NO THERMAL INVERSION AND A SOLAR WATER ABUNDANCE FOR THE HOT JUPITER HD 209458B FROM HST/WFC3 SPECTROSCOPY

Michael R. Line1,2,3,10, Kevin B. Stevenson4,11, Jacob Bean4, Jean-Michel Desert5, Jonathan J. Fortney6, Laura Kreidberg4, Nikku Madhusudhan7, Adam P. Showman8, and Hannah Diamond-Lowe9

1 NASA Ames Research Center, Moffet Field, CA 94035, USA
2 Bay Area Environmental Research Institute, 625 2nd Street, Suite 209, Petaluma, CA 94952, USA
3 School of Earth & Space Exploration, Arizona State University, USA
4 Department of Astronomy & Astrophysics, University of Chicago, 5640 S Ellis Avenue, Chicago, IL 60637, USA
5 University of Amsterdam, The Netherlands
6 Department of Astronomy & Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
7 Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK
8 Department of Planetary Sciences and Lunar and Planetary Laboratory, University of Arizona, 1629 E. University Blvd., Tucson, AZ 85721, USA
9 Department of Astronomy, Harvard Smithsonian Center for Astrophysics, 60 Garden Street, MS-10, Cambridge, MA 02138, USA

Received 2016 May 27; revised 2016 September 29; accepted 2016 September 29; published 2016 December 2

1. INTRODUCTION

Hubble Space Telescope (HST) Wide Field Camera 3 (WFC3) spectrophotometric observations of transiting exoplanets are rapidly transforming our knowledge of planetary atmospheres (Berta et al. 2012; Swain et al. 2013; Deming et al. 2013; Huitson et al. 2013; Line et al. 2013b; Mandell et al. 2013; Wakeford et al. 2013; Ehrenreich et al. 2014; Fraine et al. 2014; Knutson et al. 2014; Kreidberg et al. 2014a, 2014b, 2015; McCullough et al. 2014; Ranjan et al. 2014; Stevenson et al. 2014c; Wilkins et al. 2014; Sing et al. 2015; Evans et al. 2016). As the highest-precision and most robust spectroscopy obtainable with current instruments, WFC3 data allow us to infer the thermal structures, energy transport, molecular abundances, and cloud properties in exoplanet atmospheres. Determining these properties allows us to then begin to understand the fundamental processes occurring in planetary atmospheres over a wide range of conditions.

The thermal structures of hot Jupiter atmospheres, and in particular the question of thermal inversions, have driven much of the characterization work on these planets over the past decade (for reviews, see Burrows & Orton 2010; Madhusudhan et al. 2014b, p. 739; Crossfield et al. 2015). Thermal inversions typically occur in the presence of a high-altitude heating source, usually a strong short-wave absorber. Most solar system bodies with substantial atmospheres possess a thermal inversion (e.g., Robinson & Catling 2014), due to ozone in the case of Earth, or aerosol heating in the case of the giant planets. Hot Jupiters were theorized to have thermal inversions due to the possible presence of strong gas-phase optical absorbers (Hubeny et al. 2003; Burrows et al. 2008), for example, TiO and VO, which can persist at high temperatures ($T > 2000$ K at 1 bar; Fortney et al. 2006, 2008), or polysulfur aerosols (Zahnle et al. 2009).

Observations using five broadband channels on the Spitzer Space Telescope claimed the first detection of a thermal inversion in a hot Jupiter atmosphere, in the prototype first transiting planet HD 209458b (Burrows et al. 2007; Knutson et al. 2008), presumably confirming the presence of a high-altitude absorber, the leading candidates being TiO and VO given their presence in the gas phase at HD 209458b’s temperatures. Evidence for the inversion was based on the relatively high fluxes of the 4.5 and 5.8 μm channels compared with the 3.6 and 8.0 μm channels. These results were
challenging to interpret theoretically as, first, for plausible compositions, the contribution functions of the four \textit{Spitzer} channels greatly overlap, so it is extremely difficult to get significantly differing brightness temperatures between the channels, and second, the infrared photosphere (optical depth of unity) tends to occur at deeper layers than the layers at which most of the visible layers are absorbed. An inversion is formed because at the altitudes at which the visible radiation is strongly absorbed the infrared emissivity is low, requiring hotter temperatures to maintain radiative equilibrium (e.g., Robinson & Catling 2014). Nevertheless, Madhusudhan & Seager (2009) and Line et al. (2014), using retrieval methods, were able to fit the Knutson et al. (2008) \textit{Spitzer} data by invoking an inversion above the \textasciitilde1 bar level and large abundances of CO. The large abundance of CO is required to decouple the 4.5 \textmu m (and to a lesser extent 5.8 \textmu m) weighting functions from the rest of the channels, permitting the large temperature differences. It is unclear, however, how well the mean infrared opacity required to produce the retrieved thermal structures agrees with those derived from the abundances.

In addition to the Knutson et al. (2008) results, the secondary eclipse of HD 209458b has been observed and processed repeatedly by many different groups with a variety of different space-based (\textit{Spitzer} IRS: Richardson et al. 2007; Swain et al. 2008; \textit{Spitzer} MIPS: Deming et al. 2005; \textit{Spitzer} IRAC: Diamond-Lowe et al. 2014; Zellem et al. 2014b; Evans et al. 2015; \textit{HST} NICMOS: Swain et al. 2009; Crossfield et al. 2012) and ground-based (Richardson et al. 2003; Zellem et al. 2014a; Schwarz et al. 2015) instruments, many studies of which questioned the possible existence of the previously claimed inversion, though the robustness of some of these earlier data is questionable (e.g., Gibson et al. 2011, 2012).

As an example, Diamond-Lowe et al. (2014) reanalyzed the Knutson et al. (2008) data set combined with new data taken with more optimal observing modes at 3.6, 4.5, and 8.0 \textmu m. This resulted in a significant reduction of the 4.5 and 5.8 \textmu m fluxes (which drove the need for an inversion) and showed that a thermal inversion was no longer required to model those data. Additionally, with an entirely different method altogether, Schwarz et al. (2015), using high-dispersion spectroscopy with the CRIRES instrument combined with cross-correlation data reduction techniques, were able to rule out an inversion between 1 bar and 1 mbar. However, they also did not make a strong detection of molecular absorption indicative of a normal thermal structure.

Furthermore, evidence for inversions in other planets of similar or hotter temperatures is currently lacking. In the years since the HD 209458b observations, retrieval methods of interpreting spectra have shown that often models without temperature inversions could explain wide-band \textit{Spitzer} data as readily as models with inversions (Madhusudhan & Seager 2009; Line et al. 2014). A growing body of work has tried to explain why inversions may not occur. Spiegel et al. (2009) and Showman et al. (2009) investigated how condensation in cold traps could lead to removal of TiO and VO from the gas phase. Madhusudhan et al. (2011) suggested that the possible lack of inversions could be due to high carbon-to-oxygen (C/O) ratios. High C/O would deplete the oxygen required to build the strongly absorbing oxides. However, Line et al. (2014) and Benneke (2015) showed that there is no strong evidence for high C/O ratios over a wide range of planetary conditions—including those in which inversions are expected. An alternative hypothesis for the apparent lack of inversions, brought forth by Knutson et al. (2010), suggests that high incident UV flux could potentially photolyze high-altitude absorbers.

