



Forecasting Chemical Abundance Precision for Extragalactic Stellar Archaeology

Nathan R. Sandford¹ , Daniel R. Weisz¹ , and Yuan-Sen Ting^{2,3,4,5} ¹ Department of Astronomy, University of California Berkeley, Berkeley, CA 94720, USA; nathan_sandford@berkeley.edu² Institute for Advanced Study, Princeton, NJ 08540, USA³ Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA⁴ Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA⁵ Research School of Astronomy and Astrophysics, Australia National University, Cotter Road, ACT 2611, Canberra, Australia

Received 2020 January 29; revised 2020 May 21; accepted 2020 June 11; published 2020 July 29

Abstract

Increasingly powerful and multiplexed spectroscopic facilities promise detailed chemical abundance patterns for millions of resolved stars in galaxies beyond the Milky Way (MW). Here, we employ the Cramér–Rao lower bound (CRLB) to forecast the precision to which stellar abundances for metal-poor, low-mass stars outside the MW can be measured for 41 current (e.g., Keck, MMT, the Very Large Telescope, and the Dark Energy Spectroscopic Instrument) and planned (e.g., the Maunakea Spectroscopic Explorer, the James Webb Space Telescope (JWST), and Extremely Large Telescopes (ELTs)) spectrograph configurations. We show that moderate-resolution ($R \lesssim 5000$) spectroscopy at blue-optical wavelengths ($\lambda \lesssim 4500 \text{ \AA}$) (i) enables the recovery of two to four times as many elements as red-optical spectroscopy ($5000 \lesssim \lambda \lesssim 10000 \text{ \AA}$) at similar or higher resolutions ($R \sim 10,000$) and (ii) can constrain the abundances of several neutron-capture elements to $\lesssim 0.3$ dex. We further show that high-resolution ($R \gtrsim 20,000$), low signal-to-noise ratio ($\sim 10 \text{ pixel}^{-1}$) spectra contain rich abundance information when modeled with full spectral fitting techniques. We demonstrate that JWST/NIRSpec and ELTs can recover (i) ~ 10 and 30 elements, respectively, for metal-poor red giants throughout the Local Group and (ii) $[\text{Fe}/\text{H}]$ and $[\alpha/\text{Fe}]$ for resolved stars in galaxies out to several Mpc with modest integration times. We show that select literature abundances are within a factor of ~ 2 (or better) of our CRLBs. We suggest that, like exposure time calculators, CRLBs should be used when planning stellar spectroscopic observations. We include an open-source Python package, `Chem-I-CalC`, that allows users to compute CRLBs for spectrographs of their choosing.

Unified Astronomy Thesaurus concepts: Local Group (929); Red giant stars (1372); Astronomy data analysis (1858); Stellar abundances (1577); Fundamental parameters of stars (555); Spectroscopy (1558)

1. Introduction

Absorption features imprinted in the spectrum of a star encode its physical structure and chemical composition. In turn, the chemical composition of individual stars trace the chemistry of the interstellar medium at their birth,⁶ providing a detailed fossil record of a galaxy’s chemical evolution over cosmic time. Various enrichment processes (e.g., core-collapse and thermonuclear supernovae, stellar winds, neutron star mergers, and gas inflows) each leave a unique chemical signature on their environment, which are captured in the abundance patterns of stars observed today (Tinsley 1980). Accordingly, the spectra of resolved stars provide a wealth of information on everything from the formation histories of galaxies to detailed nuclear and quantum physics.

However, translating stellar spectra to stellar composition is a nontrivial undertaking that relies on ~ 200 years of advancement in atomic and stellar physics, astronomical instrumentation, and computational methods. The field of stellar spectroscopy and chemical abundance measurements has had a rich history since the first recorded solar spectrum by Fraunhofer (1817) and the subsequent identification of specific elemental absorption features nearly 50 years later (e.g., Kirchhoff & Bunsen 1860; Kirchhoff 1860, 1863; Huggins & Miller 1864). As chronicled in Hearnshaw (2010), it was another ~ 70 years until the first quantitative abundance measurements were made. Such measurements were only possible after breakthroughs in theoretical physics (e.g., atomic/ionization theory and stellar atmospheres), development of new instrumentation (e.g., blazed gratings, coude spectrographs, and Schmidt cameras), and substantial investment

in laboratory experiments (e.g., transition wavelengths, oscillator strengths, and opacities). Together, these advances enabled the pioneering abundance work of Payne (1925), Russell (1929), Unsöld (1938, 1942), Strömgren (1940), Aller (1942, 1946), Greenstein (1948), and Wright (1948) upon which modern stellar spectroscopy is founded.

Since the first half of the 20th century, high-resolution ($R > 10,000$) spectroscopy with broad optical wavelength coverage and high signal-to-noise ratio ($S/N; > 30 \text{ pixel}^{-1}$) has been the gold standard for measuring precise stellar atmospheric parameters and detailed chemical abundance patterns (Nissen & Gustafsson 2018). These spectra provide clean, unblended absorption features that can typically be fit with equivalent widths (EWs).⁷ At the same time, such high-resolution studies are often limited to small numbers of bright stars due to high-dispersion, low-throughput, and poor multiplexing capabilities.

In comparison, low- and medium-resolution spectrographs provide the opportunity to observe more and fainter stars, but are burdened with the cost of having (sometimes heavily) blended features that prohibit the use of conventional EW techniques.

As a means around this challenge, a number of studies have employed spectral indices for low-resolution chemical abundance measurements. One especially common index is centered around the Ca II triplet at $\sim 9000 \text{ \AA}$ (e.g., Cenarro et al. 2001a, 2001b, 2002, and references therein). In this method, the strength of a blended spectral feature (e.g., the Ca II triplet) is calibrated to abundance measurements from high-resolution studies (e.g., Olszewski et al. 1991; Rutledge et al. 1997; Carrera et al. 2013)

⁶ Modulo mixing and gravitational settling.

⁷ See Minnaert (1934) for an early discussion of EWs.

or to theoretical (i.e., *ab initio*) spectra generated from stellar atmosphere and spectrum synthesis models (Baschek 1959; Fischel 1964; Bell 1970; Bell & Branch 1976).⁸ However, spectral indices provide only bulk metal abundances (requiring assumptions of chemical abundance patterns) and are restricted to the parameter space of their calibrating stars or models (Battaglia et al. 2008; Koch et al. 2008a; Starkenburg et al. 2010).

As computational resources and stellar models continued to improve, it became possible to directly compare theoretical (*ab initio*) spectra to observed spectra on a pixel-by-pixel basis (pioneering examples include Gingerich 1969; Sneden 1973, 1974; Suntzeff 1981; Carbon et al. 1982; Leep et al. 1986, 1987; Wallerstein et al. 1987). This technique leverages the full statistical power of the many absorption lines in a spectrum, yielding precise abundance measurements without the use of EWs or spectral indices. These methods have proven powerful for the recovery of detailed abundance patterns from low- and medium-resolution spectra, which contain predominantly weak and blended absorption features.

In the last two decades, massively multiplexed stellar spectroscopic surveys (e.g., the RAdial Velocity Experiment (RAVE), Steinmetz et al. 2006; the Sloan Extension for Galactic Understanding and Exploration (SEGUE), Yanny et al. 2009; the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) survey, Luo et al. 2015; the Galactic Archaeology with HERMES (GALAH) survey, de Silva et al. 2015; the Apache Point Observatory Galactic Evolution Experiment (APOGEE), Majewski et al. 2017; and Dark Energy Spectroscopic Instrument (DESI) survey, DESI Collaboration et al. 2016a) have collected millions of spectra of Milky Way (MW) stars. Coupled with steady progress in theoretical and laboratory astrophysics, these surveys have revolutionized our ability to collect and interpret the spectra of stars (see reviews by Allende Prieto 2016; Nissen & Gustafsson 2018; Jofré et al. 2019). Importantly, they have motivated the development of novel fitting techniques designed to efficiently fit the full spectrum of many stars. Some techniques are data driven (e.g., The Cannon; Ness et al. 2015), some are trained on *ab initio* spectra (e.g., The Payne; Ting et al. 2019), and others adopt hybrid methods (e.g., The DD-Payne; Xiang et al. 2019). All employ sophisticated statistical techniques (e.g., neural networks, Bayesian inference, and/or machine learning), enabling the precise recovery of dozens of elemental abundances from both low- and high-resolution spectra in modest compute times.

However, extragalactic stellar spectroscopy has yet to experience the same tremendous gains in quantity and quality of abundance measurements as seen for spectroscopy of stars in the MW. This is primarily the result of stars in external galaxies being much fainter and thus more challenging to observe. Generally, only the few brightest stars ($m_V \lesssim 19.5$) in extragalactic systems can be observed at high resolution, even when using 10 meter class telescopes (e.g., Shetrone et al. 1998, 2001, 2003; Tolstoy et al. 2003; Fulbright et al. 2004; Venn et al. 2004; Walker et al. 2007, 2009a, 2009b, 2015a, 2015b; Koch et al. 2008a, 2008b; Aoki et al. 2009; Cohen & Huang 2009; Frebel et al. 2010, 2014, 2016; Starkenburg et al. 2013; Koch & Rich 2014; Ji et al. 2016a, 2016b, 2016c; Spencer et al. 2017; Venn et al. 2017; Spite et al. 2018; Hill et al. 2019; Theler et al. 2019).⁹

Instead, highly multiplexed low- and moderate-resolution ($R < 10,000$) spectrographs on large-aperture telescopes have become the workhorse instruments of extragalactic stellar spectroscopy (e.g., the DEep Imaging Multi-Object Spectrograph (DEIMOS); Faber et al. 2003). Over the past 20 years, tens of thousands of low- and medium-resolution spectra have been acquired for extragalactic stars. Because detailed abundance measurements were typically viewed as the purview of high-resolution spectroscopy, most of the spectra were taken for the purpose of measuring radial velocities and bulk metallicities with spectral indices (e.g., Suntzeff et al. 1993; Pont et al. 2004; Tolstoy et al. 2004; Battaglia et al. 2006; Muñoz et al. 2006; Koch et al. 2007a, 2007b, 2009; Simon & Geha 2007; Battaglia et al. 2008, 2011; Norris et al. 2008; Leaman et al. 2009; Shetrone et al. 2009; Kalirai et al. 2010; Hendricks et al. 2014; Ho et al. 2015; Simon et al. 2015, 2017; Slater et al. 2015; Martin et al. 2016a, 2016b; Swan et al. 2016; Li et al. 2017; Longeard et al. 2020).

The groundbreaking work of Kirby et al. (2009) was the first to demonstrate that precise abundances could be recovered from moderate-resolution spectra in external galaxies. Since then, the method has been further refined and applied to thousands of stars in Local Group (LG) galaxies, measuring up to ~ 10 abundances in MW satellites and ~ 5 abundances at the distance of M31 (e.g., Kirby et al. 2010, 2015a, 2015b, 2015c, 2017a, 2017b, 2018, 2020; Duggan et al. 2018; Vargas et al. 2013, 2014a, 2014b; Escala et al. 2019a, 2019b; Gilbert et al. 2019).

Currently, the field of extragalactic stellar spectroscopy (and with it, the field of extragalactic chemical evolution) is poised for enormous growth. Current and future spectroscopic facilities on large-aperture telescopes promise to increase the number of stars outside the MW with observed spectra by at least an order of magnitude. Already, existing spectrographs on 6+ meter telescopes have been used to measure abundances of over $\sim 10^4$ stars in LG dwarf galaxies and the halo of M31 (see Suda et al. 2017 and references therein) and are capable of measuring thousands more.

In the next decade, dedicated spectroscopic surveys on large telescopes (e.g., with the Prime Focus Spectrograph (PFS), Takada et al. 2014; the Maunakea Spectroscopic Explorer (MSE), MSE Science Team et al. 2019; and the Fiber-Optic Broadband Optical Spectrograph (FOBOS), Bundy et al. 2019) will homogeneously collect hundreds of thousands of resolved star spectra in external galaxies. The next decade will also bring the James Webb Space Telescope (JWST) and Extremely Large Telescopes (ELTs; e.g., the Giant Magellan Telescope (GMT), the European ELT (E-ELT), and the Thirty Meter Telescope (TMT)), which will make possible the spectroscopy of stars in the most distant, faint, and crowded environments in the LG and beyond that are inaccessible to current ground-based facilities.

To fully realize the scientific potential of upcoming massive data sets and to plan for observational campaigns further in the future, it is imperative that we can quantify what we expect to be able to measure from these spectra, and to what precision. While there exist preferred spectral wavelength regions, absorption features, and minimum S/N for abundance measurements, best practices are frequently informally passed down in the community. Comprehensive and quantitative analyses of the chemical information content of spectra given their wavelength coverage, resolution, and S/N are important planning tools, but are sparse in the literature (e.g., Caffau et al. 2013; Bedell et al. 2014; Hansen et al. 2015; Ruchti et al. 2016; Ting et al. 2017a; Feeney et al. 2019).

⁸ Similar to how EWs are calibrated for high-resolution studies.

⁹ To date, $\sim 10^4$ stars outside the MW have measured [Fe/H] from $R > 10,000$ spectroscopy, though most have only been observed over a small (~ 100 Å) range in wavelength.

In this paper, we employ *ab initio* stellar spectra and the Cramér–Rao lower bound (CRLB) to quantify the chemical information content of stellar spectra in terms of the precision (not accuracy¹⁰) to which elemental abundances can be measured. We apply this method to realistic observing conditions of metal-poor, low-mass stars outside the MW for >40 instrument configurations on current (e.g., Keck, the Large Binocular Telescope (LBT), Magellan, MMT, and the Very Large Telescope (VLT)) and future (e.g., JWST, GMT, TMT, E-ELT, and MSE) spectroscopic facilities. For this exercise, we assume the use of full-spectrum-fitting techniques and adopt many of the assumptions commonly used at present in this field (e.g., 1D local thermodynamic equilibrium (LTE) models). We note, however, that the techniques we present can readily be adapted for other choices (e.g., when large grids of non-LTE and/or 3D atmospheres become available).

The paper is organized as follows. In Section 2 we provide a technical description of the information content of spectra and how it can be quantified using CRLBs. In Section 3 we summarize the scope of stars, instruments, and observing scenarios evaluated in this work, our method of stellar spectra generation, and the assumptions that went into our CRLB calculations. We report the forecasted stellar abundance precision for current and planned spectrographs in Sections 4 and 5, respectively. We discuss the highlights and caveats of our forecasts in Section 6. In Section 7 we present Chem-I-Calculator, an open-source Python package for calculating CRLBs of spectroscopic chemical abundance measurements. We summarize our findings in Section 8 and present a number of technical details in the appendices.

2. Information Content of Spectra

In this section we introduce the notion of a spectrum’s information content and its relation to the maximal precision to which stellar labels¹¹ can be measured. We begin in Section 2.1 with a qualitative description of the factors that play a role in the degree of information contained in a stellar spectrum. This is followed by a quantitative description of the information content as represented by the CRLB in Section 2.2.

2.1. A Qualitative Description of Spectral Information

The information content of a star’s spectrum determines the precision to which we can measure its stellar labels—or more technically, how broad the stellar labels’ posteriors are. The amount of information and how constraining that information is depends on the following intrinsic and observed properties of the spectrum:

- (i) *Wavelength Coverage*: How many (and which) spectral features are included in the spectrum.
- (ii) *Wavelength Sampling*: How many wavelength pixels are measured per resolution element.
- (iii) *Spectral Resolution*: How distinct the spectral features of one label are from those of another label.
- (iv) *Flux Covariance*: How uncertain/covariant is the flux in each spectral pixel.

- (v) *Gradient Spectra*: How strongly spectral features respond to changes in the stellar labels.

