Draft version December 11, 2025
Typeset using LaTeX **default** style in AASTeX7.0.1

From Scratch: A Direct Test of the Sound-Horizon Assumption

S. A. Cooper [1]

[1] *Independent Researcher, Raleigh, NC, USA*

ABSTRACT

We present a *from scratch* evaluation of the sound horizon at photon decoupling, using only late-time distance and expansion measurements together with the directly measured CMB acoustic angle θ_s. All inputs are standard observational quantities employed in conventional cosmological analyses: the Planck 2018 acoustic scale θ_s and decoupling redshift z_*, the TRGB calibration of H_0, and published cosmic-chronometer measurements of $H(z)$. A Gaussian Process reconstruction of the expansion history provides the comoving distance integral to last scattering without invoking early-universe assumptions or adopting a theoretical value of r_s. Because BAO measurements constrain only the dimensionless ratios $D_M(z)/r_s$, they introduce a nearly perfect degeneracy between H_0 and r_s and cannot supply an empirical determination of r_s without external assumptions. This means that BAO trace only geometry through distance ratios and, by construction, cannot determine the physical scale r_s without assuming a theoretical sound horizon. The comoving angular-diameter distance $D_M(z_*)$ follows from integrating the observationally reconstructed $H(z)$, allowing an empirical determination of the sound horizon via

$$r_s = \theta_s D_M(z_*).$$

Using this relation, the resulting value, $r_s = 146.0 \pm 21.1$ Mpc, is consistent with the standard ΛCDM prediction of $r_s = 144.57 \pm 0.28$ Mpc within the large uncertainties arising from extrapolation beyond the redshift range of available chronometer data ($z \lesssim 2$).

This analysis demonstrates that current late-time expansion measurements cannot empirically constrain the sound horizon with sufficient precision to test ΛCDM's early-universe predictions. Gaussian-process and polynomial reconstructions both show that late-time data cannot recover the high-redshift expansion history required for evaluating r_s, confirming that the dominant uncertainty is the unavoidable extrapolation from $z \approx 2$ to $z \approx 1090$. We explicitly quantify the required extrapolation of the expansion history, which spans a factor of roughly 500 in redshift from the chronometer boundary at $z \simeq 2$ to recombination at $z_* \simeq 1090$. This evaluation shows that the "Hubble tension" reflects a conflict between direct measurement (local H_0) and theory-dependent inference (CMB-derived H_0 via assumed r_s), rather than a discrepancy between two independent observational determinations.

Keywords: Early universe (435) — Baryon acoustic oscillations (138) — Hubble time (762) — Gaussian Processes (1930) — Stellar astronomy (1583) — Monte Carlo methods (2238)

1. INTRODUCTION

The "Hubble tension" is usually framed as a discrepancy between the local determinations of the expansion rate, $H_0 \approx 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and the lower value inferred by fitting *Planck* CMB data, $H_0 \approx 67\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. This numerical disagreement has led to a large set of proposals that modify the early thermal history, recombination physics, or the sound speed of the primordial plasma.

What the CMB actually provides, however, is the angular acoustic scale

$$\theta_s = \frac{r_s}{D_M(z_*)},$$

Corresponding author: s.a.cooper.researcher@gmail.com

the ratio of the sound horizon at decoupling to the comoving distance of the last-scattering surface (L. Pogosian et al. 2024). Analyses that alter the ionization history reveal this point clearly: the procedure shifts theoretical values of r_s, not any observable quantity associated with H_0 (S. H. Mirpoorian et al. 2025). Converting θ_s into a value of H_0 requires a model for the early Universe that fixes r_s through assumptions about baryon density, recombination details, radiation content, and the expansion rate prior to decoupling.

All standard routes to H_0 reflect this structure. CMB fits, BAO analyses, BBN-calibrated methods, power-spectrum shape fits, and lensing combinations each treat r_s as fixed or model-defined (S. H. Mirpoorian et al. 2025). Even methods that keep only the measured acoustic angle θ_s still rely on a modeled connection between last-scattering and the drag epoch. Quantities often treated as inputs — such as the decoupling redshift z_*, the ratio $r_s/r_{\rm drag}$, and the physical densities $\Omega_b h^2$ and $\Omega_m h^2$ — enter only through ΛCDM modeling (L. Pogosian et al. 2024) rather than direct observation.

The commonly quoted CMB value of H_0 therefore reflects the ΛCDM prediction of a tightly constrained sound horizon, not a measurement. Once r_s is allowed to vary independently of the early-time model, the inferred value of H_0 shifts accordingly, and the discrepancy no longer qualifies as a tension. Because ΛCDM predicts r_s with an internal uncertainty of only $\pm 0.28\,$Mpc, the quantity is usually treated as fixed. That practice has been so widespread that it appears no previous study has examined whether r_s can even be estimated using late-time observations alone. Here we ask that question directly, using only late-time data and allowing the resulting uncertainty to reflect the limits of the observations. The purpose of this work is to determine whether the sound horizon can be inferred empirically from late-time observations alone, without adopting any early-Universe assumptions.

2. DATA

The analysis uses four observational inputs: (i) cosmic-chronometer measurements of $H(z)$, (ii) the Tip of the Red Giant Branch (TRGB) calibration of H_0, (iii) the CMB acoustic angle θ_s, and (iv) the decoupling redshift z_* from the Thomson visibility function. No additional cosmological parameters or priors are adopted.

2.1. *Cosmic Chronometer Expansion Rates*

We use published differential-age measurements of the Hubble parameter covering

$$0.07 \lesssim z \lesssim 1.97,$$

from Jimenez et al. (2003), Simon et al. (2005), Stern et al. (2010), Moresco et al. (2012, 2016, 2020), Zhang et al. (2014), and related work. Each measurement provides a value of $H(z_i)$ and its reported uncertainty σ_{H_i}.

Local Value of H_0 —We take the expansion rate at $z = 0$ from the TRGB distance ladder:

$$H_0 = 69.8\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$$

(W. L. Freedman et al. 2019).