Recently, Haynes et al. (2015) suggested the presence of an inversion along with potential evidence for TiO emission in the hot Jupiter WASP-33b by combining \textit{HST}, \textit{Spitzer}, and ground-based data; while tantalizing evidence, a bulk of the inversion evidence is driven by the ground-based \textgamma-band point and the two bluest channels in WFC3. Furthermore, WASP-33 is a delta-scuti variable, thus making the data analysis and interpretation extremely challenging.

The power of spectroscopy, compared to photometry, is that the planetary thermal emission is probed across a range of wavelengths, and hence a range of depths in the atmosphere, simultaneously. The interpretation of such observations is in principle much more straightforward than in wide-band photometry. Molecular features seen in emission can be interpreted as a telltale sign of an inverted temperature structure. However, a high signal-to-noise ratio (S/N) spectrum is required. As part of \textit{HST} GO program 13467 we have endeavored to produce such high-S/N spectral data to bring clarity to the characterization of exoplanet atmospheres.

In what follows, we describe our WFC3 observations in Section 2, followed by a global dayside atmospheric retrieval analysis Section 3, and finally we discuss implications of our observations and conclude in Section 4.

2. \textit{HST}/WFC3 OBSERVATIONS AND DATA ANALYSIS

2.1. Observations and Reduction

As part of a large \textit{HST} Treasury Program (GO 13467), we used the Wide Field Camera 3 (WFC3) to observe HD 209458b during secondary eclipse over five visits from 2014 September to December. WFC3’s G141 grism disperses light from 1.12 to 1.65 \textmu m onto its detector, which we operated in 256 \times 256 subarray mode.

Each visit comprises five \textit{HST} orbits: the first orbit achieves instrument stability and is excluded from our analysis following standard procedure (Berta et al. 2012; Deming et al. 2013; Kreidberg et al. 2014b), the second and fifth orbits determine the out-of-eclipse baseline, and the third and fourth orbits contain the secondary eclipse. Each \textit{HST} orbit begins with a direct image of HD 209458 for calibration purposes, is followed by 43 spectroscopic frames (each lasting 14.971 s in duration), and concludes with a buffer dump to maximize observing efficiency. During the acquisition of each spectroscopic frame, we utilize the bidirectional spatial scan mode at a rate of 1"/15 s\textsuperscript{-1}, thus spreading the light over \textasciitilde140 pixel rows perpendicular to the spectral dispersion. The median peak pixel count across all frames and all visits is 40,615, which is close to the recommended threshold level of \textasciitilde40,000e for optimal performance (Berta et al. 2012; Swain et al. 2013; Wilkins et al. 2014).

Our data reduction and analysis closely follow the techniques described in our previous WFC3 papers (Stevenson et al. 2014a, 2014b, 2014c). Here we provide a brief description of our methods. For the reduction of spatial scan data, we use the calibrated “ima” frames provided by the STScI Archive. Using a 2D Gaussian, we compute the centroid of the direct image from the start of each \textit{HST} orbit, calculate the trace of the equivalent stare spectra, and then determine the
field-dependent wavelength solution of the scanned spectra (Kuntzchner et al. 2009). We apply spectroscopic flat-field corrections using coefficients from the calibration file WFC3, IR-G141-flat.2.fits.

Each frame consists of three evenly spaced, nondestructive reads. We compute the difference between pairs of adjacent reads, use optimal extraction (Horne 1986) to produce two 1D spectra, and then combine their values to arrive at a single spectrum per frame. Our extraction window consists of 110 pixel rows centered on the differenced spectra and flanked by an additional 110 pixel rows for background subtraction (220 pixel rows total). This optimized configuration produces the most precise light curves. We divide the spectra into 10 spectrophotometric bins of width ~0.0525 μm and spanning 1.125–1.655 μm.

Our extracted light curves exhibit similar systematics to those noted previously for WFC3 data (e.g., Berta et al. 2012; Deming et al. 2013; Stevenson et al. 2014c). These include a ramp (or hook) at the beginning of each HST orbit and a visit-long trend. We follow a two-step process to extract HD 209458b’s emission spectrum. First, we model the band-integrated (white) light curves to compute nonanalytic, transit-removed systematic model components and to evaluate the consistency of the measured eclipse depths between visits. Second, we divide the spectroscopic light curves by our nonanalytic model components and then determine the best fit for each channel to derive wavelength-dependent eclipse depths with uncertainties. Below we discuss each of these steps in detail.

2.2. White Light-curve Fits

Our model fitting procedure consists of finding the best solution by simultaneously fitting all free parameters using a Levenberg–Marquardt minimizer and then using a Differential Evolution Markov Chain algorithm (ter Braak & Vrugt 2008) to estimate parameter uncertainties. For our final white light-curve fits, we use 10 chains and terminate after 4 × 10^5 steps when the Gelman & Rubin (1992) diagnostic for all free parameters has dropped to within 1% of unity. For more details on error estimation and convergence, see Stevenson et al. (2012). We find no evidence for correlated noise in the best-fit residuals as determined by plotting the rms versus bin size for each scan direction of each visit. We test various model components and shared parameters and then determine the best combination by comparing their Bayesian Information Criterion (BIC) values (Cornish & Littenberg 2007; Trotta 2008).

Our final white light-curve model takes the form

\[ F(\frac{\phi}{2}, \phi, \Phi) = F(\frac{\phi}{2})E(\phi)R(\phi, \Phi)H(\Phi, \frac{\phi}{2}), \]

where \( F(\frac{\phi}{2}, \phi, \Phi) \) is the modeled flux for scan direction, \( \frac{\phi}{2} \), at planet orbital phase, \( \phi \), and HST orbital phase, \( \Phi \); \( F(\frac{\phi}{2}) \) is the scan-direction-dependent out-of-eclipse system flux; \( E(\phi) \) is the eclipse model (Kreidberg 2015); \( R(\phi, \Phi) = 1 + r_{1,2}(\phi - 0.5) + r_{2,3}(\phi - 0.5)^2 \) is the time-dependent quadratic trend model component; and \( H(\Phi, \frac{\phi}{2}) = 1 - e^{-h_{0,2}b + h_{2,2}\Phi} \) is the HST orbital-phase-dependent rising exponential ramp with linear trend model component. Scan-direction- and visit-dependent free parameters include \( F(\frac{\phi}{2}) \), \( r_{1,2}, r_{2,3}, \) and \( h_{2,2} \). In our final fits, the eclipse depth, \( h_{0,2} \), and \( h_{2,2} \) free parameters have the same values for both scan directions and all visits. We fix the eclipse midpoint (\( \phi = 0.5 \)), total eclipse duration (3.06 hr), and ingress/egress times (0.42 hr; Seager & Mallén-Ornelas 2003) to precisely known from previous observations of this canonical planet (Zellem et al. 2014b).