Aspects (i)–(iv) are determined by the instrument configuration and observing conditions. Generally speaking, they set the size and quality of the spectral data set in question, modulating the availability and accessibility of the spectrum’s information. Larger wavelength coverage and higher wavelength sampling both increase the amount of information-carrying pixels contained in a spectrum. Increased spectral resolution, or resolving power ($R = \lambda/\delta\lambda$), reduces the blending of spectral features and the covariance between stellar labels. Lower flux covariance (i.e., higher S/N) increases the constraining power of informative spectral features. These various characteristics can depend on one another as well (e.g., spectral resolution and wavelength sampling affect the S/N and pixel-to-pixel flux covariance), and there are often trade-offs between them for a fixed instrument configuration or observational strategy.

The gradient spectra, aspect (v), is the most important factor in determining a star’s spectral information content. Generally speaking, it is the stellar labels that result in the largest spectral gradients that have the highest information content and therefore can be recovered to the highest precision. In a χ^2 sense, the more strongly a spectral feature responds to a change in stellar labels, the less the labels need to be offset from the true value to result in a large χ^2 value. More technically phrased: the expectation of the negative second derivative of the spectrum with respect to the stellar labels gives the Fisher information matrix (FIM), which provides a lower bound on the covariance matrix of the stellar labels as discussed in Section 2.2.

Figure 1 helps build intuition for the importance of spectral gradients. Here, we consider a moderate-resolution ($R = 6500$) *ab initio* normalized spectrum of a metal-poor ($\log Z = -1.5$) red giant branch (RGB) star¹² and the partial derivative of that spectrum with respect to Fe, Mg, and Y.¹³ This spectrum and its derivatives were generated using the ATLAS12 and *synthe* models (Kurucz 1970, 1993, 2005, 2013, 2017; Kurucz & Avrett 1981), which we describe in more detail in Section 3.2. The locations and strengths of certain features in the spectral gradient may depend on the adopted stellar atmosphere and radiative transfer models, an issue we discuss in Section 6.4.

As depicted in panel (b) of Figure 1, Fe contributes strongly to a large number of absorption features between 6500 and 9000 Å, including over 200 lines with changes of $>1\% \text{ dex}^{-1}$ and nearly 50 lines with changes of $>5\% \text{ dex}^{-1}$. The large number of information-rich lines is the reason why Fe is one of the most readily recovered elements for cool, low-mass stars.

Compared to Fe, Mg contributes to only 20 features at the $>1\% \text{ dex}^{-1}$ level and only one that is $>5\% \text{ dex}^{-1}$ (at $\lambda 8809$). As a result, it is not as well constrained as Fe. Finally, Y exhibits only three features with gradients larger than $1\% \text{ dex}^{-1}$, illustrating the challenge of recovering its abundance,

¹⁰ See Blanco-Cuaresma (2019) and Jofré et al. (2019) for investigations of the systematics present in spectroscopically derived elemental abundances.

¹¹ In this work we use “stellar labels” to broadly encompass both atmospheric parameters (e.g., effective temperature, surface gravity, and microturbulent velocity) and elemental abundances. We do not, however, include radial velocities in our analysis.

¹² It is important to remember that the gradient spectrum of a star depends on the star’s labels. Cool stars, giant stars, and metal-rich stars all have stronger gradients than hot stars, dwarf stars, and metal-poor stars, meaning that it is easier to precisely recover their stellar labels.

¹³ Unless otherwise stated, elemental abundances are assumed to be in the form of standard solar-scaled abundance ratios with respect to H i.e., $[X/H] = \log_{10}(X/H) - \log_{10}(X/H)_{\odot}$, where $(X/H)_{\odot}$ is the solar abundance ratio.

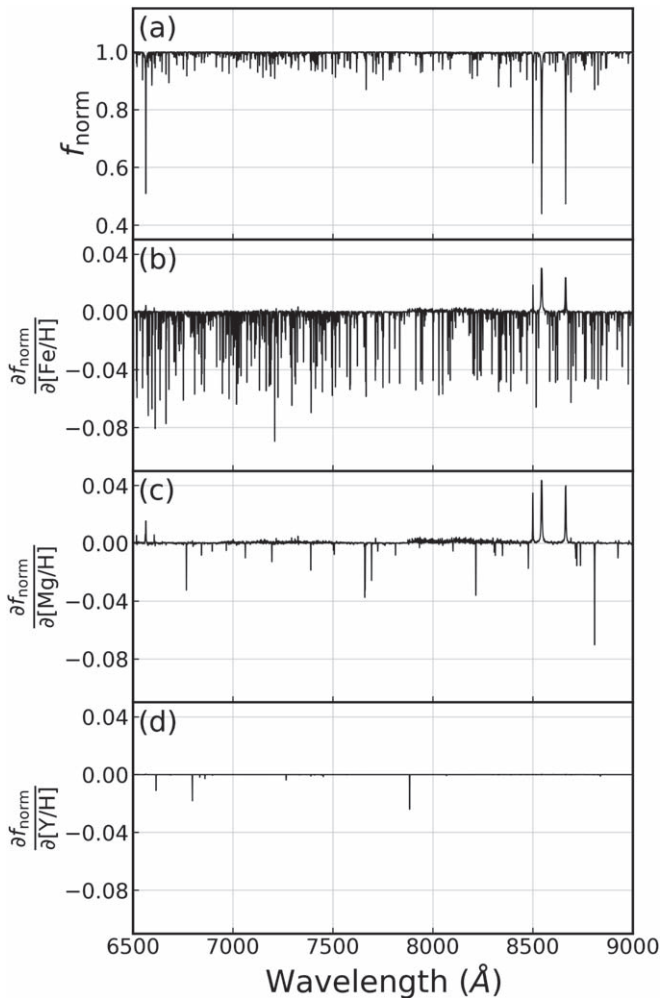


Figure 1. (a) Normalized flux of a synthetic $\log Z = -1.5$ RGB star at $R = 6500$ generated using `atlas12` and `synthe` (see Section 3.2 for model details). (b)–(d) Gradients of the normalized flux with respect to Fe, Mg, and Y, respectively. Many features in the stellar spectrum respond strongly to changes in Fe, meaning that there is considerable information about the iron abundance contained in this spectrum. Changes in Y, on the other hand, cause very weak changes in only a few lines; as a result, the Y abundance would be difficult to recover precisely from this spectrum. Strong positive gradients for Fe and Mg can be seen at the location of the Ca II triplet, which is sensitive to the number of free electrons provided by Fe, Mg, and other electron donors.

even with favorable telescope (high spectral resolution) and observational (high S/N) configurations.

Visually exploring the gradients is a particularly informative exercise. For example, there are clear peaks (i.e., positive deviations in the gradient) in the gradient spectra of Fe and Mg at ~ 8500 Å. These peaks are not due to Fe or Mg transitions, but rather to the Ca II triplet, which is sensitive to the number density of free electrons that Fe and Mg contribute. Y, unlike Fe and Mg, is not a key electron donor and thus does not yield a strong gradient at the location of the Ca II triplet. In this manner, elements that change a star’s atmospheric structure or otherwise indirectly affect the line formation of other elements may be measured—even in the absence of strong absorption features of the element in question (e.g., O can be recovered from spectra that contain few, or no, O lines due to its important role in the CNO molecular network; see Ting et al. 2018). Such measurements, however, require a high degree of

trust in the stellar atmosphere and radiative transfer models being used.

2.2. Quantifying Information Content with CRLBs

A main goal of this paper is to quantify the information content encapsulated in the gradient spectrum, modulated by commonly used instrumental setups and realistic observational considerations. To do this, we employ the CRLB (Fréchet 1943; Darmois 1945; Rao 1945; Cramer 1946), a formal metric for quantifying information content, which we now describe mathematically.

Suppose that we wish to quantify the information content of a stellar spectra observed using a spectrograph with a wavelength coverage of $\lambda_0 \leq \lambda \leq \lambda_N$, a resolving power R , and a wavelength sampling of $\Delta\lambda = \lambda/nR$, where n is the number of pixels per resolution element. Let $f_{\text{obs}}(\lambda)$ be the star’s continuum-normalized flux and Σ be the covariance matrix of the normalized flux.

To make any assessment about the information contained within this spectrum requires a model that relates the star’s physical characteristics (e.g., T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, $[\text{X}/\text{Fe}]$) to its observed spectrum. Suppose we have such a model, $f(\lambda, \theta)$, that predicts the normalized flux of a star at each wavelength, λ , given a set of stellar labels, θ . The nature of this model, whether it be data driven (e.g., Ness et al. 2015), ab initio (e.g., Ting et al. 2019), or a combination of the two (e.g., Xiang et al. 2019), is unimportant provided that it is generative (i.e., it predicts a normalized flux that mimics the observed spectrum from a set of stellar labels) and differentiable in θ (i.e., the spectrum varies smoothly as the star’s labels change).

We can then quantify the precision of our measurements by evaluating the log-likelihood of the data given our model,

$$\ln L(D|\theta) = -\frac{1}{2} \sum_{i=0}^N [(f_{\text{obs}}(\lambda_i) - f(\lambda_i, \theta))^T \times \Sigma^{-1}(f_{\text{obs}}(\lambda_i) - f(\lambda_i, \theta)) + \ln(2\pi|\Sigma|)], \quad (1)$$

for all θ (i.e., over all stellar labels).

The precision to which these labels can be recovered is given by the width of this likelihood function. In practice, however, evaluating the likelihood over a sufficiently large region of parameter space is computationally expensive (and sometimes infeasible) given the high-dimensional nature of spectral fitting.¹⁴ If one assumes priors on the stellar labels (uniform or otherwise), a Markov Chain Monte Carlo (MCMC) method, which enables more efficient sampling of the full posterior than evaluating the likelihood at a grid of labels, can be employed. However, it ultimately still succumbs to the curse of dimensionality when the simultaneous fitting of >20 elemental abundances is required. Because we require our model to be differentiable, this can be made more tractable with alternative sampling techniques like the Hamiltonian Monte Carlo (HMC) algorithm (Duane et al. 1987). Even so, this is still a very computationally expensive exercise to do for every instrument and observational combination.

A more efficient way to obtain the width of the distribution (and in turn the precision on each label) is with the CRLB. Within astrophysics, the CRLB has been used extensively in cosmological contexts (e.g., Albrecht et al. 2006; Adshead & Easther 2008;

¹⁴ Note that the number of grid points needed to fully sample the likelihood scales exponentially with the number of dimensions.

Wang 2010; Becker et al. 2012; Betoule et al. 2014; Font-Ribera et al. 2014; King et al. 2014; Eriksen & Gaztañaga 2015), but has only recently been applied to abundance measurements from full-spectrum stellar spectroscopy (Ting et al. 2016, 2017a).¹⁵

Formally, the CRLB is the highest possible precision achievable for a set of observations and can be derived from the FIM,

$$F_{\alpha\beta} = E \left[\frac{\partial^2 [-\ln L(D|\theta)]}{\partial \theta_\alpha \partial \theta_\beta} \right]_{\hat{\theta}}, \quad (2)$$

where $E[\cdot]$ denotes the expectation value, $\hat{\theta}$ is the maximum likelihood estimate, and α and β are each a specific label. In simpler terms, the FIM describes how fast the likelihood function declines for each label around the maximum likelihood point. The steeper the decline, the narrower the distribution, and the more precisely a label can be measured.

Using the Cramér–Rao inequality, this curvature can be related directly to the width of the Gaussian likelihood. Specifically, the inverse of the FIM gives the lower bound on the covariance matrix of the labels,

$$K_{\alpha\beta} \geq (F^{-1})_{\alpha\beta}, \quad (3)$$

or in terms of measurement uncertainty,

$$\sigma_\alpha \geq \sqrt{(F^{-1})_{\alpha\alpha}}. \quad (4)$$

This lower bound on the measurement uncertainty, σ_α , is the CRLB for the label α .

In order to apply CRLBs to the fitting of stellar spectra, we must make two fundamental assumptions:

- (i) The observed spectra have Gaussian noise, and the likelihood of the spectra given our model is well described by a multivariate Gaussian.
- (ii) The spectral models accurately reproduce the observed spectra (i.e., the fitting is free of systematic errors and $\hat{\theta}$ is an unbiased estimator of a star’s true labels).¹⁶

Assuming Gaussianity (i) is standard practice in the fitting of stellar spectra with $S/N > 10 \text{ pixel}^{-1}$ and enables substituting Equation (1) for the log-likelihood in Equation (2).

Though rarely strictly true, the assumption of accurate models (ii) is commonplace across all of astronomy and astrophysics. Model fidelity is a necessary assumption in all matters of parameter estimation, and so we too assume the stellar models to be correct though we know them to have flaws and oversimplifications (e.g., 1D LTE atmospheres, mixing length theory, incomplete line lists, miscalibrated oscillator strengths). It is important to remember that the CRLBs we calculate are predictions of precision, not accuracy. And while they may be challenging to achieve in practice due to various systematics (see Section 6.4 for further discussion), they nevertheless provide useful guidance for stellar abundance work (see Section 4.1.1 and Appendix D for a comparison of CRLBs with the abundance precision measured in practice).

¹⁵ Ireland (2005) first applied the CRLB formalism to stellar spectroscopy in their analysis of the limiting precision of solar emission lines. Hansen et al. (2015) later used CRLBs to quantify the precision of EW measurements of blended stellar absorption lines.

¹⁶ The CRLB can be generalized to relax the assumption that $\hat{\theta}$ is an unbiased estimator (see Appendix A), but this requires knowing the bias of $\hat{\theta}$ as a function of the stellar labels, which is beyond the scope of this paper.

Under the assumption of perfect models, we can replace $f_{\text{obs}}(\lambda_i)$ in Equation (1) with $f(\lambda_i, \hat{\theta})$, noting that $\hat{\theta}$, as an unbiased estimator, corresponds to the true stellar labels. Combined with the assumption of a multivariate Gaussian log-likelihood, we can rewrite Equation (2) in terms of the gradient spectra as

$$F_{\alpha\beta} = \left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} \right]_{\hat{\theta}}^T \Sigma^{-1} \left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\beta} \right]_{\hat{\theta}} + \frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_\alpha} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_\beta} \right), \quad (5)$$

as worked out in Kay (1993). Because in the context of stellar spectra the covariance matrix of the normalized flux, Σ , is independent of the stellar labels, the second term in Equation (5) vanishes, leaving the FIM as the quadrature sum of the gradient spectra across all wavelength pixels, weighted by the uncertainty of the normalized flux:

$$F_{\alpha\beta} = \left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} \right]_{\hat{\theta}}^T \Sigma^{-1} \left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\beta} \right]_{\hat{\theta}}. \quad (6)$$

Using this form of the FIM, we can now write the CRLB in terms of the spectral gradients as

$$\sigma_\alpha = \left(\left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} \right]_{\hat{\theta}}^T \Sigma^{-1} \left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} \right]_{\hat{\theta}} \right)^{-1/2}. \quad (7)$$

Equation (7) shows that the CRLB is sensitive to the factors that affect the information content of spectra as discussed in Section 2.1. More specifically, if the gradient of the spectrum with respect to a given label is high ($\partial f(\lambda, \theta)/\partial \theta_\alpha$ is large), then σ_α is small and more precise measurements are possible.

Similarly, having high S/N (Σ^{-1} is large) in informative regions of the spectrum will also result in small σ_α and high possible precision. Larger wavelength coverage and higher wavelength sampling mean summing over more pixels and thus higher precision, provided that the pixels are informative and not highly correlated. The importance of instrumental resolution is embedded in the matrix multiplication, where higher resolution gradients lead to deeper spectral features and less blended features, resulting in smaller covariances between stellar labels.

