightly as do the current Baryon Acoustic Oscillation data.

*Subject headings:* gravitational lensing: general — gravitational lensing: individual (B1608+656) — cosmology: observations — methods: data analysis – distance scale

## 1. INTRODUCTION

The Hubble constant ($H_0$, measured in units of $\mathrm{km \, s^{-1} \, Mpc^{-1}}$) is one of the key cosmological parameters since it sets the present age, size, and critical density of the Universe. Furthermore, a measurement of $H_0$ to better than a few percent precision provides the single most useful complement to results obtained from studies of the cosmic microwave background (CMB) for dark energy studies (e.g. Hu 2005; Riess et al. 2009).

Methods for measuring the Hubble constant include Type Ia supernovae (SNe Ia) (e.g. Riess et al. 2005; Sandage et al. 2006), the Sunyaev-Zel'dovich effect (e.g. Birkinshaw 1999; Schmidt et al. 2004; Bonamente et al. 2006), the maser distances (e.g. Herrnstein et al. 1999; Macri et al. 2006), and CMB observations (e.g. Komatsu et al. 2009). Perhaps the two most well-known recent measurements come from the *Hubble Space Telescope* (*HST*) Key Project (KP) (Freedman et al. 2001) and the Wilkinson Microwave Anisotropy Probe (WMAP). The *HST* KP measurement of $H_0$ is based on secondary distance indicators (including Type Ia supernovae, Tully-Fisher, surface brightness fluctuations, Type II supernovae, and the fundamental plane) that are calibrated using Cepheid distances to nearby galaxies with a zero point in the Large Magellanic Cloud (LMC). The resulting Hubble constant is $72 \pm 8 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ (Freedman et al. 2001). We note that the largest contributor to the systematic error from the distance ladder of which this measurement depends is the metallicity dependence of the Cepheid period-luminosity relation. More recently, Riess et al. (2009) addressed some of these systematic effects with an improved differential distance ladder using Cepheids, SNe Ia, and the maser galaxy NGC 4258, finding $H_0 = 74.2 \pm 3.6 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, a 5% local measurement of Hubble's constant.

[*] Based in part on observations made with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program GO-10158.
[1] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[2] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, PO Box 20450, MS 29, Stanford, CA 94309, USA
[3] Department of Physics, University of California, Santa Barbara, CA 93106-9530, USA
[4] Department of Physics, University of California at Davis, 1 Shields Avenue, Davis, CA 95616, USA
[5] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85741 Garching, Germany
[6] Kapteyn Astronomical Institute, P.O. Box 800, 9700AV Groningen, The Netherlands
[7] Sloan Fellow, Packard Fellow
Electronic address: suyu@astro.uni-bonn.de

The five year measurement made using WMAP temperature and polarization data is $H_0 = 71.9^{+2.6}_{-2.7}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Dunkley et al. 2009), under the assumption that the Universe is flat and that the dark energy is described by a cosmological constant (with equation of state parameter $w = -1$). The uncertainty in $H_0$ increases markedly if either of these two assumptions is relaxed, due to degeneracies with other cosmological parameters. For examples, WMAP gives $H_0 \sim 50\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ without the flatness assumption, and $H_0 = 74^{+15}_{-14}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ for a flat Universe with time-independent $w$ not fixed at $w = -1$. As $H_0$ is such an important parameter, it is essential to measure it using multiple methods. In this paper, we use a single strong gravitational lens as an independent probe of $H_0$, and explore its systematic errors and relations with other cosmological parameters to provide guidance for future studies. We will show that the single lens is competitive with those of the best current cosmographic probes. Given the current progress in measuring time delays (e.g., Vuissoz et al. 2007, 2008; Paraficz et al. 2009), the methodology in this paper should lead to substantial advances when applied to samples of gravitational lenses.

Strong gravitational lensing occurs when a source galaxy is lensed into multiple images by a galaxy lying along its line of sight. The principle of using strong gravitational lens systems with time-variable sources to measure the Hubble constant is well understood (e.g. Refsdal 1964, Schneider, Kochanek, & Wambsganss 2006). The relative time delays between the multiple images are inversely proportional to $H_0$ via a combination of angular diameter distances and depend on the lens potential (mass) distribution. We refer to the combination of angular diameter distances as the "time-delay distance". By measuring the time delays and modeling the lens potential, one can infer the value for the time-delay distance; this distance-like quantity is primarily sensitive to $H_0$ but depends also on other cosmological parameters which must be factored into the analysis. The direct measurement of the time-delay distance means that gravitational lensing is independent of distance ladders.

Despite being an elegant method, gravitational lensing has its limitations. Perhaps the most well-known is the "mass-sheet degeneracy" between $H_0$ and external convergence (Falco, Gorenstein, & Shapiro 1985). There is also a degeneracy between $H_0$ and the slope of the lens mass distribution, especially for lenses where the configuration is nearly symmetric (e.g. Wucknitz 2002). In such cases, the image positions are at approximately the same radial distance from the lens center and so the slope is poorly constrained. In both cases the remedy is to provide more information. Modeling the mass environment of the lens can, in principle, independently constrain the external convergence (e.g., Keeton & Zaludoff 2004; Fassnacht et al. 2006a; Blandford et al. in preparation); likewise, lens galaxy stellar velocity dispersion measurements (e.g., Grogin & Narayan 1996a,b; Tonry & Franx 1999; Koopmans & Treu 2002; Treu & Koopmans 2002; Barnabè & Koopmans 2007; McKean et al. 2009) and analysis of any extended images (e.g., Dye & Warren 2005; Dye et al. 2008) can constrain the mass distribution slope.

In this paper, we present a measurement of $H_0$ from the gravitational lens B1608+656. A comprehensive lensing analysis of the lens system is in a companion paper (Paper I; Suyu et al. 2009). Using the results from Paper I, we focus on the techniques to break the mass-sheet degeneracy in this paper for inferring a value of $H_0$. By combining B1608+656 and WMAP data, we derive a constraint on the dark energy parameter $w$.

The organization of the paper is as follows. In Section 2, we briefly review the theory behind using gravitational lenses to measure $H_0$, include a description of the mass-sheet degeneracy, and describe the dynamics modeling for the measured velocity dispersion. In Section 3, we outline the probability theory for combining various data sets and for including cosmological priors. In Section 4, we present the gravitational lens B1608+656 as a candidate for measuring $H_0$, and show the lens modeling results. We present the new velocity dispersion measurement and the stellar dynamics modeling in Section 5. The study of the convergence accumulated along the line of sight to B1608+656 is discussed in Section 6. The priors for our model parameters are described in Section 7. Finally, in Section 8 we combine the lensing, dynamics and external convergence analyses to break the mass-sheet degeneracy and infer $H_0$ from the B1608+656 data set. We then show how B1608+656 aids in constraining flatness and measuring $w$ when combined with WMAP, before concluding in Section 9.

Throughout this paper, we assume a $w$-CDM universe where dark energy is described by a time-independent equation of state with parameter $w = P/\rho c^2$ with present day dark energy density $\Omega_\Lambda$, and the present day matter density is $\Omega_\mathrm{m}$. Each quoted parameter estimate is the median of the appropriate one-dimensional marginalized posterior probability density function (PDF), with the quoted uncertainties showing, unless otherwise stated, the $16^\mathrm{th}$ and $84^\mathrm{th}$ percentiles (that is, the bounds of a 68% confidence interval).

## 2. MEASURING $H_0$ USING LENSING, STELLAR DYNAMICS, AND LENS ENVIRONMENT STUDIES

In this section we briefly review the theory of gravitational lensing for $H_0$ measurement (Section 2.1), describe the mass-sheet degeneracy (Section 2.2), and present the dynamics modeling (Section 2.3). Readers familiar with these subjects can proceed directly to Section 3.

### 2.1. *Theory of gravitational lensing*

For a strong lens system in an otherwise homogeneous Robertson-Walker universe, the excess time delay of an image at angular position $\vec{\theta} = (\theta_1, \theta_2)$ with corresponding source position $\vec{\beta} = (\beta_1, \beta_2)$ relative to the case of no lensing is

$$t(\vec{\theta}, \vec{\beta}) = \frac{1}{c} \frac{D_\mathrm{d} D_\mathrm{s}}{D_\mathrm{ds}} (1 + z_\mathrm{d})\, \phi(\vec{\theta}, \vec{\beta}), \qquad (1)$$

where $z_\mathrm{d}$ is the redshift of the lens, $\phi(\vec{\theta}, \vec{\beta})$ is the so-called Fermat potential, and $D_\mathrm{d}$, $D_\mathrm{s}$, and $D_\mathrm{ds}$ are, respectively, the angular diameter distance from us to the lens, from us to the source, and from the lens to the source. The Fermat potential is defined as

$$\phi(\vec{\theta}, \vec{\beta}) \equiv \left[ \frac{(\vec{\theta} - \vec{\beta})^2}{2} - \psi(\vec{\theta}) \right], \qquad (2)$$

where the first term comes from the geometric path difference as a result of the strong lens deflection, and the second term is the gravitational delay described by the lens potential $\psi(\vec{\theta})$. The scaled deflection angle of a light ray is $\vec{\alpha}(\vec{\theta}) = \vec{\nabla}\psi(\vec{\theta})$, and the lens equation that governs the deflection of light rays is $\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta})$.

The projected dimensionless surface mass density $\kappa(\vec{\theta})$ is

$$\kappa(\vec{\theta}) = \frac{1}{2}\nabla^2\psi(\vec{\theta}), \qquad (3)$$

where

$$\kappa(\vec{\theta}) = \frac{\Sigma(D_{\mathrm{d}}\vec{\theta})}{\Sigma_{\mathrm{cr}}} \qquad \text{with} \qquad \Sigma_{\mathrm{cr}} = \frac{c^2 D_{\mathrm{s}}}{4\pi G D_{\mathrm{d}} D_{\mathrm{ds}}}, \qquad (4)$$

and $\Sigma(D_{\mathrm{d}}\vec{\theta})$ is the physical projected surface mass density.

The constant coefficient in Equation (1) is proportional to the angular diameter distance and hence inversely proportional to the Hubble constant. We can thus simplify Equation (1) to the following:

$$t(\vec{\theta}, \vec{\beta}) = \frac{D_{\Delta\mathrm{t}}}{c}\,\phi(\vec{\theta}, \vec{\beta}) \qquad (5)$$

$$\propto \frac{1}{H_0}\phi(\vec{\theta}, \vec{\beta}), \qquad (6)$$

where $D_{\Delta\mathrm{t}} \equiv (1 + z_{\mathrm{d}})D_{\mathrm{d}}D_{\mathrm{s}}/D_{\mathrm{ds}}$ is referred to as the time-delay distance.

Therefore, by modeling the lens potential ($\psi(\vec{\theta})$) and the source position ($\vec{\beta}$), we can use time-delay lens systems to deduce the value of the Hubble constant, and indeed the other cosmological parameters that appear in $D_{\Delta\mathrm{t}}$. In this way, strong lensing can be seen as a kinematic probe of the universal expansion, in the same general category as Type Ia supernovae and baryonic acoustic oscillations. Since the principal dependence of $D_{\Delta\mathrm{t}}$ is on $H_0$, we continue to discuss lenses as a probe of this one parameter; however, we shall see that the other cosmological parameters play an important role in the analysis.

Gravitational lens systems with spatially extended source surface brightness distributions are of special interest since they provide additional constraints on the lens potential. However, in this case, simultaneous determinations of the source surface brightness and the lens potential are required.

## 2.2. Mass-sheet degeneracy

We now briefly describe the mass-sheet degeneracy and its relevance to this research (see e.g. Falco et al. 1985; Schneider et al. 2006, for details). As its name suggests, this is a degeneracy in the mass modeling corresponding to the addition of a mass sheet that contributes a convergence and zero shear (and a matching scaling of the original mass distribution) which leaves the predicted image positions unchanged. A circularly symmetric surface mass density distribution that is uniform interior to the line of sight is one example of such a lens. Suppose we have a lens model $\kappa_{\mathrm{model}}(\vec{\theta})$ that fits the observables of a lens system (i.e., image positions, flux ratios for point sources, and the image shapes for extended sources). A new model described by the transformation

$\kappa_{\mathrm{trans}}(\vec{\theta}) = \lambda + (1 - \lambda)\kappa_{\mathrm{model}}(\vec{\theta})$, where $\lambda$ is a constant, would also fit the lensing observables equally well. The parameter $\lambda$ corresponds physically to the convergence of the sheet. Since we might think of including exactly such a parameter to account for additional physical mass lying along the line of sight, or in the lens plane to model a nearby group or cluster, it is clear that the mass-sheet degeneracy corresponds to a degeneracy between this external convergence ($\kappa_{\mathrm{ext}}$) and the mass normalization of the lens galaxy.[9]

Despite the invariance of the image positions, shapes and relative fluxes under a mass-sheet transformation, the relative Fermat potential between the images changes according to $\Delta\phi_{\mathrm{trans}}(\vec{\theta}, \vec{\beta}_{\mathrm{trans}}) = (1 - \lambda)\Delta\phi_{\mathrm{model}}(\vec{\theta}, \vec{\beta}_{\mathrm{model}})$. Therefore, given measured relative time delays $\Delta t$, which are inversely proportional to $H_0$ and proportional to the relative Fermat potential (Equation 6), the transformed model $\kappa_{\mathrm{trans}}$ would lead to an $H_0$ that is a factor $(1 - \lambda)$ lower than that of the initial $\kappa_{\mathrm{model}}$ (for fixed $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$, and $w$). In other words, if there is physically any external convergence $\kappa_{\mathrm{ext}}$ due to the lens' local environment or mass structure along the line of sight to the lens system that is not incorporated in the lens modeling, then

$$H_0^{\mathrm{true}} = (1 - \kappa_{\mathrm{ext}})H_0^{\mathrm{model}}. \qquad (7)$$

This degeneracy is present because lensing observations only deliver relative positions and fluxes. The degeneracy can be broken, allowing us to measure $H_0$, if (i) we know the magnitude or angular size of the source in absence of lensing, (ii) we have information on the mass normalization of the lens, or (iii) we can compare the measured shear in the lens with the observed distribution of mass to calibrate $\kappa_{\mathrm{ext}}$. For most of the strong lens systems including B1608+656, case (i) does not apply, so circumventing the mass-sheet degeneracy requires the input of more information, either about the lensing galaxy itself, or its three-dimensional environment. We distinguish between two kinds of mass sheets: *internal* and *external*. Internal mass sheets, which are physically associated with the lens galaxy, are due to nearby, physically associated galaxies, groups or clusters which, crucially, affect the stellar dynamics of the lens galaxy. External mass sheets describe mass distributions that are not physically associated with the lens galaxy and, by definition, do not affect the stellar dynamics. Typically these will lie along the line of sight to the lens (Fassnacht et al. 2006a). We identify $\kappa_{\mathrm{ext}}$ as the net convergence of this external mass sheet.

Two methods for breaking the mass-sheet degeneracy are then:

    i. *Stellar dynamics of the lens galaxy.* Stellar dynamics can be used jointly with lensing to break the internal mass-sheet degeneracy by providing an estimate of the enclosed mass at a radius different from the Einstein radius, which is approximately the radius of the lensed images from the lens galaxy (e.g., Grogin & Narayan 1996a,b; Tonry &

---

[9] To be specific, the prescription that we adopt for combining the effects of many mass sheets at redshifts $z_i$ with surface mass densities $\Sigma_i$ is $\kappa_{\mathrm{ext}} = \frac{4\pi G}{c^2}\sum_i \frac{\Sigma_i(D_i\vec{\theta})D_i D_{i\mathrm{s}}}{D_{\mathrm{s}}}$.

Franx 1999; Koopmans & Treu 2002; Treu & Koopmans 2002; Barnabè & Koopmans 2007). We note that for a given stellar velocity dispersion, there is a degeneracy in the mass and the stellar orbit anisotropy (which characterizes the amount of tangential velocity dispersion relative to radial dispersion). Nonetheless, the mass-isotropy degeneracy is nearly orthogonal to the mass-sheet degeneracy, so a combination of the mass within the effective radius (from the stellar velocity dispersion) and the mass within the Einstein radius (from lensing) effectively breaks both the mass-isotropy and the internal mass-sheet degeneracies. We describe how this works within the context of our chosen mass model in Section 2.3 below.

ii. *Studying the environment and the line of sight to the lens galaxy.* Observations of the field around lens galaxies allows a rough picture of the projected mass distribution to be built up. Many lens galaxies lie in galaxy groups, which can be identified either by their spectra or, more cheaply (but less accurately), by their colors and magnitudes. By modeling the mass distribution of the groups and galaxies in the lens plane and along the line of sight to the lens galaxy, one can estimate the external convergence $\kappa_{\rm ext}$ at the redshift of the lens (e.g. Momcheva et al. 2006; Fassnacht et al. 2006a; Auger et al. 2007, and references therein). The group modeling requires (i) identification of the galaxies that belong to the group of the lens galaxy, and (ii) estimates of the group centroid and velocity dispersion. A number of recipes can be followed. For example, Keeton & Zabludoff (2004) considered two extremes: (i) the group is described by a single smooth mass distribution, and (ii) the masses are associated with individual galaxy group members with no common halo. The realistic mass distribution for a galaxy group should be somewhere between these two extremes. The experience to date is that modeling lens environments accurately is very difficult, with uncertainties of 100% typical. In Section 6, we describe an alternative approach for quantifying the external convergence in a statistical manner: ray-tracing through numerical simulations of large-scale structure (Hilbert et al. 2007). In this section we also present a first attempt at tailoring the ray-tracing results to our one line of sight, using the relative galaxy number counts in the field.

We emphasize that the mass-sheet degeneracy is simply one of the several parameter degeneracies in the lens modeling that has been given a special name. When power-laws ($\kappa \sim bR^{1-\gamma'}$, where $R$ is the radial distance from the lens center, $b$ is the normalization of the lens, and $\gamma'$ is the radial slope in the mass profile) are used to describe the lens mass distribution, one often finds a $H_0$-$\gamma'$ degeneracy in addition to the $H_0$-$b$-$\kappa_{\rm ext}$ (mass-sheet) degeneracy (for fixed $\Omega_{\rm m}$, $\Omega_\Lambda$ and $w$; more generally, $D_{\Delta {\rm t}}$ would be in place of $H_0$). These two degeneracies are of course related via $H_0$. The $H_0$-$\gamma'$ degeneracy primarily occurs in lens systems with symmetric configurations due to a lack of information on $\gamma'$. In contrast, lens systems with images spanning a range of radii or with extended images provide information on $\gamma'$ (e.g. Wucknitz et al. 2004; Dye et al. 2008), and so the $H_0$-$\gamma'$ degeneracy is broken. Nonetheless, the $H_0$-$b$-$\kappa_{\rm ext}$ degeneracy is still present unless we provide information from dynamics and lens environment studies.