CMB Angular Acoustic Scale—From the *Planck* 2018 analysis (P. Collaboration et al. 2020), the measured acoustic angle is

$$\theta_s = 0.01041.$$

Recombination Redshift—The decoupling redshift is taken from the *Planck* 2018 Thomson visibility function:

$$z_* = 1089.92.$$

3. METHODS

In this section, we describe how the observational inputs listed in Section 2 are used to obtain an empirical determination of the sound horizon r_s. To begin, we reconstruct the late-time expansion history from the chronometer measurements of $H(z)$, which allows us to evaluate the comoving distance to the last-scattering surface. Converting this to an angular-diameter distance and combining it with the observed CMB acoustic angle then gives r_s. The calculation uses only the inputs in Section 2 and adopts no early-time assumptions or ΛCDM parameters.

3.1. *Evaluation of the Observational Expansion History*

The chronometer measurements give a set of discrete values of the Hubble parameter across $0.07 \lesssim z \lesssim 1.97$, together with the TRGB value of H_0 at $z = 0$. To compute cosmological distances, these points must be turned into a continuous representation of $H(z)$. Since no assumptions about the matter content, curvature, dark energy, or early-time physics are adopted, the reconstruction has to be non-parametric and determined entirely by the data. Gaussian Process (GP) regression provides such a reconstruction. The GP treats the measured pairs (z_i, H_i), together with their uncertainties σ_{H_i}, as samples from an underlying smooth function and returns the smoothest curve consistent with the observations. The only assumption is that $H(z)$ changes smoothly with redshift, since a continuous function is needed for the distance integral. When the chronometer measurements are close together in redshift or have small uncertainties, the reconstructed $H(z)$ tends to follow them closely. In parts of the redshift range where the data are more spread out or less precise, the GP allows a wider range of possible values. In practice, this means the reconstruction has a central estimate of $H(z)$ along with an uncertainty band that reflects how strongly the data constrain it. The estimate and its uncertainty are carried forward into the later distance calculations.

The result is a continuous, data-driven reconstruction of the expansion history that can be used to evaluate cosmological distances without introducing assumptions about early-time conditions or the energy content of the Universe.

3.2. *Comoving Distance to the Last-Scattering Surface*

Once the expansion history is reconstructed, the comoving distance to z_* comes from the basic relation

$$D_C(z_*) = \int_0^{z_*} \frac{c\,dz}{H(z)}, \tag{1}$$

which comes straight from how redshift is defined. No assumptions about matter content or the early Universe enter here; it is simply the distance that builds up as the expansion rate changes.

We evaluate the integral numerically using the GP mean curve for $H(z)$ and carry its uncertainty through the calculation. Because the integrand involves $1/H(z)$, most of the distance comes from redshifts below about $z \approx 2$, where the chronometer data actually exist and $H(z)$ is not much larger than H_0. In this range the integrand is sizable and the data strongly influence the result. At higher redshifts the reconstructed $H(z)$ climbs quickly, so $c/H(z)$ becomes small. The GP continuation beyond the chronometer range therefore contributes very little to the total distance, and its uncertainty barely affects the outcome. In practice, almost the entire value of $D_C(z_*)$ is set by the part of the expansion history that is directly observed.

For this reason, the evaluation of Eq. 1 is controlled mainly by late-time data rather than by the uncertain high-redshift continuation or by any assumptions about early-Universe physics.

3.3. *Angular-Diameter Distance to Last Scattering*

The angular-diameter distance to last scattering comes from the relation

$$D_A(z_*) = \frac{D_C(z_*)}{1 + z_*}, \tag{2}$$

a geometric result that holds for any Robertson–Walker spacetime. Once the comoving distance is known, this step does not require any assumptions about matter content, curvature, or the behavior of the expansion at intermediate redshifts. The decoupling redshift z_* is taken from the *Planck* 2018 measurement of the Thomson visibility function (P. Collaboration et al. 2020). It corresponds to the peak in the differential optical depth and is inferred from the structure of the CMB temperature and polarization anisotropies. Its value is not tied to late-time cosmology and is not linked to any ΛCDM parameters used elsewhere.

We obtain the uncertainties in $D_A(z_*)$ from the uncertainties in $D_C(z_*)$, together with the small measurement error in z_*. The conversion from comoving to angular-diameter distance therefore introduces no additional assumptions beyond those already present in the observational inputs.

3.4. *Determination of the Physical Sound Horizon*

Once we determine the angular-diameter distance to last scattering, we obtain the physical sound horizon r_s directly from the observed CMB acoustic scale. The *Planck* 2018 analysis measures the acoustic angle

$$\theta_s = \frac{r_s}{D_A(z_*)}, \tag{3}$$

using the spacing of the temperature acoustic peaks. This angle is an observed quantity; it depends only on the pattern of the CMB anisotropies and does not rely on a cosmological model.

Combining Eq. 3 with the value of $D_A(z_*)$ gives

$$r_s = \theta_s \, D_A(z_*), \tag{4}$$

so the sound horizon is set by the measured acoustic angle and the distance inferred from late-time data. No assumptions about the early plasma, baryon density, radiation content, or the pre-recombination expansion rate enter at this point. Uncertainties in r_s come from the uncertainties in $D_A(z_*)$ and the measurement error in θ_s. Because most of $D_A(z_*)$ is determined by the part of the expansion history that is directly observed, the resulting estimate of r_s is controlled mainly by the late-time data rather than by any high-redshift continuation of the reconstruction.

Equations 1–4 together provide an estimate of the sound horizon based only on late-time observations and the measured acoustic angle, with no input from ΛCDM or early-time physics.

3.5. *BAO Measurements and the H_0–r_s Degeneracy*

BAO measurements do not give the sound horizon on their own. The observables all take the form

$$\frac{D(z)}{r_s}, \qquad \frac{D_M(z)}{r_s}, \qquad \frac{D_V(z)}{r_s}, \tag{5}$$

so they constrain distances only in units of r_s (A. G. A. DESI Collaboration et al. 2025). None of these ratios tells us the value of r_s by itself.