An analytic description of the resolution dependence of the CRLBs is presented in Ting et al. (2017a), which we summarize here:

- (i) The rms depth per pixel (and information) of an absorption feature in the gradient spectrum scales as R .
- (ii) For fixed exposure time and stellar flux, the S/N scales as $R^{-1/2}$ due to Poisson statistics.
- (iii) For a fixed number of detector pixels, the wavelength range scales as $1/R$. Assuming that absorption features are evenly distributed in wavelength space, the information content scales as $R^{-1/2}$ because information adds in quadrature.
- (iv) Together, the simple arguments in (i)–(iii) show that to first order, the stellar label precision is independent of spectral resolving power.

We add to this analytic description that, similar to (iii), the information content scales as $n^{-1/2}$, where n is the number of independent pixels per resolution element. In the extreme case where all n pixels in a resolution element are 100% correlated, the CRLB will be \sqrt{n} larger than if the pixels were entirely uncorrelated. We present a more detailed exploration of the effects of sampling and pixel-to-pixel correlation on the CRLBs in Appendix C.

For a given spectral model (i.e., 1D LTE, as we employ in this work, or 3D non-LTE when they become widely available), forecasting abundance precision is reduced to a matter of calculating derivatives and multiplying matrices. Furthermore, because most spectra have thousands, if not tens of thousands, of pixels, the central limit theorem can be used to show that the CRLB becomes theoretically attainable (i.e., Equation (3) becomes an equality if all assumptions hold). CRLBs are thus an incredibly valuable tool for efficiently exploring the possible precision of a large number of instrumental and observational scenarios when the high dimensionality of the problem makes more rigorous sampling techniques costly or unfeasible.

2.2.1. Incorporating Prior Information

In many cases, there may be additional knowledge of the star's properties beyond the spectra in hand. For example, in an extragalactic context, we may know the distance to the star's host galaxy quite well and/or we may have photometry of the star. Such information can give external constraints on the luminosity, surface gravity, temperature, and even metallicity of a star, and can be used to improve the spectral fitting process. We now demonstrate how this information can be included in the CRLB calculation.

While the CRLB was initially derived in a frequentist context, a Bayesian equivalent of the CRLB can be formulated for application to scenarios in which prior information on the stellar labels is available. This is done by replacing the log-likelihood in Equation (1) with the full Bayesian probability:

$$\ln P(\theta|D) = \ln \Pi(\theta) + \ln L(D|\theta), \quad (8)$$

where $\Pi(\theta)$ is the prior on the stellar labels. This results in the following equation for the Bayesian FIM:

$$F_{\text{Bayes}} = F_{\text{spec}} + F_{\text{prior}} \quad (9)$$

Appendix A of Echeverria et al. (2016) presents a detailed derivations of Equation (9).

The first term on the right-hand side of the equation is the standard spectral gradient FIM found previously (Equation (6)). The second term on the right-hand side of the equation is the FIM of the prior and encapsulates the additional information included in the prior. It can be shown that for Gaussian priors with standard deviation $\sigma_{\text{prior},\alpha}$ for each stellar label, the prior FIM is the diagonal matrix

$$F_{\text{prior},\alpha\alpha} = \left(\frac{1}{\sigma_{\text{prior},\alpha}} \right)^2. \quad (10)$$

As a result, we can write the Bayesian CRLB of a stellar label, α , with Gaussian priors as

$$\sigma_{\text{Bayes},\alpha} = \sqrt{(F_{\text{spec}} + F_{\text{prior}})_{\alpha\alpha}^{-1}} \quad (11)$$

$$= \left(\left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} \right]_{\hat{\theta}}^T \Sigma^{-1} \left[\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} \right]_{\hat{\theta}} + \frac{1}{\sigma_{\text{prior},\alpha}^2} \right)^{-1/2}. \quad (12)$$

As a check, we note that in the case of weak priors or strongly informative data, the CRLBs approach the value predicted by Equation (7), while in the case of strong priors or uninformative data, the CRLBs approach the standard deviation of the priors.

2.2.2. Combining Information from Multiple Spectra

The CRLB can also be applied to the context in which multiple disjoint spectra of the same star exist across different wavelength ranges and resolutions, but are to be fit together. Such cases commonly arise for multiarmed spectrographs (e.g., the Low Resolution Imaging Spectrometer (LRIS) on Keck, Multi-Object Double Spectrographs (MODS) on the LBT, and DESI) and for echelle spectrographs, which observe multiple discrete orders of the stellar spectrum (e.g., GIRAFFE on the VLT).

Replacing the log-likelihood in Equation (2) with the sum of the log-likelihoods for each spectra and following through the previous derivation (Equations (5)–(7)) reveals that the relevant FIM for the joint fitting is simply the sum of the individual spectra's FIM. This is equivalent to concatenating the gradient spectra and covariance matrices of each observation together and using these combined quantities in Equation (7). This can be done for arbitrary combinations of stellar spectra provided that the covariance of overlapping wavelength ranges is properly accounted for (as done in Czekala et al. 2015), otherwise the number of independent information-carrying pixels is artificially inflated.

3. Methods

In this section, we outline our process of generating synthetic stellar spectral gradients and using them to compute CRLBs for a variety of stars, observing scenarios, and spectrographs. We begin by describing the nonexhaustive scope of instruments (Section 3.1.2) and stellar targets (Section 3.1.1) considered in this work. In Section 3.1.3, we describe the determination of realistic S/N estimates for each spectrograph and stellar target. Lastly, we walk through our methodology for generating gradient spectra in Section 3.2. The technical details of the matrix multiplication and inversion used to calculate the CRLBs can be found in Appendix B.

3.1. Observational Scope

While the CRLB is broadly applicable to the entire field of resolved star spectroscopy, we choose to focus this work on forecasting the precision possible for spectroscopy of stars outside of the MW. In general, this limits the scope of this work to large-aperture ground- and space-based telescopes observing faint, metal-poor RGB stars at low and moderate resolution ($R < 10,000$). In the rest of this section, we describe in detail our choice of targets, instruments, and observing conditions.

3.1.1. Properties of Reference Stars

In this work, we limit our analysis to the stars predominantly accessible to spectroscopic campaigns of extragalactic stellar populations: metal-poor RGB stars. We also consider how the CRLBs vary from this fiducial star along several axes, including apparent magnitude, metallicity, and evolutionary

Table 1
Stellar Labels of the Stars Considered in This Work

| Phase | M_V | T_{eff} (K) | $\log g$ | v_{turb} (km s $^{-1}$) | $\log Z$ |
|------------|-------------|----------------------|------------|-----------------------------------|-------------|
| RGB | -0.5 | 4200 | 1.5 | 2.0 | -0.5 |
| RGB | -0.5 | 4530 | 1.7 | 1.9 | -1.0 |
| RGB | -0.5 | 4750 | 1.8 | 1.9 | -1.5 |
| RGB | -0.5 | 4920 | 1.9 | 1.9 | -2.0 |
| RGB | -0.5 | 5050 | 1.9 | 1.9 | -2.5 |
| MSTO | 3.5 | 6650 | 4.1 | 1.2 | -1.5 |
| TRGB | -2.5 | 4070 | 0.5 | 2.3 | -1.5 |

Note. The row in bold designates the fiducial stellar reference used throughout this study. All stars have solar abundance patterns. T_{eff} and $\log g$ are determined from MIST isochrones given the star’s age (10 Gyr), metallicity, and absolute magnitude. v_{turb} is found using the scaling relationship presented in Holtzman et al. (2015). For $\log Z = -1.5$, $M_V = -0.5$ corresponds to a star roughly halfway up the RGB; for more metal-poor stars, the same magnitude corresponds to stars lower on the RGB, closer to the main-sequence turn-off (see Figure 2).

phase as described below. The stellar labels used for these reference stars can be found in Table 1. Their position on the Kiel and Hertzsprung–Russell diagrams can be seen in Figure 2.

For each of the stellar targets considered in this work, we determine the effective temperature and surface gravity of the star using an isochrone from the MESA Isochrones and Stellar Tracks (MIST) project corresponding to the star’s age, metallicity, and absolute magnitude (Paxton et al. 2011, 2013, 2015; Choi et al. 2016; Dotter 2016). As was done in Ting et al. (2017a), we assume a microturbulent velocity for each star using the relationship between microturbulent velocity and surface gravity found by Holtzman et al. (2015):

$$v_{\text{turb}} = 2.478 - 0.325(\log g) \text{ km s}^{-1} \quad (13)$$

Fiducial Star—We adopt as our fiducial stellar reference a star that is roughly halfway up the RGB with a V -band absolute magnitude of $M_{V,\text{Vega}} = -0.5$ ($M_{g,\text{AB}} \sim -0.2$). This choice splits the difference between the brighter but rarer stars at the tip of the RGB (TRGB) and the more numerous but fainter main-sequence turn-off (MSTO) stars. Furthermore, we assume that this fiducial star is 10 Gyr old, has a metallicity of $\log(Z/Z_{\odot}) = -1.5$, and has solar abundance patterns.

Apparent Magnitude—As can be seen from Equation (7), the CRLB scales inversely proportional to the S/N of the spectrum. We consider our fiducial star with apparent magnitudes $m_V = 18, 19.5,$ and 21 , but at fixed stellar evolutionary phase, to avoid conflating the effects of S/N and the star’s atmospheric parameters. This amounts to observing an identical star at distances of $\sim 50, 100,$ and 200 kpc, which are typical distances to nearby MW satellites. When not evaluating the effects of S/N on the chemical abundance precision, we assume the star is located at a distance of 100 kpc ($m_V = 19.5$).

Metallicity—We also investigate how the the information content of an RGB star’s spectrum changes as its metallicity decreases from $\log Z = -1.5$ to -2.5 . Because the shape of the RGB changes as a function of metallicity, we make this comparison at fixed M_V instead of at fixed evolutionary phase. As a result, the lower-metallicity stars considered in this work are located farther down the RGB (i.e., have higher effective temperature and surface gravity; see Figure 2).

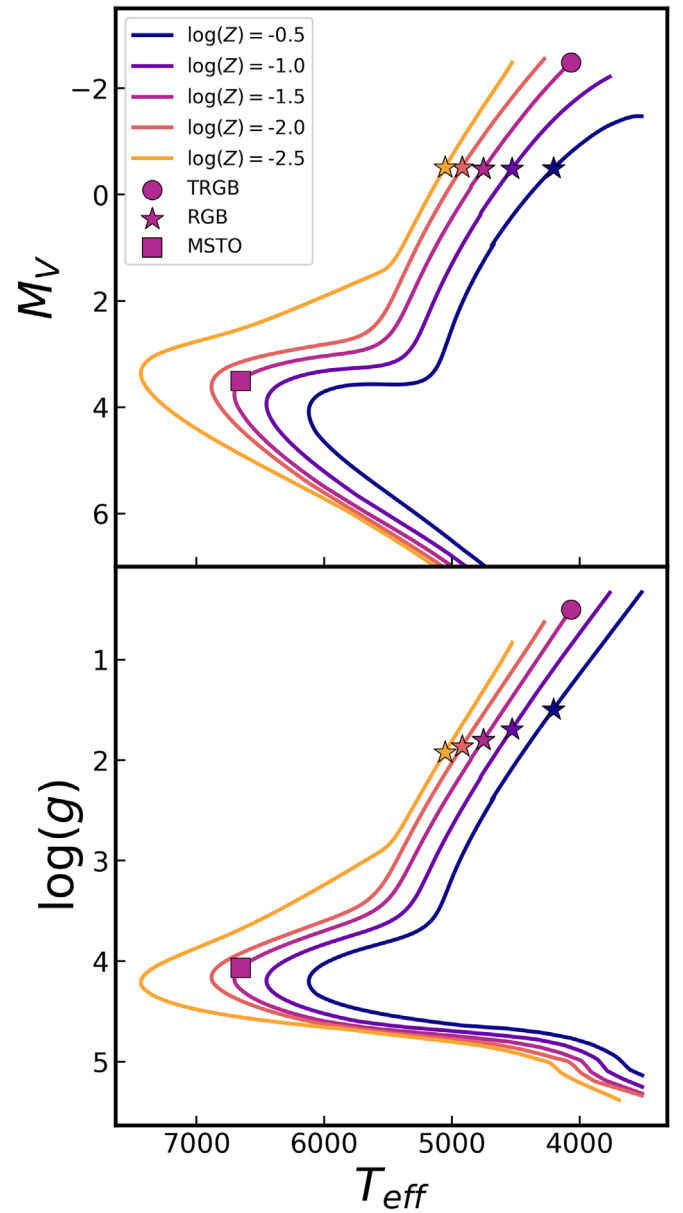


Figure 2. Hertzsprung–Russell (top) and Kiel (bottom) diagrams of the seven reference stars considered in this work (see Table 1). Shapes denote stellar evolutionary phase and colors denote metallicity. The five RGB stars of differing metallicity were chosen to have the same V -band absolute magnitude and thus lie on slightly different portions of the RGB. Solid lines are MIST isochrones of a 10 Gyr old main sequence and red giant branch.

Evolutionary Phase—To isolate the effect of stellar evolutionary phase on the chemical abundance precision, we compare the CRLBs of our fiducial RGB star to that of an MSTO or RGB star of the same metallicity and apparent brightness.

3.1.2. Instruments

Because the stars we consider in this work are so faint ($m_V = 19.5$), we limit our forecasts to instruments, both existing and planned, that can efficiently acquire spectra with modest S/N ($>15 \text{ pixel}^{-1}$) in reasonable amounts of time (<1 night).

In practice, this includes instruments on ground-based telescopes with >5 meter apertures and large-aperture space telescopes. This excludes most of the spectrographs responsible

for large MW surveys (e.g., RAVE, Steinmetz et al. 2006; SEGUE, Yanny et al. 2009; LAMOST, Luo et al. 2015; GALAH, de Silva et al. 2015; and APOGEE, Majewski et al. 2017) and most spectrographs with very high resolving powers ($R > 50,000$). We do not include any instruments with very low resolving powers ($R < 1000$), though there is reason to believe that the information content accessible to very low-resolution grism spectroscopy is still considerable (Bailer-Jones 2000).

Lastly, the line lists¹⁷ we use to generate synthetic spectra are limited in extent to wavelengths between 3000 Å and 1.8 μm. As such, we exclude instruments observing in the ultraviolet (UV) and infrared (IR) despite the significant chemical information that these wavelength regimes contain (e.g., García Pérez et al. 2016; Roederer 2019; Ting et al. 2019).

Even with the aforementioned restrictions, the list of spectrographs already on sky suitable for extragalactic stellar spectroscopy is extensive. As shown in Table 2, we consider 12 existing spectrographs at five world-class observing facilities as well as 9 spectrographs that will be coming online within the next decade. Each of these instruments features numerous choices of observing modes, dispersive elements, and other specifications. This flexibility enables a broad range of science, but makes an exhaustive evaluation of each observing configuration infeasible. Instead, we consider only the setups that we believe most relevant to acquiring precise chemical abundances in extragalactic stellar populations for a total of 41 configurations.¹⁸ For each observational setup, we attempt to use realistic wavelength coverage, wavelength sampling, and resolving power as reported either in literature or in design documents.

Despite an extensive literature search, not all pertinent spectrograph details were readily available, and we had to make some assumptions. For example, for several instruments, the number of pixels per resolution element could not be found; in these cases we adopt a fiducial wavelength sampling of 3 pixels/FWHM as assumed in Ting et al. (2017a). For multiobject spectrographs (MOSs), we assume the nominal wavelength coverage for a star observed in the center of the instrument’s field of view and ignore the variations in wavelength coverage incurred for off-center stars. Additionally, most instruments have wavelength-dependent resolving powers, usually decreasing toward the blue. The manner in which the resolving power changes across the spectrum, known as the line-spread function (LSF), depends on the star’s position in the slit and can vary from slit to slit. For simplicity, we assume all instruments have a fixed LSF with a resolution approximately equal to the average across the entire spectrum.