### 2.3. *Stellar dynamics modeling*

In order to model the velocity dispersion of the stars in the lens galaxy, we need a model for the local gravitational potential well in which those stars are orbiting. This potential is due to both the mass distribution of the lens galaxy, and also the "internal mass sheet" due to neighboring groups and galaxies physically associated with the lens, as described in the previous subsection. Recent studies such as the Sloan Lens ACS Survey (SLACS) found that the sum of these internal components can be well-described by a power law (e.g. Treu et al. 2006; Koopmans et al. 2006; Gavazzi et al. 2007; Koopmans et al. 2009). With this in mind, we assume that the total (lens plus sheet) mass density distribution is spherically symmetric and of the form

$$\rho_{\rm local} = \rho_0 \left(\frac{r_0}{r}\right)^{\gamma'}, \qquad (8)$$

where $\gamma'$ is the logarithmic slope of the effective lens density profile, and $\rho_0 r_0^{\gamma'}$ is the normalization of the mass distribution that is determined quite precisely by the lensing, up to a small offset contributed by the external convergence $\kappa_{\rm ext}$. This normalization can be expressed in terms of observable or inferrable quantities as we show below.

By integrating $\rho_{\rm local}$ within a cylinder with radius given by the Einstein radius $R_{\rm Ein}$, we find

$$M_{\rm local} = 4\pi \int_0^\infty dz \int_0^{R_{\rm Ein}} \rho_0 r_0^{\gamma'} \frac{s\,ds}{(s^2 + z^2)^{\gamma'/2}} \qquad (9)$$

$$= -\rho_0 r_0^{\gamma'} \frac{\pi^{3/2}\Gamma(\frac{\gamma'-3}{2})R_{\rm Ein}^{3-\gamma'}}{\Gamma(\frac{\gamma'}{2})}. \qquad (10)$$

However, the mass responsible for creating an Einstein ring is a combination of this local mass and the external mass contributed along the line of sight, so the mass contained within the Einstein ring is

$$M_{\rm Ein} = M_{\rm local} + M_{\rm ext} \qquad (11)$$

where $M_{\rm Ein}$ is the mass enclosed within the Einstein radius $R_{\rm Ein}$ that would be inferred from lensing,[10] given by

$$M_{\rm Ein} = \pi R_{\rm Ein}^2 \Sigma_{\rm cr}, \qquad (12)$$

and $M_{\rm ext}$ is the mass contribution from $\kappa_{\rm ext}$,

$$M_{\rm ext} = \pi R_{\rm Ein}^2 \kappa_{\rm ext} \Sigma_{\rm cr}. \qquad (13)$$

Combining Equations (10), (11), (12), and (13), we find

$$\rho_0 r_0^{\gamma'} = (\kappa_{\rm ext} - 1)\Sigma_{\rm cr} R_{\rm Ein}^{\gamma'-1} \frac{\Gamma(\frac{\gamma'}{2})}{\pi^{1/2}\Gamma(\frac{\gamma'-3}{2})}. \qquad (14)$$

Substituting this in Equation (8), we obtain

$$\rho_{\rm local} = (\kappa_{\rm ext} - 1)\Sigma_{\rm cr} R_{\rm Ein}^{\gamma'-1} \frac{\Gamma(\frac{\gamma'}{2})}{\pi^{1/2}\Gamma(\frac{\gamma'-3}{2})} \frac{1}{r^{\gamma'}}. \qquad (15)$$

---

[10] By definition, $R_{\rm Ein}$ is the radius within which the total mean convergence is unity.

Spherical Jeans modeling can then be employed to infer the line-of-sight velocity dispersion, $\sigma^{\mathrm{P}}(\gamma', \kappa_{\mathrm{ext}}, \beta_{\mathrm{ani}}, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, w)$, from $\rho_{\mathrm{local}}$ by assuming a model for the stellar distribution $\rho_*$ (e.g., Binney & Tremaine 1987). Here, $\beta_{\mathrm{ani}}$ is a general anisotropy term that can be expressed in terms of an anisotropy radius parameters for the stellar velocity ellipsoid, $r_{\mathrm{ani}}$, in the Osipkov-Merritt formulation (Osipkov 1979; Merritt 1985):

$$\beta_{\mathrm{ani}} = \frac{r^2}{r_{\mathrm{ani}}^2 + r^2}, \tag{16}$$

where $r_{\mathrm{ani}} = 0$ is pure radial orbits and $r_{\mathrm{ani}} \rightarrow \infty$ is isotropic with equal radial and tangential velocity dispersions. The dependence of $\sigma^{\mathrm{P}}$ on $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$, and $w$ enters through $\Sigma_{\mathrm{cr}}$ and the physical scale radius of the stellar distribution, but the dependence on $H_0$ drops out.

In the next section, we present the probability theory for obtaining posterior probability distribution of $H_0$ by combining the lensing, dynamics and lens environment studies.

## 3. PROBABILITY THEORY

We aim to obtain an expression for the posterior probability distribution of cosmological parameters $H_0$, $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$, and $w$ given the various independent data sets of B1608+656.

### 3.1. *Notations for joint modeling of data sets*

We introduce notations for the observed data and the model parameters that will be used throughout the rest of this paper.

We have three independent data sets for B1608+656: the time delay measurements from the radio observations of the four lensed images A, B, C and D (Fassnacht et al. 1999, 2002), *HST* Advanced Camera for Surveys (ACS) observations associated with program 10158 (PI:Fassnacht; Suyu et al. 2009), and the stellar velocity dispersion measurement of the primary lens galaxy G1 (see Section 5). Let $\Delta t$ be the time delay measurements of images A, C and D relative to image B, $d$ be the data vector of the lensed image surface brightness measurements of the gravitational lensed image, and $\sigma$ be the stellar velocity dispersion measurement of the lens galaxy.

As shown in Section 2.1, information on $H_0$, $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$, and $w$ comes primarily from the relative time delays between the images, which is a product of the time-delay distance $D_{\Delta \mathrm{t}}$ and the Fermat potential difference. The Fermat potential is determined by the lens potential and the source position that is given by the lens equation. Therefore, the first step is to model the lens system using the observed lensed image $d$. In order to model the lens mass distribution using the extended source information, we need to model the point-spread function (PSF) $\mathbf{B}$, image covariance matrix $\mathbf{C_D}$, lens galaxy light $l$, and dust $\mathbf{K}$ (if present) (e.g. Suyu et al. 2009). We collectively denote these discrete models associated with the lensed image processing as $M_{\mathrm{D}} = \{\mathbf{B}, \mathbf{C_D}, l, \mathbf{K}\}$. We explored a representative subspace of models $M_{\mathrm{D}}$ in Paper I, using the Bayesian evidence from the ACS data analysis to quantify the appropriateness of each model tested. Given a particular image processing model, we can infer the parameters of the lens potential and the source surface

brightness distribution from the ACS data $d$. The data models are denoted by $M_j = M_2, \ldots, M_{11}$ for Models 2–11 in Paper I.

The lens potential can be simply parametrized by, for example, a singular power-law ellipsoid (SPLE) with a few parameters such as the centroid, ellipticity, position angle, Einstein radius, and radial slope. There is no need to include an external convergence parameter in the mass distribution during the lens modeling since we cannot determine it due to the mass-sheet degeneracy. Instead, we explicitly incorporate the external convergence in the Fermat potential later on, taking into account the interplay among this parameter, the slope, and the normalization of the effective lens mass distribution. We collectively label all the parameters of the simply-parametrized model by $\eta$, except for the radial slope $\gamma'$.

Alternatively, the lens potential can be described on a grid of pixels, especially when the source galaxy is spatially extended (which provides additional constraints on the lens potential). We focus on this case; in particular, we decompose the lens potential into an initial simply-parametrized SPLE model $\psi_0(\gamma', \eta)$ and grid-based potential corrections denoted by the vector $\delta\psi$. The final potential, which is on the same grid of pixels as the corrections, is $\psi = \psi_0(\gamma', \eta) + \delta\psi$, where $\psi_0(\gamma', \eta)$ is the vector of initial potential values evaluated at the grid points. Furthermore, we also describe the extended source surface brightness distribution on a (different) grid of pixels by the vector $s$. The determination of the source surface brightness distribution given the lens potential model is a regularized linear inversion. The strength and form of the regularization are denoted by $\lambda$ and $\mathbf{g}$, respectively. The procedure for obtaining the pixelated potential corrections and the corresponding extended source surface brightness distribution is iterative and is described in detail in Paper I.

The resulting lens potential allows us to compute the Fermat potential $\phi$ at each image position, up to a factor of $(1 - \kappa_{\mathrm{ext}})$. Combining the Fermat potential with a value of $D_{\Delta \mathrm{t}}$ computed given the cosmological parameters $\{H_0, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, w\}$ provides us with predicted values of the image time delays, $\Delta t^{\mathrm{P}}$.

The dynamics modeling of the galaxy is set up such that the profile is consistent with the lens modeling, as described in Section 2.3. The predicted velocity dispersion is dependent on six parameters: 1) the effective lens mass distribution profile slope $\gamma'$, 2) the external convergence $\kappa_{\mathrm{ext}}$, 3) the anisotropy radius $r_{\mathrm{ani}}$, and then the cosmological parameters 4) $\Omega_{\mathrm{m}}$, 5) $\Omega_{\Lambda}$, and 6) $w$.

By combining lensing, dynamics, and lens environment studies, we can break the $D_{\Delta \mathrm{t}}$-$\kappa_{\mathrm{ext}}$ degeneracy to obtain a probability distribution for the cosmological parameters $\{H_0, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, w\}$ given the data sets. In the inference, we assume that the redshifts of the lens and source galaxies are known exactly for the computation of $D_{\Delta \mathrm{t}}$. This is approximately true for B1608+656, which has spectroscopic measurements for the redshifts (Myers et al. 1995; Fassnacht et al. 1996) — an uncertainty of 0.0003 on the redshifts translates to $< 0.2\%$ in time-delay distance, and hence $H_0$ for fixed $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$ and $w$. By imposing sensible priors on $\{H_0, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, w\}$ from other independent experiments such as WMAP's 5-year result (WMAP5), we can marginalize the distribution to

obtain the posterior probability distribution for $H_0$.

## 3.2. Constraining cosmological parameters

In this section, we describe the probability theory for inferring cosmological parameters from the B1608+656 data sets.

Our goal is to obtain the posterior PDF for the model parameters $\boldsymbol{\xi}$ given the three independent data sets $\{\boldsymbol{\Delta t}, \boldsymbol{d}, \sigma\}$:

$$P(\boldsymbol{\xi}|\boldsymbol{\Delta t}, \boldsymbol{d}, \sigma) \propto P(\boldsymbol{\Delta t}|\boldsymbol{\xi})P(\boldsymbol{d}|\boldsymbol{\xi})P(\sigma|\boldsymbol{\xi})P(\boldsymbol{\xi}), \quad (17)$$

where the parameters $\boldsymbol{\xi}$ consist of all the model parameters for obtaining the predicted data sets described in Section 3.1: $\gamma'$, $\kappa_{\mathrm{ext}}$, $\boldsymbol{\eta}$, $\boldsymbol{\delta\psi}$, $\boldsymbol{s}$, $\boldsymbol{M}_{\mathrm{D}}$, $r_{\mathrm{ani}}$, $H_0$, $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$, $w$. For notational simplicity, we denote the cosmological parameters as $\boldsymbol{\pi} = \{H_0, \Omega_{\mathrm{m}}, \Omega_{\Lambda}, w\}$. In Equation (17), the dependence on $z_{\mathrm{s}}$ and $z_{\mathrm{d}}$ are implicit.

To obtain the PDF of cosmological parameters $\boldsymbol{\pi}$, we marginalize Equation (17) over all parameters apart from $\boldsymbol{\pi}$:

$$P(\boldsymbol{\pi}|\boldsymbol{\Delta t}, \boldsymbol{d}, \sigma) \propto \int \mathrm{d}\gamma' \, \mathrm{d}\kappa_{\mathrm{ext}} \, \mathrm{d}\boldsymbol{\eta} \, \mathrm{d}\boldsymbol{\delta\psi} \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}\boldsymbol{M}_{\mathrm{D}} \, \mathrm{d}r_{\mathrm{ani}} \cdot$$

$$\overbrace{P(\boldsymbol{\Delta t}|\boldsymbol{\xi})P(\boldsymbol{d}|\boldsymbol{\xi})P(\sigma|\boldsymbol{\xi})}^{\text{likelihood}} \cdot$$

$$\underbrace{P(\boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{\eta}, \boldsymbol{\delta\psi}, \boldsymbol{s}, \boldsymbol{M}_{\mathrm{D}}, r_{\mathrm{ani}})}_{\text{prior}}. \quad (18)$$

In the following subsection, we discuss each of the three terms in the joint likelihood function in turn.

## 3.3. Likelihoods

Each of the three likelihoods in Equation (18) generally depends only on a subset of the parameters $\boldsymbol{\xi}$. Specifically, dropping independences, we have $P(\boldsymbol{\Delta t}|\boldsymbol{\xi}) = P(\boldsymbol{\Delta t}|\boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{\eta}, \boldsymbol{\delta\psi}, \boldsymbol{M}_{\mathrm{D}})$, $P(\boldsymbol{d}|\boldsymbol{\xi}) = P(\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}, \boldsymbol{s}, \boldsymbol{M}_{\mathrm{D}})$, and $P(\sigma|\boldsymbol{\xi}) = P(\Omega_{\mathrm{m}}, \Omega_{\Lambda}, w, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}})$.

For B1608+656, we can simplify and drop independences further in the time delay likelihood $P(\boldsymbol{\Delta t}|\boldsymbol{\xi})$ by expressing the relative Fermat potential (relative to image B for the images A, C and D) as

$$\Delta\phi(\gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}}) = (1 - \kappa_{\mathrm{ext}})q(\gamma', \boldsymbol{M}_{\mathrm{D}}), \quad (19)$$

and writing the $i^{\mathrm{th}}$ (AB, CB or DB) predicted time delay as

$$\Delta t_i^{\mathrm{P}} = \frac{1}{c} D_{\Delta t}(z_{\mathrm{d}}, z_{\mathrm{s}}, \boldsymbol{\pi}) \cdot \Delta\phi_i(\gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}}) \quad (20)$$

(see Appendix A for details). The resulting likelihood is

$$P(\boldsymbol{\Delta t}|z_{\mathrm{d}}, z_{\mathrm{s}}, \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}})$$
$$= \prod_{i=1}^{3} P(\Delta t_i|z_{\mathrm{d}}, z_{\mathrm{s}}, \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}}), \quad (21)$$

where we assume that the three time delay measurements are independent, and that each $P(\Delta t_i|z_{\mathrm{d}}, z_{\mathrm{s}}, \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}})$ is given by the PDF in Fassnacht et al. (2002).

The pixelated lens potential and source surface brightness reconstruction allows us to compute

$$P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}_{\mathrm{MP}}, \boldsymbol{M}_{\mathrm{D}}) = \int \mathrm{d}\boldsymbol{s} \, P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}_{\mathrm{MP}}, \boldsymbol{s}, \boldsymbol{M}_{\mathrm{D}}) \cdot$$
$$P(\boldsymbol{s}|\lambda, \mathbf{g}), \quad (22)$$

by marginalizing out the source surface brightness $\boldsymbol{s}$. The most probable potential correction, $\boldsymbol{\delta\psi}_{\mathrm{MP}}$, is the result of the pixelated potential reconstruction method. The likelihood for the lens parameters, $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}_{\mathrm{MP}}, \boldsymbol{M}_{\mathrm{D}})$, is also the Bayesian evidence of the source surface brightness reconstruction. Part of the marginalization in Equation (18) can be simplified via

$$\int \mathrm{d}\boldsymbol{\delta\psi} \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}\boldsymbol{M}_{\mathrm{D}} \, P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}, \boldsymbol{s}, \boldsymbol{M}_{\mathrm{D}}) \cdot$$

$$P(\boldsymbol{s}|\lambda, \mathbf{g})P(\boldsymbol{M}_{\mathrm{D}})P(\boldsymbol{\delta\psi})$$
$$\propto\sim P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5), \quad (23)$$

under various assumptions stated in Appendix A that are either justified in Paper I or will be shown to be valid in Section 4.2. In essence, we find that the ACS data models that give acceptable fits are all equally probable within their errors, making conditioning on $\boldsymbol{M}_5$ (i.e., setting $\boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5$, where $\boldsymbol{M}_5$ is Model 5 in Paper I for the lensed image processing) approximately equivalent to marginalizing over all models $\boldsymbol{M}_{\mathrm{D}}$.

Furthermore, we can marginalize out the parameters of the smooth lens model $\boldsymbol{\eta}$ separately:

$$P(\gamma'|\boldsymbol{d}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5) \propto \int \mathrm{d}\boldsymbol{\eta} \, P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5) \cdot$$
$$P_{\mathrm{no \, ACS}}(\gamma') \, P(\boldsymbol{\eta}). \quad (24)$$

(See Appendix A for details of the assumptions involved.) We see that the resulting PDF, $P(\gamma'|\boldsymbol{d}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$, can itself be treated as a prior on the slope $\gamma'$. Without the ACS data $\boldsymbol{d}$, this distribution will default to the lower level prior $P_{\mathrm{no \, ACS}}(\gamma')$. For the rest of this section we refer only to the generic prior $P(\gamma')$, keeping in mind that this distribution may or may not include the information from the ACS data. This will allow us to isolate the influence of the ACS data on the final results, when we compare the PDF in Equation (24) with some alternative choices of $P(\gamma')$.

For the velocity dispersion likelihood, the predicted velocity dispersion $\sigma^{\mathrm{P}}$ as a function of the parameters described in Section 3.1 is

$$\sigma^{\mathrm{P}} = \sigma^{\mathrm{P}}(\Omega_{\mathrm{m}}, \Omega_{\Lambda}, w, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}}|z_{\mathrm{d}}, z_{\mathrm{s}}, r_{\mathrm{eff}}, R_{\mathrm{Ein}}), \quad (25)$$

where the effective radius, $r_{\mathrm{eff}}$, the Einstein radius, $R_{\mathrm{Ein}}$, and the mass enclosed within the Einstein radius, $M_{\mathrm{Ein}}$, are fixed. The effective radius is fixed by observations, and $R_{\mathrm{Ein}}$ and $M_{\mathrm{Ein}}$ are the quantities that lensing delivers robustly. The uncertainty in the dynamics modeling due to the error associated with $r_{\mathrm{eff}}$, $R_{\mathrm{Ein}}$ and $M_{\mathrm{Ein}}$ is negligible compared to the uncertainties associated with $\kappa_{\mathrm{ext}}$. The likelihood function for $\sigma$ is a Gaussian:

$$P(\sigma|\Omega_{\mathrm{m}}, \Omega_{\Lambda}, w, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}})$$
$$= \frac{1}{\sqrt{2\pi\sigma_\sigma^2}} \exp\left[-\frac{(\sigma - \sigma^{\mathrm{P}})^2}{2\sigma_\sigma^2}\right]. \quad (26)$$

Finally then, we have the following simplified version of Equation (18), where the posterior PDF has been successfully compartmentalized into manageable pieces:

$$P(\boldsymbol{\pi}|\boldsymbol{\Delta t}, \boldsymbol{d}, \sigma) \propto \int \mathrm{d}\gamma' \, \mathrm{d}\kappa_{\mathrm{ext}} \, \mathrm{d}r_{\mathrm{ani}} \cdot$$

$$P(\boldsymbol{\Delta t}|z_{\mathrm{d}}, z_{\mathrm{s}}, \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5) \cdot$$
$$P(\sigma|\Omega_{\mathrm{m}}, \Omega_{\Lambda}, w, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}}) \cdot$$
$$P(\gamma') \, P(\kappa_{\mathrm{ext}}) \, P(r_{\mathrm{ani}}) \, P(\boldsymbol{\pi}). \quad (27)$$

Sections 4 to 7 address the specific forms of the likelihoods and the priors in Equation (27). In particular, in the next section, we focus on the lens modeling of B1608+656 which will justify the assumptions mentioned above and provide both the time delay likelihood and the ACS $P(\gamma')$ prior.