To test the repeatability of our measurements, we apply our white light-curve model to the data from each visit and report the best-fit eclipse depths and \( \chi^2 \) values in Table 1. Despite having a relatively good fit (\( \chi^2 = 1.18 \)) and contrary to the other visits, the fourth visit contains no discernible secondary eclipse (6 ± 35 ppm; Figure 1). As noted previously (Stevenson et al. 2014c), there is a strong correlation between the eclipse depth and the quadratic term in \( R(\phi, \Phi) \). As a test, we include data from the latter half of the first HST orbit and derive a deeper eclipse depth (54 ppm) and shallower \( r_{2,3} \) values (the other free parameters are unaltered). We conclude that this parameter degeneracy is the most likely source of the fourth visit’s eclipse depth inconsistency. As an additional test, we

<table>
<thead>
<tr>
<th>Visit #</th>
<th>Observation Date</th>
<th>Eclipse Depth (ppm)</th>
<th>( \chi^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2014 Sep 16</td>
<td>113 ± 36</td>
<td>1.11</td>
</tr>
<tr>
<td>2</td>
<td>2014 Oct 18</td>
<td>121 ± 35</td>
<td>0.87</td>
</tr>
<tr>
<td>3</td>
<td>2014 Nov 15</td>
<td>50 ± 28</td>
<td>1.43</td>
</tr>
<tr>
<td>4</td>
<td>2014 Dec 10</td>
<td>6 ± 35</td>
<td>1.18</td>
</tr>
<tr>
<td>5</td>
<td>2014 Dec 31</td>
<td>122 ± 34</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Note. a Adopted value.
explored the impact of both linear and quadratic visit-long model components on the Visit 3 eclipse depth. Fitting only visit 3 while excluding the first orbit, we obtain eclipse depths of 62 and 283 ppm for the quadratic and linear models, respectively. Adding the last half of the first orbit increases the eclipse depth from 62 to 119 ppm for the quadratic model, which is consistent with visits 1, 2, and 5, but also increases the reduced chi-square value from 1.39 to 1.64. Many WFC3 primary transit analyses emphasize obtaining robust relative transit depths; however, secondary eclipse analyses rely on the absolute depth to determine the planet’s temperature. This places additional emphasis on correctly handling the white light-curve systematics.

When we exclude the first *HST* orbit and include all of the visits in a joint fit, the best-fit model favors a shared eclipse depth of 82 ± 15 ppm. Adding data from the latter half of the first *HST* orbit to all visits increases the shared eclipse depth to 97 ± 15 ppm. This value is comparable to a fit that excludes the first *HST* orbit and the fourth visit (99 ± 17 ppm). We choose to adopt this final value as our estimate of the true white light-curve eclipse depth (Table 1). Figure 2 depicts systematics-removed, binned data (excluding visit 4) with a best-fit light-curve model from our joint fit.

### 2.3. Spectroscopic Light-curve Fits

We apply the Divide-White technique (Stevenson et al. 2014a) to remove the wavelength-independent systematics from the spectroscopic light curves. This is accomplished by dividing the spectroscopic light curves from each visit by their corresponding, transit-removed white light curves, which serve as nonanalytic models of the wavelength-independent systematics. Our final spectroscopic light-curve model takes the form

$$F(\lambda, \phi) = F_t(\lambda, \phi) E(\lambda, \phi) R(\lambda, \phi) W(\phi),$$

(2)

where $F(\lambda, \phi)$ (the wavelength-dependent modeled flux) no longer depends on *HST*'s orbital phase, the definitions for $F_t(\lambda, \phi)$ and $E(\lambda, \phi)$ are unchanged, $R(\lambda, \phi) = 1 + r_{\lambda,1}(\phi - 0.5)$ is now a time-dependent linear trend with no scan direction dependence on $r_{\lambda,1}$, and $W(\phi)$ is the nonanalytic, transit-removed white-light-curve model. The simplicity of this model relative to Equation (1) highlights the effectiveness of using the Divide-White technique to analyze *HST*/WFC3 data. We show in Figure 3 the residual rms versus bin size for the spectroscopic light-curve fits, which demonstrates that we have successfully removed time-dependent systematics.

In Table 2, we provide the spectroscopic eclipse depths from a joint fit encompassing visits 1, 2, 3, and 5. Adding the fourth visit effectively reduces the eclipse depths by an average of ~17 ppm and imparts no additional wavelength modulation. The listed eclipse depth uncertainties are relative values; therefore, a wavelength-independent offset (Table 1) should be applied when fitting atmospheric models that consider additional data. In this case, a Gaussian prior of width 17 ppm (the white light-curve eclipse depth uncertainty) is appropriate. Table 2 also lists mean $\chi^2$ values for each channel; the overall mean is 1.04 times the photon plus read-noise limit. In the next section we interpret these measurements within an atmospheric retrieval framework.

### Table 2

<table>
<thead>
<tr>
<th>Wavelength (μm)</th>
<th>Eclipse Depth (ppm)</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.125–1.178</td>
<td>43 ± 16</td>
<td>0.98</td>
</tr>
<tr>
<td>1.178–1.230</td>
<td>97 ± 15</td>
<td>1.01</td>
</tr>
<tr>
<td>1.230–1.282</td>
<td>115 ± 15</td>
<td>1.02</td>
</tr>
<tr>
<td>1.282–1.335</td>
<td>109 ± 15</td>
<td>1.13</td>
</tr>
<tr>
<td>1.335–1.388</td>
<td>80 ± 15</td>
<td>1.00</td>
</tr>
<tr>
<td>1.388–1.440</td>
<td>43 ± 15</td>
<td>0.96</td>
</tr>
<tr>
<td>1.440–1.492</td>
<td>86 ± 15</td>
<td>0.84</td>
</tr>
<tr>
<td>1.492–1.545</td>
<td>107 ± 17</td>
<td>1.11</td>
</tr>
<tr>
<td>1.545–1.598</td>
<td>167 ± 17</td>
<td>1.20</td>
</tr>
<tr>
<td>1.598–1.650</td>
<td>170 ± 18</td>
<td>1.12</td>
</tr>
</tbody>
</table>

The first number is slightly different from the value listed in Table 1 (50 ppm) because, for this test, we did not perform a joint fit with all of the visits.
3. ATMOSPHERIC RETRIEVAL ANALYSIS

Our new WFC3 measurements, combined with Spitzer broadband photometry\(^\text{13}\) (Diamond-Lowe et al. 2014), are interpreted within the context of an atmospheric retrieval framework. In what follows we first describe our nominal model setup (Section 3.1) and results (Section 3.2). We also test the robustness of our nominal model assumptions by exploring a different temperature profile parameterization (Section 3.3), the inclusion of a dayside cloud (Section 3.4), and the impact of different Spitzer data reduction methods (Section 3.5). Finally, we explore the chemical plausibility of our retrieved abundances through a new method, chemical retrieval-on-retrieval (Section 3.6).