Lastly, while we do compare and contrast the forecasted precision of these instruments, we emphasize that the “best” instrument is largely of a science-dependent nature. There are numerous trade-offs between field of view and multiplexing (see Table 3), radial velocity precision, and detailed chemical abundance measurements. Balancing them is a matter of their relative importance to the science at hand.

¹⁷ <http://kurucz.harvard.edu/>

¹⁸ This list is extensive but far from complete. We encourage readers interested in spectrographs not listed in Table 2 to calculate their own chemical abundance precision using the Chem-I-CalC Python package detailed in Section 7.

3.1.3. Observing Conditions and Integration Time

We assume the the flux covariance, Σ , is due entirely to photon noise and thus is a function solely of exposure time, instrument throughput, observing conditions, and the star’s brightness, ignoring any uncertainty introduced by imperfect data reduction or continuum normalization.¹⁹ Whenever possible, we use the exposure time calculator (ETC) specific to each instrument listed in Table 4. This allows us to adopt a flux covariance as specific as possible to each facility and accordingly compute realistic CRLBs. For instruments that do not have public ETCs, we scale the S/N from a similar instrument according to

$$S/N \propto D(nR)^{-1/2}, \quad (14)$$

where D is the effective aperture of the telescope, R is the instrument’s resolving power, and n is the instrument’s wavelength sampling.

For our S/N calculations we assume an airmass of 1.1 and a seeing of 0.75" (or as close to these values as possible with each ETC). We assume read-noise is negligible such that the S/N of a single one-hour exposure is the same as that of four 15 minute exposures stacked together.

Because not all ETCs provide the same stellar spectral energy distribution (SED), we use a K0I, K2V, or K0V spectral template (in preferential order when provided) to best match the SED of our fiducial RGB star. Additionally, we use a K0V spectral template for the RGB reference stars with $\log Z \leq -1.5$ and a K5V spectral template for the RGB stars with $\log Z > -1.5$. For the $\log Z = -1.5$ MSTO and TRGB reference stars we use G5V and K5III/K5V stellar templates respectively.

Once calculated by the ETC, the S/N is interpolated onto the same wavelength grid as the stellar spectra corresponding to that instrument’s resolving power, spectral sampling, and wavelength range.

Because most spectrographs are designed to slightly oversample the spectrum (≥ 3 pixels/FWHM), adjacent pixels are not completely uncorrelated, though most stellar abundance studies treat them as such (see however Czekala et al. 2015). For simplicity, we also assume no correlations between adjacent wavelength pixels so that we can write the covariance matrix of the normalized flux, Σ , as the diagonal matrix

$$\Sigma = \begin{bmatrix} \sigma^2(\lambda_1) & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \sigma^2(\lambda_N) \end{bmatrix}, \quad (15)$$

where $\sigma^2(\lambda_i) = (S/N)^{-2}$ is the variance in each pixel. A more accurate treatment of the pixel-to-pixel covariance would effectively reduce the number of independent information-carrying pixels in the spectrum, increasing the CRLB slightly—recall that the CRLB is proportional to $n^{-1/2}$, where n is the number of independent pixels per resolution element. A more in-depth analysis of pixel correlation and wavelength sampling is presented in Appendix C.

The large variety of resolving powers included in this work means that a universal “observing strategy” cannot be applied to all instruments. Instead, we consider separate observing setups for a fiducial spectrograph, low- and medium-resolution spectrographs ($R < 10,000$), high-resolution spectrographs

¹⁹ Reliably determining the (pseudo-)continuum in practice is challenging and is a potential source of systematic errors (see Section 6.4). However, self-consistently normalizing both the observed and model spectra can mitigate these systematics. Evaluating these effects is beyond the scope of this paper.

Table 2
Spectroscopic Configurations Used in This Work

| Telescope/Instrument | Spectroscopic Configuration | Wavelength Range (Å) | R ($\lambda/\Delta\lambda$) | Sampling (Pixels/FWHM) | Aperture (m) | Section | Reference |
|------------------------------|-----------------------------|----------------------|---------------------------------|------------------------|--------------|---------|-----------|
| Existing Instruments | | | | | | | |
| Keck II/DEIMOS ^a | 1200G | 6500–9000 | 6500 | 4 | 10.0 | 4.1 | [1] |
| | 1200B | 4000–6400 | 4000 | 4 | 10.0 | 4.2.1 | [1] |
| | 600ZD | 4100–9000 | 2500 | 5 | 10.0 | 4.2.1 | [1] |
| | 900ZD | 4000–7200 | 2500 | 5 | 10.0 | 4.2.1 | [1] |
| Keck I/LRIS ^a | 600/4000 | 3900–5500 | 1800 | 4 | 10.0 | 4.2.1 | [2] |
| | 1200/7500 | 7700–9000 | 4000 | 5 | 10.0 | 4.2.1 | [2] |
| Keck I/HIRES ^b | B5 Decker | 3900–8350 | 49000 | 3 | 10.0 | 4.3.1 | [3] |
| | C5 Decker | 3900–8350 | 35000 | 3 | 10.0 | 4.3.1 | [3] |
| LBT/MODS ^a | Blue Arm | 3200–5500 | 1850 | 4 | 11.8 | 4.2.2 | [4] |
| | Red Arm | 5500–10500 | 2300 | 4 | 11.8 | 4.2.2 | [4] |
| Magellan/MIKE ^a | Blue (1"0 slit) | 3500–5000 | 28000 | 4 | 6.5 | 4.3.1 | [5] |
| | Red (1"0 slit) | 5000–10000 | 22000 | 3 | 6.5 | 4.3.1 | [5] |
| Magellan/M2FS ^d | HiRes | 5130–5185 | 18000 | 3 ^c | 6.5 | 4.3.2 | [6] |
| | MedRes | 5100–5315 | 10000 | 3 ^c | 6.5 | 4.3.2 | [6] |
| MMT/Hectochelle ^d | RV31 | 5150–5300 | 32000 | 2 | 6.5 | 4.3.2 | [7] |
| MMT/Hectospec ^a | 270 mm ⁻¹ | 3900–9200 | 1500 | 5 | 6.5 | 4.2.2 | [8] |
| | 600 mm ⁻¹ | 5300–7800 | 5000 | 5 | 6.5 | 4.2.2 | [8] |
| MMT/Binospec ^a | 270 mm ⁻¹ | 3900–9200 | 1300 | 4 | 6.5 | 4.2.2 | [9] |
| | 600 mm ⁻¹ | 4500–7000 | 2700 | 3 | 6.5 | 4.2.2 | [9] |
| | 1000 mm ⁻¹ | 3900–5400 | 3900 | 3 | 6.5 | 4.2.2 | [9] |
| VLT/MUSE ^c | Nominal | 4800–9300 | 2500 | 3 ^c | 8.2 | 4.2.2 | [10] |
| VLT/X-SHOOTER ^b | UVB (0"8 slits) | 3000–5500 | 6700 | 5 | 8.2 | 4.3.1 | [11] |
| | VIS (0"7 slits) | 5500–10200 | 11400 | 4 | 8.2 | 4.3.1 | [11] |
| | NIR (0"9 slits) | 10200–18000 | 5600 | 4 | 8.2 | 4.3.1 | [11] |
| VLT/FLAMES-UVES ^f | r580 | 4800–6800 | 40000 | 5 | 8.2 | 4.3.1 | [12] |
| VLT/FLAMES- | LR8 | 4200–11000 | 6500 | 3 ^c | 8.2 | 4.3.2 | [13] |
| VLT/GIRAFFE ^d | HR10 | 5340–5620 | 19800 | 3 ^c | 8.2 | 4.3.2 | [13] |
| | HR13 | 6120–6400 | 22500 | 3 ^c | 8.2 | 4.3.2 | [13] |
| | HR14A | 6400–6620 | 28800 | 3 ^c | 8.2 | 4.3.2 | [13] |
| | HR15 | 6620–6960 | 19300 | 3 ^c | 8.2 | 4.3.2 | [13] |
| Future Instruments | | | | | | | |
| JWST/NIRSpec ^a | G140M/F070LP | 7000–12700 | 1000 | 3 ^c | 6.5 | 5.1 | [14] |
| | G140M/F100LP | 9700–18400 | 1000 | 3 ^c | 6.5 | 5.1 | [14] |
| | G140H/F070LP | 8100–12700 | 2700 | 3 ^c | 6.5 | 5.1 | [14] |
| | G140H/F100LP | 9700–18200 | 2700 | 3 ^c | 6.5 | 5.1 | [14] |
| GMT/GMACS ^a | Blue Arm (LR) | 3200–5500 | 1000 | 3 | 24.5 | 5.2 | [15] |
| | Blue Arm (MR) | 3700–5500 | 2500 | 3 | 24.5 | 5.2 | [15] |
| | Blue Arm (HR) | 4200–5000 | 5000 | 3 | 24.5 | 5.2 | [15] |
| | Red Arm (LR) | 5500–10000 | 1000 | 3 | 24.5 | 5.2 | [15] |
| | Red Arm (MR) | 6100–8900 | 2500 | 3 | 24.5 | 5.2 | [15] |
| | Red Arm (HR) | 6700–8300 | 5000 | 3 | 24.5 | 5.2 | [15] |
| GMT/G-CLEF ^a | Med Res | 3000–9000 | 35000 | 3 | 24.5 | 5.2 | [16] |
| TMT/WFOS ^a | B1210 | 3100–5500 | 1500 | 3 ^c | 30.0 | 5.2 | [17] |
| | B2479 | 3300–4750 | 3200 | 3 ^c | 30.0 | 5.2 | [17] |
| | B3600 | 3250–4100 | 5000 | 3 ^c | 30.0 | 5.2 | [17] |
| | R680 | 5500–10000 | 1500 | 3 ^c | 30.0 | 5.2 | [17] |
| | R1392 | 5850–8400 | 3200 | 3 ^c | 30.0 | 5.2 | [17] |
| | R2052 | 5750–7250 | 5000 | 3 ^c | 30.0 | 5.2 | [17] |
| E-ELT/MOSAIC ^a | HMM-Vis | 4500–8000 | 5000 | 4 | 39.0 | 5.2 | [18] |
| | HMM-NIR | 8000–18000 | 5000 | 3 | 39.0 | 5.2 | [18] |
| Subaru/PFS ^a | Blue Arm | 3800–6300 | 2300 | 4 | 8.2 | 5.3 | [19] |
| | Red Arm (LR) | 6300–9400 | 3000 | 4 | 8.2 | 5.3 | [19] |
| | Red Arm (MR) | 7100–8850 | 5000 | 4 | 8.2 | 5.3 | [19] |
| | NIR Arm | 9400–12600 | 4300 | 4 | 8.2 | 5.3 | [19] |
| MSE ^a | Blue Arm (MR) | 3900–5000 | 5000 | 3 | 11.3 | 5.3 | [20] |

Table 2
(Continued)

| Telescope/Instrument | Spectroscopic Configuration | Wavelength Range (Å) | R ($\lambda/\Delta\lambda$) | Sampling (Pixels/FWHM) | Aperture (m) | Section | Reference |
|--------------------------|-----------------------------|----------------------|---------------------------------|------------------------|--------------|------------|-----------|
| | Green Arm (MR) | 5750–6900 | 5000 | 3 | 11.3 | 5.3 | [20] |
| | Red Arm (MR) | 7370–9000 | 5000 | 3 | 11.3 | 5.3 | [20] |
| | All Arms (LR) | 3600–13000 | 3000 | 3 | 11.3 | 5.3 | [20] |
| Keck/FOBOS ^a | Proposed | 3100–10000 | 3500 | 6 | 10.0 | 5.3 | [21] |
| LAMOST ^d | | 3700–9000 | 1800 | 3 ^c | 4.0 | Appendix D | [22] |
| Mayall/DESI ^d | Blue Arm | 3600–5550 | 2500 | 3 | 4.0 | Appendix F | [23] |
| | Red Arm | 5550–6560 | 3500 | 3 | 4.0 | Appendix F | [23] |
| | Infrared Arm | 6560–9800 | 4500 | 3 | 4.0 | Appendix F | [23] |

Notes. This table lists the spectroscopic configurations we adopt for computing the chemical abundance precision as well as the section in which those precisions are presented. For each instrument, we adopt a constant resolution and number of pixels per resolution element across the wavelength range indicated. The instruments listed here span a large range in wavelength coverage (3200 Å–1.8 μm), resolving powers (1000 < R < 49,000), and instrument designs.

^a Low-/Medium-Resolution Multi-Object Spectrograph.

^b Single-Slit Multi-Order Echelle Spectrograph.

^c Sampling information was not found so a nominal value of 3 pixels/FWHM is assumed.

^d Multi-Object Single-Order Echelle Spectrograph.

^e Integral Field Unit Spectrograph.

^f Multi-Object Multi-Order Echelle Spectrograph.

References. [1] Faber et al. (2003), [2] Oke et al. (1995), [3] Vogt et al. (1994), [4] Pogge et al. (2010), [5] Bernstein et al. (2003), [6] Mateo et al. (2012), [7] Szentgyorgyi et al. (2011), [8] Fabricant et al. (2005), [9] Fabricant et al. (2019), [10] Bacon et al. (2010), [11] Vernet et al. (2011), [12] Dekker et al. (2000), [13] Pasquini et al. (2002), [14] Bagnasco et al. (2007), [15] DePoy et al. (2012), [16] Szentgyorgyi et al. (2016), [17] Pazder et al. (2006), [18] Jagourel et al. (2018), [19] Tamura et al. (2018), [20] MSE Science Team et al. (2019), [21] Bundy et al. (2019), [22] Cui et al. (2012), [23] DESI Collaboration et al. (2016a).

($R > 10,000$), and JWST/NIRSpec, which we describe below. A summary of all of the relevant assumptions used in the S/N calculation of each instrument is contained in Table 4.

Fiducial Spectrograph—To investigate the effects of exposure time, object brightness, and stellar evolutionary phase and metallicity, we adopt the 1200G grating on Keck/DEIMOS as our fiducial spectroscopic setup. We consider 1, 3, and 6 hr integration times and stars with $m_V = 18, 19.5$, and 21. For comparisons of metallicity and stellar evolutionary phase, we hold the integration time and apparent magnitude fixed at 1 hr and $m_V = 19.5$, respectively.

Low- and Medium-resolution Spectrographs—For spectrographs with $R < 10,000$, we consider the baseline observing strategy to be 1 hr of integration of our fiducial $m_V = 19.5$ RGB star. This is generally sufficient for spectrographs on 6+ meter telescopes to achieve $S/N > 15 \text{ pixel}^{-1}$ across the optical spectrum. In this category, we include DEIMOS, LRIS, and FOBOS on Keck; Hectospec and Binospec on the MMT; the Multi Unit Spectroscopic Explorer (MUSE) on the VLT; LBT/MODS on the LBT; PFS on Subaru; MSE; the Multi-object Astronomical and Cosmological Spectrograph (GMACS) on the GMT; the Wide Field Optical Spectrometer (WFOS) on the TMT; and the Multi-Object Spectrograph (MOSAIC) on the E-ELT.