## 4. LENS MODEL OF B1608+656

The quadruple-image gravitational lens B1608+656 was discovered in the Cosmic Lens All-Sky Survey (CLASS) (Myers et al. 1995; Browne et al. 2003; Myers et al. 2003). Figure 1 is an image of B1608+656, showing the spatially extended source surface brightness distribution (with lensed images labeled by A, B, C, and D) and two interacting galaxy lenses (labeled by G1 and G2). The redshifts of the source and the lens galaxies are, respectively, $z_s = 1.394$ (Fassnacht et al. 1996) and $z_d = 0.6304$ (Myers et al. 1995).[11] We note that the lens galaxies are in a group with all galaxy members in the group lie within $\pm 300\,\mathrm{km\,s^{-1}}$ of the mean redshift (Fassnacht et al. 2006a). Thus, even a conservative limit of $300\,\mathrm{km\,s^{-1}}$ for the peculiar velocity of B1608+656 relative to the Hubble flow would only change $D_{\Delta t}$ by 0.5%. As we will see, this is not significant compared to the systematic error associated with $\kappa_{\mathrm{ext}}$. This system is special in that the three relative time delays between the four images were measured accurately with errors of only a few percent: $\Delta t_{\mathrm{AB}} = 31.5^{+2.0}_{-1.0}$ days, $\Delta t_{\mathrm{CB}} = 36.0^{+1.5}_{-1.5}$ days, and $\Delta t_{\mathrm{DB}} = 77.0^{+2.0}_{-1.0}$ days (Fassnacht et al. 1999, 2002). The additional constraints on the lens potential from the extended source analysis and the accurately measured time delays between the images make B1608+656 a good candidate to measure $H_0$ with few-percent precision. However, the presence of dust and interacting galaxy lenses (visible in Figure 1) complicate this system. In Paper I, we presented a comprehensive analysis that took into account the extended source surface brightness distribution, interacting galaxy lenses, and the presence of dust for reconstructing the lens potential. In the following subsections, we summarize the data analysis and lens modeling from Paper I, and present the resulting Bayesian evidence values (needed in Equation (24)) from the lens modeling.

### 4.1. Summary of observations, data analysis, and lens modeling in Paper I

Deep *HST* ACS observations on B1608+656 in F606W and F814W filters were taken specifically to obtain high signal-to-noise ratio images of the lensed source emission.

In Paper I, we investigated a representative sample of PSF, dust, and lens galaxy light models in order to extract the Einstein ring for the lens modeling. Table 1 lists the various PSF and dust models, and we refer the readers to Paper I for details of each model.

The resulting dust-corrected, galaxy-subtracted F814W image allowed us to model both the lens potential and source surface brightness on grids of pixels based on an iterative and perturbative potential reconstruction scheme. This method requires an initial guess potential model that would ideally be close to the true model. In Paper I, we adopt the SPLE1+D (isotropic) model from

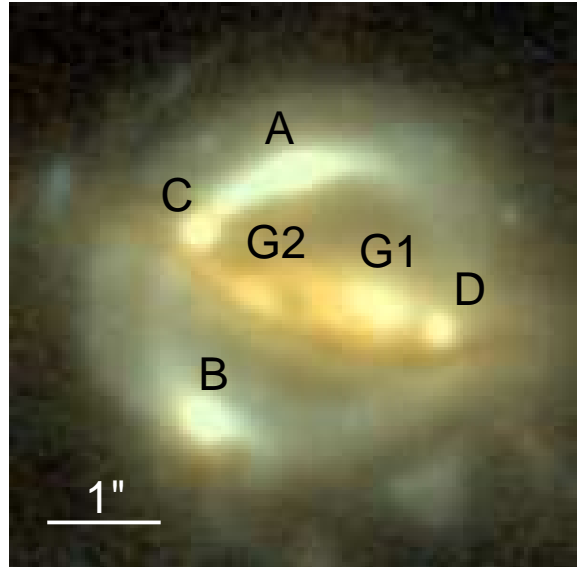[11] We assume that the redshift of G2 is the same as G1.



FIG. 1.— *HST* ACS image of B1608+656 from 11 orbits in F814W and 9 orbits in F606W. North is up and east is left. The lensed images of the source galaxy are labeled by A, B, C, and D, and the two lens galaxies are G1 and G2. 1 arcsec corresponds to approximately 7 kpc at the redshift of the lens.

Koopmans et al. (2003) as the initial model, which is the most up-to-date, simply-parametrized model combining both lensing and stellar dynamics. In the current paper, we additionally investigate the dependence on the initial model by describing the lens galaxies as SPLE models for a range of slopes ($\gamma' = 1.5, 1.6, \ldots, 2.5$). Contrary to the SPLE1+D (isotropic) model, the parameters for the SPLE models with variable slopes are constrained by lensing data only, without the velocity dispersion measurement.

The source reconstruction provides a value for the Bayesian evidence, $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}, \boldsymbol{M}_{\mathrm{D}})$, which can be used for model comparison (where model refers to the PSF, dust, lens galaxy light, and lens potential model). The reconstructed lens potential (after the pixelated corrections $\boldsymbol{\delta\psi}$) for each data model (PSF, dust, lens galaxy light) leads to three estimates of the Fermat potential differences between the image positions. These are presented in the next subsection for the representative set of PSF, dust, lens galaxy light, and pixelated potential model.

### 4.2. Lens modeling results

In Paper I, we successfully used a pixelated reconstruction method to model small deviations from a smooth lens potential model of B1608+656. The resulting source surface brightness distribution is well-localized, and the most probable potential correction $\boldsymbol{\delta\psi}_{\mathrm{MP}}$ has angular structure approximately following a $\cos\phi$ mode with amplitude $\sim 2\%$. The $\cos 2\phi$ mode, which could mimic an additional external shear or lens mass distribution ellipticity, has a lower amplitude still, indicating that the smooth model of Koopmans et al. (2003) — which includes an external shear of $\simeq 0.08$ — is giving an adequate account of the extended image light distribution. This was the main result of Paper I. The key ingredient in the ACS prior for the lens density profile slope parameter $\gamma'$ (Equation (24)) coming from this analysis is

the likelihood $P(\boldsymbol{d}|\gamma', \boldsymbol{M}_D)$. For a particular choice of slope $\gamma'$ and data model $\boldsymbol{M}_D$, this is just the evidence value resulting from the Paper I reconstruction. In this section, our objective is to use the results of this analysis to obtain $P(\gamma'|\boldsymbol{d})$ and $\Delta\phi(\gamma', \kappa_{\text{ext}})$, marginalizing over a representative sample of data models.

### 4.2.1. *Marginalization of the data model*

Table 1 shows the results of the pixelated potential reconstruction at fixed density slope in the initial smooth lens potential model, for various data models $\boldsymbol{M}_D$. Specifically, we used the SPLE1+D (isotropic) model in Koopmans et al. (2003) with $\gamma' = 2.05$. The uncertainties in the log evidence in Table 1 were estimated as $\sim 0.03 \times 10^4$ for the log evidence values before potential correction, and $\sim 0.05 \times 10^4$ for the log evidence values after potential correction.

We see a clear division between models with high and low evidence values, the two groups being separated by a very large factor in probability. Assuming that all the data models $\boldsymbol{M}_D$ are equally probable *a priori*, the contribution to the marginalized distribution $P(\boldsymbol{\pi}|\Delta\boldsymbol{t}, \boldsymbol{d}, \sigma)$ (Equation (18)) from these lower-evidence models will be negligible.

The physical difference between these evidence-ranked data models is in the dust correction: the 2-band dust models are found to be less probable than the 3-band dust models. It is useful to quantify the systematic error that would occur with the use of 2-band dust models (which was avoided from the evidence ranking) in terms of the $H_0$ value implied by the system. For this simple error estimation we use Equation (5) and assert $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$ and zero external convergence, as a fiducial reference cosmology (Koopmans et al. 2003). The implied Hubble constants are shown in the final four columns of Table 1. We see that the disfavored use of the 2-band dust maps would have led to values of $H_0$ some 15% lower than that inferred from the 3-band maps.

We note that the evidence values of each of the 3-band dust map models $\boldsymbol{M}_D$ are the same within their uncertainties. We can also see that for good data models, specifically $\boldsymbol{M}_D = \boldsymbol{M}_5$, the three $H_0$ values have low scatter: these lens models are internally self-consistent. Furthermore, the scatter between the values for the different good data models is also low: the high evidence data models consistently return the same Hubble constant. This is the basis for the approximations (in Section 3.3 and Appendix A) that the likelihood $P(\Delta\boldsymbol{t}|\boldsymbol{\xi})$ is effectively constant with the 3-band dust map models $\boldsymbol{M}_D$. Assuming that we have indeed obtained the optimal set of $\boldsymbol{M}_D$, we can approximate the likelihoods in Equations (24) and (27) as being evaluated for model $\boldsymbol{M}_5$.

### 4.2.2. *Effects of the potential corrections*

Having approximately marginalized out $\boldsymbol{M}_D$ by conditioning on $\boldsymbol{M}_5$, we now consider the impact of the potential corrections discussed in Paper I. In particular, we seek the likelihood for the density profile slope parameter $\gamma'$, $P(\boldsymbol{d}|\gamma' = \gamma'_i, \boldsymbol{\eta}, \boldsymbol{\delta\psi}_{\text{MP}}, \boldsymbol{M}_D = \boldsymbol{M}_5)$. We characterize this function on a grid of slope values in the range of $\gamma' = 1.5, 1.6 \ldots, 2.5$, first re-optimizing the parameters of the smooth lens model, and then computing the

source reconstruction evidences both with and without potential correction. These are tabulated in Table 2. We again compute the Fermat potential differences and implied Hubble constant values as before.

The spread of the three implied $H_0$ values at fixed density slope is again small: we conclude that the internal self-consistency of the lens model depends on the data model but not $\gamma'$. The table also shows that the smooth SPLE model provides a good estimate of the relative Fermat potentials. Indeed, this was the principal conclusion of Paper I. The relative thickness of the arcs is sensitive to the SPLE density profile slope $\gamma'$, as can be seen in the first two columns of Table 2: the evidence clearly favors $\gamma' \simeq 2.05$, as previously found by Koopmans et al. (2003). Indeed, exponentiating this gives quite a sharply peaked function, which we return to below.

How is the potential correction then affecting the model? In Table 2 we can see that the corrected potential leads to nearly the same evidence value ($P(\boldsymbol{d}|\gamma' = \gamma'_i, \boldsymbol{\eta}, \boldsymbol{\delta\psi}_{\text{MP}}, \boldsymbol{M}_D = \boldsymbol{M}_5)$) for a wide range of underlying density slopes, and yet barely changes the relative Fermat potential values. The unchanging nature of the Fermat potential is due to the curvature type of regularization on the potential corrections suppressing the addition of mass within the potential reconstruction annulus. From Kochanek (2002), the relative Fermat potential depends only on the mean surface mass density enclosed in the annulus between the images, to first order in $\delta R/\langle R \rangle$, where $\delta R$ is the difference in the radial distance of the image locations from the effective center of the lens galaxies and $\langle R \rangle$ is the mean radius of the images. The mean surface mass density depends on the slope of the initial SPLE model (hence the trend we see in relative Fermat potential in the left-hand side of Table 2), but not on the potential corrections due to the curvature regularization imposed. Therefore, to first order in $\delta R/\langle R \rangle$, the Fermat potential depends only indirectly on $\gamma'$ via the mean surface mass density. The second order term is very small — it has a prefactor of $1/12$ and for B1608+656, $(\delta R/\langle R \rangle)^2 \sim 0.1$. Therefore, for good and self-consistent data models, the potential corrections $\boldsymbol{\delta\psi}_{\text{MP}}$ do not change the Fermat potential significantly.

The right-hand side of Table 2, where a wide range of initial slope values provide good fits to the data, is therefore effectively a manifestation of the mass-sheet degeneracy. One can understand the effect of the potential corrections as making *local corrections to the effective density profile slope in order to fit the ACS data*. The change in slope by the pixelated corrections would create a deficit/surplus of mass in the annulus, which the pixelated potential corrections then add/subtract back into the annulus in the form of a constant mass sheet to (i) enforce the prior (no net addition of mass within annulus) and (ii) continue to fit the arcs equally well.

We conclude that the value of the potential correction analysis is in demonstrating that the double SPLE model for B1608+656 is, despite the system's complexity, a good model for the high fidelity *HST* data. The corrections are small in magnitude ($\simeq 2\%$ relative to the initial SPLE model), and the inclusion of the $\boldsymbol{\delta\psi}$ neither significantly reduces the dispersion in implied $H_0$ values between the image pairs, nor alters the rank order of the data models. We therefore use the information

on the slope of the initial SPLE model from the ACS data *without potential corrections,* thus using the information on the relative thickness of the lensed extended images clearly present. How we derive our estimate for $P(\boldsymbol{d}|\gamma', \boldsymbol{M}_{\mathrm{D}})$ from column 2 of Table 2 is described next.

### 4.2.3. *The ACS posterior PDF for $\gamma'$*

In the previous section, we explored the *HST* data constraints on the slope parameter, optimizing the other parameters of the SPLE lens model at each step. To characterize properly $P(\gamma'|\boldsymbol{d}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$ in Equation (24), we would need to marginalize over all lens parameters $\boldsymbol{\eta}$ instead. However, as we shall now see, this optimization approximation is actually a good one and is certainly the most tractable solution due to the high dimensionality of the problem (16 parameters to describe G1, G2 and external shear). Direct sampling in the 16-dimensional parameter space of $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5) P_{\mathrm{no\,ACS}}(\gamma') P(\boldsymbol{\eta})$ in Equation (24) via, for example, Markov chain Monte Carlo (MCMC) techniques using the extended source information is not feasible on a reasonable time scale. Importance sampling of the prior PDF from the radio data of image positions and fluxes $(P_{\mathrm{no\,ACS}}(\gamma', \boldsymbol{\eta}) = P_{\mathrm{no\,ACS}}(\gamma', \boldsymbol{\eta}|\mathrm{radio}))$ by weighing the samples by $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$ is difficult since $\gamma'$ is effectively unconstrained by the radio data (the $\chi^2$ changes by $\lesssim 1$ in the slope range between 1.5 and 2.5).[12]

It is precisely the unconstrained nature of the $\gamma'$ parameter that makes the optimization approximation so good. The "tube" of $\gamma'$-degeneracy traversing the 16-dimensional parameter space dominates the uncertainties in the parameters. We thus assume that the tube of $\gamma'$-degeneracy has negligible thickness (a degeneracy curve), and use $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$ to break the degeneracy. Specifically, we use the radio observations, *HST* Near Infrared Camera and Multi-Object Spectrometer 1 (NICMOS) images (Proposal 7422; PI:Readhead), and time delay data to obtain the best-fitting $\hat{\boldsymbol{\eta}}$ for a given $\gamma'{=}\gamma_i'$ (assuming $\Omega_{\mathrm{m}} = 0.3$, $\Omega_{\Lambda} = 0.7$ and $w = -1$ in using the time delay data), and compute the corresponding $P(\boldsymbol{d}|\gamma_i', \hat{\boldsymbol{\eta}}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$. These are the listed evidence values in the second column of Table 2 for the various $\gamma_i'$ values. The time delay data are included because the predicted relative Fermat potential among the image pairs using the radio and NICMOS data are otherwise inconsistent with one another. The optimized parameters from only the radio and NICMOS data lead to $\chi^2 \sim 600$ for just the time delay data; including the time delay data reduces the time delay $\chi^2$ to $\sim 1$ with only a mild increase in the radio and NICMOS $\chi^2$ of $\sim 6$. We "undo" the inclusion of the time delay data (so that we do not use the time delay data twice in the importance sampling of Equation (27)) by subtracting the log likelihood of the time delay from the log likelihood of $\boldsymbol{d}$; the effect is negligible since the latter is $\sim 10^4$ higher in magnitude.

Our thin degeneracy tube assumption implies that $P(\boldsymbol{d}|\gamma') \simeq P(\boldsymbol{d}|\gamma', \hat{\boldsymbol{\eta}})$, such that the posterior PDF for the slope is $P(\gamma'|\boldsymbol{d}) \propto P(\boldsymbol{d}|\gamma') P_{\mathrm{no\,ACS}}(\gamma')$. Assigning a uniform prior (i.e., $P_{\mathrm{no\,ACS}}(\gamma')$ is constant), we arrive at the result that our desired PDF is just the exponentiation of the log evidence in column 2 of Table 2. Fitting

these log evidences with the following quadratic function,

$$\log P(\gamma'|\boldsymbol{d}) = C - \frac{(\gamma' - \gamma_0')^2}{2\sigma_{\gamma'}^2}, \qquad (28)$$

we obtain the following best-fit parameter values: $\gamma_0' = 2.081 \pm 0.027$, $\sigma_{\gamma'} = 0.0091 \pm 0.0008$, and $C = (1.60 \pm 0.01) \times 10^4$. While the PDF width $\sigma_{\gamma'}$ is very small, the centroid is not well determined. Adding $\sigma_{\gamma'}$ and the uncertainty in $\gamma_0'$ in quadrature, we finally approximate $P(\gamma'|\boldsymbol{d})$ with a Gaussian centered on 2.08 with standard deviation 0.03. This provides the prior on $\gamma'$ from the ACS data (in Equation (27)).