3.1. Atmospheric Parameterization

We derive a representative of the emission portion of the CHIMERA retrieval suite (Line et al. 2013a, 2014; Diamond-Lowe et al. 2014; Kreidberg et al. 2014a; Stevenson et al. 2014c) to interpret the full dayside emission data of HD 209458b. The CHIMERA emission forward model computes the upwelling disk-integrated thermal emission given a temperature structure and molecular abundances and has been validated against other models in two publications (Line et al. 2013a; Morley et al. 2015). This forward model is coupled with the powerful pymultinest routine (Buchner et al. 2014), a python wrapper to the commonly used multinest nested sampling algorithm (Skilling 2006; Feroz et al. 2009), to perform the parameter estimation and Bayesian evidence computation.

The temperature–pressure (TP) profile is parameterized (with five parameters) using the analytic approximation from Parmenier & Guillot (2014; see Line et al. 2013a for implementation details). The parameterization is flexible enough to permit a wide range of thermal structures, including inversions, while also resulting in physically realistic profiles consistent with radiative energy balance as the analytic approximation enforces radiative equilibrium in its derivation (see also Hansen 2008; Guillot 2010; Robinson & Catling 2012; Heng et al. 2014). For this investigation we retrieve seven thermochemically plausible gases that absorb over the wavelengths of interest: H₂O, CH₄, CO, CO₂, NH₃, HCN, and C₂H₂ (e.g., Madhusudhan et al. 2011), with uniform-in-log mixing ratio priors spanning from −12 to 0. Molecular hydrogen and helium (in solar proportions) are assumed to comprise the remaining gas such that all species sum to unity. We use the absorption cross-section database described in Freedman et al. (2014). For this analysis we make no a priori assumption about the realism of our retrieved abundances, but will later show that they are consistent with expectations from thermochemical equilibrium, suggesting that both the data and model reasonably reflect reality. All emission model results presented in this work assume cloud-free atmospheres unless otherwise stated. We also assume a free parameter offset between the WFC3 and Spitzer data to account for uncertainties in the absolute calibration in the white light transit depths, though the impact of this parameter is minimal in our nominal retrieval (found to be consistent with no offset, 3 ± 10 ppm).

\(^{13}\) While there are numerous sources of data sets available (HST NICMOS, Spitzer IRS, MPIS, NASA IRTF, etc.), we chose to keep it simple by combining a reasonably new reduction of mid-infrared space-based Spitzer data with our new WFC3 observations. A more comprehensive analysis exploring the impact of the many possible data sets and/or combinations thereof is beyond the scope of this paper.

3.2. Fiducial Results

Figures 4 and 5 summarize our nominal retrieval results under the aforementioned model assumptions, denoted as the FULL model scenario. The spread in the models and fits suggests that the model and the data are well matched (median chi-squared per data point of 1.03), much like our previous results on WASP-43b, suggesting that our atmospheric parameterization is adequate; hence, parameter estimates from this data–model combination have meaning. It comes as no surprise that the presence of a deep water vapor absorption feature in the WFC3 bandpass results in a monotonically decreasing TP profile over the planet’s infrared photosphere (1 bar–1 mbar). Water is the only robustly constrained molecular absorber, with an abundance constraint of ±1 order of magnitude nearly centered about solar composition (68% confidence interval: 4.0 × 10⁻⁵–5.4 × 10⁻³). We can only retrieve upper limits or long unconstrained tails (CO₂) on the other gases. The resulting abundance distributions (Figure 5) appear physically plausible (with the exception of CO₂; more below), or at least they encompass physically plausible molecular combinations and are consistent with or near solar abundance to within 3σ.

In order to further demonstrate the lack of need for a thermal inversion and to identify which molecules are actually “detected” in this spectrum, we perform a standard nested Bayesian model comparison using the full Bayesian evidence calculation from pymultinest (Trotta 2008; Benneke & Seager 2013; Swain et al. 2014; Kreidberg et al. 2015; Waldmann et al. 2015a, 2015b). The Bayesian evidence is what the commonly used BIC approximates via a truncated Laplace approximation (Cornish & Littenberg 2007; Trotta 2008). A parameter is deemed “necessary” or “detected” if the improvement in likelihoods from the addition of that given parameter (or parameters) outweighs the increase in prior volume (Chapter 3.5, Gregory 2005). This is a straightforward method for determining whether or not a gas is detected as we simply remove that gas from the model and rerun the retrieval to compute the new evidence. The ratio of the evidences of two models (usually a FULL model with all parameters compared with subsets of that model) is the Bayes factor, which can be directly converted into a confidence interval/detection significance (Equation (27), Table 2 in Trotta 2008, reproduced in Benneke & Seager 2013).

Table 3 summarizes the detection significances of the various nested models. We find that water is detected at 6.2σ confidence. This detection, as with our WASP-43b WFC3 observations (Kreidberg et al. 2014b), is strongly driven by the 1.4 μm water feature covered by WFC3. The reduced gases, expected to occur in cooler planets (CH₄, NH₃) or high carbon-to-oxygen ratio atmospheres (HCN, C₂H₂), are not detected, and in fact their inclusion is penalized due to the increase in prior volume without the accompanying improvement in fit quality, consistent with retrieving only upper limits on their abundances. CO+CO₂ are detected at 4.1σ confidence. This detection is largely driven by the 4.5 μm IRAC observation. We cannot decipher the difference between CO and CO₂ as they have nearly overlapping features over the 4.5 μm bandpass; thus, the removal of one is compensated by the other.

To compute the “detection” (or lack thereof) of a thermal inversion, we perform a separate retrieval in which the prior five-parameter analytic TP profile forces TP profiles that have zero to positive temperature gradients with increasing altitude.
In other words, we restrict the prior search space to exclude monotonically decreasing profiles. We have not removed or added parameters; we simply reduced the prior volume by restricting the prior ranges. Upon comparing the Bayesian evidence of the two models, we find that a monotonically decreasing TP profile is required at the 7.7σ level.

In summary, we have robustly detected absorption due to water at or near what is expected from solar elemental abundances (±1 dex), a detection of some combination of CO and CO$_2$, and strong evidence for a monotonically decreasing temperature profile over the pressure levels robustly probed by the observations (1 bar–1 mbar). In the following sections we will explore how various model and data assumptions will impact these conclusions.

3.3. Impact of TP Profile Parameterization

Given that one of the controversies surrounding the dayside spectrum of HD 209458b is whether or not it possesses a stratospHERIC thermal inversion, it is thus prudent to explore how our thermal profile parameterization could influence our conclusion. In the nominal model, described above, we used an analytic five-parameter gray radiative equilibrium description for the thermal profile. Here we use a much simpler profile with far fewer assumptions (similar to Schwarz et al. 2015). The atmosphere is divided up into three regions: a deep isothermal region defined by a deep temperature parameter (similar to Madhusudhan & Seager 2009) and two “$T$ linear-in-log($P$)” regions defined by a slope and two log-pressure free parameters defining the break between the three regions, for a total of five parameters. Continuity between the regions is enforced. This simple parameterization does not abide by any physical rules (e.g., radiative equilibrium) and is strictly guided by the data and parameter priors. All other aforementioned model assumptions remain the same.