The GMACS ETC provides two sample settings, each of which assume a constant $\delta\lambda$ across both the blue and red channels, resulting in wavelength-dependent resolutions. We choose the higher resolution setting ($\Delta\lambda = 1.4$) and scale the S/N at each pixel according to $S/N \propto R^{-1/2}$ to match the constant resolving power we are attempting to emulate. Because ETCs do not yet exist for MOSAIC, we scale the S/N from GMACS for MOSAIC’s visual high multiplex mode (HMM-Vis) and from NIRSpec for MOSAIC’s near-infrared high multiplex mode (HMM-NIR) according to Equation (14).²⁰

²⁰ By using the ETC of space-based NIRSpec for MOSAIC (HMM-NIR), we ignore a number of telluric features that affect observations in the NIR.

High-resolution Spectrographs—Due to the higher dispersion and generally lower throughput of high-resolution spectrographs, a single hour of integration is insufficient to achieve adequate S/N ($> 15 \text{ pixel}^{-1}$) for an $m_V = 19.5$ RGB star. Instead we consider an integration of 6 hr (~ 1 night of observing). Instruments in this category include the High Resolution Echelle Spectrometer (HIRES) on Keck; the Magellan Inamori Kyocera Echelle (MIKE) on Magellan and the Michigan/Magellan Fiber System (M2FS) on Magellan; Hectospec on MMT; X-SHOOTER,²¹ GIRAFFE, and the Ultraviolet and Visual Echelle Spectrograph (UVES) on the VLT; and the GMT Consortium Large Earth Finder (G-CLEF) on GMT. M2FS and Hectospec do not have public ETCs so we scale the average S/N from the GIRAFFE HR10 ETC according to Equation (14) and assume the S/N is roughly constant over the short wavelength range observed by these instruments.

JWST/NIRSpec—The strength of JWST/NIRSpec is its high sensitivity and high angular resolution. The most likely use case will be to acquire spectra in distant and/or crowded environments, which may require longer integration times than our fiducial 1 hr setup for ground-based, low-resolution instruments. Thus, for JWST only, we adopt a 6 hr of integration on an $m_V = 21$ TRGB star.²² This scenario is chosen to mimic the observation of bright stars in the disk of M31 or in a galaxy at the edge of the Local Group.

Beyond 1 Mpc—To investigate the distance to which JWST/NIRSpec and GMT/GMACS (as a representative ELT) can provide useful chemical measurements, we additionally hold the exposure time constant at 6 hr and systematically decrease the apparent magnitude of our target TRGB star from

²¹ Despite the more moderate resolution of the X-SHOOTER UVB and NIR arms, we include X-SHOOTER with the other high-resolution spectrographs due to its higher resolution VIS arm and single-slit echelle design.

²² Specifically we assume three exposures each of which includes one integration of 170 groups (subintegrations) for a total exposure time of 6 hr 5 minutes, and 35 s.

Table 3
Field of View and Multiplexing of Instruments

| Telescope/Instrument | Field of View | N_{slits} or N_{fibers} |
|------------------------|---------------|---|
| Keck II/DEIMOS | 16' × 4'0 | 100 |
| Keck I/LRIS | 6'0 × 7'8 | 40 |
| Keck I/HIRESr | ... | 1 |
| Magellan/MIKEr | ... | 1 |
| Magellan/M2FS | 30'0 | 250 |
| MMT/Hectochelle | 1'0 | 240 |
| MMT/Hectospec | 1'0 | 300 |
| MMT/Binospec | 16'0 × 15'0 | 150 |
| VLT/MUSE | 1'0 × 1'0 | ... |
| VLT/X-SHOOTER | ... | 1 |
| VLT/FLAMES-UVES | 25'0 | 8 |
| VLT/FLAMES-GIRAFFE | 25'0 | 130 |
| LBT/MODS | 6'0 × 6'0 | 50 |
| JWST/NIRSpec | 3'0 × 3'0 | 100 |
| Mayall/DESI | 2'8 | 5000 |
| Subaru/PFS | 1'3 | 2400 |
| MSE | 9'5 | 3250 |
| Keck/FOBOS | 20'0 | 1800 |
| GMT/GMACS | 7'4 | 100 |
| GMT/GMACS+MANIFEST | 20' | 100 s |
| GMT/G-CLEF+MANIFEST | 20' | 40 |
| TMT/WFOS | 4'2 × 9'6 | 600 |
| E-ELT/MOSAIC (HMM-Vis) | 6'0 | 200 |
| E-ELT/MOSAIC (HMM-NIR) | 6'0 | 100 |

Note. N_{slits} (N_{fibers}) is the approximate number of slits (fibers) that an instrument can handle in a single pointing. This can be used as a rough estimate for the number of stars a spectrograph can observe simultaneously. In practice, of course, not all slits/fibers can be placed on stars because some may be required for guiding, alignment, or sky-subtraction, while others may go unused simply due to the distribution of stars in the field. Single numbers for the field of view (FoV) indicate the FoV's diameter, while pairs of numbers indicate the approximate rectangular dimensions of the FoV. For single-slit spectrographs, the FoV is irrelevant for resolved star spectroscopy. As an IFU, the multiplexing of MUSE depends on the density of stars in the field and the source extraction method employed.

$m_V = 21\text{--}26$. This corresponds to observing a TRGB star at distances between 0.5 and 5 Mpc.

3.2. Gradient Spectra

Ab initio spectra are generated using the same method as described in Ting et al. (2017a). Briefly, we first compute 1D LTE model atmospheres using the `atlas12` code maintained by R. Kurucz (Kurucz 1970, 1993, 2005, 2013, 2017; Kurucz & Avrett 1981). We adopt solar abundances from Asplund et al. (2009) and assume the standard mixing length theory with a mixing length of 1.25 and no overshooting for convection.²³ We then evaluate spectra for these atmospheres at a nominal resolution of $R = 300,000$ using the `synthe` radiative transfer code (also maintained by R. Kurucz). The spectrum is then continuum normalized using the theoretical continuum from `synthe`.²⁴ These high-resolution, normalized spectra are then subsequently convolved down to the average resolution of the

²³ We note that these are not identical assumptions to those made in the MIST isochrones used in Section 3.1.1. This may have a small impact on the consistency of the bolometric magnitudes of the reference stars but should not otherwise affect the results presented in this paper.

²⁴ Again, the use of imperfectly continuum-normalized spectra here should not dramatically change the results of this work as long as all spectra are self-consistently normalized.

relevant instrument (assuming a uniform Gaussian LSF) and finally subsampled onto a wavelength grid with $\Delta\lambda/nR$, where n is the number of pixels per resolution element.

To calculate stellar spectral gradients for each label, we generate a grid of 200 mock spectra, each with one of 100 stellar labels offset from the star's reference labels (see Table 1) by

$$\begin{aligned}\Delta T_{\text{eff}} &= \pm 50 \text{ K}, \\ \Delta \log g &= \pm 0.1, \\ \Delta v_{\text{turb}} &= \pm 0.1 \text{ km s}^{-1}, \text{ or} \\ \Delta [X/H] &= \pm 0.05,\end{aligned}$$

where X refers to elements with atomic numbers between 3 and 99. These step sizes are chosen to be small enough such that the spectral response to each label change is approximately linear, but large enough that the spectral responses remain dominant over numerical noise ($>0.1\%$). For each spectrum in which the abundance of an element is changed, the hydrogen mass fraction is renormalized to compensate, while the helium mass fraction remains constant.²⁵

As in Ting et al. (2017a), we reevaluate the atmospheric structure whenever a stellar label is varied. While more computationally expensive, this is not only essential to capture the response of the spectrum with respect to the atmospheric parameters (i.e., T_{eff} , $\log g$, and v_{micro}), but is also important for certain elemental abundances that have substantial impact on the star's atmospheric structure (see Ting et al. 2016 for details). For example, Mg and Fe are both major electron donors in the atmospheres of cool stars and affect the absorption features of many other elements (Figure 1). While not necessary for all elemental abundances (e.g., Y, which contributes negligibly to the atmosphere's structure), we nevertheless recompute the stellar atmosphere in all cases for consistency.

The final step is to calculate the gradients via the finite difference method. In past work, Ting et al. (2017a) calculated an asymmetric approximation of the gradient of the spectrum with respect to each stellar label by considering the difference of the reference spectrum and the spectra with offsets in that label. In this work, we use a symmetric approximation of the gradient, using the two spectra offset positively and negatively from the reference spectra as we find it yields a more accurate instantaneous derivative at the location of the reference labels. Thus, the gradient of the spectrum with respect to each stellar label, α , evaluated at the reference point θ is

$$\frac{\partial f(\lambda, \theta)}{\partial \theta_\alpha} = \frac{f(\lambda, \theta + \Delta\theta_\alpha) - f(\lambda, \theta - \Delta\theta_\alpha)}{2\Delta\theta_\alpha}. \quad (16)$$

3.3. Summary of Assumptions

For reference, we provide a list of the simplifying assumptions employed throughout our methods. This does not include any assumptions inherent to the derivation of the CRLBs in Section 2.2.

Stellar Model Assumptions:

1. `atlas12` stellar atmosphere model (1D LTE; mixing length of 1.25; no overshoot for convection).

²⁵ We opt not to calculate gradients with respect to the helium fraction, but recognize that this may be of relevance to abundance measurements of hot ($T_{\text{eff}} > 8500 \text{ K}$) stars in globular clusters or other environments where light-element variations are common (see review by Bastian & Lardo 2018 and references therein).

Table 4
ETC Configurations Used in This Work

| Instrument | m_V | t_{exp} (hr) | Airmass | Seeing | Slit Width/ Fiber Diameter | Spatial \times Spectral Binning | Stellar Template | ETC |
|----------------------------|------------------|--------------------------|---------|--------|-------------------------------|--------------------------------------|---------------------|----------------------------------|
| DEIMOS | 18.0, 19.5, 21.0 | 1, 3, 6 | 1.1 | 0".75 | 0".75 | 1 \times 1 | G5V, K0V, K5V | 1 |
| LRIS ^a | 19.5 | 1 | 1.1 | 0".75 | 0".70 | 1 \times 1 | K0V | 2 |
| HIRESr (B5/C5) | 19.5 | 6 | 1.1 | 0".75 | 0".86/1".10 | 2 \times 2 | K0V | 3 |
| MIKE | 19.5 | 6 | 1.1 | 0".75 | 1".00 | 3 \times 1 | K0V | 4 |
| M2FS | 19.5 | 6 | 1.1 | 0".75 | 1".20 | 2 \times 2 | K2V | 5 ^b |
| Hectochelle | 19.5 | 6 | 1.1 | 0".75 | 1".00 | 2 \times 3 | K2V | 5 ^b |
| Hectospec | 19.5 | 1 | 1.1 | 0".75 | 1".5 | 1 \times 1 | K0V | 6 |
| Binospic | 19.5 | 1 | 1.1 | 0".75 | 1".0 | 1 \times 1 | K0V | 6 |
| MUSE | 19.5 | 1 | 1.1 | 0".80 | ... ^c | (3 \times 3) \times 1 | K2V | 7 |
| X-SHOOTER (UVB/VIS/NIR) | 19.5 | 6 | 1.1 | 0".75 | 0".80/0".70/ 0".90 | 1 \times 1 | K2V | 8 |
| UVES | 19.5 | 6 | 1.1 | 0".80 | 1".00 | 1 \times 1 | K2V | 9 |
| GIRAFFE | 19.5 | 6 | 1.1 | 0".75 | 1".20 | 1 \times 1 | K2V | 5 |
| MODS | 19.5 | 1 | 1.1 | 0".75 | 0".70 | 1 \times 1 | K2V | 10 |
| NIRSpec | 21.0–26.0 | 6 | ... | ... | 0".2 | 1 \times 1 | K5III | 11 |
| PFS | 19.5 | 1 | 1.1 | 0".75 | 1".05 | 1 \times 1 | K2V | 12 |
| MSE | 19.5 | 1 | 1.0 | 0".75 | 0".80 | 1 \times 1 | K2V | 13 |
| FOBOS | 19.5 | 1 | 1.1 | 0".75 | 0".80 | 1 \times 1 | K2V | 14 |
| GMACS | 19.5, 21.0–26.0 | 1, 6 | 1.1 | 0".75 | 0".70 | 4 \times 4 | K0V, K5V | 15 |
| WFOS | 19.5 | 1 | 1.1 | 0".75 | 0".75 | 1 \times 1 | K0V | 14 |
| MOSAIC (NIR/Vis) | 19.5 | 1 | 1.1 | 0".75 | 0".80/0".60 | 1 \times 1 | K0I/V | 10 ^b /14 ^b |
| G-CLEF | 19.5 | 6 | 1.0 | 0".79 | 0".70 | 6 \times 9 | K2V | 16 |

Notes. Exposure times are chosen to mimic realistic observing strategies for each instrument. Multiple apparent magnitudes, exposure times, and stellar templates are used with the fiducial 1200G grating on the Keck/DEIMOS spectrograph to investigate their effects on chemical abundance precision. Stellar templates are chosen to best match the stellar energy distribution of the relevant reference star. (1) DEIMOS ETC: http://etc.ucolick.org/web_s2n/deimos, (2) LRIS ETC: http://etc.ucolick.org/web_s2n/lris, (3) HIRES ETC: http://etc.ucolick.org/web_s2n/hires, (4) LCO ETC: http://www.lco.cl/scripts/lcoetc/lcoetc_sspec.html, (5) GIRAFFE ETC: <https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=GIRAFFE+INS.MODE=spectro>, (6) SAO ETC v0.5: <http://hopper.si.edu/etc/cgi/TEST/sao-etc>, (7) MUSE ETC: eso.org/observing/etc/bin/gen/form?INS.NAME=MUSE+INS.MODE=swspectr, (8) X-SHOOTER ETC: <https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=X-SHOOTER+INS.MODE=spectro>, (9) UVES ETC: <https://www.eso.org/observing/etc/bin/gen/form?INS.NAME=UVES+INS.MODE=FLAMES>, (10) MODS Instrumental Sensitivity: http://www.astronomy.ohio-state.edu/MODS/ObsTools/Docs/MODS1_InstSens.pdf, (11) JWST ETC: <https://jwst.etc.stsci.edu/> (workbooks available upon request.), (12) PFS ETC and Spectrum Simulator: https://github.com/Subaru-PFS/spt_ExposureTimeCalculator, (13) MSE ETC: <http://etc-dev.cfht.hawaii.edu/mse/>, (14) FOBOS/WFOS ETC: <https://github.com/Keck-FOBOS/enyo>, (15) GMACS ETC v2.0: http://instrumentation.tamu.edu/etc_gmacs/, (16) G-CLEF ETC: <http://gclef.cfa.harvard.edu/etc/>.

^a The LRIS ETC does not include the 1200/7500 grating throughput so the 1200/9000 grating throughput is used in its place.

^b S/N adapted from the ETC of a similar instrument according to Equation (14).

^c As an IFU, MUSE does not have a definite fiber or slit size on the sky.

2. `synthe` radiative transfer code.
3. Perfectly normalized spectra.
4. MIST stellar isochrones.
5. Solar abundance patterns.
6. Holtzman et al. (2015) empirical relationship between surface gravity and microturbulent velocity.
7. SED approximated by a K0I, K2V, or K0V spectral template.

Instrument Assumptions:

1. Gaussian LSF constant with wavelength.
2. Nominal wavelength sampling of 3 pixels/FWHM adopted when unknown.
3. No correlations between adjacent pixels.
4. Negligible read noise.
5. Same instrument throughput when scaling the S/N using Equation (14).

4. Forecasted Precision of Existing Instruments

Having established how to calculate CRLBs, we are adequately positioned to forecast the chemical abundance precision of existing instruments. With an emphasis on extragalactic stellar

spectroscopy, we begin with a thorough analysis of our fiducial instrument setup: the 1200G grating on Keck/DEIMOS. We then proceed to forecast the precision of other low- and moderate-resolution MOSs on large ground-based telescopes, emphasizing those with wavelength coverage bluer than 5000 Å. Finally, we investigate the capability of low-S/N, high-resolution spectroscopy for precise abundance measurements. With the exception of the analysis in Section 4.1.4, we assume uniform priors on all stellar labels throughout this section.