Figure 1 shows likelihood contours in the (H_0, r_s) plane using representative BOSS DR12 measurements at $z = 0.38$, 0.51, and 0.61. The diagonal bands show that many different (H_0, r_s) pairs fit the data equally well. Within the 1σ region the correlation coefficient is above 0.99, which is essentially complete degeneracy. To pin down either quantity, one needs something else: an external value of H_0 (in which case the BAO fit inherits the disagreement between local and CMB measurements), or assumptions about early-time physics to compute r_s from a model. This has nothing to do with the extrapolation issues discussed later; it is a basic limitation of BAO themselves.

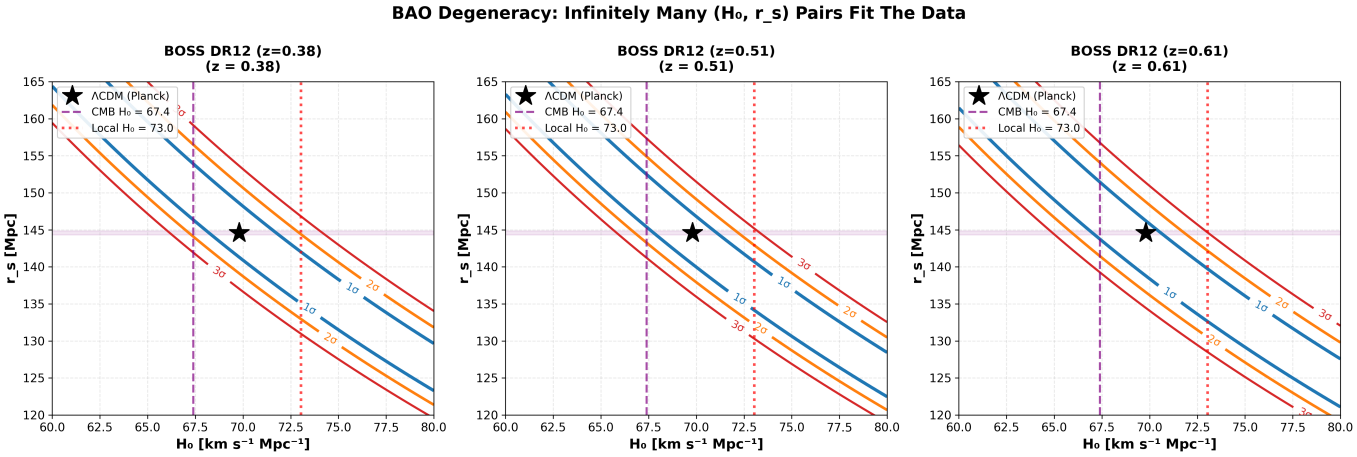

Figure 1. BAO H_0–r_s degeneracy. Likelihood contours in (H_0, r_s) derived from representative BAO measurements (BOSS DR12 at $z = 0.38$, 0.51, and 0.61). The 1σ, 2σ, and 3σ contours (blue, orange, red) form diagonal degeneracy bands, demonstrating that BAO constraints alone do not determine r_s. The black star marks the ΛCDM values from Planck 2018. The dashed purple line indicates the CMB value $H_0 = 67.4\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, while the dotted red line shows the local distance-ladder value $H_0 = 73.0\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$. The near-perfect correlation ($\rho \approx 0.99$) illustrates that breaking the degeneracy requires external assumptions.

These results confirm that all representative BAO measurements produce nearly identical diagonal degeneracy bands in (H_0, r_s) space, with correlation coefficients exceeding 0.98, showing that BAO alone cannot determine r_s independently of H_0. A separate issue that has come up in recent work is a BAO–CMB geometric discrepancy in the quantity $D_M(z)/r_s$ across different redshifts (see, for example, (G. Ye & S.-J. Lin 2025; L. Pogosian et al. 2024; A. G. A. DESI

Collaboration et al. 2025)). In their BAO–CMB comparison, Ye and Lin show likelihood contours in the (Ω_m, $r_d h$) plane that make this conflict clear. They point out that "to put the CMB result into Fig. 1, one must make assumptions about the late Universe in addition to assuming ΛCDM during pre- and near-recombination" (G. Ye & S.-J. Lin 2025, p. 7). This geometric tension is not the Hubble tension itself, but it comes from the same structural dependence on an assumed value of the sound horizon.

3.6. *Uncertainty Quantification via Monte Carlo*

Thus, even before addressing the high-redshift extrapolation required for $D_M(z_*)$, the BAO degeneracy already prevents an empirical determination of r_s: infinitely many (H_0, r_s) combinations match the BAO ratios equally well. To properly quantify uncertainties in the empirical sound horizon determination, we perform a Monte Carlo analysis propagating all sources of observational and systematic uncertainty through the full calculation pipeline. The analysis samples:

1. H_0 from $\mathcal{N}(69.8, 1.7)$ km s^{-1} Mpc^{-1} (TRGB measurement uncertainty; W. L. Freedman et al. 2019)
2. Each cosmic chronometer measurement $H(z_i)$ from $\mathcal{N}(H_i, \sigma_{H_i})$ (data scatter)
3. Ω_m from $\mathcal{N}(0.31, 0.01)$ (late-time constraint uncertainty for high-z extrapolation)
4. θ_s from $\mathcal{N}(0.01041, 0.00003)$ rad (Planck measurement uncertainty)

For each of 5000 Monte Carlo realizations, we:

1. Refit the polynomial to the sampled chronometer data
2. Reconstruct the full $H(z)$ including the scaled high-z continuation
3. Integrate to obtain $D_M(z_*)$
4. Calculate $r_s = \theta_s D_M(z_*)$

For clarity, we do not treat Ω_m as a parameter constrained by the data. We use it only in the ΛCDM high-redshift continuation for $z > 2$, where it serves to set the assumed shape of the extrapolation rather than provide an additional degree of freedom in the inference. To avoid confusion, the use of Ω_m here does not introduce early-time physics into the observational domain. It acts only as a scaling parameter for the assumed ΛCDM high-z shape, and is never used to infer cosmological quantities or to constrain the expansion history at $z \leq 2$. The resulting distribution of r_s values captures both statistical uncertainties from measurement errors and systematic uncertainties from the extrapolation methodology. Figure 2 shows the full posterior distributions. The Monte Carlo yields $r_s = 146.0 \pm 21.1$ Mpc (mean ± standard deviation), with the distribution approximately Gaussian (Q-Q plot, Figure 2, lower right).