Figure 6 compares the results from this alternate TP profile parameterization with our nominal results from Section 3.2. In general, the conclusions are unchanged: water is observed to be at or near solar composition, upper limits obtained for CH$_4$, NH$_3$, HCN, and C$_2$H$_2$, and there is similar temperature gradient over the photosphere (1 bar–1 mbar). In fact, upon inspecting the spectra in Figure 6, we find virtually no difference between the fiducial model fits and the simple TP profile fits. There are a

---

14 By robustly, we mean where most of the integrated area of the weighting functions lies. There is some very weak contribution at lower pressure levels in the broad 4.5 μm IRAC bandpass.
Table 3

<table>
<thead>
<tr>
<th>Scenario (Included Gases)</th>
<th>ln B</th>
<th>Det. Sig.(σ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FULL (all gases)</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Molecules other than water (H2O)</td>
<td>3.96</td>
<td>3.3</td>
</tr>
<tr>
<td>CO+CO2, (H2O, CH4, NH3, HCN, C2H2)</td>
<td>6.90</td>
<td>4.1</td>
</tr>
<tr>
<td>CH4+NH3, (H2O, CO, CO2, HCN, C2H2)</td>
<td>−0.92</td>
<td>undefined</td>
</tr>
<tr>
<td>HCN+C2H2, (H2O, CH4, CO, CO2, NH3)</td>
<td>−0.71</td>
<td>undefined</td>
</tr>
<tr>
<td>H2O, (CH4, CO, CO2, NH3, HCN, C2H2)</td>
<td>17.10</td>
<td>6.2</td>
</tr>
<tr>
<td>Monotonically decreasing T (all gases)</td>
<td>27.10</td>
<td>7.7</td>
</tr>
</tbody>
</table>

Note. The scenarios indicate which molecules/setup we are detecting. Water, CO+CO2, and a decreasing temperature profile are detected at high confidence. The full compliment of gases includes H2O, CH4, CO, CO2, NH3, HCN, and C2H2.

few subtle differences, however: CO is now favored over CO2 (though both are still largely unconstrained), the decreasing TP profile extends to slightly lower pressures and is better constrained, and the water abundance constraint is 30% smaller (±0.7 orders of magnitude). The change in the CO/CO2 abundance distributions is likely due to the slightly steeper temperature gradient over the 4.5 μm bandpass resulting from the simple TP profile parameterization, though solar composition abundances are captured well within their 68% confidence interval. Again, any inference about these molecules is entirely driven by a single Spitzer point and is thus highly sensitive to any model assumptions. This is because the spectral slopes over the 4.5 μm bandpass are quite steep. Changing the abundance of both increases the depth of the “V”-shaped absorption feature, but each has a different effect on the spectral slopes. CO more strongly affects the blue edge of the bandpass, whereas CO2 affects the red edge. The temperature gradient affects the slope at both edges. The integrated flux is dependent on the depth and both the red and blue slopes, making for a complicated degeneracy. Any resulting conclusions about the CO/CO2 abundances are therefore not robust. There also appears to be an “inversion” above the 1 mbar pressure level. This turnover in temperature is largely outside the photosphere—where the data are probing—and is largely driven by the prior on the temperature gradient in this region of this particular parameterization. Care must be taken when interpreting temperature profile results outside of the photosphere. For instance, our current data would not be able to rule out the presence of a thermosphere inversion, which occurs at a few microbars and lower pressures, driven by upper atmospheric energetics (e.g., Yelle 2004).

3.4. Impact of Dayside Clouds

The impact of clouds on dayside emission spectra has largely been ignored, given that there has not yet been any observational evidence to justify otherwise. This is surprising since it is widely accepted (both observationally and theoretically) that clouds strongly impact the spectra of similar-temperature brown dwarfs (e.g., Burrows 2001; Kirkpatrick 2005; Marley & Robinson 2015), along with the mounting evidence for clouds in transmission spectra and in reflected light curves from Kepler (e.g., Demory et al. 2013; Esteves et al. 2015; Shporer & Hu 2015). While clouds can dominate transmission spectra due to the enhanced path length through the atmosphere (Fortney 2005), dayside inhomogeneities (or patchiness) driven by circulation (Parmentier et al. 2016) could weaken the effect of clouds if they are present. The appearance of the strong water absorption feature over the WFC3 bandpass suggests that the impact of clouds on the dayside emission spectrum of HD 209458b is minimal. However, to be thorough, we test this hypothesis by introducing a simple, gray, nonscattering cloud in emission with a mass opacity, κc, given by

$$\kappa_c(P) = \kappa_{P_0} \left( \frac{P}{P_0} \right)$$

where $P_0$ is the cloud base pressure, κ_{P_0} is the specific absorption coefficient (area/mass) at the base of the cloud, and $\alpha$ is the cloud fill factor index (Burrows et al. 2006).

Figure 7 summarizes the results. As anticipated, we find that the cloud has negligible impact on the retrieved abundances and thermal structure, though the retrieved cloudy TP profile is ever so slightly shifted “upward” in pressure and has greater uncertainty in the deep atmosphere. The Bayes factor between the cloudy model and clear model (0.15—extremely insignificantly) suggests that a cloud is not required or justified. The histogram inset in Figure 7 shows the integrated cloud optical depth¹⁵ over the pressure levels probed by the WFC3 water band for 1000 randomly drawn samples. Most of the optical

¹⁵ The photosphere for each sample occurs at a different pressure level. Therefore, to compute the integrated cloud optical depth, we convolved the cloud optical depth profiles with the normalized 1.256 (continuum) and 1.414 (water band) μm contribution functions. This is a rough approximation to the integrated optical depth over the pressure levels probed by the WFC3 observations.
and molecular abundances. The temperature gradients over the photosphere
are largely consistent, though the uncertainties are smaller within the
simple TP parameterization. The turnover in the TP parameterization above the 1 mbar level (outside most of the photosphere in the “simple TP parameterization” is largely

Figure 6. Impact of the temperature profile parameterization on the retrieval results. Plot structure similar to Figure 4. In all panels, red corresponds to the nominal
Parmentier & Guillot (2014) analytic radiative equilibrium parameterization described in Section 3.1, and blue to the “simple” TP parameterization described in
Section 3.3. The spectra resulting from each TP profile parameterization are nearly indistinguishable, but there are noticeable differences in the retrieved TP profiles and molecular abundances. The temperature gradients over the photosphere (∼1–0.001 bars) are largely consistent, though the uncertainties are smaller within the
simple TP parameterization. The turnover in the TP profile above the 1 mbar level (outside most of the photosphere) in the “simple TP parameterization” is largely
driven by the prior on the slope in that region. Finally, the abundances are largely consistent with the exception of CO and CO2. See Section 3.3 for details.