The deep ACS data therefore allow a significant improvement to the previous measurement in Koopmans et al. (2003) of $\gamma' = 1.99 \pm 0.20$, which was based on the radio data and the NICMOS ring. Coincidentally, our $\gamma' = 2.08 \pm 0.03$ is identical, apart from the spread, to the measurement from SLACS of $\gamma' = 2.08 \pm 0.2$ that was based on a sample of massive elliptical lenses (Koopmans et al. 2009). We note that our measurement is not the first percent-level determination of a strong lens density profile slope. Wucknitz et al. (2004) used high precision astrometric measurements from VLBI data to constrain the $\gamma'$ parameter in B0218+357 to be $1.96 \pm 0.02$ (where we have transformed their $\beta$ into our notation). However, they did not use exactly the same model as we do here (instead working with combinations of isothermal elliptical potentials and neglecting external convergence). Dye & Warren (2005) measured the power-law slope of the lens galaxy in the Einstein ring system 0047-2808 to be $\gamma' = 2.11 \pm 0.04$ based on the extended image constraints. More recently, Dye et al. (2008) determined the power-law slope of the extremely massive and luminous lens galaxy in the Cosmic Horseshoe Einstein ring system J1004+4112 to be $\gamma' = 1.96 \pm 0.02$.

### 4.2.4. *Predicted relative Fermat potentials*

In order to be able to calculate the time delay likelihood function, $P(\Delta \boldsymbol{t}|z_{\mathrm{d}}, z_{\mathrm{s}}, \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, \boldsymbol{M}_{\mathrm{D}})$, at any value of the slope $\gamma'$, we need to interpolate the Fermat potential differences given in Table 2. In fact, these data give us the function $q(\gamma')$ to insert into Equation (19): we can do the interpolation at $\kappa_{\mathrm{ext}} = 0.0$ and then rescale by $(1 - \kappa_{\mathrm{ext}})$ without loss of generality.

For each of the image pairs, we fit the relative Fermat potential difference as a third-order polynomial function of $\gamma'$ using the values we have at the discrete points $\gamma_i'$ for the SPLE models in the table. Recall that the SPLE model provides an unbiased estimate of the relative Fermat potential, and that the various top data models $\boldsymbol{M}_{\mathrm{D}}$ gave consistent estimates. Thus, the polynomial fit gives the function $q(\gamma', \boldsymbol{M}_{\mathrm{D}})$ in Equation (19). The third-order polynomial fit leads to residuals $(= (\Delta \phi_i - \Delta \phi^{\mathrm{poly}})/(\Delta \phi_i))$ of $< 1\%$ for all image pairs at all slope points in Table 2 except for $\gamma_i' = 1.7$, which has residuals of $\sim 2\%$.

### 5. BREAKING THE MASS-SHEET DEGENERACY: STELLAR DYNAMICS

In this section, we present the observations and data reduction for measuring the velocity dispersion $\sigma$ of G1 in B1608+656. This measurement appears as the likelihood function given in Equation (26) above.

---

[12] We set $\gamma_{\mathrm{G2}}' = \gamma_{\mathrm{G1}}' = \gamma'$ since the slope of G2 is ill-constrained (Koopmans et al. 2003).

TABLE 1
LOG EVIDENCE VALUES AND RELATIVE FERMAT POTENTIAL VALUES BEFORE AND AFTER THE PIXELATED POTENTIAL RECONSTRUCTION FOR
VARIOUS DATA MODELS WITH THE SPLE1+D (ISOTROPIC) INITIAL MODEL

| Data Model | | | Initial Potential | Corrected Potential | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | PSF | dust | $\log P$ $(\times 10^4)$ | $\log P$ $(\times 10^4)$ | $\Delta\phi^{AB}$ | $\Delta\phi^{CB}$ | $\Delta\phi^{DB}$ | $H_0^{AB}$ | $H_0^{CB}$ | $H_0^{DB}$ | $\bar{H}_0$ |
| 5 | B1 | 3-band | 1.56 | 1.77 | 0.244 | 0.279 | 0.575 | 78.1 | 78.1 | 75.1 | $77.1 \pm 1.7$ |
| 9 | C | B1/3-band | 1.56 | 1.76 | 0.240 | 0.280 | 0.563 | 76.7 | 78.3 | 73.5 | $76.2 \pm 2.4$ |
| 3 | C | 3-band | 1.60 | 1.76 | 0.243 | 0.277 | 0.570 | 77.6 | 77.5 | 74.4 | $76.5 \pm 1.8$ |
| 2 | drz | 3-band | 1.48 | 1.75 | 0.238 | 0.278 | 0.548 | 76.0 | 77.7 | 71.6 | $75.1 \pm 3.1$ |
| 7 | B2 | 3-band | 1.55 | 1.75 | 0.237 | 0.274 | 0.571 | 75.7 | 76.7 | 74.6 | $75.7 \pm 1.0$ |
| 11 | B1 | no dust | 1.27 | 1.72 | 0.229 | 0.263 | 0.576 | 73.2 | 73.6 | 75.3 | $74.0 \pm 1.1$ |
| 10 | B1 | C/2-band | 1.36 | 1.61 | 0.193 | 0.227 | 0.565 | 61.8 | 63.5 | 73.8 | $66.4 \pm 6.4$ |
| 4 | C | 2-band | 1.40 | 1.58 | 0.199 | 0.234 | 0.560 | 63.6 | 65.6 | 73.1 | $67.4 \pm 5.0$ |
| 6 | B1 | 2-band | 1.10 | 1.41 | 0.196 | 0.226 | 0.559 | 62.5 | 63.2 | 73.0 | $66.2 \pm 5.8$ |
| 8 | B2 | 2-band | 1.23 | 1.40 | 0.201 | 0.234 | 0.556 | 64.3 | 65.4 | 72.7 | $67.4 \pm 4.5$ |
| | | | | $\Delta\phi$ and $H_0$ values from initial SPLE1+D (isotropic) | | | | | | | |
| | | | | | 0.243 | 0.271 | 0.575 | 77.7 | 75.8 | 75.1 | $76.2 \pm 1.3$ |

Notes — The uncertainties in the log evidence before and after the potential corrections are $\sim 0.03 \times 10^4$ and $\sim 0.05 \times 10^4$, respectively. The relative Fermat potentials are in units of arcsec$^2$, and the $H_0$ values are in units of km s$^{-1}$ Mpc$^{-1}$. The $\bar{H}_0$ values are the mean and standard deviation from the mean of the three estimates obtained using the initial/corrected potential and the three time delays, without taking into account the uncertainties associated with the time delays. These $H_0$ values assume $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $w = -1$, and are listed purely to aid the digestion of the $\Delta\phi$ values. The full analysis for obtaining the probability distribution for the cosmological parameters is described in Section 8.

TABLE 2
LOG EVIDENCE VALUE BEFORE AND AFTER THE PIXELATED POTENTIAL RECONSTRUCTION FOR INITIAL MODELS WITH VARIOUS SLOPE USING
PSF-B1 AND THE 3-BAND DUST MAP ($\boldsymbol{M}_D$=MODEL 5)

| | Initial Potential | | | | | | | | Corrected Potential | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma'$ | $\log P$ $(\times 10^4)$ | $\Delta\phi^{AB}$ | $\Delta\phi^{CB}$ | $\Delta\phi^{DB}$ | $H_0^{AB}$ | $H_0^{CB}$ | $H_0^{DB}$ | $\bar{H}_0$ | $\log P$ $(\times 10^4)$ | $\Delta\phi^{AB}$ | $\Delta\phi^{CB}$ | $\Delta\phi^{DB}$ | $H_0^{AB}$ | $H_0^{CB}$ | $H_0^{DB}$ | $\bar{H}_0$ |
| 1.5 | 1.38 | 0.125 | 0.139 | 0.287 | 40.2 | 39.0 | 37.6 | $38.9 \pm 1.3$ | 1.73 | 0.130 | 0.143 | 0.290 | 41.7 | 40.2 | 38.0 | $39.9 \pm 1.9$ |
| 1.6 | 1.48 | 0.147 | 0.163 | 0.338 | 47.2 | 45.8 | 44.3 | $45.7 \pm 1.4$ | 1.77 | 0.150 | 0.170 | 0.349 | 48.1 | 47.6 | 45.6 | $47.1 \pm 1.3$ |
| 1.7 | 1.52 | 0.174 | 0.193 | 0.403 | 55.5 | 54.0 | 52.7 | $54.0 \pm 1.4$ | 1.75 | 0.178 | 0.201 | 0.417 | 57.0 | 56.2 | 54.5 | $55.9 \pm 1.3$ |
| 1.8 | 1.54 | 0.190 | 0.211 | 0.442 | 60.8 | 59.1 | 57.7 | $59.2 \pm 1.5$ | 1.77 | 0.194 | 0.215 | 0.457 | 61.9 | 60.2 | 59.7 | $60.7 \pm 1.2$ |
| 1.9 | 1.58 | 0.210 | 0.234 | 0.491 | 67.1 | 65.4 | 64.1 | $65.6 \pm 1.4$ | 1.76 | 0.210 | 0.237 | 0.510 | 67.3 | 66.4 | 66.6 | $66.8 \pm 0.5$ |
| 2.0 | 1.60 | 0.229 | 0.256 | 0.540 | 73.3 | 71.6 | 70.5 | $71.8 \pm 1.3$ | 1.79 | 0.231 | 0.261 | 0.549 | 73.8 | 73.0 | 71.7 | $72.9 \pm 1.1$ |
| 2.1 | 1.60 | 0.247 | 0.276 | 0.586 | 79.0 | 77.3 | 76.6 | $77.6 \pm 1.2$ | 1.79 | 0.250 | 0.287 | 0.606 | 80.0 | 80.1 | 79.1 | $79.8 \pm 0.5$ |
| 2.2 | 1.58 | 0.264 | 0.296 | 0.632 | 84.5 | 82.8 | 82.6 | $83.3 \pm 1.0$ | 1.77 | 0.258 | 0.299 | 0.648 | 82.5 | 83.7 | 84.6 | $83.7 \pm 1.1$ |
| 2.3 | 1.57 | 0.281 | 0.315 | 0.676 | 89.8 | 88.0 | 88.3 | $88.7 \pm 0.9$ | 1.79 | 0.267 | 0.311 | 0.678 | 85.3 | 86.9 | 88.5 | $86.9 \pm 1.6$ |
| 2.4 | 1.55 | 0.297 | 0.332 | 0.720 | 94.8 | 92.8 | 94.0 | $93.9 \pm 1.0$ | 1.79 | 0.299 | 0.344 | 0.738 | 95.6 | 96.3 | 96.4 | $96.2 \pm 0.4$ |
| 2.5 | 1.49 | 0.312 | 0.348 | 0.763 | 99.8 | 97.4 | 99.6 | $98.9 \pm 1.3$ | 1.78 | 0.311 | 0.357 | 0.759 | 99.4 | 99.7 | 99.1 | $99.5 \pm 0.3$ |

Notes — notation and uncertainties are the same as those described in the notes for Table 1.

### 5.1. Observations

We have obtained a high signal-to-noise spectrum of B1608+656 using the Low-Resolution Imaging Spectrometer (LRIS; Oke et al. 1995) on Keck 1. The data were obtained from the red side of the spectrograph on 12 June 2007 using the 831/8200 grating with the D680 dichroic in place. A slit mask was employed to obtain simultaneously spectra for two additional strong lenses in the field (Fassnacht et al. 2006b) and to continue to probe the structure along the line of sight to the lens (Fassnacht et al. 2006a). The night was clear with a nominal seeing of 0″9, and 10 exposures of 1800s and one exposure of 600s were obtained for a total exposure time of 18600s.

Each exposure was reduced individually using a custom pipeline (see Auger et al. 2008, for details) that performs a single resampling of the spectra onto a constant wavelength grid; the same wavelength grid was used for all exposures to avoid resampling the spectra when combining them, and an output pixel scale of 0.915 Å was used to match the dispersion of the 831/8200 grating. Individual spectra were extracted from an aperture 0″84 wide (corresponding to 4 pixels on the LRIS red side) centered on the peak of the flux of the lensing galaxy G1. The size of the aperture was chosen to avoid contamination from the spectrum of G2 while maximizing the total flux for an improved signal-to-noise ratio. The extracted spectra were combined by clipping the extreme points at each wavelength and taking the variance-weighted sum of the remaining data points. The same extraction and coaddition scheme was performed for a sky aperture to determine the resolution of the output co-added spectrum; we find the resolution to be R = 2560, corresponding to $\sigma_{obs} = 49.7$ km s$^{-1}$. The signal-to-noise ratio per pixel of the final spectrum is $\sim 60$.

### 5.2. Velocity dispersion measurement

We use a Python-based implementation of the velocity-dispersion code from van der Marel (1994), with one important modification. Our implementation allows for a linear sum of template spectra to be modeled using a
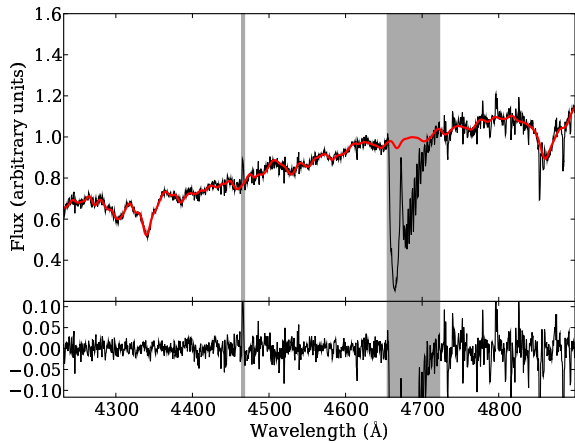
FIG. 2.— The LRIS spectrum of B1608+656 (black line) with a model generated from all 9 INDO-US templates and a 9th order continuum overplotted (red line). The gray shaded areas were not included in the fit, and the lower panel shows the fit residuals. The spectrum and our modeling suggest a central velocity dispersion of $\sigma = 260 \pm 15 \, \mathrm{km \, s^{-1}}$, including systematic errors.

bounded variable least squares solver with the constraint that each template must have a non-negative coefficient. We use a set of templates from the INDO-US stellar library containing spectra for a set of seven K and G giants with a variety of temperatures and spectra for an F2 and an A0 giant. These templates of early-type stars are particularly important for B1608+656, which has a post-starburst spectrum (Myers et al. 1995).

We perform our modeling over a wide range of wavelength intervals and find a stable solution over a variety of spectral features; we therefore choose to use the rest-frame range from 4200 Å to 4900 Å for our fit. The INDO-US templates have a constant-wavelength resolution of 1.2 Å which corresponds to $\sigma_{\mathrm{template}} = 33.6 \, \mathrm{km \, s^{-1}}$ over this wavelength range. We iterate over a range of template combinations and polynomial continuum orders and find that a variety of solutions that vary around $260 \, \mathrm{km \, s^{-1}}$ with a spread of about $13 \, \mathrm{km \, s^{-1}}$ and statistical uncertainties of $7.7 \, \mathrm{km \, s^{-1}}$ (see Figure 2). We therefore adopt a velocity dispersion of $\sigma = 260 \pm 15 \, \mathrm{km \, s^{-1}}$, with the error incorporating the systematic template mismatch and the statistical error for the models. This agrees with the previous measurement of $\sigma_{\mathrm{ap}} = 247 \pm 35 \, \mathrm{km \, s^{-1}}$ by Koopmans et al. (2003) with a significant reduction in the uncertainties, though we note that the two velocity dispersions have been measured in slightly different apertures.

## 6. BREAKING THE MASS-SHEET DEGENERACY: LENS ENVIRONMENT

In this section, we outline two approaches for quantifying the prior probability distributions of the external mass sheet $\kappa_{\mathrm{ext}}$. Computing this quantity such that Equation (7) holds true is not a trivial matter. The non-linearity of strong lensing means that the surface mass density at a given angular position in successive redshift planes between the observer and the source cannot simply be scaled by the appropriate distance ratios and summed: Rather, the deflection angles (which can be large) need to be taken into account when calculating the distortion matrices (which contain and define the external convergence and shear), leading us towards a ray-tracing approach (Hilbert et al. 2009). Detailed in-
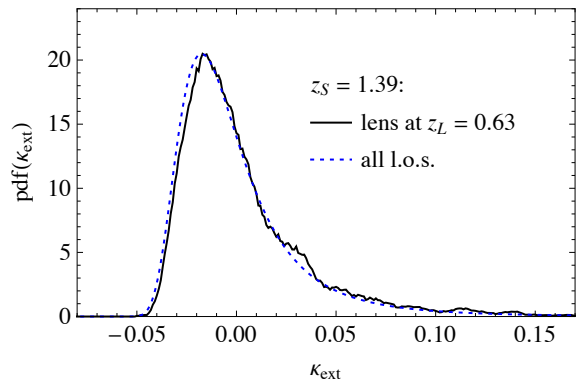


FIG. 3.— Probability distribution for the external convergence $\kappa_{\mathrm{ext}}$ along strongly lensed lines of sight from the Millennium Simulation for the lens redshift $z_L$ and source redshifts $z_S$ of B1608+656 (solid line) compared to the convergence distribution for all lines of sight (dotted line).

vestigation of the ray paths down the B1608+656 light cone is beyond the scope of this paper, and we defer it to a later work (Blandford et al. in preparation). In this section we use the statistics of B1608+656-like fields in numerical simulations to derive a PDF for $\kappa_{\mathrm{ext}}$.

### 6.1. Ray-tracing through the Millennium Simulation

Following Hilbert et al. (2007), we use the multiple-lens-plane algorithm to trace rays through the Millennium Simulation (MS; Springel et al. 2005), one of the largest N-body simulations of cosmic structure formation.[13] We then identify lines of sight where strong lensing by matter structures at $z_{\mathrm{d}} = 0.63$ occurs for sources at $z_{\mathrm{s}} = 1.39$. The convergence along these lines of sight is estimated by summing the projected matter density on the lens planes weighted for a source at $z_{\mathrm{s}} = 1.39$[9] along the ray trajectory. By excluding the primary lens plane at $z_{\mathrm{d}} = 0.63$ that causes the strong lensing, the constructed convergence is truly external to the lens and is due to the line-of-sight contributions only. By sampling many lines of sight, we obtain an estimate for the probability density function of $\kappa_{\mathrm{ext}}$ from simulations. We denote this as the "MS" prior on $\kappa_{\mathrm{ext}}$.

Figure 3 shows the predicted amount of external convergence constructed using $6.4 \times 10^8$ lines of sight (with and without strong lenses) to sources at $z_{\mathrm{s}} = 1.39$: of these, $8.0 \times 10^3$ lines of sight contain strong lenses. For both curves, the mean $\kappa_{\mathrm{ext}}$ is consistent with zero with a spread of $\sim 0.04$.

How should we interpret this distribution? According to its definition, $\kappa_{\mathrm{ext}}$ *could* have contributions from galaxies on the primary lens plane that do not affect the dynamics. Neglecting these contributions (effectively assuming that the lens is an isolated galaxy) might lead to an underestimate of $\kappa_{\mathrm{ext}}$, since most lenses are massive galaxies that often live in over-dense environments like galaxy groups and clusters.[14] However, if the local contribution to the external convergence is accounted for

---

[13] The details of the ray-tracing algorithm are described in Hilbert et al. (2009). The methods for sampling lines of sight, identifying strong lensing events, and calculating the convergence are described in Hilbert et al. (2007). Note that we also include a stellar component in the ray-tracing as described in Hilbert et al. (2008).