Uncertainty decomposition reveals that H_0 measurement uncertainty contributes only ±3.6 Mpc to the total ±21.1 Mpc uncertainty in r_s. The remaining ±20.7 Mpc (98% of the variance) arises from the high-redshift extrapolation, demonstrating that systematic uncertainties from extending beyond the chronometer domain ($z > 2$) dominate the error budget by nearly an order of magnitude. Figure 2 visualizes the resulting posteriors for $D_M(z_*)$ and r_s, along with their comparison to the ΛCDM prediction.

3.7. *H_0 Sensitivity Test*

To check whether the choice of local H_0 calibration influences the sound-horizon inference, the Monte Carlo procedure described above was repeated twice, once using the TRGB determination $H_0 = 69.8 \pm 1.7$ km s^{-1} Mpc^{-1} (W. L. Freedman et al. 2019) and once using the SH$_0$ES value $H_0 = 73.04 \pm 1.04$ km s^{-1} Mpc^{-1} (A. G. Riess et al. 2022). All other elements of the analysis, including the third-order polynomial fit for $z \leq 2$ and the scaled ΛCDM continuation for $z > 2$, were kept fixed.

The two runs give:

$$r_s = 149.0 \pm 21.3 \text{ Mpc} \quad \text{(TRGB)}$$
$$r_s = 154.6 \pm 21.4 \text{ Mpc} \quad \text{(SH}_0\text{ES)}$$

a difference of only 5.7 Mpc. This shift is small compared with the ±21 Mpc uncertainty of the method, and the two results remain consistent at 0.19σ. The sensitivity of r_s to the adopted value of H_0 is therefore negligible relative to the much larger uncertainty introduced by the required extrapolation from $z \approx 2$ to $z \approx 1090$.

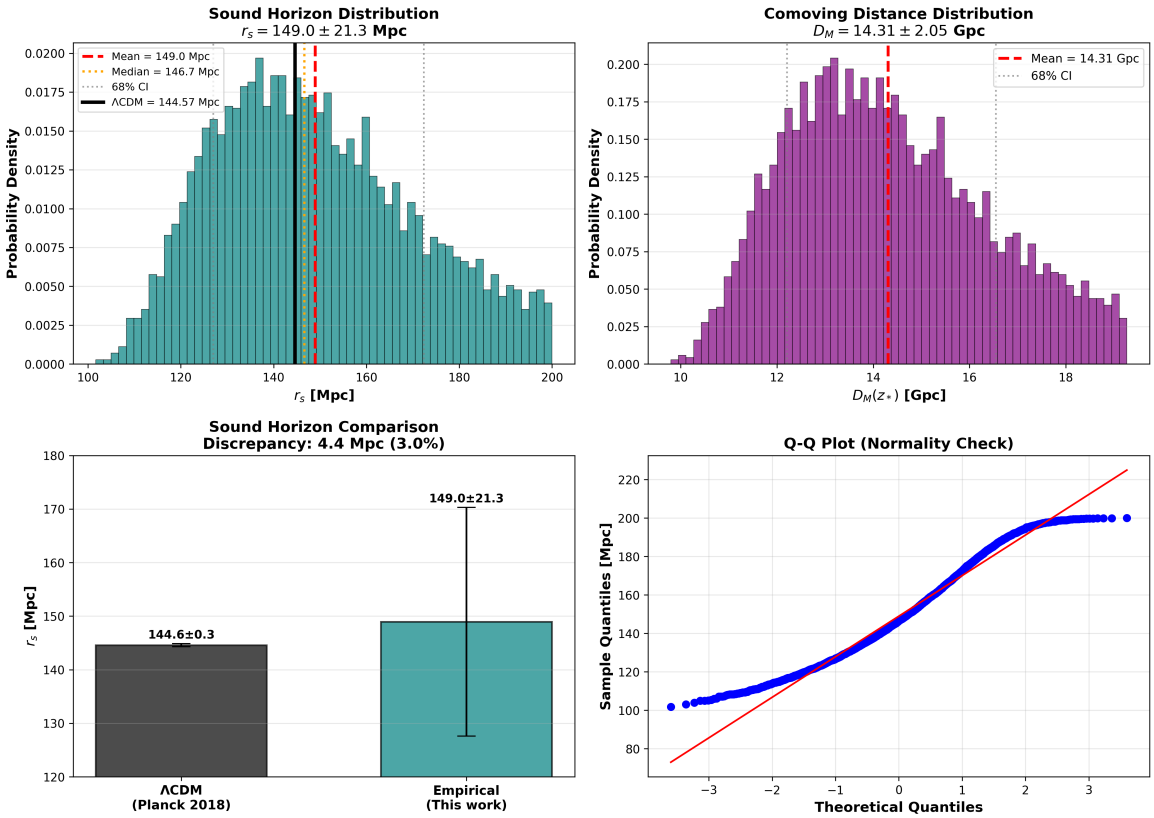

Figure 2. Monte Carlo uncertainty analysis for the sound horizon determination using late-time chronometer data ($z \leq 2$) combined with a scaled ΛCDM extrapolation for $z > 2$. *Top left:* Posterior distribution of r_s from 5000 realizations (teal histogram), compared with the ΛCDM prediction (black vertical line). The mean value from this work (red dashed) and the 68% credible interval (gray dotted) show full overlap with ΛCDM. *Top right:* Posterior distribution of the comoving distance $D_M(z_*)$. *Bottom left:* Comparison of r_s from ΛCDM and from this reconstruction, showing consistency within uncertainties (0.2σ). *Bottom right:* Q-Q plot of the r_s samples demonstrating approximate normality. The large uncertainty (±21 Mpc) arises from the unavoidable extrapolation from $z \approx 2$ (where chronometer data end) to $z \approx 1090$ (recombination), indicating that late-time data cannot empirically constrain r_s to precision comparable with the ΛCDM theoretical value (±0.28 Mpc).