3.5. Impact of Spitzer Data Reduction Technique

The robustness of published Spitzer results has been called
into question on multiple occasions (Hansen et al. 2014 [and comparison of references therein]; Diamond-Lowe et al. 2014; Lanotte et al. 2014), and photometric flux values for many planets have changed by several sigma. As such, we feel that it is necessary to explore how the differences in published
dayside photometry impact our results. As an illustrative example, we compare the retrieved quantities when using the
Spitzer photometry presented in Diamond-Lowe et al. (2014)
(and what we use as our nominal) with those resulting from the
photometry published in Evans et al. (2015) (Figure 8). The measured eclipse depths from Diamond-Lowe et al. (2014) and
Evans et al. (2015) are consistent (to within 1σ), but the latter
report larger uncertainties in all channels. This is typical of
Gaussian processes (GP, used in Evans et al. 2015) relative to
other methods, as determined in a recent comparison of Spitzer
data reduction techniques (Ingalls et al. 2016). GP also ranked
lower in repeatability and reliability compared to BLISS
mapping (Stevenson et al. 2012), which was the systematic
correction technique used by Diamond-Lowe et al. (2014). For these reasons, we adopt the eclipse depths reported by
Diamond-Lowe et al. (2014) to derive our nominal results.

The subtle differences in the two reductions are reflected in the fits, as the amplitude of the CO/CO2 absorption feature at 4.5 μm is shallower with the Evans et al. (2015) data. These
differences in general have a small effect on the retrieved
quantities. The largest noticeable difference, however, is in the
shape of the CO2 histogram, which has “filled” out more at low
abundances. The detailed shape of the water histogram changed
slightly, with the median shifting by 5% and the 68% confidence
interval widening by 20% (in log(H2O)). This increased water
uncertainty is due to the increased temperature uncertainty in the
TP profile around ∼100 mbar, where the in-water WFC3
contribution functions partially overlap with the 4.5 μm Spitzer
contribution function. We also experimented with the full-phase-
derived 4.5 μm eclipse depth from Zellem et al. (2014b). The
eclipse depth falls between the Diamond-Lowe et al. (2014) and
Evans et al. (2015) measurements and thus results in retrieved
values (not shown) that fall in between those derived from the
Diamond-Lowe et al. (2014) and Evans et al. (2015) data. This
exercise illustrates the complex sensitivity of the retrieved
quantities to the observations, especially when the constraint on a
given species (e.g., CO, CO2) is entirely dependent on a single

16 However, the GP analysis of Evans et al. (2016) was able to produce consistent results for all epochs, whereas the Diamond-Lowe et al. (2014) paper
discarded the 2005 and 2010b 4.5 μm results.
photometric datum. Nevertheless, the main conclusions of this paper (water abundance consistent with solar and no thermal inversion) are unchanged no matter which Spitzer data are used.

3.6. Elemental Abundance Analysis: Chemical Retrieval-on-retrieval

In order to more rigorously determine the derived elemental abundances from the molecular abundances (parameterized through the metallicity, [M/H], and C/O), we introduce a new analysis method, chemical retrieval-on-retrieval. The chemical retrieval-on-retrieval approach straddles the line between the classic "fully data driven" (Madhusudhan & Seager 2009; Barstow et al. 2013; Lee et al. 2012; Line et al. 2012, 2014; Benneke & Seager 2013) and "a priori imposition of physical processes" (Fraine et al. 2014; Knutson et al. 2014; Kreidberg et al. 2015; Benneke 2015). This allows us both to obtain the full range of possible molecular combinations that can explain the data and to address a posteriori what regions of that parameter space are plausible. We note that this is the approach taken within the solar system community: molecular abundances (or their profiles) are directly retrieved from the data (whether in situ or remotely) and then compared to chemical models (e.g., Greathouse et al. 2005; Moses et al. 2005; Visscher et al. 2010; Orton et al. 2014).

The chemical retrieval-on-retrieval is a two-step process (as its name implies). First, we retrieve the molecular abundances and TP profile directly from the spectroscopic/photometric data, making no assumptions about the relationship between the molecular abundances (Section 3.2). We then treat the retrieved marginalized mixing ratio posterior distributions (from Section 3.2) as independent data points to which we fit a thermochemical equilibrium model (again using pymultinest). From a statistical perspective, this is very much how one treats wavelength-dependent eclipse depths as independent data points when performing a classic atmospheric retrieval. The thermochemical model (NASA CEA2; Gordon & McBride 1994; Line et al. 2011; Moses et al. 2011; Kreidberg et al. 2015; Mollière et al. 2015) computes the molecular mixing ratios along a given pressure–temperature profile given [M/H] and C/O, which are the retrieved parameters.

We consider all species (~2000) in the CEA2 thermodynamic library that contain H, He, C, N, O, S, P, Fe, Ti, V, Na, and K and assume pure equilibrium (no rainout). Rather than assume Gaussian distributions for the mixing ratio data (e.g., as one does when using chi-squared), we evaluate the probability density of each thermochemically computed mixing ratio with the normalized histograms themselves, through interpolation. When computing the net log-likelihood for multiple gases, we simply sum the log of their probability densities together.\textsuperscript{17} We propagate the uncertainty in the retrieved TP profile from the atmospheric retrieval into the retrieved metallicity and C/O by drawing a random TP profile from the TP posterior at each pymultinest likelihood evaluation. This ignores the

\textsuperscript{17}This process is the equivalent of computing chi-square for normal distributions.
correlations between the elemental abundances and the TP profile; however, this is acceptable as the marginalized molecular abundances contain that information. We integrate out vertical profile information by column-averaging the thermochemically derived mixing ratio profiles over the infrared photosphere (3 bars–1 mbar), as low-resolution spectra are largely insensitive to vertical profile information (e.g., Lee et al. 2012; Barstow et al. 2013).

Within this framework, we investigate how the constraints on each molecular species contribute to our knowledge of the metallicity and C/O. We first ask how well we can constrain the metallicity and C/O from the retrieved water abundance alone, followed by water combined with methane, and finally water combined with methane and carbon dioxide. Figure 9 summarizes the results. The left panel shows the metallicity and C/O constraints obtained for each of the three scenarios. The right panel shows the mixing ratios derived from the retrieved metallicity and C/O under thermochemical equilibrium (colored histograms), within the photosphere, compared with the mixing ratios retrieved directly from the data (thick black histograms).

When using only water to constrain the elemental abundances, we find a metallicity range of 0.06–9.8 × solar.18 There still remains, at 3σ, a possibility of higher metallicities. We also find, unsurprisingly, that the water histogram derived from the chemical retrieval is nearly identical to that retrieved from the data, that is, the chemical model is providing a “good fit” to the mixing ratio “data.” Because there are no carbon species fit for in this scenario, there is a small trickle of probability beyond C/O = 1. In the predicted mixing ratio distributions this low-probability high C/O results in a low-probability tail of appreciable amounts of HCN and C2H2. However, when we include the upper limit derived from the methane mixing ratio, the high C/O tail is eliminated, also eliminating the HCN and C2H2 possibilities. Water and methane together put a firm upper limit on the C/O ratio, but offer little to constrain any lower limit since there is no lower limit to the methane mixing ratio retrieved from the spectrum. The retrieved water and methane distributions are consistent with the predicted distributions when fit together, suggesting that the retrieved water and methane mixing ratios are physically plausible.