[14] It is beyond the scope of this paper to quantify this contribution from our ray-tracing simulations. This would require modeling the lenses and their environment in a way that allows one to split

in the lensing plus dynamics modeling (as discussed in Fassnacht et al. 2006a), then the MS PDF will give an accurate uncertainty in the inferred Hubble constant after marginalization.

Indeed, what the MS PDF also verifies is that *on average* the contribution to the external convergence at a strong lens from line-of-sight structures is almost the same as that for a random line of sight, namely zero. The MS prior therefore suggests that ensembles of *isolated* strong lenses will yield estimates of cosmological parameters that are not strongly biased by line-of-sight structures. The PDF in Figure 3 gives us an idea of by how much individual lenses' line-of-sight $\kappa_{ext}$ values vary, and hence an estimate of the uncertainty on $H_0$ due to this structure. In the absence of any other information, we can assign the Millennium Simulation PDF as a prior on $\kappa_{ext}$ in order to limit the possible values of external convergence to those likely to occur. This assignment has the effect of adding an additional uncertainty of $\sim 0.04$ in $\kappa_{ext}$, with no systematic shift in $\kappa_{ext}$.

### 6.2. Combining galaxy density observations with ray-tracing simulations

The prior discussed in the preceding section does not take into account any information about the environment of B1608+656. Here, we combine knowledge of the lens environment with ray-tracing to obtain a more informative prior on the external convergence.

Fassnacht et al. (2009) compared galaxy number counts in fields around strong galaxy lenses, including B1608+656, with number counts in random fields and in the COSMOS field. Among other measures, they used the number of galaxies with apparent magnitude $18.5 \leq m_{F814W} < 24.5$ in the F814W filter band in apertures of $45''$ radius (300 kpc at the redshift of B1608+656) to quantify the galaxy number density $n_{gal}$ projected along lines of sight. They found that the distribution of $n_{gal}$ for lines of sight containing strong lenses is not very different from that for random lines of sight. However, B1608+656 lies along a line of sight with a galaxy density $n_{gal}$ that is about twice the mean over random lines of sight, $\langle n_{gal} \rangle$. A positive $\kappa_{ext}$ bias can arise through Poissonian fluctuations that are present in the number of groups along the line of sight in the *observed* sample of strong lenses.

We can use this measurement of galaxy number density in the B1608+656 field to generate a more informative prior PDF for $\kappa_{ext}$. As for the MS prior in the previous section, we use the ray-tracing through the MS together with the semi-analytic galaxy model of De Lucia & Blaizot (2007) to quantify the expected external convergence $\kappa_{ext}$ for lines of sight with a given *relative overdensity* $n_{gal}/\langle n_{gal} \rangle$. Dividing out the absolute number of galaxies in the field accounts for differences due to the particular set of cosmological parameters used by the Millennium Simulation and inaccuracies in the galaxy model: We assume that differences in the relative overdensity between the MS cosmology and the true one are small.

We generate 32 simulated fields of $4 \times 4 \deg^2$ on the sky containing the positions and apparent magnitudes[15] of
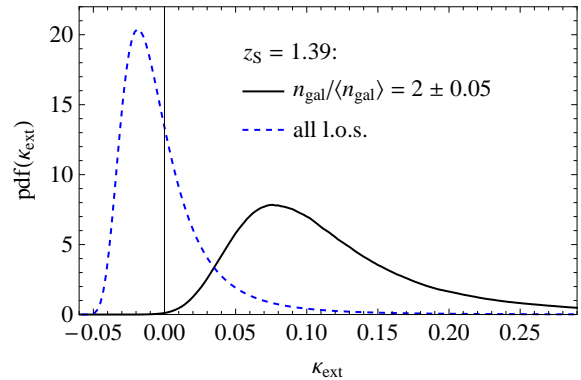


FIG. 4.— Probability distribution for the external convergence $\kappa_{ext}$ obtained from combining results of galaxy number counts around B1608+656 with results from ray-tracing through the Millennium Simulation. Compared are the distribution along lines of sight with a relative galaxy number density $n_{gal}/\langle n_{gal} \rangle = 2.00 \pm 0.05$ (solid line) to the distribution along all lines of sight (dotted line).

the model galaxies at redshifts $0 < z < 5.2$ together with maps of the convergence $\kappa$ to source redshift $z_s = 1.39$. The galaxy positions and magnitudes in the simulated fields are converted into maps of the galaxy density $n_{gal}$. We then select all lines of sight with relative overdensity $1.95 \leq n_{gal}/\langle n_{gal} \rangle < 2.05$ and compute the distribution of the convergence along these lines of sight. The resulting convergence distribution (shown in Figure 4) is then used as prior distribution for the external convergence $\kappa_{ext}$, which we denote as the "OBS" (observations and MS) prior.

The convergence computed in this way is not strictly speaking external convergence, since (i) we do not subtract any contribution from any primary strong lens, (ii) we take all lines of sight and not just those to strong lenses. We are instead building on one of the results of the previous section and assume that the distribution of external convergences is very similar to the distribution of convergences along random lines of sight.

Where this approach becomes inappropriate is where a ray passes close to a galaxy center, and is hence associated with a very large convergence. Assuming such a line of sight as foreground/background for a strong lens galaxy essentially creates a lens system with two or more strong deflectors. These sightlines correspond to compound lenses such as SDSS J0946+1006 (Gavazzi et al. 2008), but not to B1608+656. However, the tail of high convergence values does not pose a problem here: as we will see in Section 8.1 below, the high external convergence is rejected by the dynamics modeling. We expect the mean and width of the PDF in Figure 4 to represent well the possible values of $\kappa_{ext}$ for a field that is over-dense in galaxy number by a factor of two.

Our OBS $\kappa_{ext}$ distribution agrees with earlier estimates from Fassnacht et al. (2006a), who identified and modeled the 4 groups along the line of sight to B1608+656 using various mass assignment recipes. In both approaches, we and Fassnacht et al. (2006a) are concerned primarily with extracting information on the external convergence and not the external shear. If we were to estimate the ex-

---

the mass distribution into a part that is accounted for by the lens model (and constrained by lensing and dynamics data) and a part that acts as external convergence.

[15] The model galaxy catalogs do not provide F814W magnitudes.

We simply approximate $m_{F814W}$ by combining SDSS $i$-band and $z$-band magnitudes to get $m_{F814W} = x_i m_i + (1 - x_i) m_z$ with $x_i = 0.5$. We have checked that our results do not depend strongly on $x_i \in [0, 1]$.

ternal convergence by assigning masses and redshifts to all objects in the B1608+656 field, and then ray tracing through the resulting model mass distribution, the external shear as required in the strong lens modeling would serve as an important calibrator for the external convergence. Such a procedure is beyond the scope of this paper, and we defer it to a future publication (Blandford et al., in preparation). However, we do find (by computing the distribution of external shears in MS fields with different external convergences) that the magnitude of the external shear required by the strong lens modeling ($\gamma_{\mathrm{ext}} \simeq 0.075$) is consistent with the external shear amplitude predicted in the OBS scenario for the B1608+656 field.

### 6.3. *The influence on lens modeling.*

As already remarked, the description of ray propagation in an inhomogeneous cosmology is quite subtle. The matter (dark plus baryonic) density is partitioned between virialized structures (galaxies, groups and clusters) and a depleted background medium. Any structures sufficiently close to the line of sight will imprint convergence and shear onto a ray congruence. Meanwhile the background medium will contribute less Ricci focusing than would be present in a homogeneous, flat universe and will diminish the net convergence.

As the foregoing discussion makes clear, the line of sight to B1608+656 is unusual and we know quite a lot about the photometry and redshifts of the intervening galaxies. It is therefore possible, in principle, to make a refined estimate of the external convergence and shear and to compare the former with the simulations discussed above and the latter with the shear inferred in the lens model described in Paper I. In this way, the shear, again in principle, can be used to calibrate $\kappa_{\mathrm{ext}}$.

There is a second complication that must be addressed. Matter inhomogeneities in front of G1 and G2 distort the image of the primary lens as well as the multiple images of the source. Inhomogeneities behind the lens contribute further distortion in the images of the source. In a more accurate approach, these effects should be taken into account explicitly in the construction of the lens model, while here we are subsuming them in a single correction factor $\kappa_{\mathrm{ext}}$. The way that the resulting corrections affect the inference of a value for $H_0$ turns out to be quite complex. However, it appears that in the particular case of B1608+656, the error that is incurred does not contribute significantly to our quoted errors.

These matters will be discussed in a forthcoming publication.

### 7. PRIORS FOR MODEL PARAMETERS $\boldsymbol{\xi}$

A key goal of this work is to quantify the impact of the most serious systematic errors associated with using time-delay lenses for cosmography. Our approach is to characterize these errors as nuisance parameters, and then investigate the effects of various choices of prior PDF on the inference of cosmological parameters. To this end, we use either well motivated priors based on the results of Section 4, Section 6 and other independent studies, or, for contrast, uniform (maximally ignorant) prior PDFs. We now describe our choices for each parameter in turn.

- $P(\boldsymbol{\pi})$. We consider a set of four cosmological parameters, $\boldsymbol{\pi} = \{H_0, \Omega_{\mathrm{m}}, \Omega_\Lambda, w\}$. We then assign the following four different joint prior PDFs:

  **K03:** uniform prior on $H_0$ between 0 and 150 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{m}} = 0.3$, $\Omega_\Lambda = 0.7$, and $w = -1$. This is the cosmology that was assumed in Koopmans et al. (2003) (the most recent $H_0$ measurement from B1608+656 before this work), and is the cosmology that is typically assumed in the literature for measuring $H_0$ from time-delay lenses. This form of prior allows us to compare our $H_0$ to earlier work.

  **UNIFORM** priors on all four cosmological parameters, with either the $w = -1$ or the flatness ($\Omega_{\mathrm{m}} = 1 - \Omega_\Lambda$) constraint imposed. These priors allow us to quantify the information in the B1608+656 data set as conservatively as possible.

  **WMAP5:** WMAP 5 year data set posterior PDF for $\{H_0, \Omega_{\mathrm{m}}, \Omega_\Lambda, w\}$, assuming either $w = -1$ or a flat geometry. This allows us to constrain either flatness or $w$ by combining B1608+656 with WMAP.

  **WBS:** Joint posterior PDF for $\{H_0, \Omega_\Lambda, w\}$ with a flat geometry, given the WMAP5 data in combination with compendia of baryonic acoustic oscillations (BAO) and supernovae (SN) data sets. This allows us to quantify the gain in precision made when incorporating B1608+656 into the current global analysis.

  The last two priors are defined by the Markov chains provided by the WMAP team[16] based on the analysis performed by Dunkley et al. (2009) and Komatsu et al. (2009).

- $P(\gamma')$. We consider three different prior PDFs for the density profile slope. In the first two priors, we ignore the B1608+656 ACS data (i.e., dropping $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$ in Equation (24)); these first two are controls, to allow the assessment of the amount of information contained in the ACS data.

  **Uniform:** a maximally ignorant prior PDF, defined in the range $1.5 \leq \gamma' \leq 2.5$.

  **SLACS:** This is a Gaussian prior based on the result from the SLACS project: $\gamma' = 2.08 \pm 0.2$ (Koopmans et al. 2009). This was derived from a sample of low-redshift massive elliptical lenses, studied with combined strong lens and stellar dynamics modeling. We note that this was obtained without considering the presence of any external convergence $\kappa_{\mathrm{ext}}$. However, Treu et al. (2009) find that the environmental effects in the SLACS lenses are smaller than their measurement errors and are typically undetected. Since SLACS lenses do

---

[16] http://lambda.gsfc.nasa.gov

not require an external shear in the modeling, typical $\kappa_{\rm ext}$ values for these lenses are expected to be small. Only in a few extreme cases does the $\kappa_{\rm ext}$ reach values of order 0.05–0.10. Therefore, we take directly the prior on the slope from SLACS lenses without corrections for $\kappa_{\rm ext}$.

**ACS:** This prior is the PDF $P(\gamma'|\boldsymbol{d}, \boldsymbol{M}_{\rm D} = \boldsymbol{M}_5)$ obtained from the analysis of the ACS image of B1608+656 in Section 4.2.

- $P(\boldsymbol{\eta})$. As described in Section 4.2, we use the radio observations and the NICMOS F160W images of B1608+656 to constrain the smooth lens model parameters $\boldsymbol{\eta}$ for a given slope $\gamma'$. The posterior PDF from this analysis forms the prior PDF for the current work.

- $P(\kappa_{\rm ext})$. We consider three forms of prior for the external convergence:

  **Uniform** between $-0.25$ and $+0.25$: again, such a maximally ignorant prior, again to provide contrast.

  **MS:** from the strong lenses in the MS, discussed in Section 6.1.

  **OBS:** from the galaxy number counts in the field of B1608+656 and the MS, discussed in Section 6.2.

- $P(r_{\rm ani})$. For the lens galaxy stellar orbit radial anisotropy parameter $r_{\rm ani}$, we simply assign a uniform prior between $0.5r_{\rm eff}$ and $5r_{\rm eff}$, where $r_{\rm eff}$ is the effective radius for the velocity dispersion measurement. The inner cutoff is motivated by observations (e.g., Kronawitter et al. 2000) and radial instability arguments (e.g., Merritt & Aguilar 1985; Stiavelli & Sparke 1991), while the outer cutoff is for computational simplicity (the model velocity dispersion changes by a negligible amount between $r_{\rm ani} = 5r_{\rm eff}$ and $r_{\rm ani} \to \infty$).

These priors are summarized in Table 3.

## 8. INFERENCE OF $H_0$ AND DARK ENERGY PARAMETERS FROM B1608+656

In this section we present the results of the analysis outlined in Section 3, putting together all the likelihood functions and prior PDFs described in Sections 4 to 7. We obtain $P(\boldsymbol{\pi}|\boldsymbol{\Delta t}, \boldsymbol{d}, \sigma)$ by importance sampling, using the two likelihoods in Equation (27) as the weights for the various priors on $\gamma'$, $\kappa_{\rm ext}$, $r_{\rm ani}$, and $\boldsymbol{\pi}$ listed in Table 3 (see Appendix A.2 for details). Focusing first on the systematic errors now quantified as the nuisance parameters $\gamma'$ and $\kappa_{\rm ext}$, we gradually increase the complexity of the cosmological model to probe the full space of parameters.

For each possible combination of the priors on the parameters in Table 3, we generate 96000 samples of $\gamma'$, $\kappa_{\rm ext}$, $r_{\rm ani}$, and $\boldsymbol{\pi}$ to characterize the prior probability distribution. We also have two types of stellar distribution functions, Hernquist and Jaffe, for modeling the stellar velocity dispersion; we find that the two different types of stellar distribution function produce nearly identical PDFs for the cosmological parameters. Since the priors on the parameters play a greater role than does the

choice of stellar dynamics model, we focus only on the Hernquist stellar distribution function for the remainder of the section.

### 8.1. *Exploring the degeneracies among $H_0$, $\gamma'$ and $\kappa_{\rm ext}$*

To investigate the impact of our limited knowledge of the lens density profile slope $\gamma'$ and external convergence $\kappa_{\rm ext}$, we first fix the cosmological parameters $\Omega_{\rm m}$, $\Omega_\Lambda$ and $w$ according to the K03 prior. This allows us a simplified view of the problem, and also a comparison with previous work that used this rather restrictive prior.

We first assign the OBS prior for $\kappa_{\rm ext}$, and look at the effect of the various choices of density profile slope priors. The left-hand panel in Figure 5 shows the marginalized posterior PDF for $H_0$ for the three different priors for $\gamma'$ given in Table 3. From this graph, we see that the SLACS prior gives a similar estimate of $H_0$ as the uniform prior with a negligible increase in precision. The ACS prior lowers $H_0$ relative to that of the SLACS and uniform priors, and improves the precision in $H_0$ to 4.4%. Overall, the impact of the prior on $\gamma'$ is relatively low in the sense that, even with a uniform prior on $\gamma'$, $H_0$ is still constrained to 7% (taking $H_0 = 70.6$ as our reference value). For the remainder of this paper, we assign the ACS prior.

As expected, the prior for $\kappa_{\rm ext}$ has a greater effect, shown in the right-hand panel of Figure 5. Taking the maximally informative OBS prior as our default, we see that relaxing this to the MS prior causes an increase in inferred $H_0$ value of some $6\,{\rm km\,s^{-1}\,Mpc^{-1}}$, and relaxing further to a uniform prior increases it by $12\,{\rm km\,s^{-1}\,Mpc^{-1}}$. The precision in $H_0$ also drops by more than a factor of two from the OBS prior to the uniform prior. Our knowledge of $\kappa_{\rm ext}$ is therefore limiting the inference of $H_0$.

We note that the stellar dynamics contain a significant amount of information on $H_0$. The stellar dynamics effectively constrain $\kappa_{\rm ext}$ and $\gamma'$ to an approximately linear relation, where an increase in $\kappa_{\rm ext}$ requires a steepening of the slope in order to keep the predicted velocity dispersion the same. Therefore, for a fixed range of $\gamma'$ values, the modeling of the stellar dynamics would only permit a corresponding range of $\kappa_{\rm ext}$ values. Specifically, without dynamics as constraints, we find $H_0 = 68.1^{+3.7}_{-6.4}\,{\rm km\,s^{-1}\,Mpc^{-1}}$ for the ACS and OBS priors. The lower bound on $H_0$ is somewhat weakened by the high tail of the OBS $\kappa_{\rm ext}$ distribution. On the other hand, this high tail is rejected by the use of the dynamics data, hence our resulting tight constraint on $H_0$.

To summarize, using all available information on B1608+656 and the ACS and OBS priors gives $H_0 = 70.6 \pm 3.1\,{\rm km\,s^{-1}\,Mpc^{-1}}$, a precision of 4.4%. We interpret Figure 5 as evidence that we are approaching saturation in the information we have on the lens model for B1608+656: the mass model is now so well constrained that the inference of cosmological parameters from this system is limited by our knowledge of the lens environment. We now explore this joint inference in more detail, first putting it in some historical context.