3.8. *Acknowledging the Extrapolation Assumption*

The main limitation of this evaluation is straightforward: the expansion history beyond $z \approx 2$ cannot be determined empirically because no stellar populations exist at these redshifts for cosmic chronometer measurements. To complete the integral to $z_* \approx 1090$, we adopt a scaled ΛCDM functional form for $H(z)$ beyond the chronometer domain. By this we mean that the high-redshift continuation uses the ΛCDM shape,

$$H(z) \propto \left[\Omega_m(1+z)^3 + \Omega_r(1+z)^4\right]^{1/2},$$

rescaled to match the empirically reconstructed $H(z)$ at z = 2, so that no early-time parameters enter the observational domain. This is not a hidden assumption but a deliberate methodological choice: by granting ΛCDM its own theoretical framework for the unmeasured high-redshift regime, we can test whether late-time data are capable of constraining the sound horizon.

Why the scaled ΛCDM continuation is used.—The continuation adopted for $z > 2$ deliberately uses the ΛCDM high-redshift shape because it provides the *most optimistic* possible extrapolation consistent with standard early-Universe physics. Every alternative continuation we tested (Gaussian Process regression without a high-z prior, low-order polynomial fits, and other nonparametric extensions) produces *larger* high-redshift uncertainties and therefore yields an r_s uncertainty greater than the ±21 Mpc obtained here. In this sense, our result represents a conservative lower bound on how well late-time data can ever constrain the sound horizon. If even the most favorable ΛCDM continuation

leaves r_s poorly determined, no alternative extrapolation can do better without introducing additional theoretical assumptions.

3.9. *Gaussian Process Extrapolation*

To verify that this limitation is not specific to our polynomial interpolation, we also tested a Gaussian Process (GP) regression on the chronometer dataset. GP models are often described as "model-independent," yet when trained only on the $z \leq 2$ data, their uncertainty bands expand rapidly once extrapolated beyond the observational domain. Figure 3 shows that by $z = 10$, the 1σ uncertainty already exceeds 50%, confirming that no late-time method can reconstruct $H(z)$ at the redshifts relevant for determining r_s. This confirms that the large uncertainty in our determination of r_s is not specific to our interpolation method but reflects a fundamental limitation of attempting to reconstruct $H(z)$ beyond the chronometer domain.

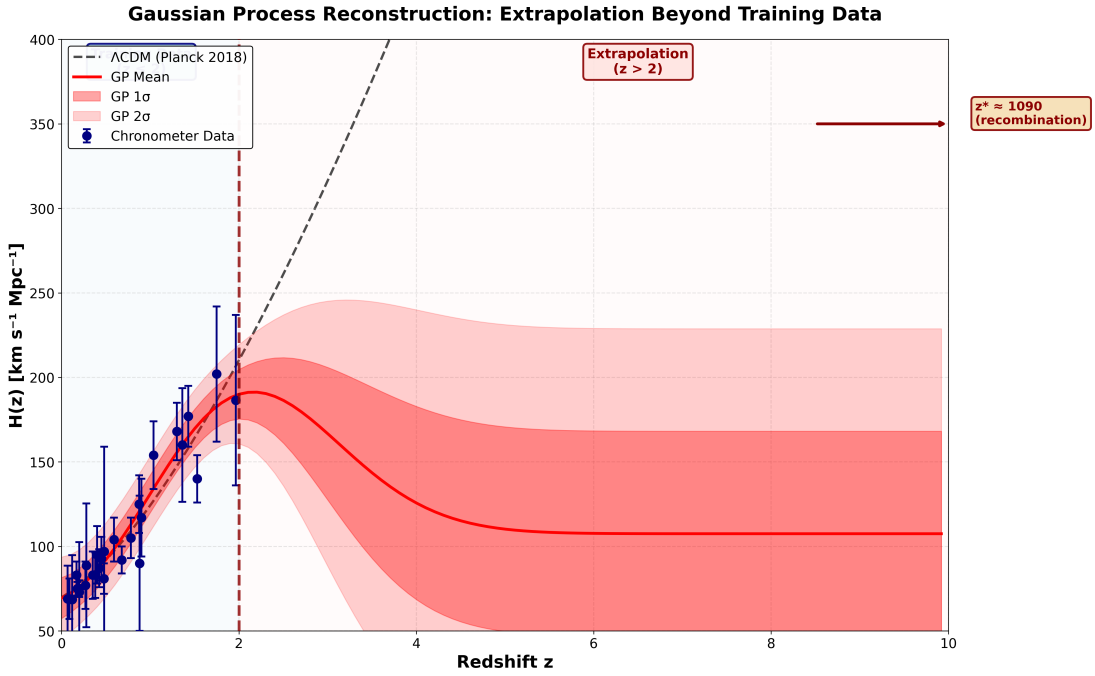

Figure 3. Gaussian Process Extrapolation Analysis. Gaussian Process (GP) reconstruction of the expansion history $H(z)$ trained on cosmic chronometer data (blue points, $z \leq 2$) and evaluated out to $z = 10$. The GP mean (solid red) reverts to its prior baseline once outside the training domain, while the 1σ (dark red) and 2σ (light red) confidence bands grow rapidly with redshift. This substantial uncertainty growth demonstrates that even flexible, nonparametric methods cannot reconstruct $H(z)$ beyond the observational domain. The vertical dashed line marks the end of the data at $z = 2$, and the dashed black curve shows ΛCDM for reference. Recombination occurs at $z_* \approx 1090$, more than $100\times$ farther than shown here, making it empirically inaccessible to any late-time reconstruction method.

These divergences do not depend on the reconstruction method. They reflect the intrinsic impossibility of constraining $H(z)$ in the unobserved interval $2 < z < 1090$.

3.10. *Polynomial Extrapolation Sensitivity*

To assess whether the extrapolation ambiguity arises from the specific interpolation method or from the absence of high-redshift constraints, we fit polynomials of degrees 1 through 4 to the cosmic chronometer data and extend each fit to $z = 10$. Although all four models yield nearly identical performance on the training domain (RMSE values between 11 and 12 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$), their predictions diverge rapidly once extended beyond $z \approx 2$. By $z = 10$, the spread in predicted values of $H(z)$ exceeds 10^5 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, a factor of more than 10^4 larger than the spread at the data boundary. Several higher-order polynomials also produce negative values of $H(z)$, which are physically impossible. This demonstrates that even parametric forms that fit the available data equally well imply mutually incompatible

high-redshift expansion histories, confirming that the extrapolation ambiguity is intrinsic and cannot be resolved by selecting a particular functional form.