4.1. D1200G: A Fiducial Example

Though designed with galaxy spectra in mind, the DEIMOS spectrograph on the 10 meter Keck telescope has been critical to our understanding of the resolved stellar populations and chemical evolution of dwarf galaxies. Over the past two decades, observational campaigns with DEIMOS have measured spectra of nearly 10,000 stars in roughly 60 Local Group dwarf galaxies and the halo of M31 (e.g., Chapman et al. 2005; Martin et al. 2007; Simon & Geha 2007; Kirby et al. 2010; Collins et al. 2013; Vargas et al. 2014a, 2014b; Martin et al. 2016b, 2016a; Kirby et al. 2018). The majority of these observations have been made with the 1200G grating centered at 7000 Å (see Table 2 for

details). We will refer to this observational setup as D1200G throughout this work.

In the years immediately following the commissioning of DEIMOS, its primary scientific application was the measurement of radial velocities (e.g., Chapman et al. 2005; Martin et al. 2007; Simon & Geha 2007). Stellar chemistry was often a secondary goal, particularly because high-resolution spectroscopy was often assumed to be necessary for any reliable abundance determinations (see Tolstoy et al. 2009 and references therein). Kirby et al. (2009) demonstrated that the D1200G setup on Keck (and medium-resolution spectroscopy more generally) could be used to recover accurate abundances. Since then, D1200G has become a predominant observing mode for resolved star abundance measurements in dwarf galaxies, making it an excellent fiducial setup for our CRLB calculations.

For this exercise we consider 1, 3, and 6 hr of integration on our fiducial $[\text{Fe}/\text{H}] = -1.5$ RGB with apparent magnitudes of $m_V = 18, 19.5,$ and 21.0 (or equivalently at 50, 100, and 200 kpc). The S/N in each case is calculated using the public ETC according to the configurations in Table 4.

The CRLBs for D1200G are displayed in Figure 3. Throughout this work, we report precisions for solar-scaled relative abundances with respect to hydrogen (i.e., $\sigma[\text{X}/\text{H}]$).²⁶ We consider $\sigma_{\text{CRLB}} = 0.3$ dex to be the worst precision that still enables useful science and thus restrict our analysis to those that can be recovered to this precision or better. We forecast that one hour on D1200G is sufficient to measure 13 elements to better than 0.3 dex in RGB stars out to 50 kpc, 10 elements out to 100 kpc, and 3 elements out to 200 kpc.

As expected from the many features seen in the gradient spectrum (Figure 1(b)), the Fe abundance is recovered to the highest precision. The many strong (and weak) Fe lines included in the D1200G spectrum lead to a precision of 0.02 dex at 50 kpc and to better than 0.2 dex at 200 kpc in only 1 hr of integration. Ni and Si are also precisely recovered due to their numerous features (~ 40 lines with gradients $> 1\%$ dex⁻¹) in the red optical. The high precision possible for Ca, however, is predominantly a result of the very strong Ca II triplet²⁷ at $\lambda\lambda 8498, 8542,$ and 8662 . Meanwhile, elements like Y have only a few weak lines within the D1200G wavelength range (Figure 1(d)) and are thus only recoverable in nearby stars.

Longer exposures provide better S/N, allowing for more precise measurements of more abundances. For a 3 hr observation, the number of elements measured to < 0.3 dex increases to 20, 11, and 7 for RGB stars at 50, 100, and 200 kpc, respectively. For a nearby 18th mag star, the S/N is sufficient (~ 150 pixel⁻¹) to measure elements with only weak signatures in the spectrum. For example, C and N can be recovered from broad, weak CN molecular features between 7000 and 9000 Å. Cu can be measured from two weak ($\sim 1\%$ dex⁻¹) absorption lines at $\lambda\lambda 7935, 8095$. Similarly, elements like La, Mn, O, and Eu have no more than 10 absorption lines with gradients $> 0.5\%$ dex⁻¹ and only 1 or 2 lines with gradients $> 1\%$ dex⁻¹. However, given the high S/N of these observations, they can nevertheless be recovered to a precision of < 0.3 dex.

²⁶ The precision of abundances with respect to Fe (i.e., $\sigma[\text{X}/\text{Fe}]$) can be found by adding $\sigma[\text{X}/\text{H}]$ and $\sigma[\text{Fe}/\text{H}]$ in quadrature.

²⁷ We note that the Ca II triplet is produced in the chromosphere of stars and is subject to substantial non-LTE effects, especially at low metallicities and so must be treated with caution in practice (Jorgensen et al. 1992; Mashonkina et al. 2007; Starkenburg et al. 2010).

At six hours of integration, the S/N is approximately 200, 75, and 30 pixel⁻¹ for RGB stars at 50, 100, and 200 kpc respectively. This enables the recovery of 22, 13, and 9 elements to better than 0.3 dex for these stars. Only after 6 hr of exposures are the weak Y lines enough to measure its abundance out to 100 kpc. These extra three hours of integration are necessary to measure Nd and V in the 18th mag RGB from roughly a dozen very weak lines with gradients $< 0.5\%$ dex⁻¹.

In Figure 3, we also include the spectroscopic precision on the atmospheric parameters T_{eff} , $\log g$, and v_{micro} . With the continuum shape removed from our spectrum, the effective temperature can only be constrained by its impact on atomic and molecular transitions as seen in absorption features. Compared to changes in abundance, the effect of T_{eff} on absorption lines is quite weak ($\sim 2\%$ per 100 K for H α and $< 1\%$ per 100 K for most other lines), but because it manifests in thousands of lines across the D1200G wavelength coverage, it nonetheless allows for T_{eff} to be recovered to better than 100 K in most of the scenarios considered here. In contrast to T_{eff} , changes in $\log g$ affect fewer lines, but much more strongly. H α and the Ca II triplet are notable lines sensitive to the surface gravity in the red optical. The microturbulent velocity lies somewhere between T_{eff} and $\log g$, moderately impacting ($1\% - 4\%$ per km s⁻¹) ~ 50 absorption features across the spectrum.

4.1.1. Comparison to Literature Precision

Our CRLBs formally represent the best achievable abundance precision via full spectral fitting, not necessarily what is obtained in practice (due to imperfect models, variable LSFs, masked or obscured features, etc.). It is therefore useful to compare our CRLB estimates to published abundance precisions from full spectral fitting to get a sense of how close current abundance measurements get to our predictions.

For an illustrative comparison, we select abundances measured by Kirby et al. (2018), who use a full spectral fitting technique (as opposed to EWs) for RGB stars in Local Group galaxies (Kirby et al. 2009). Because of the large variety in stellar targets and spectral quality, we make several cuts to the Kirby et al. (2018) sample in order to fairly compare the reported precision and our CRLBs. First, we consider only stars with T_{eff} between 4500 and 5000 K, $\log g$ between 1.7 and 1.9, and $[\text{Fe}/\text{H}]$ between -2.0 and -1.0 . Second, we consider only stars that were observed to $35 \text{ \AA}^{-1} < \text{S/N} < 65 \text{ \AA}^{-1}$, which corresponds to roughly the mean S/N of a 1 hr exposure of a 19.5 mag star. These cuts leave the reported abundance precision of 33 stars.

Before we make a direct comparison, we modify our CRLB calculation to closely adhere to the choices made by Kirby et al. (2018). For example, $\log g$ and v_{micro} are not fit via spectroscopy, but held fixed at values determined by the star's photometry. This can lead to more precise recovery of abundances by removing their covariances with these labels. Similarly, only Fe, Ca, Ni, Si, Ti, Co, Mg, and Cr are fit, while all other abundances are fixed at the solar abundance value. These are not unreasonable assumptions because the information content of the spectra is dominated by these elements, and $\log g$ is typically better constrained with photometry than spectroscopy in extragalactic contexts where the distance is well constrained. We mimic this analysis by adopting a delta

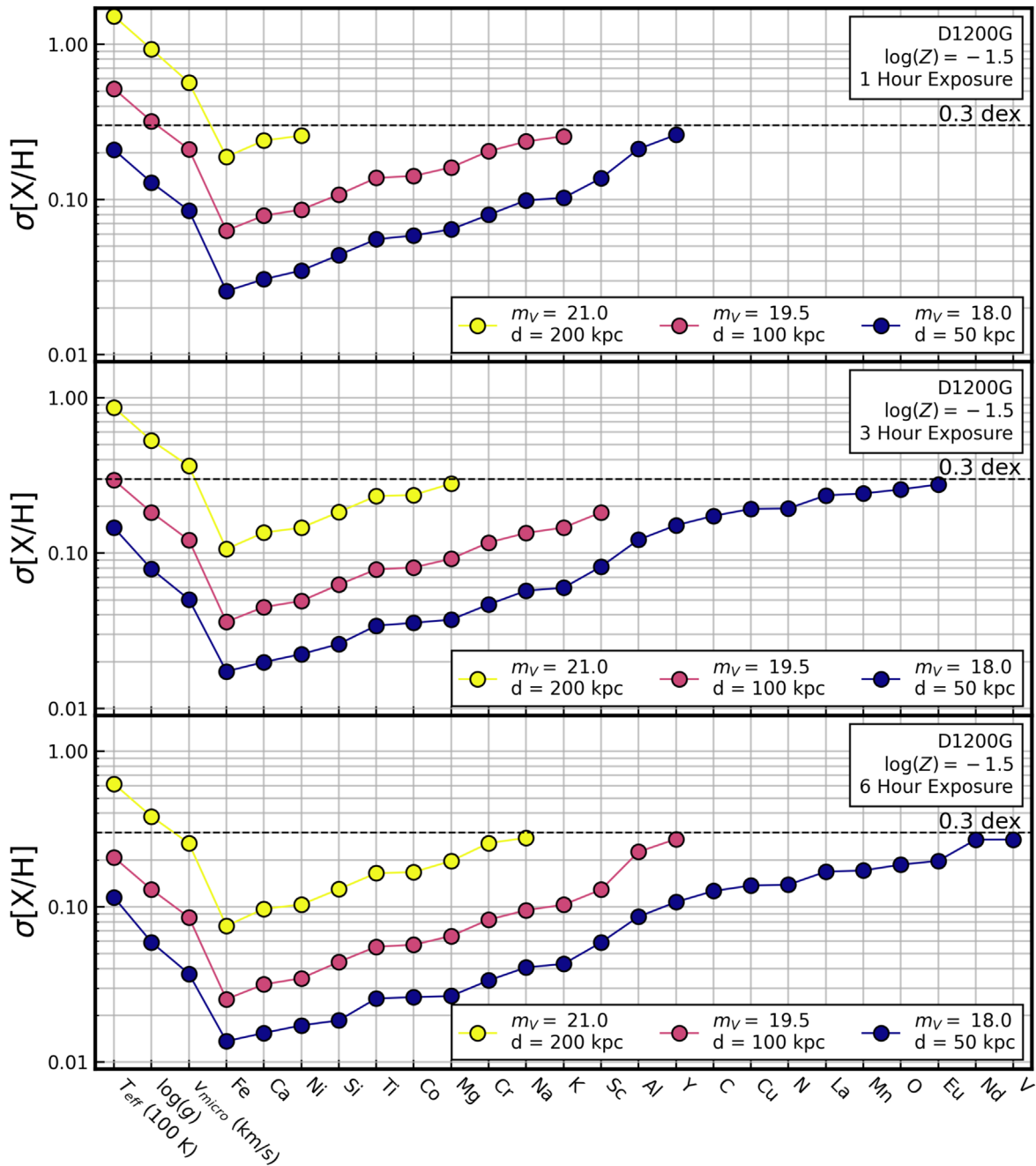


Figure 3. CRLBs for 1, 3, and 6 hr exposures (top, middle, and bottom, respectively) of a $\log Z = -1.5$, $M_V = -0.5$ RGB star (see Table 1) using the 1200G grating on Keck/DEIMOS (see Table 2). Each panel includes the CRLBs for the RGB star located at a distance of 50, 100, and 200 kpc. The elements are ordered by decreasing precision up to 0.3 dex.

function prior on all stellar labels that are not fit for by Kirby et al. (2018).

In addition, Kirby et al. (2018) mask a handful of specific spectral regions that are contaminated by poorly modeled lines or strong telluric absorption features. Following Kirby et al. (2008) we mask 13 spectral regions including notable spectral features such as the Ca II triplet ($\lambda\lambda 6498, 8542, 8662$) and the Mg I $\lambda 8807$ line.

It is worth noting that there are several aspects of the method used by Kirby et al. (2018) that we cannot account for. First, they adopt a different set of stellar models and line lists than we do, albeit with similar 1D LTE assumptions (e.g., ATLAS9 vs. ATLAS12; see Kirby et al. 2010). Second, they fit stellar labels

iteratively by looping through the labels and fitting each individually while holding the rest constant until convergence is achieved. It is possible that this approach may ignore some covariances between labels that are expected when all labels are fit simultaneously as assumed by the CRLB. Third, the specific wavelength coverage of each spectrum varies from the nominal depending on the star’s location on DEIMOS’s detector.

Lastly, we note that the chemical abundance uncertainties reported by Kirby et al. (2018) include both a statistical and systematic uncertainty component added in quadrature. Because CRLBs are purely a measure of statistical precision and not accuracy, we subtract out in quadrature the systematic

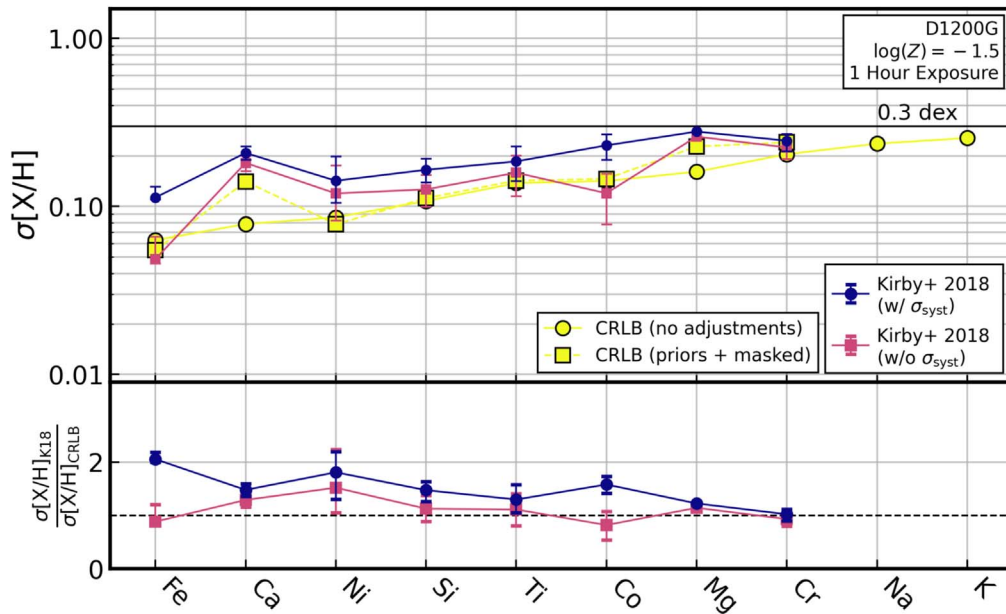


Figure 4. (Top) D1200G CRLBs for a 1 hr exposure of a 19.5 mag $\log Z = -1.5$ RGB star overplotted with the uncertainties of abundances for 35 comparable RGB stars reported by Kirby et al. (2018). The CRLBs represented by squares and dashed lines are calculated by fixing the same stellar labels and masking the same spectral features as Kirby et al. (2018), while the CRLBs represented by circles and solid lines are the same as those presented in Figure 3. Literature uncertainties include a systematic uncertainty and are only provided for stars with uncertainties less than 0.3 dex. Uncertainties for atmospheric parameters T_{eff} , $\log g$, and v_{turb} are not provided. Kirby et al. (2018) did not measure [Na/Fe] or [K/Fe] abundances and therefore have no uncertainties to report for those elements. (Bottom) The ratio of the reported precision to the CRLBs that mimic the analysis techniques of Kirby et al. (2018). Measurement precisions for most elements are within a factor of 2 larger than the CRLBs.

component (of order 0.2 dex for Co and 0.1 dex for all other elements) to make a better one-to-one comparison with the literature uncertainties.