### 8.2. *Comparison with other lensing $H_0$ results*

What improvement in the measurement of $H_0$ do we gain from our new observations of B1608+656? The most

| $P(\gamma')$ | uniform ($1.5 \leq \gamma' \leq 2.5$) | SLACS ($\gamma' = 2.08 \pm 0.2$) | ACS ($\gamma' = 2.08 \pm 0.03$) |
|---|---|---|---|
| $P(\kappa_{ext})$ | uniform ($-0.25 \leq \kappa_{ext} \leq 0.25$) | MS (Millennium Simulations; Figure 3) | OBS (Observations and MS; Figure 4) |
| $P(r_{ani})$ | uniform ($0.5 r_{eff} \leq r_{ani} \leq 5 r_{eff}$) | | |
| $P(\boldsymbol{\pi})$ | K03 ($\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$, uniform $H_0 \in [0, 150]$ km s$^{-1}$ Mpc$^{-1}$) | UNIFORMopen ($w = -1$, $\Omega_m$ and $\Omega_\Lambda$ uniform $\in [0,1]$, uniform $H_0 \in [0, 150]$ km s$^{-1}$ Mpc$^{-1}$) | UNIFORMw ($\Omega_\Lambda = 1 - \Omega_m$ uniform $\in [0,1]$, uniform $w \in [-2.5, 0.5]$, uniform $H_0 \in [0, 150]$ km s$^{-1}$ Mpc$^{-1}$) |
| | WMAPopen (WMAP5 with $w = -1$) | WMAPw (WMAP5 with flatness and time-independent $w$) | WBSw (WMAP5 + BAO + SN with flatness and time-independent $w$) |

Notes — The K03 entry for $P(\boldsymbol{\pi})$ is the same prior as in Koopmans et al. (2003). This is also the most common cosmology prior assumed in previous studies of time-delay lenses.
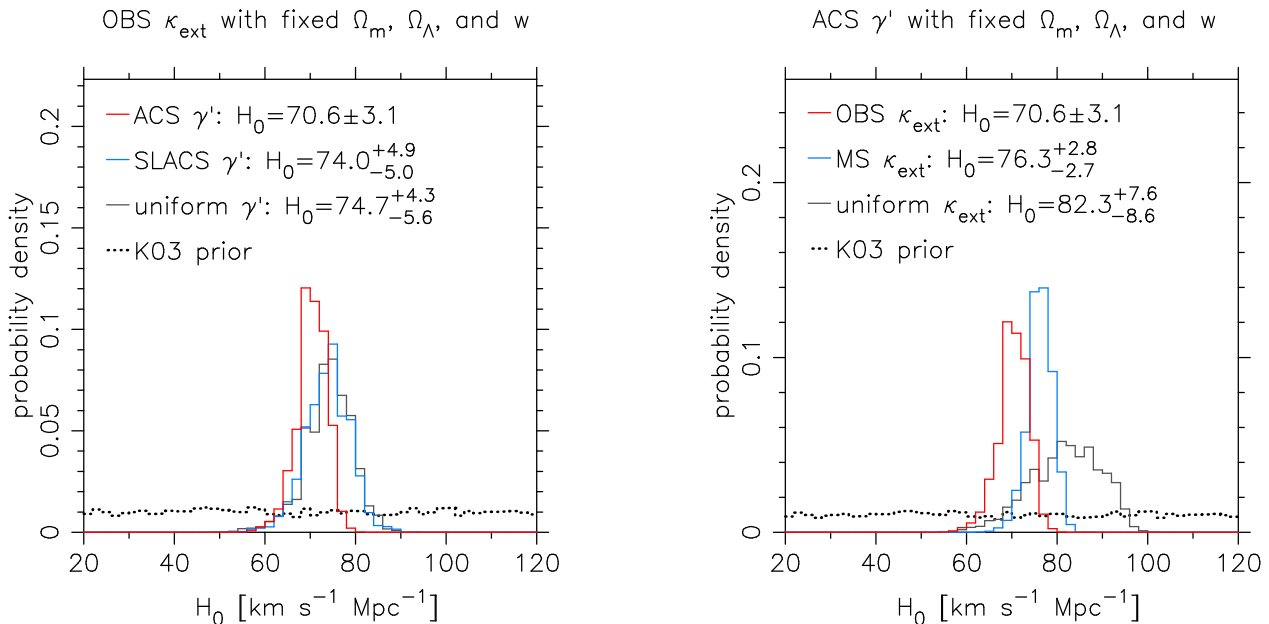


FIG. 5.— Left: the marginalized posterior PDF for $H_0$ assuming K03 cosmology and OBS $\kappa_{ext}$ priors. Right: the marginalized posterior PDF for $H_0$ assuming K03 cosmology and ACS $\gamma'$ priors. The prior uncertainty in external convergence determines the precision of the inferred Hubble constant.

recent measurement before this work by Koopmans et al. (2003) was $H_0 = 75^{+7}_{-6}$ km s$^{-1}$ Mpc$^{-1}$. This result was based on a joint lensing and dynamics modeling using the radio data, shape of the Einstein ring from the NIC-MOS images and the earlier less precise velocity dispersion measurement. Our improved analysis using the deep ACS images and the newly measured velocity dispersion reduce the uncertainty by more than a factor of two, even with the inclusion of the systematic error due to the external convergence that was previously neglected. We attribute our lower $H_0$ value to our incorporation of the realistically-skewed OBS $\kappa_{ext}$.

Let us now compare our $H_0$ measurement based on the K03 cosmology to several recent measurements (within the past five years) from other time-delay lenses. Most analyses assumed $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ — we point out explicitly the few that did not. In B0218+357, Wucknitz et al. (2004) measured $H_0 = 78 \pm 6$ km s$^{-1}$ Mpc$^{-1}$ ($2\sigma$) by modeling this two-image lens system with isothermal elliptical potentials (and effectively measuring $\gamma'$, see Section 4.2.3) but neglecting external convergence. York et al. (2005) refined this using the centroid position of the spiral lens galaxy based on $HST$/ACS observations as a constraint; depending on the spiral arm mask-

ing, they found $H_0 = 70 \pm 5$ km s$^{-1}$ Mpc$^{-1}$ (unmasked) and $H_0 = 61 \pm 7$ km s$^{-1}$ Mpc$^{-1}$ (masked) (both with $2\sigma$ errors). In the two-image FBQ 0951+2635, Jakobsson et al. (2005) obtained $H_0 = 60^{+9}_{-7}$ (random, $1\sigma$) $\pm 2$ (systematic) km s$^{-1}$ Mpc$^{-1}$ for a singular isothermal ellipsoid model and $H_0 = 63^{+9}_{-7}$ (random, $1\sigma$) $\pm 1$ (systematic) km s$^{-1}$ Mpc$^{-1}$ for a constant mass-to-light ratio model, again ignoring external convergence. In the two-image quasar system SDSS J1650+4251, Vuissoz et al. (2007) found $H_0 = 51.7^{+4.0}_{-3.0}$ km s$^{-1}$ Mpc$^{-1}$ assuming a singular isothermal sphere and constant external shear for the lens model. More general lens models considered by these authors (e.g. including lens ellipticity, or using a de Vaucouleurs density profile) were found to be underconstrained. In the two-image quasar system SDSS J1206+4332, Paraficz et al. (2009) found $H_0 = 73^{+3}_{-4}$ km s$^{-1}$ Mpc$^{-1}$ using singular isothermal ellipsoids or spheres to describe the three lens galaxies. (Photometry was used to place additional constraints on the lens parameters.) Recently, Fadely et al. (2009) modeled the gravitational lens Q0957+561 using four different dark matter density profiles, each with a stellar component. The lens is embedded in a clus-

ter, and the authors constrained the corresponding mass sheet using the results of a weak lensing analysis by Nakajima et al. (2009). Assuming a flat universe with $\Omega_{\rm m} = 0.274$ and cosmological constant $\Omega_\Lambda = 0.726$, they found $H_0 = 85^{+14}_{-13}\,{\rm km\,s^{-1}\,Mpc^{-1}}$, where the principle uncertainties were due to the weakly constrained stellar mass-to-light ratio (a manifestation of the radial profile degeneracy in the lens model). Imposing constraints from stellar population synthesis models led to $H_0 = 79.3^{+6.7}_{-8.5}\,{\rm km\,s^{-1}\,Mpc^{-1}}$.[17]

In a nutshell, most of the recent $H_0$ measurements from individual systems assumed isothermal profiles, and neglected the effects of both $\gamma'$ and $\kappa_{\rm ext}$: we interpret the significant variation between the $H_0$ estimates in the recent literature as being due to these model limitations. In contrast, our B1608+656 analysis explicitly incorporates the uncertainties due to our lack of knowledge of both $\gamma'$ and $\kappa_{\rm ext}$. In fact, a spread of $\sim 0.2$ in $\gamma'$ around 2.0 would give a spread of $\sim 40\%$ in $H_0$ for the cases where isothermal lenses are assumed (Wucknitz 2002). These in turn are set by a lack of information on the systems, either because only two images are formed, or the extended source galaxy is not observed.

Other groups have looked to improve the constraints on $H_0$ by combining several lenses together in a joint analysis. Using a sample of 10 time-delay lenses, Saha et al. (2006) measured $H_0 = 72^{+8}_{-11}\,{\rm km\,s^{-1}\,Mpc^{-1}}$ by modeling the lens' convergence distributions on a grid and using the point image positions of the lenses as constraints (the PixeLens method). Coles (2008) improved on the method and obtained $H_0 = 71^{+6}_{-8}\,{\rm km\,s^{-1}\,Mpc^{-1}}$ while addressing more clearly their prior assumptions. Oguri (2007) used a sample of 16 time-delay lenses to constrain $H_0 = 68\pm6({\rm stat.})\pm8({\rm syst.})\,{\rm km\,s^{-1}\,Mpc^{-1}}$ (for $\Omega_{\rm m} = 0.24$ and $\Omega_\Lambda = 0.76$; see footnote 17) by employing a statistical approach based on the image configurations. By simultaneously modeling SDSS J1206+4332 with four other systems using PixeLens, Paraficz et al. (2009) derive $H_0 = 61.5^{+8}_{-4}\,{\rm km\,s^{-1}\,Mpc^{-1}}$. The larger quoted error bars on these ensemble estimates are perhaps a reflection of the paucity of information available for each lens, as discussed above. All four analyses effectively assume that the ensemble external convergence distribution has zero mean, which may not be accurate: for example, Oguri (2007) constructed a sample for which external convergence could be neglected, and then incorporated this into the systematic error budget. Furthermore, Oguri (2007) imposed a Gaussian prior on the slope of $\gamma' = 2.00\pm0.15$, and the PixeLens method's priors on $\kappa$ may well implicitly impose constraints on $\gamma$ that are similar to the prior in Oguri (2007); these priors on the slope may not be appropriate for individual systems in the ensembles.

In contrast, our measurement of $\gamma'$ from the ACS data means that our results are independent of external priors on $\gamma'$. In fact, our detailed study of the single well-observed lens B1608+656, even incorporating the effects of $\kappa_{\rm ext}$, constrains $H_0$ better than the studies using ensembles of lenses. Our claim is that our analysis of the systematic effects in B1608+656 — explicitly including density profile slope and external convergence as nuisance parameters — is one of the most extensive on a single lens, and is rewarded with one of the most accurate measurements of $H_0$ from time-delay lenses.

### 8.3. *Relaxing the K03 prior*

As we described in Section 2, strong lens time delays enable a measurement of a cosmological distance-like quantity, $D_{\Delta{\rm t}} \equiv (1 + z_{\rm d})D_{\rm d}D_{\rm s}/D_{\rm ds}$. While there is some slight further dependence on cosmology in the stellar dynamics modeling, we expect this particular distance combination to be well constrained by the system. To illustrate this, we plot in Figure 6 the PDF for $D_{\Delta{\rm t}}$ with and without the constraints from B1608+656, for various choices of the cosmological parameter prior PDF. Specifically, we show the effect of relaxing the prior on $\Omega_{\rm m}$, $\Omega_\Lambda$ and $w$ from the K03 delta function to the two types of uniform distributions detailed in Table 3: "UNIFORMopen" and "UNIFORMw". We see that all of these distributions predict the same uninformative prior for $D_{\Delta{\rm t}}$, and that the B1608+656 posterior PDFs are correspondingly similar. With the OBS and ACS priors for $\kappa_{\rm ext}$ and $\gamma'$, we estimate $D_{\Delta{\rm t}} \simeq (5.16^{+0.29}_{-0.24}) \times 10^3\,{\rm Mpc}$, a precision of $\sim 5\%$. The difference between the $D_{\Delta{\rm t}}$ estimates among the three priors shown is $\lesssim 2\%$.
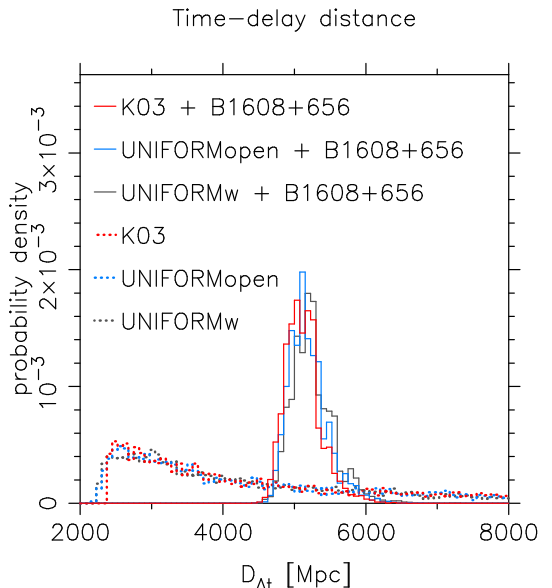


FIG. 6.— PDFs for $D_{\Delta{\rm t}}$, showing the B1608+656 posterior constraints on $D_{\Delta{\rm t}}$ (solid) given assorted uniform priors for the cosmological parameters (dotted, labeled). See the text for a full description of these various priors. In this figure we assign the ACS and OBS priors for $\gamma'$ and $\kappa_{\rm ext}$. B1608+656 provides tight constraints on $D_{\Delta{\rm t}}$, which translates into information about $\Omega_{\rm m}$, $\Omega_\Lambda$ and $w$ as well as $H_0$.

Figure 6 suggests that a simple Gaussian approximation for the product of the B1608+656 likelihood function, marginalized over the OBS and ACS priors, is an appropriate compression of our results. We find that

$$P(D_{\Delta{\rm t}}|H_0, \Omega_{\rm m}, \Omega_\Lambda, w) \simeq$$
$$\frac{1}{\sqrt{2\pi\sigma_{\rm D}^2}} \cdot \exp\left[-\frac{(D_{\Delta{\rm t}} - D_0)^2}{2\sigma_{\rm D}^2}\right], \qquad (29)$$

where $D_0 = 5226\,{\rm Mpc}$ and $\sigma_{\rm D} = 206\,{\rm Mpc}$, accurately reproduces the cosmological parameter inferences: for example, Hubble's constant is recovered to 2.3%.

[17] The corresponding $H_0$ for the K03 cosmology is within $\sim 0.1\%$ of the listed values.

| | | |
|---|---|---|
| WMAP5[a,b] | $-0.285 < \Omega_k < 0.010$ | 15% |
| WMAP5 + $HST$ KP[b,c] | $-0.052 < \Omega_k < 0.013$ | 3.3% |
| WMAP5 + SN[b,d] | $-0.032 < \Omega_k < 0.008$ | 2.0% |
| WMAP5 + BAO[b,e] | $-0.017 < \Omega_k < 0.007$ | 1.2% |
| **WMAP5 + B1608** | $\mathbf{-0.031 < \Omega_k < 0.009}$ | **2.0%** |

The third column gives the "precision," quantified as half the 95% confidence interval in $(1.0 - \Omega_k)$, as a percentage. [a] http://lambda.gsfc.nasa.gov [b] Komatsu et al. (2009). [c] Freedman et al. (2001). [d] Based on the "union" SN samples compiled by Kowalski et al. (2008). [e] Percival et al. (2007).

### 8.4. Constraints on $\Omega_m$ and $\Omega_\Lambda$

Based on the construction of $D_{\Delta t}$, we expect strong lens time delays to be more sensitive to $H_0$ than the other three cosmological parameters. This is shown in Figure 7, where we consider $w = -1$ uniform cosmological prior "UNIFORMopen" and plot the marginalized B1608+656 posterior PDF to show the influence of the lensing data (blue lines). While there is a slight dependence on $\Omega_\Lambda$, we see that the B1608+656 data do indeed primarily constrain $H_0$. In contrast, we plot the posterior PDF from the analysis of the 5 year WMAP data set (red lines, Dunkley et al. 2009). With no constraint on the curvature of space, the CMB data provides only a weak prior on $H_0$, which is highly degenerate with $\Omega_m$ and $\Omega_\Lambda$. Importance sampling the WMAP MCMC chains with the B1608+656 likelihood, we obtain the joint posterior PDF, plotted in black.

Strong lens time delays are an example of a kinematic cosmological probe, i.e., one that is sensitive to the geometry and expansion rate of the Universe, but not to dynamical assumptions about the the growth of structure in the Universe. In Table 4, we compare the B1608+656 data set to a number of other kinematic probes from the literature. The WMAP data constrain the angular diameter distance to the last scattering surface; these other data sets effectively provide a second distance estimate that breaks the degeneracy between $H_0$ and the curvature of space. In the B1608+656 case, we constrain $\Omega_k$ to be $-0.005^{+0.014}_{-0.026}$ (95% CL). We can see that in terms of constraining the curvature parameter, B1608+656 is more informative than the $HST$ Key Project $H_0$ measurement, and is comparable to the current Type Ia SNe data set.

Figure 7 also shows the primary nuisance parameter, $\kappa_{ext}$. When B1608+656 and the WMAP data are combined, the PDF for $\kappa_{ext}$ shifts and tightens very slightly, as we expect from the discussion in Section 8.1. If we relax the OBS prior on $\kappa_{ext}$ to uniform, then we obtain $-0.032 < \Omega_k < 0.021$ (95% CL), which is still tighter than the $HST$ KP constraints.

### 8.5. Constraints on dark energy

As noted by many authors, the degeneracy-breaking shown in the previous subsection can be recast as a mechanism for constraining the equation of state of dark energy, $w$. If we assert a precisely flat geometry for the Universe, as motivated by the inflationary scenario, we can spend our available information on constraining $w$ instead. Figure 8 shows the marginalized posterior PDF

for the cosmological parameter $H_0$, $\Omega_\Lambda = 1 - \Omega_m$ and $w$, along with the nuisance parameter $\kappa_{ext}$, again comparing the B1608+656 constraints with uniform and WMAP priors, and the WMAP constraints alone. With the WMAP data alone, $w$ is strongly degenerate with $H_0$ and $\Omega_\Lambda$. Including B1608+656, which mainly provides constraints on $H_0$, the $H_0$-$w$-$\Omega_\Lambda$ degeneracy is partly broken. The resulting marginalized distribution gives $w = -0.94^{+0.17}_{-0.19}$, consistent with a cosmological constant. The corresponding value of Hubble's constant is $H_0 = 69.7^{+4.9}_{-5.0}\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

We summarize our inferences of $H_0$ and $w$ in this variable-$w$ model in Table 5, comparing to a similar set of alternative kinematic probes referred to in the previous section. We see that, combining with the WMAP 5 year data set and marginalizing over all other parameters, the B1608+656 data set provides a measurement of Hubble's constant with an uncertainty of 6.9%, with the equation of state parameter simultaneously constrained to 18%. This level of precision is better than that available from the $HST$ KP and is competitive to the current BAO measurements.