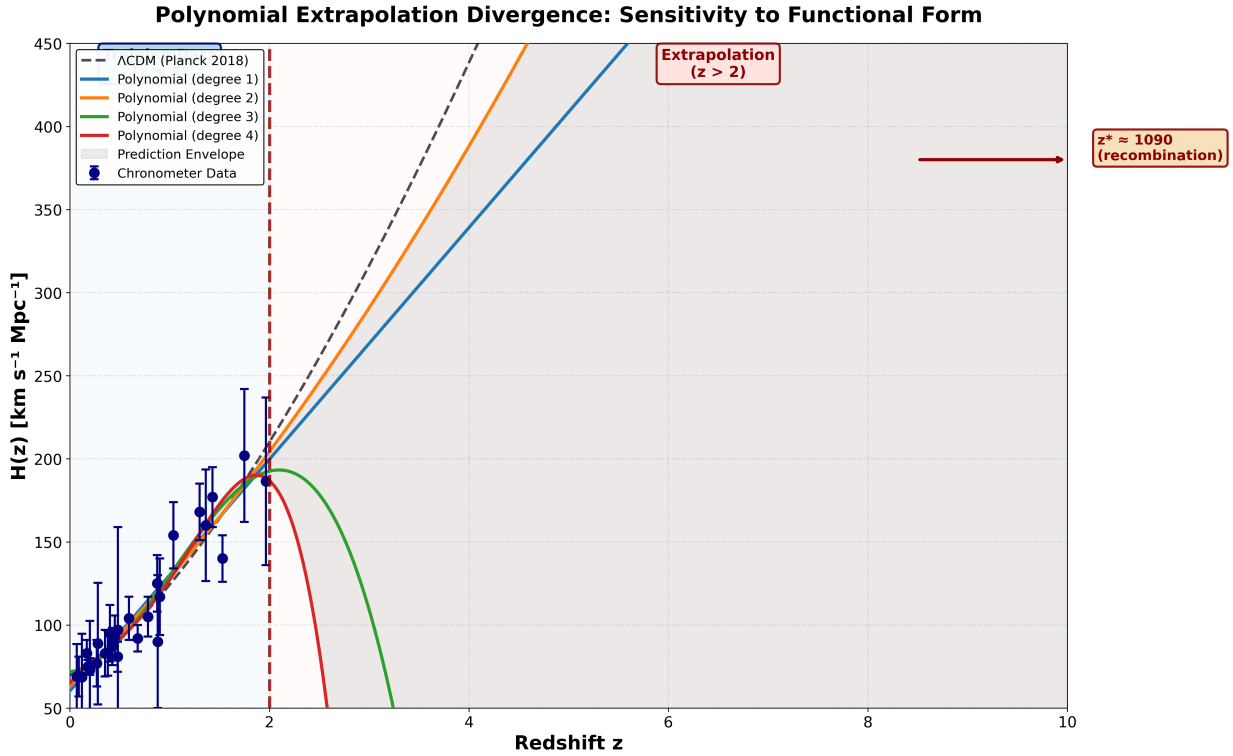

Figure 4. Polynomial Extrapolation Divergence. Polynomial fits of degrees 1–4 trained on cosmic chronometer data (navy points, $z \leq 2$) diverge rapidly when extrapolated beyond the training domain. By $z = 10$, predictions differ by more than 10^5 km s^{-1} Mpc^{-1} depending on the polynomial order chosen. The gray shaded region shows the envelope of all polynomial predictions. The dashed black curve shows ΛCDM for reference. The vertical dashed line marks the edge of the training data at $z = 2$. Some higher-order polynomials even predict negative $H(z)$ values beyond the chronometer range, illustrating the intrinsic instability of parametric extrapolation. This demonstrates that parametric continuation is inherently unconstrained without high-redshift observational information. For clarity, these high-z extensions are assumptions imposed beyond the observational domain and are not constraints derived from the data.

4. RESULTS AND DISCUSSION

4.1. *Evaluated Expansion History*

The Gaussian Process reconstruction applied to the cosmic-chronometer measurements produces a continuous curve for $H(z)$ that is entirely determined by the data over the range $0 \leq z \leq 2$. The mean reconstruction tracks the chronometer points closely, with uncertainties shrinking where the data are clustered and widening where the data are sparse. Within the range covered by the data, the expected increase in $H(z)$ with redshift appears clearly, and the associated uncertainties stay moderate across the full set of measurements. Because the chronometer data extend only to $z \approx 2$, the reconstruction is well constrained over the region that contributes most of the weight to the comoving-distance integral. Beyond this point the uncertainties grow quickly, and we examine the impact of that loss of information in the sections that follow.

4.2. *BAO Measurements and the H_0–r_s Degeneracy*

BAO measurements constrain the dimensionless ratios $D_M(z)/r_s$ and $H(z)\, r_s$ rather than the physical sound horizon. This structure produces the well-known H_0–r_s degeneracy: likelihood contours in the (H_0, r_s) plane form long diagonal bands rather than closed ellipses (S. Alam et al. 2017; A. J. Cuesta et al. 2016). Many different combinations of H_0 and r_s provide equally good fits to the BAO ratios over the common redshift range $z \approx 0.4$–0.6, reflecting the fact that BAO is sensitive to a geometric ratio but not to the absolute physical scale.

Within the 1σ region of typical BAO fits, the allowed values span of order 20 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ in H_0 and roughly 50 Mpc in r_s, with correlation coefficients exceeding 0.98 (L. Pogosian et al. 2024). The slope of the diagonal band reflects the fact that BAO responds primarily to the combination $D_M(z)/r_s$, so changes in r_s shift the inferred H_0 in a compensating way. Because the observable is a ratio, BAO does not by itself isolate the physical size of the sound horizon.