Figure 4 shows the reported precision of the 33 stars from Kirby et al. (2018) plotted with our D1200G CRLBs—both with and without adjustments to match their specific analysis. We find that the abundances reported by Kirby et al. (2018) are within a factor of ~ 2 of our corresponding CRLBs. The precisions reported for Fe (0.05 dex), Co (0.12 dex), and Cr (0.22 dex) are slightly less than our predicted precisions (0.06, 0.14, and 0.20 dex respectively). This may be due to a slight overestimation of the systematic uncertainty on these labels or the underestimation of label degeneracies as a result of the iterative fitting. The reported precision for Co, Mg, and Cr, are likely skewed to higher precision because only abundances recovered to better than 0.3 dex are reported, leaving only eight stars with Co abundances, one star with Mg abundances, and six stars with Cr abundances.

The biggest difference between the CRLBs calculated previously and those calculated to mimic the analysis of Kirby et al. (2018) is in the forecasted uncertainty of Ca and Mg, which increased from 0.07 and 0.16 dex to 0.14 and 0.22 dex, respectively. This is the result of masking strong lines for these elements, which are both highly informative but challenging to model correctly. Fixing $\log g$ would have considerably improved the precision for Ca had the Ca I triplet not been masked due to the feature’s strong dependence on surface gravity. Instead, it only very slightly increases the precision of Fe and Ni from 0.06 and 0.09 dex to 0.05 and 0.08 dex, respectively, but otherwise does not change the CRLB substantially. From this comparison, we can see the importance of folding in these effects to our ability to estimate the expected precision.

While the reported uncertainty for most elements is slightly higher than the CRLB, it is encouraging to see them within a

factor of ~ 2 . There are several reasons why poorer precision in practice could be expected. Examples include poor model fidelity, imperfect calibrations, and masked or lost spectral regions (see Section 6.4 for further discussion). While future comparisons with abundance precisions from full-spectrum fitting are necessary to more completely understand the prospects of achieving the CRLB in practice, this comparison with D1200G illustrates that the CRLBs at least provide a realistic benchmark for spectroscopic abundance precision. In Appendix D, we perform an analogous comparison with LAMOST and find similar agreement between our CRLBs and the literature abundance precision.

4.1.2. CRLBs versus $[Fe/H]$

We now consider how the CRLB changes as a function of metallicity. To do this we compare the CRLBs for RGB stars with $\log Z = -0.5, -1.0, -1.5, -2.0,$ and -2.5 . In order to achieve similar observing conditions for each star, we make comparisons at fixed m_V instead of at fixed stellar phase (or fixed location on the RGB; see Figure 2). As a result of the RGB isochrone’s metallicity-dependent morphology, T_{eff} and $\log g$ for these stars are all slightly different with more metal-poor stars having higher T_{eff} and $\log g$ (Table 1). The S/Ns for these stars are calculated for our fiducial observation of a 1 hr exposure of a star at 100 kpc ($m_V = 19.5$) and the configurations summarized in Table 4.

The CRLBs for the stars of various metallicity are plotted in Figure 5. As expected, the achievable abundance precision decreases toward lower metallicity as there are fewer and weaker absorption features. However, the dependence of precision with metallicity is not uniform across all elements. For example, the precision of Fe steadily decreases from ~ 0.03 dex to ~ 0.1 dex as the metallicity decreases from \log

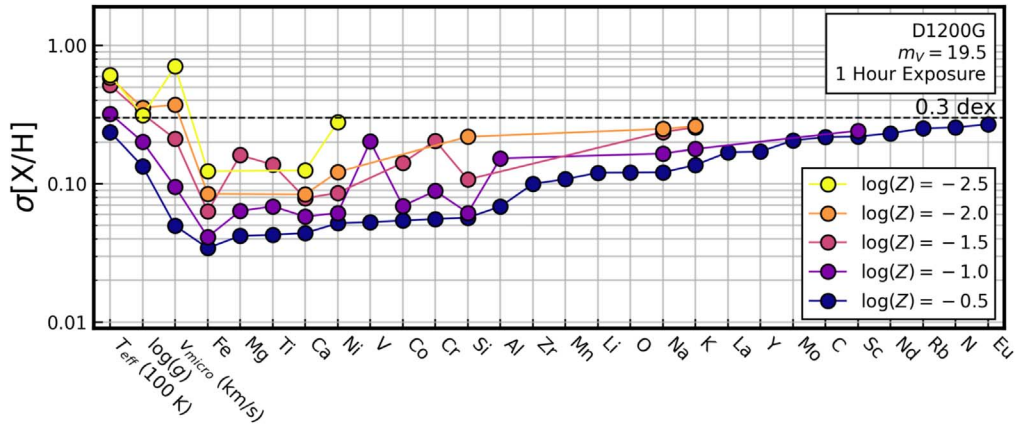


Figure 5. D1200G CRLBs for a 1 hr exposure of RGB stars with metallicities of $\log Z = -0.5, -1.0, -1.5, -2.0,$ and -2.5 at a distance of 100 kpc ($m_V = 19.5$). Table 1 lists the atmosphere parameters for each star. In general, abundance recovery is less precise for lower-metallicity stars due to weaker absorption features.

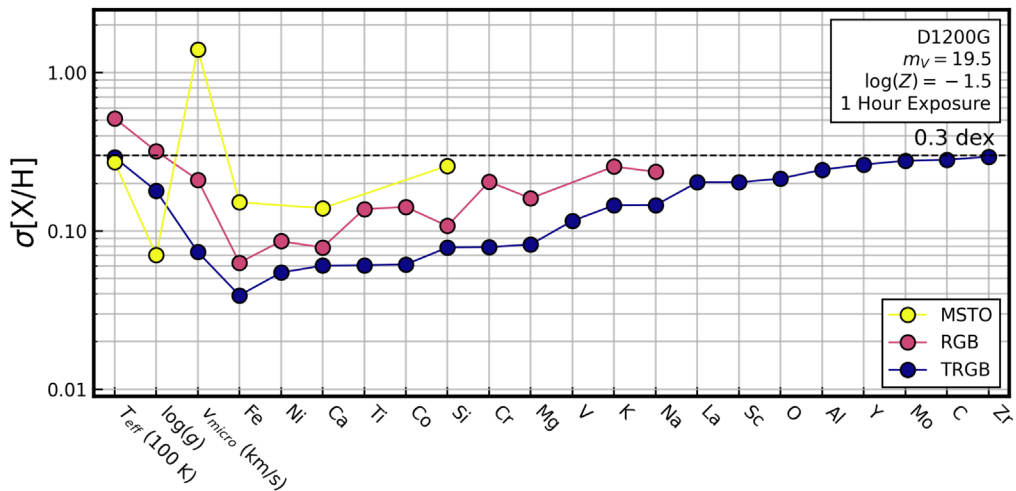


Figure 6. D1200G CRLBs for a 1 hr exposure of $\log Z = -0.5, m_V = 19.5$ MSTO, RGB, and TRGB stars. The atmosphere parameters for each star can be found in Table 1. At low metallicities (such as $\log Z = -0.5$), abundance recovery is more precise for cool giants due to stronger absorption features and less precise for hot subgiants, which have weaker absorption features.

$Z = -0.5$ to -2.5 . The precision of V, however, decreases dramatically from ~ 0.05 dex to ~ 0.2 dex between $\log Z = -0.5$ and -1.0 as a result of its absorption features being strongly temperature dependent. At even lower metallicities (and slightly higher T_{eff}), V features are nearly entirely absent.

Below $\log Z = -0.5$, the CRLBs for T_{eff} and $\log g$ remain constant, or even improve. This seemingly counterintuitive result is due to increasingly prominent Paschen lines redward of 8200 \AA with increasing temperature. These lines are very sensitive to the star’s T_{eff} and $\log g$, allowing for precise measurements of these atmospheric parameters despite the lower metallicities.

4.1.3. CRLBs versus Stellar Phase

Just as a star’s spectral gradients vary as a function of metallicity, it also varies as a function of atmospheric structure (i.e., $\log g, T_{\text{eff}}$, and v_{micro}). As a result, we expect the achievable abundance precision at varying stellar phases to be different even at fixed metallicity and apparent magnitude. While we focus our analysis on a typical RGB star, stars from the MSTO to the tip of the red giant branch (TRGB) are also targets of extragalactic studies.

Here, we consider the CRLBs for the $\log Z = -0.5$ RGB star considered previously with that of an MSTO and TRGB star at the same metallicity (see Table 1). We once more consider a 1 hr integration of a $m_V = 19.5$ star with the relevant ETC configuration in Table 4.

The CRLBs of each of these stellar phases are plotted in Figure 6, illustrating that the chemical abundance precision is best for TRGB stars and worst for MSTO stars (all other things being equal). While only 3 elements can be measured to better than 0.3 dex from the spectrum of the MSTO star, 10 elements can be measured to this precision in the RGB star, and 19 in the TRGB star. For a fixed element the precision is roughly two times better for the TRGB star than the RGB star and another two times better than the MSTO star.

These differences are expected because the absorption features of hot subgiants are significantly weaker than for cool giants. This is especially true for elements like C, N, and O, which are measured primarily from molecular features that are pronounced in TRGB stars but practically nonexistent in MSTO stars. Similarly, Fe, Si, Mg, Al, and other elements whose abundances affect a star’s atmospheric structure leave a larger signature in cool, low surface gravity stars than hot, high surface gravity stars.

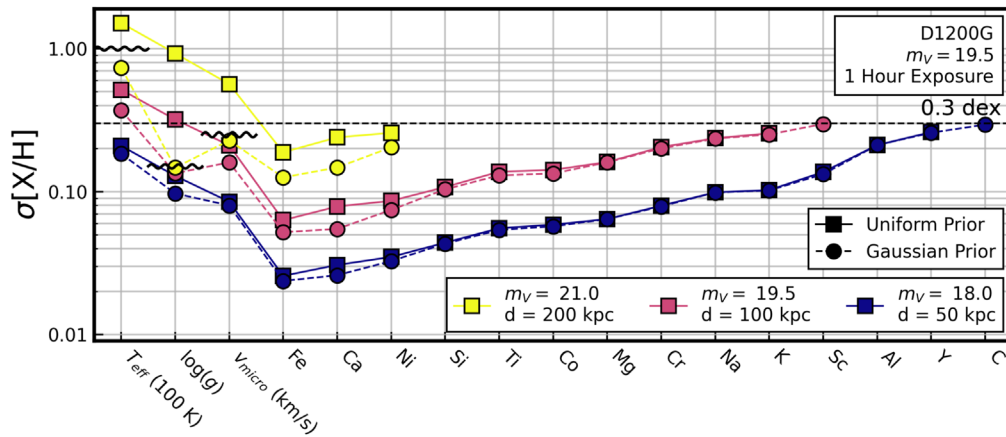


Figure 7. Same as the top panel of Figure 3 but also including the Bayesian CRLBs assuming $\sigma_{T_{\text{eff,prior}}} = 100$ K, $\sigma_{\log g, \text{prior}} = 0.15$ dex, and $\sigma_{v_{\text{micro,prior}}} = 0.25$ km s⁻¹ (dashed lines). The black wavy lines mark the priors on T_{eff} and $\log g$. In addition to better constrained T_{eff} and $\log g$, the inclusion of priors also improves the precision of abundance determinations, particularly at lower S/N.

Recovering T_{eff} and $\log g$, on the other hand, can be done more precisely in MSTO stars, due to the strong dependence of the Paschen lines on the star’s atmospheric parameters.

4.1.4. CRLBs with Priors

For stars with secure distances (as members of external galaxies typically are), photometry can be used to constrain T_{eff} and $\log g$ to roughly ± 100 K and ± 0.15 dex, respectively (Kirby et al. 2009; Casagrande et al. 2011; Heiter et al. 2015). Knowledge of $\log g$ and Equation (13) can also constrain v_{micro} to roughly ± 0.25 km s⁻¹ (Holtzman et al. 2015). We can incorporate these photometric estimates as priors on our spectroscopically determined labels as shown in Section 2.2.1. To do so we adopt Gaussian priors on these parameter with standard deviations equal to their photometric uncertainties. We once more consider a 1 hr observation of our fiducial $\log Z = -1.5$ RGB star at 50, 100, and 200 kpc.

Figure 7 shows the results of the CRLBs assuming Gaussian priors. For reference, we include the CRLBs from Figure 3 (top), which assume uniform priors.

For the highest S/N case (at 50 kpc; S/N ~ 75 pixel⁻¹), the precision on T_{eff} and $\log g$ from D1200G spectroscopy alone is significantly better than the priors. The priors therefore contribute negligible additional information, and the CRLBs only minimally improve.

However, in the lowest S/N case (at 200 kpc; S/N ~ 10 pixel⁻¹), T_{eff} , $\log g$, and v_{micro} are substantially less constrained by the spectroscopy compared to the priors and so nearly all of the information about these stellar labels are coming from the prior. As a result, use of these priors improve the precision of T_{eff} , $\log g$, and v_{micro} by factors of 2–6 compared to the uniform prior case.

In addition, because spectral gradients of T_{eff} and $\log g$ are covariant with the spectral gradients of elements like Fe, Ca, and Ni, priors that better constrain T_{eff} and $\log g$ also lead to improved precision on these chemical abundances. For example, in the case of our faintest star, the Fe, Ca, and Ni abundance precision improves by $\sim 50\%$ when Gaussian priors on T_{eff} and $\log g$ are included. We expect the inclusion of photometric priors to have more impact when the spectral gradients of different labels are more covariant (i.e., for low-resolution spectra with heavily blended lines and spectra with very limited wavelength coverage and few absorption lines).

4.2. Low- and Medium-resolution MOS

All other things being equal, high-resolution spectra would be preferable for abundance measurements, as fewer lines are blended, which results in fewer coupled abundance determinations. Unfortunately, as described in Section 4.3, high-resolution spectrographs are typically limited to the brightest extragalactic stars due to their high spectral dispersion, relatively low throughput, and limited multiplexing capabilities. As a result, it is not possible at present to efficiently observe large numbers of extragalactic resolved stars with broad wavelength coverage and $R > 10,000$ spectroscopy.

Low- and medium-resolution MOSs, on the other hand, provide high multiplexing capabilities, increased throughput, and broad wavelength coverage, enabling them to achieve modest S/N of many faint stars simultaneously in distant systems. Furthermore, as we will show, wavelengths bluer than ~ 5000 Å—even at low resolution—are incredibly rich in absorption features, especially for the cool low-mass giants typically observed outside the MW.

Historically, low- and moderate-resolution blue-optical spectra have not been favored for abundance determinations due to the challenge in identifying the continuum and substantial blending of lines (Ting et al. 2017a). However, in recent years, advances in spectral fitting techniques have led to large improvements in abundance recovery from low-resolution blue-optical spectra. Notably, Ting et al. (2017b) and Xiang et al. (2019) have shown that it is possible to measure 16 + elements of ~ 6 million MW stars from $R \sim 1800$ LAMOST spectroscopy with a wavelength coverage of 3700–9000 Å. While the small aperture of LAMOST (1.75 m) precludes it from abundance measurements of most stars outside the MW, there are a handful of MOS already in commission that provide similar resolving power and wavelength coverage on 6+ meter telescopes (e.g., Keck/LRIS, LBT/MODS, and MMT/Hectospec). In the following sections, we quantify the potential of these facilities for chemical abundance measurements outside the MW.