Our results are consistent with the results from all the other probes listed. This is not a trivial statement: combining each data set with the WMAP 5 year prior allows us not only to quantify the relative constraining power of each one, it also retains the possibility of detecting inconsistencies between data sets. As it is, it appears that all the kinematic probes listed are in agreement within their quoted uncertainties. Some tension might be present if the supernovae and B1608+656 were considered separately from a combination of local $HST$ $H_0$ measurements and BAO constraints, but we have no compelling reason to make such a division. As the statistical errors associated with each probe are decreased, other inconsistencies may arise: we might expect there to always be a need for careful pairwise data set combinations.

Finally then, we incorporate B1608+656 into a global analysis of cosmological data sets. As an example, we importance sample from the WBSw prior PDF; this is the joint posterior PDF from the joint analysis of WMAP5, BAO and SN data. This prior is already very tight, characterized by a median and 68% confidence limits of $H_0 = 70.3^{+1.6}_{-1.5}$. When we include information from B1608+656 with the ACS $\gamma'$ and OBS $\kappa_{ext}$ priors, we obtain $H_0 = 70.4^{+1.5}_{-1.4}$, a slight shift in centroid and 6% reduction in the confidence interval. This is good as it shows global consistency in the WMAP5, BAO, SN and B1608+656 data sets.

### 8.6. Future prospects

In this paper, we have studied a single strong gravitational lens, B1608+656, investigating in depth the various model parameter degeneracies and systematic effects. At present, B1608+656 remains the only strong lens system with (i) time delay measurements with errors of only a few percent, and (ii) extended source surface brightness distribution for accurate lens modeling; as we have shown, these two properties together enable the careful study and the resulting tight constraint on $H_0$.

Table 5 shows that even this one system provides competitive accuracy on $H_0$ and $w$ for a single kinematic probe, especially when we consider that all the other experiments involved averaging together many independent
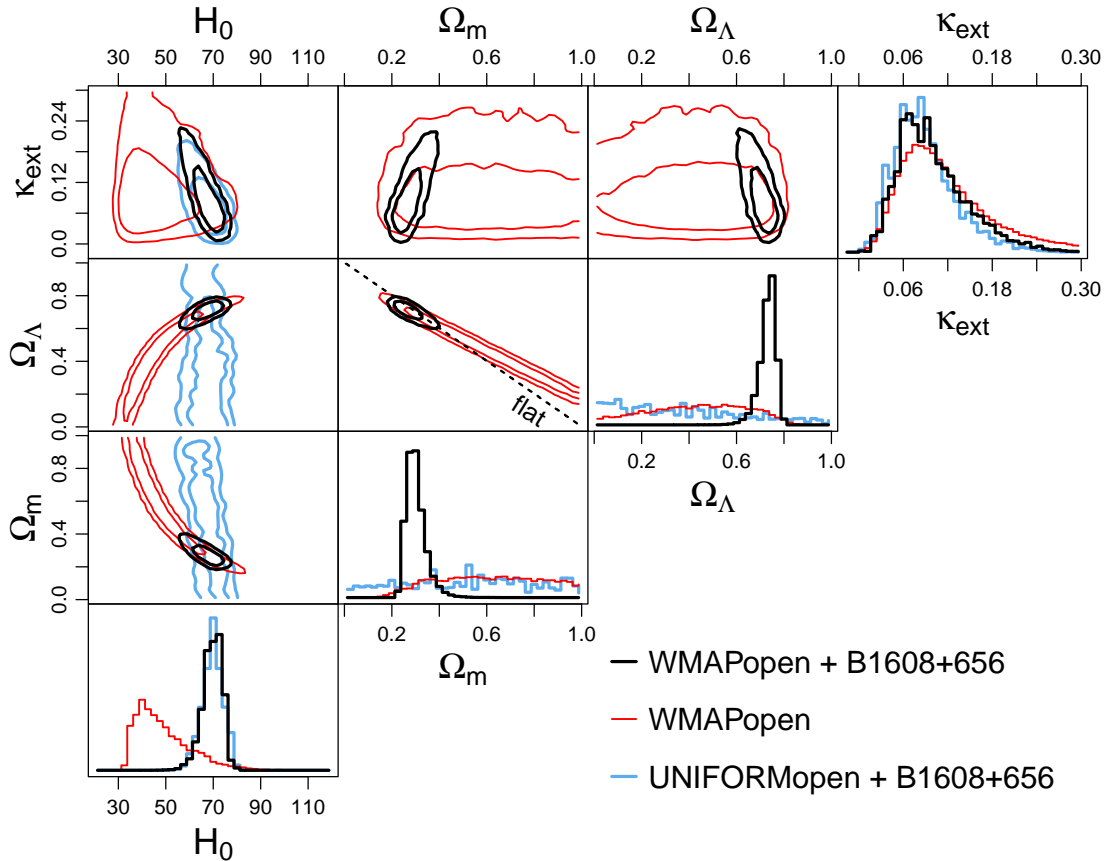
FIG. 7.— The B1608+656 marginalized posterior PDF for $H_0$, $\Omega_{\mathrm{m}}$, $\Omega_{\Lambda}$ and $\kappa_{\mathrm{ext}}$ in a $w = -1$ cosmological model and assuming ACS $\gamma'$ and OBS $\kappa_{\mathrm{ext}}$ priors; contours are 68% and 95% confidence levels. The three sets of colored contours correspond to three different prior/data set combinations. Blue: B1608+656 constraints, given the UNIFORMopen prior; red: the prior provided by the WMAP 5 year data set alone; black: the joint constraints from combining WMAP and B1608+656. The blue contours in the $\Omega_{\mathrm{m}}$ and $\Omega_{\Lambda}$ columns are omitted since they would show almost no constraints, as indicated by the diagonal panels.

TABLE 5
DARK ENERGY CONSTRAINTS FROM WMAP5 COMBINED WITH VARIOUS DATA SETS, ASSUMING FLAT GEOMETRY.

|  | $H_0/\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ |  | $w$ |  |
| --- | --- | --- | --- | --- |
| WMAP5[a,b] | $74^{+15}_{-14}$ | 20% | $-1.06^{+0.41}_{-0.42}$ | 42% |
| WMAP5+$HST$ KP[a,b,c] | $72.1^{+7.4}_{-7.6}$ | 10% | $-1.01^{+0.23}_{-0.22}$ | 23% |
| WMAP5+SN[a,b,d] | $69.4^{+1.6}_{-1.7}$ | 2.3% | $-0.977^{+0.065}_{-0.064}$ | 6.5% |
| WMAP5+BAO[a,b,e] | $73.9^{+4.7}_{-4.8}$ | 6.6% | $-1.15^{+0.21}_{-0.22}$ | 22% |
| WMAP5+Riess[f] | $74.2 \pm 3.6$[g] | 5.0% | $-1.12 \pm 0.12$ | 12% |
| **WMAP5+B1608** | $\mathbf{69.7^{+4.9}_{-5.0}}$ | **6.9%** | $\mathbf{-0.94^{+0.17}_{-0.19}}$ | **18%** |

The "precisions" in the third and fifth columns are defined as half the 68% confidence interval, as a percentage of either 72 for $H_0$ or -1.0 for $w$.   [a] http://lambda.gsfc.nasa.gov   [b] Komatsu et al. (2009). The $H_0$ estimate was taken from the previously listed website.   [c] Freedman et al. (2001).   [d] Based on the "union" SN samples compiled by Kowalski et al. (2008).   [e] Percival et al. (2007).   [f] Riess et al. (2009).   [g] not marginalized over other cosmological parameters.

distance measurements. What should we expect from extending this study to many more lenses? As we showed in Section 8.1, if the data are good enough to constrain the density profile slope to a few percent, the accuracy of the cosmological parameter inference is limited, as it is in B1608+656, by our knowledge of the lens environment, $\kappa_{\mathrm{ext}}$.

However, we also outlined in Section 6 how using information from numerical simulations and the photometry

in the field can be used to constrain this nuisance parameter and yield an unbiased estimate of $H_0$. Furthermore, as discussed in Section 8.1, stellar dynamics provides significant amount of information on $\kappa_{\mathrm{ext}}$ by limiting its permissible range of values. While we, and also Treu et al. (2009) and Fassnacht et al. (2009), discuss how the line-of-sight contributions to $\kappa_{\mathrm{ext}}$ should average to zero over many lens systems, lens galaxies — like all massive galaxies — tend to live in locally overdense environ-
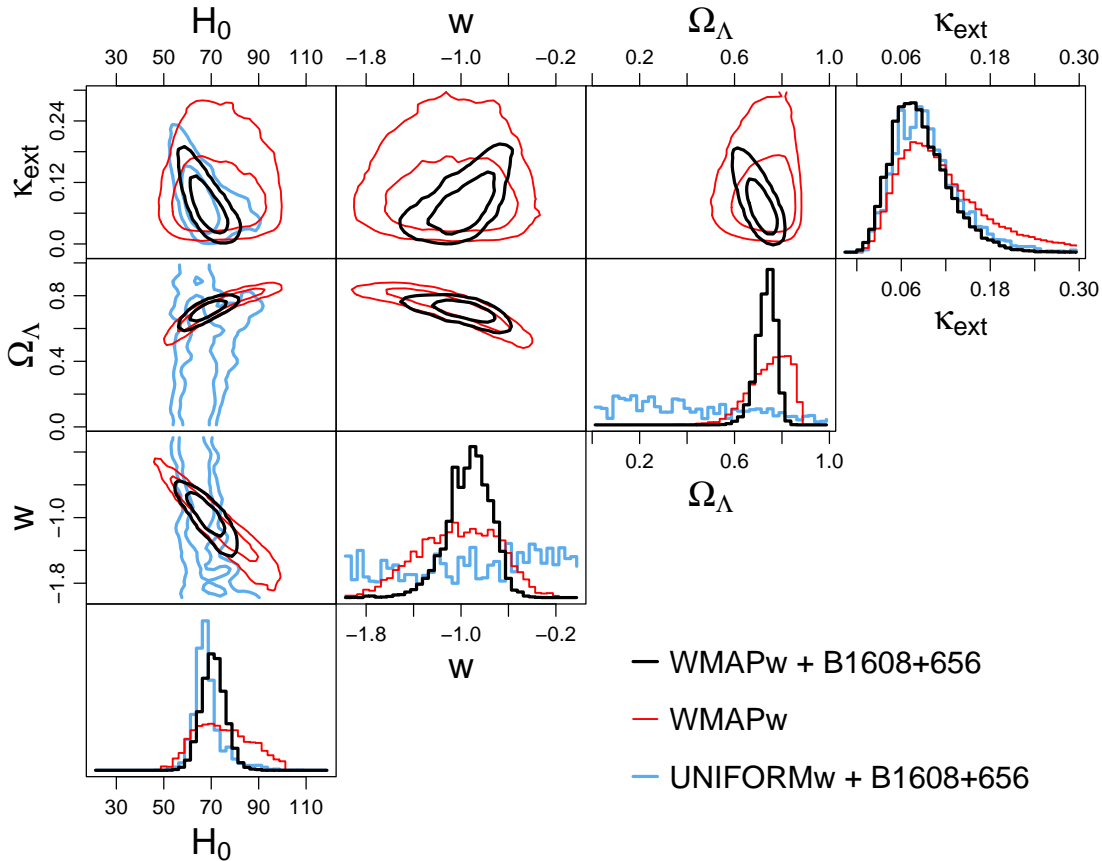
FIG. 8.— The B1608+656 marginalized posterior PDF for $H_0$, $\Omega_\Lambda$, $w$ and $\kappa_{ext}$ in a flat cosmological model, again assuming ACS $\gamma'$ and OBS $\kappa_{ext}$ priors; contours are 68% and 95% confidence levels. The three sets of colored contours correspond to three different prior/data set combinations. Blue: B1608+656 constraints, given the UNIFORMw prior; red: the prior provided by the WMAP 5 year data set alone; black: the joint constraints from combining WMAP and B1608+656. The blue contours in the $\Omega_m$ and $\Omega_\Lambda$ columns are omitted since they would show almost no constraints, as indicated by the diagonal panels.

ments, such that the local contribution to $\kappa_{ext}$ would be non-zero. Careful studies of the lens environments (e.g. Momcheva et al. 2006; Fassnacht et al. 2006a; Blandford et al. in preparation) and of N-body simulations with gas physics to determine this local contribution to $\kappa_{ext}$ will be crucial for obtaining $H_0$ from a large sample of lenses. If we are able to average together $N$ systems we should, in principle, be able to reduce our uncertainty by $\sqrt{N}$. In practice, the accuracy of the combination procedure will sooner be limited by the systematic uncertainty in the shape and centroid of the assumed $\kappa_{ext}$ distribution: investigating the properties of this distribution is perhaps the most urgent topic for further work. Likewise, if the density profile slope cannot be constrained for each time-delay lens individually, the details of the prior PDF assigned for $\gamma'$ will become important as the ensemble grows.

In the near future, cadenced surveys such as those planned with the Large Synoptic Survey Telescope (LSST) and being undertaken by the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) will discover large numbers of time-delay lenses, prompting us to consider performing analyses such as the one described here on hundreds of lens systems. In prac-

tice, obtaining data of the quality we have presented here for hundreds of suitable lenses will pose a significant observational challenge. Nevertheless, Dobke et al. (2009), Coe & Moustakas (2009) and Oguri & Marshall (in preparation) investigate constraints on cosmological parameters based on large samples of time-delay lenses. In particular, Coe & Moustakas (2009) suggest that, in terms of raw precision and in combination with a prior PDF from Planck, an LSST ensemble could reach subpercent level precision in $H_0$, and constrain $w$ to 3% or better, provided that the systematic effects such as $\kappa_{ext}$ are under control. Our work has already addressed some of these systematic effects, and will provide a basis for future analysis of large samples of time-delay lenses and lens environment studies.

## 9. CONCLUSIONS

We have studied the well-observed gravitational lens B1608+656 and used it to infer the values of cosmological parameters; we outlined and followed a Bayesian approach for combining three data sets: $HST$/ACS imaging, stellar velocity dispersion measurement, and the time delays between the multiple images. Diagnosing the principal systematic effects, we included two nuisance parameters into the data model to account for them, assign-

ing well-motivated prior PDFs and marginalizing over them. We draw the following conclusions:

- We find that the *HST*/ACS images constrain the density profile slope parameter $\gamma' = 2.08 \pm 0.03$, which we propagate through the cosmological parameter inference as a prior PDF. Relaxing this prior to a uniform distribution degrades the precision on $H_0$ from 4.4% to 7.0%; the SLACS intrinsic profile slope parameter distribution is not significantly more informative than the uniform prior.

- With the ACS prior for $\gamma'$, we find that the inferred cosmological parameters is dominated by the the external convergence $\kappa_{\rm ext}$. Ray-tracing through the Millennium Simulation gives a PDF for $\kappa_{\rm ext}$ due to line-of-sight contributions that has zero mean and width $\sim 0.04$, while using the galaxy number counts in the B1608+656 field in conjunction with the MS gives $\kappa_{\rm ext} = 0.10^{+0.08}_{-0.05}$.

Using our most informative priors on the two nuisance parameters, we arrive at the following cosmographic inferences:

- In the K03 cosmology ($\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$, $w = -1$, and uniform $H_0$), we obtain from the B1608+656 data set $H_0 = 70.6 \pm 3.1 \, {\rm km \, s^{-1} \, Mpc^{-1}}$ (68% CL). The 4.4% error includes both statistical and dominant systematic uncertainties, through the marginalization described above. This is a significant improvement to the earlier measurement of $H_0 = 75^{+7}_{-6} \, {\rm km \, s^{-1} \, Mpc^{-1}}$ by Koopmans et al. (2003).

- Time-delay lenses are sensitive primarily to $H_0$ but are weakly dependent on other cosmological parameters; the lensing measurement of $H_0$ is robust and useful for studying dark energy when combined with other cosmological probes. We find that for B1608+656 the cosmographic information can be summarized as a Gaussian probability distribution for the time-delay distance $D_{\Delta \rm t}$, centered on 5226 Mpc and with width 206 Mpc.

- In a $\Lambda$-CDM cosmology (with $w = -1$), the B1608+656 data set breaks the degeneracy between $\Omega_{\rm m}$ and $\Omega_\Lambda$ in the WMAP 5 year data set, and constrains the curvature parameter to be zero to 2.0% (95% CL), a level of precision similar to those afforded by the current Type Ia SNe sample.

- B1608+656 in combination with the WMAP 5 year data set, assuming flatness and allowing (a time-independent) $w$ to vary, gives $H_0 = 69.7^{+4.9}_{-5.0} \, {\rm km \, s^{-1} \, Mpc^{-1}}$ and $w = -0.94^{+0.17}_{-0.19}$ (68% CL).

These are significant improvements to the WMAP5 only constraints of $H_0 = 74^{+15}_{-14} \, {\rm km \, s^{-1} \, Mpc^{-1}}$ and $w = -1.06^{+0.41}_{-0.42}$. B1608+656 is as competitive as the current BAO data in determining $w$ when combined with WMAP5.

Our detailed analysis of B1608+656 provides the framework for using large samples of time-delay lenses as cosmological probes in the near future. We anticipate the local contribution to $\kappa_{\rm ext}$, which would not average away with a large sample of lenses, being the dominant residual systematic error. Several lens environment studies to circumvent this are underway; with the effects from $\kappa_{\rm ext}$ accurately modeled, future samples of time-delay gravitational lenses should be a competitive cosmological probe.