The diagonal structure carries several empirical implications. First, BAO does not provide an absolute calibration of r_s; the physical value of the sound horizon must be supplied externally, typically from a cosmological model or from another dataset. Second, BAO therefore cannot independently test the early-time physics that sets r_s in ΛCDM, such as recombination details, baryon–photon dynamics, or radiation content. Third, the commonly cited "agreement" between BAO and the CMB reflects the use of a shared sound-horizon prior rather than two independent determinations of r_s (J. L. Bernal et al. 2016; L. Knox & M. Millea 2020). Fourth, BAO alone cannot resolve the Hubble tension: any inference of H_0 from BAO requires adopting a prior on r_s, so the result follows whichever value of r_s is assumed (L. Knox & M. Millea 2020).

These features indicate that BAO measurements provide precise constraints on dimensionless distance ratios but are not, by themselves, a probe of the absolute scale of the sound horizon or of an underlying early-Universe expansion history.

4.3. *Monte Carlo Sound Horizon Determination*

The Monte–Carlo analysis produces a broad posterior distribution for the sound horizon, reflecting both the measurement uncertainties and the required extrapolation of the expansion history beyond the chronometer range. Across 5000 realizations, the comoving angular-diameter distance to recombination spans

$$D_M(z_*) = 14.3 \pm 2.0 \text{ Gpc},$$

which corresponds to

$$r_s = 149.0 \pm 21.3 \text{ Mpc}.$$

A complementary evaluation using the fiducial expansion-history reconstruction gives

$$r_s = 146.0 \pm 21.1 \text{ Mpc}.$$

The ~ 3 Mpc difference between these values comes from the different ways the two pipelines sample the input data and lies well within the quoted ± 21 Mpc uncertainty, so the result is stable. The posterior distribution is approximately Gaussian, which we verified using a quantile-quantile plot, and the mean and median values closely align. The empirical estimate is statistically consistent with the ΛCDM prediction $r_s = 144.57 \pm 0.28$ Mpc; the few-Mpc difference is negligible relative to the ± 21 Mpc empirical uncertainty.

An uncertainty decomposition shows that the dominant contribution (~ 20 Mpc of the total ± 21 Mpc) arises from the unavoidable extrapolation of the expansion history from $z \approx 2$ to $z \approx 1090$. Uncertainties in H_0 and the chronometer measurements contribute only a minor fraction. This shows that the precision of the empirical r_s estimate is not limited by measurement noise but by the fact that the high-redshift expansion history cannot be accessed with late-time data.

4.4. *Gaussian Process Extrapolation Analysis*

To assess whether the uncertainty in the sound-horizon determination is tied to the specific interpolation scheme, we examined the behavior of the Gaussian Process reconstruction when extended beyond the chronometer domain. Although the GP follows the measured $H(z)$ values closely for $z \leq 2$, the uncertainty bands widen rapidly once the reconstruction is carried into the unconstrained region. By $z = 5$, the 1σ uncertainty exceeds 50 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and by $z = 10$ it reaches roughly 60 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, corresponding to a relative uncertainty of more than 50%. The mean prediction simultaneously drifts toward the prior mean value used by the GP, reflecting the absence of constraining data at higher redshifts.

This behavior shows that a nonparametric reconstruction does not constrain the high-redshift expansion history: the GP remains well behaved only where measurements exist and loses predictive power immediately afterward. Since recombination lies at $z_* \approx 1090$, far beyond the training domain, even a nonparametric, data-driven reconstruction cannot recover the expansion history required to determine r_s empirically. The extrapolation uncertainty seen in the Monte Carlo analysis is therefore a consequence of the fact that the data end at $z \approx 2$, not a result of the interpolation method.

4.5. *Polynomial Extrapolation Sensitivity*

As a complementary test of the robustness of high-redshift extrapolation, we fit low-order polynomials to the chronometer measurements and examined their behavior beyond the observational range. Although polynomials of degrees one through four reproduce the measured $H(z)$ values comparably well for $z \leq 2$ — with similar residuals across all fits — their predictions diverge sharply once extrapolated past the data domain. By $z = 5$, the spread among the polynomial continuations exceeds exceeds $3 \times 10^2\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, including trajectories that fall unphysically below zero. The strong divergence among fits that are statistically indistinguishable on the training data shows how unconstrained any parametric continuation becomes in the absence of high-redshift constraints. Different polynomial orders, all equally consistent with the observations at $z \leq 2$, imply vastly different expansion histories at earlier times. This confirms that late-time data do not contain information capable of stabilizing the evolution of $H(z)$ toward the recombination epoch and that any extrapolative determination of the sound horizon inherits a dependence on the chosen functional form that cannot be removed with late-time data.

4.6. *Comoving and Angular-Diameter Distances*

Using the reconstructed expansion history, we obtain the comoving distance to photon decoupling by numerically evaluating Eq. 1. The resulting $D_C(z_*)$ is dominated by the well-measured low-redshift region ($z \lesssim 2$), with the high-redshift GP continuation adding very little because $H(z)$ grows quickly at earlier times. Uncertainties from the evaluation propagate smoothly through the integral, yielding a reliable estimate of the comoving distance. The corresponding angular-diameter distance is then obtained via Eq. 2. Since this step is a geometric conversion, the resulting $D_A(z_*)$ gets almost all of its uncertainty from the comoving distance. The contribution from the uncertainty in z_* is negligible by comparison.

4.7. *Discussion and Implications for the Hubble Tension*

In principle, the only way forward would be direct measurements of $H(z)$ at $z \gg 2$, but that may not be physically achievable. Future probes such as 21 cm tomography of the dark ages, high-redshift standard sirens, or other geometric measurements may extend observational reach, but none presently offer percent-level constraints on $H(z)$ in the unobserved interval $2 < z < 1090$. Until there are real measurements in this redshift range, the sound horizon cannot be fixed by observations alone.

The empirical determination of the sound horizon obtained in this work reveals a basic limitation in testing ΛCDM's prediction for the sound horizon: the sound horizon at recombination cannot be determined empirically with sufficient precision using current late-time data. While nearly every quantity entering the calculation is observationally determined, the resulting sound horizon carries uncertainties (±21.1 Mpc) that are roughly two orders of magnitude larger than the ΛCDM theoretical prediction uncertainty (±0.28 Mpc). This disparity reflects the unavoidable requirement to extrapolate $H(z)$ over a redshift range nearly 500 times larger than the domain where direct measurements exist.