The predicted CO mixing ratios from the water and water +methane scenarios are well within the spectrally retrieved CO upper limit. However, we do find that that most of the chemically predicted CO2 probability is significantly below the retrieved CO2 distribution. This suggests that the chemically retrieved metallicity and C/O, while self-consistently able to explain all the other spectrally retrieved molecular species, are unable to adequately explain the CO2 abundance. Finally, when including the retrieved CO2 mixing ratio, with methane and water, we find that a higher metallicity (1.3–123 × solar) is

---

18 This is a more accurate reflection of the true atmospheric metallicity than simply comparing the retrieved water abundance to what the water abundance should be at solar composition in thermochemical equilibrium, as this method takes into account the degeneracy between the C/O and metallicity on the water abundance. Lower C/O values require a lower metallicity to give the same water abundance, and vice versa.
required, which is unsurprising since CO2 is strongly sensitive to metallicity. The requirement to fit the relatively high CO2 abundance via high metallicity also pulls the chemically predicted water abundance to higher values, to the point of not being a particularly good fit to the spectrally retrieved water abundance distribution. In short, the spectrally retrieved CO2 abundance and water abundances are not particularly consistent with each other.

We note that we obtained a similarly high CO2 abundance for our WASP-43b dayside results (Kreidberg et al. 2014a) when using the 3.6 and 4.5 μm Spitzer data. Obviously it is unphysical for CO2 abundances to be larger than the CO abundance for solar metallicities in thermochemical equilibrium (Prinn & Barshay 1977; Heng & Lyons 2016). Perhaps the unusually high CO2 abundance is because we are missing some chemical process that enhances CO2 beyond thermochemical equilibrium. This is unlikely, however, as the most likely processes to do so are chemical quenching via vertical mixing, or photochemistry. Both of these processes are unlikely to significantly enhance the CO2 abundance to this level in an atmosphere of these temperatures (Cooper & Showman et al. 2006; Moses et al. 2011, Figure 3). Dynamical-chemical effects are also unlikely to significantly impact the abundances (Cooper & Showman et al. 2006; Agundez et al. 2014). We cannot rule out that there could be some additional unknown chemical mechanism that we cannot think of. More likely, however, is that the retrieved CO2 abundance is not robust. This is supported by the high degree of sensitivity of the CO2 posterior to our choice of TP profile parameterization or Spitzer data, something to consider for future analyses. We note in an independent analysis (Feng et al. 2016, submitted) with synthetic data of a similar observational setup (WFC3+Spitzer IRAC 3.6 and 4.5 μm channels) that the retrieved CO2 abundance is highly sensitive to the particular random noise draw of the 4.5 μm point. Again, this suggests that spurious abundances derived from a single broadband measurement should be taken with a grain of salt. The nonlinearities of infrared radiative transfer, combined with ultralow spectral resolution, conspire to produce unphysical solutions. We are thus more inclined to believe the chemically retrieved elemental abundances derived via the other molecular species, which are less impacted by our assumptions.

To summarize, we find that the chemical retrieval-on-retrieval method produces self-consistent chemical abundances, in agreement with most of our spectrally retrieved abundances at 0.06–9.8 × solar and C/O < 1. The method is not able to self-consistently reproduce the spectrally retrieved CO2. However, we question the robustness of the spectrally retrieved CO2 abundances, which are largely derived from a single broadband measurement and are therefore highly sensitive to our model assumptions.

4. DISCUSSION AND CONCLUSIONS

We present the first high-precision HST WFC3 measurements of the dayside emission of the canonical hot Jupiter HD 209458b. These measurements achieve an ultrahigh, photon-noise-limited precision of 15 ppm, providing one of the most robust detections of water (6.2σ) in a dayside emission spectra. The deep water feature, combined with four broadband Spitzer observations, suggests a monotonically decreasing temperature profile at 7.7σ confidence, firmly ruling out the presence of a thermal inversion between 1 bar and 1 mbar. We tested the robustness of our atmospheric inferences by exploring multiple model assumptions, such as the impact of the temperature profile parameterization, influence of a gray dayside cloud, and impact of the source of Spitzer photometry. We found that the decreasing temperature profile and water abundance were largely invariant to these assumptions, but that the retrieved CO and CO2 abundances were strongly assumption dependent, owing to their dependence on a single broadband Spitzer measurement (4.5 μm).

In Figure 10 we compare the two versions of the retrieved TP profile to two profiles from a self-consistent radiative
in terms of the temperature gradient over the regions probed by the temperature profile. The measured temperature profile (red curve; Fortney et al. 2008) is cooler than the self-consistent TP profile (black curve; Fortney 2005; Fortney et al. 2008), as the cooler profile corresponds to full redistribution of energy, whereas the hotter profile corresponds to a dayside-only reradiation. The retrieved profiles fall between the two, suggesting a redistribution that falls in between, broadly consistent with previous phase-curve observations (Zellem et al. 2014b). Overall, the agreement is fairly good in terms of the temperature gradient over the regions probed by the observations (1 bar–1 mbar).

Figure 10. Comparison of retrieved temperature profiles (under both temperature profile parameterizations; red and blue) with self-consistent radiative-convective-thermochemical equilibrium models (black curves; Fortney et al. 2008). The cooler self-consistent TP profile corresponds to full redistribution, whereas the hotter profile corresponds to dayside-only reradiation. The retrieved profiles fall between the two, suggesting a redistribution that falls in between, broadly consistent with previous phase-curve observations (Zellem et al. 2014b). Overall, the agreement is fairly good in terms of the temperature gradient over the regions probed by the observations (1 bar–1 mbar).

equilibrium model (Fortney 2005; Fortney et al. 2008). The hotter black profile assumes only redistribution of energy over the dayside. The cooler profile is a planet-wide average. These self-consistent models include equilibrium chemistry and solar abundances and yield Bond albedos of 0.08. The retrievals yield a profile intermediate between the radiative equilibrium cases, suggesting either a larger Bond albedo for the planet or modest energy redistribution to the nightside. The very low geometric albedo from MOST for this planet (Rowe et al. 2008) suggests a low Bond albedo, which favors the explanation that energy is being advected to the nightside, as seen in the 4.5 μm light curve from Zellem et al. (2014b).

This relatively “cool” dayside profile is in contrast to our findings for WASP-33b from Stevenson et al. (2014c, Figure S5). For that planet we found a profile consistent with little energy loss to the nightside, which was in agreement with the low nightside flux seen in the full-orbit phase curve. The two planets have nearly equal equilibrium temperatures, but other significant differences: a 4.3 times longer rotation period for HD 209458b (assuming tidal locking), and at 4.8 times larger surface gravity for WASP-43b. These differences could be important in understanding the energy balance and redistribution of the two planets.