REFERENCES

Auger, M. W., Fassnacht, C. D., Abrahamse, A. L., Lubin, L. M., & Squires, G. K. 2007, AJ, 134, 668
Auger, M. W., Fassnacht, C. D., Wong, K. C., Thompson, D., Matthews, K., & Soifer, B. T. 2008, ApJ, 673, 778
Barnabè, M. & Koopmans, L. V. E. 2007, ApJ, 666, 726
Binney, J. & Tremaine, S. 1987, Galactic dynamics (Princeton, NJ, Princeton University Press, 1987, 747 p.)
Birkinshaw, M. 1999, Phys. Rep., 310, 97
Bonamente, M., Joy, M. K., LaRoque, S. J., Carlstrom, J. E., Reese, E. D., & Dawson, K. S. 2006, ApJ, 647, 25

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., Myers, S. T., Fassnacht, C. D., Koopmans, L. V. E., Marlow, D. R., Norbury, M., Rusin, D., Sykes, C. M., Biggs, A. D., Blandford, R. D., de Bruyn, A. G., Chae, K.-H., Helbig, P., King, L. J., McKean, J. P., Pearson, T. J., Phillips, P. M., Readhead, A. C. S., Xanthopoulos, E., & York, T. 2003, MNRAS, 341, 13
Coe, D. & Moustakas, L. 2009, ArXiv e-prints (0906.4108)
Coles, J. 2008, ApJ, 679, 17
De Lucia, G. & Blaizot, J. 2007, MNRAS, 375, 2
Dobke, B. M., King, L. J., Fassnacht, C. D., & Auger, M. W. 2009, MNRAS, 397, 311

Dunkley, J., Komatsu, E., Nolta, M. R., Spergel, D. N., Larson, D., Hinshaw, G., Page, L., Bennett, C. L., Gold, B., Jarosik, N., Weiland, J. L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., & Wright, E. L. 2009, ApJS, 180, 306

Dye, S., Evans, N. W., Belokurov, V., Warren, S. J., & Hewett, P. 2008, MNRAS, 388, 384

Dye, S. & Warren, S. J. 2005, ApJ, 623, 31

Fadely, R., Keeton, C. R., Nakajima, R., & Bernstein, G. M. 2009, ArXiv e-prints (0909.1807)

Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, ApJ, 289, L1

Fassnacht, C. D., Gal, R. R., Lubin, L. M., McKean, J. P., Squires, G. K., & Readhead, A. C. S. 2006a, ApJ, 642, 30

Fassnacht, C. D., Koopmans, L. V. E., & Wong, K. C. 2009, ArXiv e-prints (0909.4301)

Fassnacht, C. D., McKean, J. P., Koopmans, L. V. E., Treu, T., Blandford, R. D., Auger, M. W., Jeltema, T. E., Lubin, L. M., Margoniner, V. E., & Wittman, D. 2006b, ApJ, 651, 667

Fassnacht, C. D., Pearson, T. J., Readhead, A. C. S., Browne, I. W. A., Koopmans, L. V. E., Myers, S. T., & Wilkinson, P. N. 1999, ApJ, 527, 498

Fassnacht, C. D., Womble, D. S., Neugebauer, G., Browne, I. W. A., Readhead, A. C. S., Matthews, K., & Pearson, T. J. 1996, ApJ, 460, L103

Fassnacht, C. D., Xanthopoulos, E., Koopmans, L. V. E., & Rusin, D. 2002, ApJ, 581, 823

Freedman, W. L., Madore, B. F., Gibson, B. K., Ferrarese, L., Kelson, D. D., Sakai, S., Mould, J. R., Kennicutt, Jr., R. C., Ford, H. C., Graham, J. A., Huchra, J. P., Hughes, S. M. G., Illingworth, G. D., Macri, L. M., & Stetson, P. B. 2001, ApJ, 553, 47

Gavazzi, R., Treu, T., Koopmans, L. V. E., Bolton, A. S., Moustakas, L. A., Burles, S., & Marshall, P. J. 2008, ApJ, 677, 1046

Gavazzi, R., Treu, T., Rhodes, J. D., Koopmans, L. V. E., Bolton, A. S., Burles, S., Massey, R. J., & Moustakas, L. A. 2007, ApJ, 667, 176

Grogin, N. A. & Narayan, R. 1996a, ApJ, 464, 92

—. 1996b, ApJ, 473, 570

Herrnstein, J. R., Moran, J. M., Greenhill, L. J., Diamond, P. J., Inoue, M., Nakai, N., Miyoshi, M., Henkel, C., & Riess, A. 1999, Nature, 400, 539

Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. 2009, A&A, 499, 31

Hilbert, S., White, S. D. M., Hartlap, J., & Schneider, P. 2007, MNRAS, 382, 121

—. 2008, MNRAS, 386, 1845

Hu, W. 2005, 339, 215

Jakobsson, P., Hjorth, J., Burud, I., Letawe, G., Lidman, C., & Courbin, F. 2005, A&A, 431, 103

Keeton, C. R. & Zabludoff, A. I. 2004, ApJ, 612, 660

Kochanek, C. S. 2002, ApJ, 578, 25

Komatsu, E., Dunkley, J., Nolta, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Meyer, S. S., Tucker, G. S., Weiland, J. L., Wollack, E., & Wright, E. L. 2009, ApJS, 180, 330

Koopmans, L. V. E., Bolton, A., Treu, T., Czoske, O., Auger, M. W., Barnabè, M., Vegetti, S., Gavazzi, R., Moustakas, L. A., & Burles, S. 2009, ApJ, 703, L51

Koopmans, L. V. E. & Treu, T. 2002, ApJ, 568, L5

Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599

Koopmans, L. V. E., Treu, T., Fassnacht, C. D., Blandford, R. D., & Surpi, G. 2003, ApJ, 599, 70

Kowalski, M., Rubin, D., Aldering, G., Agostinho, R. J., Amadon, A., Amanullah, R., Balland, C., Barbary, K., Blanc, G., Challis, P. J., Conley, A., Connolly, N. V., Covarrubias, R., Dawson, K. S., Deustua, S. E., Ellis, R., Fabbro, S., Fadeyev, V., Fan, X., Farris, B., Folatelli, G., Frye, B. L., Garavini, G., Gates, E. L., Germany, L., Goldhaber, G., Goldman, B., Goobar, A., Groom, D. E., Haissinski, J., Hardin, D., Hook, I., Kent, S., Kim, A. G., Knop, R. A., Lidman, C., Linder, E. V., Mendez, J., Meyers, J., Miller, G. J., Moniez, M., Mourão, A. M., Newberg, H., Nobili, S., Nugent, P. E., Pain, R., Perdereau, O., Perlmutter, S., Phillips, M. M., Prasad, V., Quimby, R., Regnault, N., Rich, J., Rubenstein, E. P., Ruiz-Lapuente, P., Santos, F. D., Schaefer, B. E., Schommer, R. A., Smith, R. C., Soderberg, A. M., Spadafora, A. L., Strolger, L.-G., Strovink, M., Suntzeff, N. B., Suzuki, N., Thomas, R. C., Walton, N. A., Wang, L., Wood-Vasey, W. M., & Yun, J. L. 2008, ApJ, 686, 749

Kronawitter, A., Saglia, R. P., Gerhard, O., & Bender, R. 2000, A&AS, 144, 53

Lewis, A. & Bridle, S. 2002, Phys. Rev. D, 66, 103511

Macri, L. M., Stanek, K. Z., Bersier, D., Greenhill, L. J., & Reid, M. J. 2006, ApJ, 652, 1133

McKean, J. P., Auger, M. W., Koopmans, L. V. E., Vegetti, S., Czoske, O., Fassnacht, C. D., Treu, T., More, A., & Kocevski, D. D. 2009, ArXiv e-prints (0910.1133)

Merritt, D. 1985, AJ, 90, 1027

Merritt, D. & Aguilar, L. A. 1985, MNRAS, 217, 787

Momcheva, I., Williams, K., Keeton, C., & Zabludoff, A. 2006, ApJ, 641, 169

Myers, S. T., Fassnacht, C. D., Djorgovski, S. G., Blandford, R. D., Matthews, K., Neugebauer, G., Pearson, T. J., Readhead, A. C. S., Smith, J. D., Thompson, D. J., Womble, D. S., Browne, I. W. A., Wilkinson, P. N., Nair, S., Jackson, N., Snellen, I. A. G., Miley, G. K., de Bruyn, A. G., & Schilizzi, R. T. 1995, ApJ, 447, L5

Myers, S. T., Jackson, N. J., Browne, I. W. A., de Bruyn, A. G., Pearson, T. J., Readhead, A. C. S., Wilkinson, P. N., Biggs, A. D., Blandford, R. D., Fassnacht, C. D., Koopmans, L. V. E., Marlow, D. R., McKean, J. P., Norbury, M. A., Phillips, P. M., Rusin, D., Shepherd, M. C., & Sykes, C. M. 2003, MNRAS, 341, 1

Nakajima, R., Bernstein, G. M., Fadely, R., Keeton, C. R., & Schrabback, T. 2009, ApJ, 697, 1793

Oguri, M. 2007, ApJ, 660, 1

Oke, J. B., Cohen, J. G., Carr, M., Cromer, J., Dingizian, A., Harris, F. H., Labrecque, S., Lucinio, R., Schaal, W., Epps, H., & Miller, J. 1995, PASP, 107, 375

Osipkov, L. P. 1979, Pis ma Astronomicheskii Zhurnal, 5, 77

Paraficz, D., Hjorth, J., & Elíasdóttir, Á. 2009, A&A, 499, 395

Percival, W. J., Cole, S., Eisenstein, D. J., Nichol, R. C., Peacock, J. A., Pope, A. C., & Szalay, A. S. 2007, MNRAS, 381, 1053

Refsdal, S. 1964, MNRAS, 128, 307

Riess, A. G., Li, W., Stetson, P. B., Filippenko, A. V., Jha, S., Kirshner, R. P., Challis, P. M., Garnavich, P. M., & Chornock, R. 2005, ApJ, 627, 579

Riess, A. G., Macri, L., Casertano, S., Sosey, M., Lampeitl, H., Ferguson, H. C., Filippenko, A. V., Jha, S. W., Li, W., Chornock, R., & Sarkar, D. 2009, ApJ, 699, 539

Saha, P., Coles, J., Macciò, A. V., & Williams, L. L. R. 2006, ApJ, 650, L17

Sandage, A., Tammann, G. A., Saha, A., Reindl, B., Macchetto, F. D., & Panagia, N. 2006, ApJ, 653, 843

Schmidt, R. W., Allen, S. W., & Fabian, A. C. 2004, MNRAS, 352, 1413

Schneider, P., Kochanek, C. S., & Wambsganss, J. 2006, Gravitational Lensing: Strong, Weak and Micro (Springer)

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., & Pearce, F. 2005, Nature, 435, 629

Stiavelli, M. & Sparke, L. S. 1991, ApJ, 382, 466

Suyu, S. H., Marshall, P. J., Blandford, R. D., Fassnacht, C. D., Koopmans, L. V. E., McKean, J. P., & Treu, T. 2009, ApJ, 691, 277

Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, MNRAS, 371, 983

Tonry, J. L. & Franx, M. 1999, ApJ, 515, 512

Treu, T., Gavazzi, R., Gorecki, A., Marshall, P. J., Koopmans, L. V. E., Bolton, A. S., Moustakas, L. A., & Burles, S. 2009, ApJ, 690, 670
Treu, T., Koopmans, L. V., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 640, 662
Treu, T. & Koopmans, L. V. E. 2002, MNRAS, 337, L6
van der Marel, R. P. 1994, MNRAS, 270, 271
Vuissoz, C., Courbin, F., Sluse, D., Meylan, G., Chantry, V., Eulaers, E., Morgan, C., Eyler, M. E., Kochanek, C. S., Coles, J., Saha, P., Magain, P., & Falco, E. E. 2008, A&A, 488, 481

Vuissoz, C., Courbin, F., Sluse, D., Meylan, G., Ibrahimov, M., Asfandiyarov, I., Stoops, E., Eigenbrod, A., Le Guillou, L., van Winckel, H., & Magain, P. 2007, A&A, 464, 845
Wucknitz, O. 2002, MNRAS, 332, 951
Wucknitz, O., Biggs, A. D., & Browne, I. W. A. 2004, MNRAS, 349, 14
York, T., Jackson, N., Browne, I. W. A., Wucknitz, O., & Skelton, J. E. 2005, MNRAS, 357, 124

<div align="center">APPENDIX</div>

<div align="center">A. PROBABILITY THEORY FOR MEASURING COSMOLOGICAL PARAMETERS</div>

In this appendix, we describe how we derive the expressions for the likelihoods of the time delay and the ACS data sets stated in Section 3.3. We also provide the details on the sampling techniques for calculating the posterior probability density of cosmological parameters.

### A.1. *Simplification of the likelihoods*

For B1608+656, we can simplify the marginalization of the likelihoods $P(\mathbf{\Delta t}|\boldsymbol{\xi})$ and $P(\boldsymbol{d}|\boldsymbol{\xi})$ in Equation (18) based on the following facts, which are either from Paper I or shown in Section 4:

- from Paper I, the top $\boldsymbol{M}_{\mathrm{D}}$ models led to equal evidence values $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}_{\mathrm{MP}}, \boldsymbol{M}_{\mathrm{D}})$ (within the uncertainties).

- in Section 4, the likelihood function $P(\mathbf{\Delta t}|\boldsymbol{\xi})$ is approximately constant for these top $\boldsymbol{M}_{\mathrm{D}}$ models for a given cosmology.

- in Section 4, for good data models $\boldsymbol{M}_{\mathrm{D}}$ of B1608+656, the potential corrections $\boldsymbol{\delta\psi}$ do not change significantly the predicted values of the Fermat potential, i.e., the simply-parametrized SPLE initial model provides an unbiased estimator for the Fermat potential.

- the likelihood $P(\sigma|\boldsymbol{\xi})$ is also constant for the various $\boldsymbol{M}_{\mathrm{D}}$ because the dynamics modeling is independent of the lensed image processing models $\boldsymbol{M}_{\mathrm{D}}$.

- simulations suggest that the potential corrections $\boldsymbol{\delta\psi}$ are sharply peaked about the most probable values $\boldsymbol{\delta\psi}_{\mathrm{MP}}$.

With the above results, the Fermat potential can be more easily computed from the SPLE model: there is a strong correlation between the $\Delta\phi$ and $\gamma'$, and we obtain the relation between $\Delta\phi$ and $\gamma'$ by evaluating them at several discrete $\gamma'$ values and interpolating between them. For notational simplicity, we consequently drop the nearly-true independences of $\Delta\phi$ on $\boldsymbol{\eta}$ and $\boldsymbol{\delta\psi}$. In computing the predicted $\Delta\phi$, the average source position of the four mapped (via the lens equation) image positions on the source plane is used. Denoting the dependence of $\Delta\phi$ on $\gamma'$ for $\kappa_{\mathrm{ext}} = 0$ and a given $\boldsymbol{M}_{\mathrm{D}}$ as $q(\gamma', \boldsymbol{M}_{\mathrm{D}})$ and using the mass-sheet degeneracy relation for the dependence of $\Delta\phi$ on $\kappa_{\mathrm{ext}}$, we obtain Equations (19) through (21) for the likelihood of the time delay data.[18]

For the likelihood of the ACS data, we work with the SPLE potential model on the basis that the Fermat potential is insensitive to the potential corrections and the potential corrections only slightly alter the ranking of $\boldsymbol{M}_{\mathrm{D}}$ models. We can choose the $\boldsymbol{M}_{\mathrm{D}}$ to be one of the top models, say $\boldsymbol{M}_5$ (since the normalization in $P(\boldsymbol{\pi}|\mathbf{\Delta t}, \boldsymbol{d}, \sigma)$ is irrelevant), and drop the dependence on $\boldsymbol{\delta\psi}$ in the likelihood of the ACS data to simplify part of the integrand in Equation (18):

$$\int \mathrm{d}\boldsymbol{\delta\psi}\ \mathrm{d}\boldsymbol{s}\ \mathrm{d}\boldsymbol{M}_{\mathrm{D}}\ P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{\delta\psi}, \boldsymbol{s}, \boldsymbol{M}_{\mathrm{D}}) \cdot$$
$$P(\boldsymbol{s}|\lambda, \mathbf{g})P(\boldsymbol{M}_{\mathrm{D}})P(\boldsymbol{\delta\psi})$$
$$\propto \sim \int \mathrm{d}\boldsymbol{s}\ P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{s}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)P(\boldsymbol{s}|\lambda, \mathbf{g})$$
$$= P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5), \tag{A1}$$

which is also Equation (23). In deriving the above equation, we assume that the representative set of models $\boldsymbol{M}_{\mathrm{D}}$ obtained in Suyu et al. (2009) are equally probable a priori (i.e., the prior $P(\boldsymbol{M}_{\mathrm{D}})$ is constant). For the priors $P(\boldsymbol{s}|\lambda, \mathbf{g})$ and $P(\boldsymbol{\delta\psi})$, we use quadratic forms of the regularizing function. Specifically, we try zeroth-order, gradient and curvature forms for $P(\boldsymbol{s}|\lambda, \mathbf{g})$, and the curvature form for $P(\boldsymbol{\delta\psi})$, as described in Suyu et al. (2006) and Suyu et al. (2009). As a reminder, the quantity $P(\boldsymbol{d}|\gamma', \boldsymbol{\eta}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$ is the Bayesian evidence from source reconstruction given the lens model parameters $\{\gamma', \boldsymbol{\eta}\}$ and the data model $\boldsymbol{M}_5$; this evidence value is calculable based on Suyu et al. (2009).

---

[18] The Fermat potential is independent of cosmological parameters because we work in terms of the scaled lens potential in angular (arcsecond) units. Cosmological parameters and redshifts are only needed for deriving physical quantities of the lens system, such as the one-dimensional velocity dispersion (calculable from the Einstein radius) of the lens, mass of the lens, and the physical extent of the lens/source.

### A.2. *Importance sampling*

In practice, we can incorporate the various likelihoods in Equation (27) by importance sampling the prior distribution (see e.g. Lewis & Bridle 2002, for an introduction). This is a method for calculating integrals over a PDF $P_2$ when all we have is samples drawn from some other PDF $P_1$. Consider the expectation value of a parameter $x$:

$$\langle x \rangle_2 = \int x \cdot P_2(x)\, dx, \tag{A2}$$

$$= \int x \frac{P_2(x)}{P_1(x)} \cdot P_1(x)\, dx. \tag{A3}$$

The process of weighting the samples from $P_1$ by the ratio $P_2(x)/P_1(x)$ is called importance sampling. It works most efficiently when $P_1$ and $P_2$ are quite similar, and fails if $P_1$ is zero-valued over some of the range of $P_2$, or if the sampling of $P_1$ is too sparse.

In our case, we would like to calculate integrals over, for example, $P_2 = P(\boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}} | \boldsymbol{\Delta t}, \boldsymbol{d}, \sigma)$, while the prior is written simply $P_1 = P(\boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}} | \boldsymbol{d})$ (recall that $\boldsymbol{d}$ is used to provide a prior on $\gamma'$). Using Bayes' theorem, we can write

$$\begin{aligned} P(\boldsymbol{\pi}, \gamma', &\kappa_{\mathrm{ext}}, r_{\mathrm{ani}} | \boldsymbol{\Delta t}, \boldsymbol{d}, \sigma) \propto \\ &P(\boldsymbol{\Delta t}, \sigma | \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}}) P(\boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}} | \boldsymbol{d}), \\ \mathrm{i.e.} \quad &P_2 \propto P(\boldsymbol{\Delta t}, \sigma | \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}}) P_1. \end{aligned} \tag{A4}$$

From this we can see that the weight we must attach to each sample from the prior is just the value of the likelihood $P(\boldsymbol{\Delta t}, \sigma | \boldsymbol{\pi}, \gamma', \kappa_{\mathrm{ext}}, r_{\mathrm{ani}})$. Note that these weights can be rescaled by an arbitrary factor, which can be important in retaining numerical stability.

We apply this technique to perform the marginalization in Equation (27). Specifically, we have samples of $P(\boldsymbol{\pi})$, $P(\gamma' | \boldsymbol{d}, \boldsymbol{M}_{\mathrm{D}} = \boldsymbol{M}_5)$, $P(\kappa_{\mathrm{ext}})$ and $P(r_{\mathrm{ani}})$, and employ importance sampling to obtain $P(\boldsymbol{\pi} | \boldsymbol{\Delta t}, \boldsymbol{d}, \sigma)$.