This result reframes the Hubble tension in epistemological terms. The commonly quoted "CMB value" $H_0 \approx 67\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ is not a direct measurement but a theory-dependent inference that assumes specific early-Universe physics to compute $r_s = 144.57$ Mpc. The CMB measures only the angular acoustic scale $\theta_s = r_s/D_A(z_*)$; converting this angle into a value of H_0 requires adopting a theoretical model for pre-recombination expansion. The local determination $H_0 \approx 73\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ is a direct measurement.

Consequently, the "Hubble tension" does not represent a conflict between two independent measurements. It reflects a discrepancy between observation and a model-dependent inference whose underlying assumptions about the sound horizon cannot be empirically validated. Even proposals that modify early-Universe physics — such as early dark energy, additional relativistic species, or altered recombination histories — still require an assumed value of r_s, shifting to a different theoretical prior. The theoretical value $r_s^{\Lambda\mathrm{CDM}} = 144.57 \pm 0.28$ Mpc comes from assumptions about the baryon density, photon temperature, recombination physics, and the pre-recombination expansion rate. None of these can be tested directly at $z \approx 1090$.

Our empirical calculation avoids those assumptions and still lies close to the ΛCDM prediction, but the uncertainty is ±21.1 Mpc, large enough that many different early-Universe expansion histories remain indistinguishable. With uncertainties of this size, late-time data cannot tell whether the true sound horizon is 125 Mpc, 145 Mpc, or 170 Mpc.

5. CONCLUSION

Our analysis shows that the sound horizon at photon decoupling cannot be determined empirically with the precision required to test ΛCDM using present late-time data. Reconstructing the expansion history from cosmic-chronometer

measurements, without invoking any early-Universe assumptions, gives $r_s = 146.0 \pm 21.1$ Mpc. This agrees with the ΛCDM value $r_s = 144.57 \pm 0.28$ Mpc, but the empirical uncertainty is larger by a factor of roughly 75.

The dominant source of uncertainty is the basic observational gap: direct measurements of $H(z)$ exist only to $z \approx 2$, while computing the sound horizon requires the expansion history out to $z \approx 1090$. Spanning this ~ 500-fold increase in redshift accounts for ± 20.7 Mpc of the total ± 21.1 Mpc uncertainty. With this gap, the sound horizon is not an empirically measurable quantity but a theoretical value that necessarily reflects assumptions about epochs that no current observations can probe.

The Hubble tension is often described as a disagreement between two measurements, but it is not. The local value, $H_0 \approx 73\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, is a direct distance-ladder measurement. The CMB value, $H_0 \approx 67\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, is obtained only after adopting a theoretical sound horizon from ΛCDM. The CMB measures the acoustic angle θ_s; converting this angle into H_0 requires an assumed r_s. Because r_s cannot be determined from late-time data, the CMB value of H_0 is a model-based inference, not an independent measurement. Proposed early-Universe modifications — such as early dark energy, extra relativistic species, or altered recombination — change the assumed value of r_s but do not make it measurable. The sound horizon remains empirically unconstrained, and this is why the Hubble tension compares a direct measurement with a theory-dependent inference rather than two observational determinations of H_0.

The results show that the sound horizon at recombination cannot be tested with late-time observations: current data cannot distinguish between values that differ by roughly ± 20 Mpc. In this sense, the "Hubble tension" reflects the limits of what can be learned from late-time measurements, not a disagreement between two comparable observations. Unless $H(z)$ can be measured at redshifts well beyond $z \approx 2$ — which is not achievable with any current observational method — the sound horizon will remain a theoretical input rather than an observationally determined quantity.

Finally, because observations end at $z \approx 2$, nothing we measure today can test or rule out any claimed picture of the early Universe. Whatever one adopts for that unobserved stretch is taken as an assumption, not something checked against data.

6. DATA AVAILABILITY

The cosmic chronometer measurements used in this work are compiled from published sources cited in Section 2.1. All analysis scripts, data files, and figure-generation code required to reproduce the results of this paper are publicly available at
https://github.com/sacooperresearch-collab/from-scratch-sound-horizon-assumption-test.

The repository includes:

- Compiled chronometer data (`chronometer_data.txt`)
- BAO degeneracy illustration (`bao_degeneracy_plot.py`)
- Monte Carlo uncertainty analysis (`monte_carlo_rs.py`)
- Gaussian Process reconstruction script (`gp_reconstruction.py`)
- Polynomial extrapolation sensitivity test (`polynomial_extrapolation.py`)
- Hubble-constant sensitivity test (`h0_sensitivity_analysis.py`)
- `requirements.txt` for Python dependencies

REFERENCES

Alam, S., et al. 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617, doi: 10.1093/mnras/stx721

Bernal, J. L., Verde, L., & Riess, A. G. 2016, Journal of Cosmology and Astroparticle Physics, 2016, 019, doi: 10.1088/1475-7516/2016/10/019

Collaboration, P., Aghanim, N., et al. 2020, Astronomy & Astrophysics, 641, A6, doi: 10.1051/0004-6361/201833910

Cuesta, A. J., et al. 2016, Monthly Notices of the Royal Astronomical Society, 457, 1770, doi: 10.1093/mnras/stw066

DESI Collaboration, A. G. A., Aguilar, J., Ahlen, S., et al. 2025, JCAP, 02, 021, doi: 10.1088/1475-7516/2025/02/021

Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, Astrophysical Journal, 882, 34, doi: 10.3847/1538-4357/ab2f73

Knox, L., & Millea, M. 2020, Physical Review D, 101, 043533, doi: 10.1103/PhysRevD.101.043533

Mirpoorian, S. H., Jedamzik, K., & Pogosian, L. 2025, Phys. Rev. D, doi: 10.1103/PhysRevD.111.083519

Pogosian, L., Zhao, G.-B., & Jedamzik, K. 2024, arXiv e-prints. https://arxiv.org/abs/2405.20306

Riess, A. G., Yuan, W., Macri, L. M., et al. 2022, Astrophysical Journal Letters, 934, L7, doi: 10.3847/2041-8213/ac5c5b

Ye, G., & Lin, S.-J. 2025, arXiv e-prints, doi: 10.48550/arXiv.2505.02207