As discussed in Section 1, the claimed detection of a thermal inversion relatively deep in HD 209458b’s infrared photosphere launched an active area of research to look for thermal inversions in other planets, search for the short-wave absorber responsible for causing them, and elucidate the phenomenon theoretically. We had previously suggested that the thermal inversion hypothesis for this prototype was not correct based on a reevaluation of the Spitzer photometry that the original inference was based on. Now we show definitively that HD 209458b does not have a thermal inversion in its infrared photosphere using high-precision spectroscopy. The existence of similar thermal inversions in other conventional hot Jupiters (i.e., those with similar temperatures to HD 209458b, say, 800 K < \( T_{eq} \) < 2000 K), which all have lower-quality data, should now be questioned. Parmentier et al. (2016) suggest that thermal inversions should not exist at \( T < 1900 \) K due to a deep CaTiO3 cold trap, removing titanium from the atmosphere. The recent possible detection of a thermal inversion for WASP-33b (Haynes et al. 2015) is not inconsistent with this statement, as this planet exhibits an average dayside brightness temperature of 3000 K, which means that it is in a very different physical and chemical regime than more typical hot Jupiters like HD 209458b.

In retrospect, the thermal inversions proposed for hot Jupiters were always difficult to understand. Perhaps the most serious issue for the thermal inversion hypothesis is that the Spitzer photometric bands typically used to discriminate for inversions are expected to probe similar pressure levels for solar-like composition atmospheres in chemical equilibrium (Burrows et al. 2007, 2008; Showman et al. 2009). Therefore, the altitude and absorption strength of the short-wave observer had to be carefully tuned to give the large temperature difference over a small region of the atmosphere that was needed to match the data. The original leading candidate chemical species (TiO and VO) are now thought to not be present in the dayside atmospheres of conventional hot Jupiters because of the strong vertical mixing to keep these heavy molecules aloft at low pressures (Spiegel et al. 2009) and/or the rainout of larger condensate droplets on the cooler nightside (Parmentier et al. 2013). Furthermore, forced inclusion of these species in general circulation models (GCMs) could not reproduce the data anyway (Showman et al. 2009).

In addition to studying the thermal structure of HD 209458b, we also introduced a new elemental abundance analysis, chemical retrieval-on-retrieval, where we fit a posteriori the retrieved molecular abundances with a thermochemical model to determine the atmospheric metallicity and carbon-to-oxygen ratio. From this analysis we find that the measured water abundance from our spectrum of HD 209458b is consistent with 0.06–9.8 \( \times \) the solar metallicity, and we firmly rule out C/O values greater than 1. The metallicity derived from the water abundance measurement (from both the chemical retrieval-on-retrieval method and comparison of the retrieved water abundance to solar) is consistent with the trend of increasing atmospheric metallicity for decreasing planet mass that is observed in our solar system (see Figure 11). This result extends the agreement that is seen between the solar system trend and exoplanet atmosphere abundances, which was first investigated by Kreidberg et al. (2014a). The C/O upper limit of 1 butresses the similarly strict limits that were recently obtained for HD 209458b and other hot Jupiters from the interpretation of transit transmission spectra while accounting for clouds (Benneke 2015; Kreidberg et al. 2015).

Our solar water abundance determination for HD 209458b is inconsistent with the strongly subsolar (0.007–0.05 \( \times \) solar) abundance interpretation of the planet’s WFC3 transmission spectrum by Madhusudhan et al. (2014a). However, Madhusudhan et al. (2014a) did not account for clouds in their retrieval. We note that the original Deming et al. (2013)
analysis suggested that a cloud is likely what is muddling the water feature amplitude. Benneke (2015) and our own unpublished analysis have shown that neglecting clouds in the interpretation of the planet’s WFC3 transmission spectrum yields spuriously low and precise abundances because clouds are truncating the water absorption feature (e.g., Iyer et al. 2016) compared to the expectation for a clear atmosphere. The water abundance we derive from the dayside emission spectrum of HD 209458b is consistent with, but more precise than, the water abundance derived from the transmission spectrum by Benneke (2015). The consistency between water abundance values derived from dayside emission and transmission spectra is expected for hot Jupiters on thermochemical grounds (Moses et al. 2011) and was previously seen for the only other planet for which this comparison has been made (WASP-43b; Kreidberg et al. 2014a). We also find that our retrieved molecular abundances are thermochemically self-consistent, and thus “physical,” with the exception of CO₂, which is best explained with metallicities >10× solar. However, again, the retrieved CO₂ abundance is strongly model and data dependent, suggesting that robust inference of CO₂ is not yet possible.

Of the seven planets with published HST WFC3 emission spectra (WASP-12b, Swain et al. 2013; WASP-4b, TrES-3b, Ranjan et al. 2014; HD 189733b, Crouzet et al. 2014; WASP-43b, Kreidberg et al. 2014a; WASP-33b, Haynes et al. 2015), only two, WASP-43b and HD 209458b (this work), report a robust detection of water over the WFC3 bandpass. Detecting or not detecting water has significant implications for the nature of the thermal structure and/or composition. Over the WFC3 bandpass alone, it is difficult to determine whether or not the lack of a water absorption feature is due to isothermal atmospheres (or nearly isothermal), an intrinsic depletion of water, due to either high C/O or low metallicity, or an optically thick dayside cloud at high altitudes. This ambiguity emphasizes the need to push for higher-precision emission observations to determine the presence, or lack thereof, of water absorption over a wide range of stellar compositions (C/O, metallicity) and planetary equilibrium temperatures. Furthermore, by combining emission and transmission observations (e.g., Kreidberg et al. 2014a), we can begin to break the temperature profile—water–cloud degeneracy.

These observations were made under the GO Treasury Program 13467 with the NASA/ESA Hubble Space Telescope at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under the contract NAS 5-26555. M.R.L. acknowledges support provided by NASA through Hubble Fellowship grant 51362 awarded by the Space Telescope Science Institute. K.B.S. acknowledges support from the NASA Exoplanet Science Institute Sagan Postdoctoral Fellowship. J.L.B. acknowledges support from the David and Lucile Packard Foundation. We also thank Dan Foreman-Mackey for making the corner.py plotting routine available to the public and Johannes Buchner for making pymultinest publicly available. We also thank Vivien Parmentier, Kevin Heng, Tom Evans, Drake Deming, Adam Burrows, and Heather Knutson for useful discussion and comments on the manuscript. We thank Drake Deming, Peter McCullough, Adam Burrows, Sara Seager, David Charboneau, and Derek Homeier for being co-investigators on the HST observing proposal. Finally, we thank Roxana Lupu for support with the molecular absorption cross sections.

REFERENCES

Benneke, B. 2015, arXiv:1504.07655
Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, RVMP, 73, 719
Burrows, A., & Orton, G. 2010, Exoplanets, 419
Cornish, N. J., & Littenberg, T. B. 2007, PhRvD, 76, 083006