4.2.1. Blue-optical MOS on Keck

On the Keck/DEIMOS spectrograph there are several options that provide access to wavelengths bluer than 5000 Å. As listed in Table 2, the 900ZD, 600ZD, and 1200B

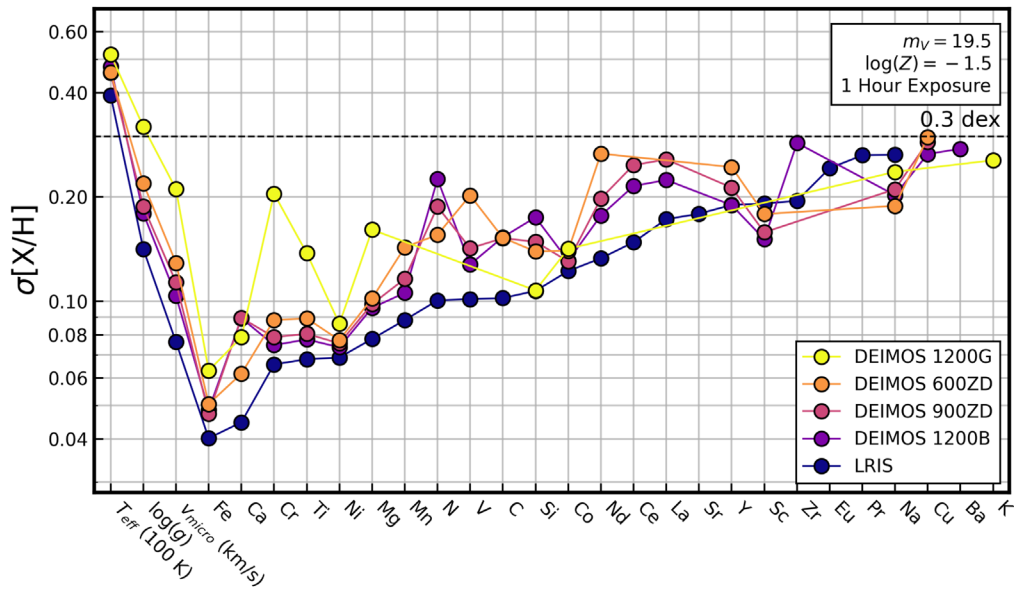


Figure 8. Comparison of CRLBs for several MOS setups on Keck/DEIMOS and Keck/LRIS assuming a 1 hr exposure of a $\log Z = -1.5$, $M_V = -0.5$ RGB star at 100 kpc. The LRIS setup includes the spectral coverage of both its blue and red channels. The elements are ordered by decreasing precision as forecasted for LRIS up to 0.3 dex. The CRLB for D1200G is the same as shown previously in Figures 3(top), 5, and 7.

gratings all provide bluer wavelength coverage, but slightly lower resolution, compared to the D1200G setup. These gratings have already enabled abundance determinations not possible from red-optical spectroscopy, such as the measurement of α elements in the M31 halo (Escala et al. 2019b) and Ba in several dwarf galaxies (Duggan et al. 2018). The 1200B grating is a recent addition to DEIMOS’s grating collection and has not been used to measure stellar abundances at the time of this paper’s writing.

In addition to DEIMOS, the Keck telescopes also host the LRIS MOS, which operates using separate red and blue channels. The 600/4000 grism on the blue arm boasts impressive blue throughput compared to DEIMOS gratings²⁸, while the 1200/7500 grating on the red arm provides coverage around the Ca II triplet (Table 2). While LRIS has only ever been used for very limited stellar abundance determinations (Shetrone et al. 2009; Lai et al. 2011), it is nonetheless a promising instrument, particularly given the demonstrated success of LAMOST.

To quantify the information content accessible in the blue optical by these instrumental setups, we calculate their CRLBs given a 1 hr exposure of our fiducial $\log Z = -1.5$ RGB star at 100 kpc and the relevant ETC configurations for each instrument from Table 4.

The forecasted abundance precision for each element is presented in Figure 8. Despite their lower resolving powers, instruments with bluer wavelength coverage provide more precise measurements of more elements than D1200G. For example, the 1200B grating on DEIMOS and the 600/4000 + 1200/7500 LRIS setup enable the recovery of 21 and 22 elements respectively to better than 0.3 dex—about twice that from comparable red-optical spectroscopy at fixed integration time and stellar type. This includes eight r- and s-process elements (Y, Ce, La, Zr, Ba, Sr, Pr, and Eu), which have most, if not all of their absorption features at wavelengths shorter

than 5000 Å and are thus largely inaccessible to D1200G and other longer wavelength spectrographs. Information about C and N comes primarily from C₂, CH, and CN absorption bands between 4000 and 5000 Å and to a lesser extent from CN bands between 7000 and 9000 Å.

D1200G does provides comparable or better precision for Fe, Ni, Si, and Co, which have many lines at wavelengths longer than ~ 6500 Å, as well as for Ca, Na, and K, which have strong features in the red optical.²⁹

LRIS’s improved precision is due to a combination of its exceptional throughput down to 3900 Å and the additional wavelength coverage provided by its red arm.³⁰ However, it is important to remember that LRIS has roughly half the field of view and half the multiplexing as DEIMOS (Table 3), meaning that it may ultimately be less efficient for some elements, when the number of stars is included in the calculation.

As a reminder, the DEIMOS 600ZD and 900ZD gratings and the LRIS 1200/7500 grating all oversample their spectra with 5 pixels/FWHM. If the pixels in these spectra are not completely independent as we assume here, the CRLBs we present may be slightly more precise than would be expected in practice (see Section 6.4.3).

4.2.2. Blue-optical MOS on Other Telescopes

We now turn our attention to blue-sensitive instruments on facilities other than the Keck Telescopes, which include MODS on the LBT, MUSE on the VLT, and Hectospec and Binospec on the MMT.

MODS, like LRIS, operates at low resolution ($R \sim 2000$) across the optical spectrum with a red and a blue arm, and modest multiplexing (Tables 2 and 3). Other than a recent study on a chemically peculiar ultra metal-poor star in the dwarf

²⁸ 25% at 4500 Å compared to 13% for DEIMOS 1200B and 4% for DEIMOS 1200G.

²⁹ The Ca II triplet at $\lambda\lambda 8498, 8542, 8662$, the Na I doublet at $\lambda\lambda 8185, 8197$, and the K I doublet at $\lambda\lambda 7667, 7701$, respectively.

³⁰ Though LRIS does lose considerable information for Sc, Na, Cu, Ba, and K in the gap between its red and blue coverage. This can be mitigated to a degree by carefully choosing the dichroic and grating angle employed.

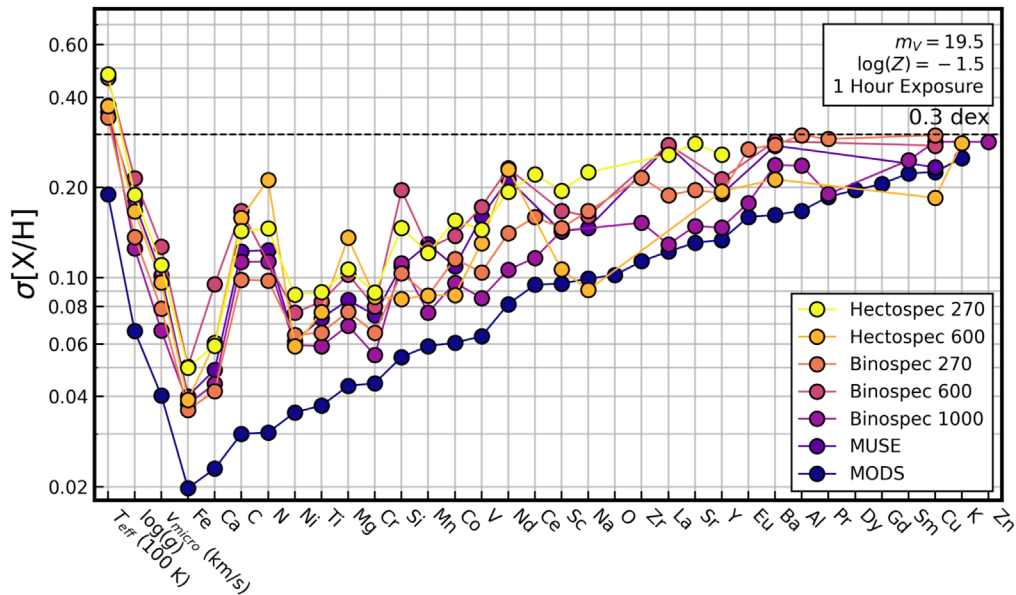


Figure 9. Same as Figure 8 but for LBT/MODS, MMT/Hectospec, and MMT/Binospec. Elements are ordered by the precision forecasted for LBT/MODS up to 0.3 dex.

galaxy Canes Venatici I (Yoon et al. 2019), MODS has not been utilized for stellar chemical abundance measurements.

While MUSE is not technically an MOS but rather an integral field unit (IFU), it can nonetheless be used effectively for low-resolution multi-object resolved star spectroscopy in crowded fields. MUSE has already been used to conduct several campaigns for both stellar radial velocity and chemical abundance measurements in globular clusters (e.g., Husser et al. 2016; Kamann et al. 2016, 2018; Latour et al. 2019), in dwarf galaxies (e.g., Voggel et al. 2016; Alfaro-Cuello et al. 2019; Evans et al. 2019), and in NGC 300 (Roth et al. 2018; McLeod et al. 2020).

Hectospec, in comparison to MODS, MUSE, and the spectrographs on Keck, has a very large field of view ($1^\circ \times 1^\circ$), which makes it a powerful instrument for spectroscopic observations of very extended stellar populations. For example, Carlin et al. (2009) used Hectospec to measure the kinematics and bulk metallicity of stars in the disrupted MW dwarf galaxy Boötes III. Binospec is a new, complementary MOS to Hectospec with very high throughput, but a significantly smaller field of view and a more limited multiplexing capability (Table 3). Both Hectospec and Binospec have a number of gratings that allow for a range in wavelength coverage and resolving power. We examine a few setups we consider to be most applicable to extragalactic stellar spectroscopy (see Table 2 for specifics).

Figure 9 shows the CRLBs for our fiducial RGB star ($\log Z = -1.5$, $m_V = 19.5$) and a 1 hr exposure. For these observing conditions, MODS is forecasted to recover up to 30 individual elements to better than 0.3 dex. MODS’s precision can be attributed to two key factors: its large, nearly 12 meter effective aperture and its throughput below 4000 Å, which together achieve S/N of $>40 \text{ pixel}^{-1}$ down to 4000 Å and $>10 \text{ pixel}^{-1}$ down to 3500 Å. As discussed in Section 4.2.1, these regions become increasingly information rich due to the high densities and strengths of absorption features of many elements.

There are a few specific elements that are worth examining in more detail. Just as with the blue-optimized spectrographs on Keck, the constraints on C and N abundances come predominantly

from absorption bands at wavelengths bluer than 5000 Å and (to a lesser extent) between 8000 Å and $1 \mu\text{m}$. MODS’s sensitivity across both of these ranges leads exceptional recovery of C and N compared to the other instruments analyzed here. MUSE and the 600 gratings of Hectospec and Binospec do not push nearly as blue (or red) and thus recover C and N abundances less precisely or not at all. While the 270 grating on Hectospec and the 270 and 1000 gratings on Binospec do include most of the blue carbon features, they miss most of the blue nitrogen features and (with the exception of the 270 grating on Binospec) achieve an S/N in this region roughly half that of MODS. As a result they also do not recover C and N as precisely as MODS.

In addition to C and N, MODS is also able to recover O to better than 0.2 dex because of strong OH absorption features below 3500 Å and the important role of O in the CNO molecular network (Ting et al. 2018).

Again, it is worth highlighting the precision capable of these blue-optimized spectrographs for heavy r- and s-process elements Nd, Ce, Zr, La, Sr, Y, Eu, Ba, Pr, Dy, Gd, and Sm (in order of decreasing precision for MODS). In addition to those seen in Figure 8, the ability to recover Nd, Dy, Gd, and Sm is the direct result of MODS blue sensitivity (discussed further in Section 6.1). A few of these are recoverable by MUSE, Hectospec, or Binospec, but measurement is made more difficult due to lower S/N and smaller wavelength coverage.

Given the smaller light-collecting power of MMT, it is reasonable that Hectospec and Binospec are forecasted to recover fewer elemental abundances and at larger uncertainties. It is nonetheless still interesting to look at them in greater detail and compare the various Hectospec and Binospec settings. Generally Binospec’s higher throughput leads to higher precision measurements, but this of course comes with a diminished field of view and fewer fibers for stars.

Similarly, the increased abundance precision of MODS, MUSE, and other Keck spectrographs is also modulated by much reduced fields of view. The choice between these instruments then ultimately comes down to weighing the

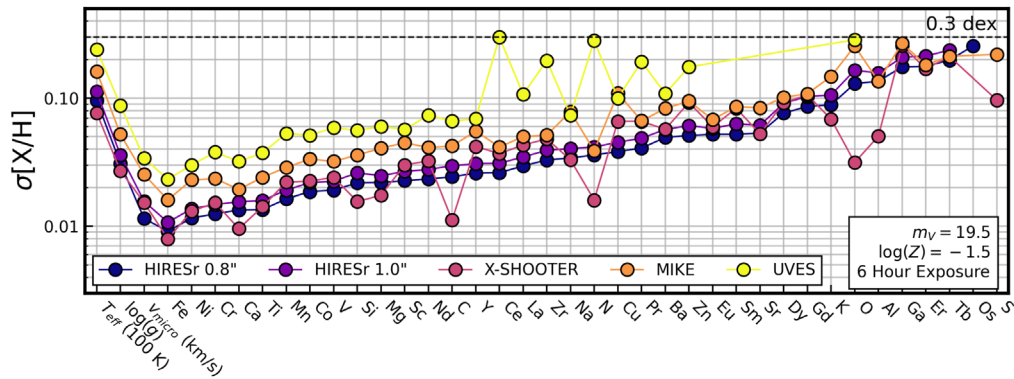


Figure 10. Comparison of CRLBs for high-resolution single-slit echelle spectrographs Keck/HIRES, Magellan/MIKE, and VLT/X-SHOOTER assuming a 6 hr exposure of a $\log Z = -1.5$, $M_V = -0.5$ RGB star at 100 kpc. The elements are ordered by decreasing precision as forecasted for HIRES up to 0.3 dex. The CRLBs suggest that even at low S/N (~ 15 – 20), the chemical information content of high-resolution spectra is considerable.

importance of detailed abundance patterns versus the importance of a large sample size to the desired science.

We remind the reader that the Hectospec configurations oversample their spectra with 5 pixels/FWHM. If the pixels in these spectra are not completely independent as we assume here, the CRLBs we present may be slightly more precise than would be expected in practice (see Section 6.4.3).

4.3. Low-S/N, High-resolution Spectroscopy

In this section, we consider two classes of high-resolution spectrographs: single-slit echelle spectrographs and multiplexed single-order spectrographs.

4.3.1. High Resolution, Single Slit

High-resolution spectroscopic observations of stars provide precise radial velocities and are the gold standard for chemical abundance determinations. Because high-resolution spectroscopy provides spectra with fewer blended absorption features, spectral abundance determinations preferentially use clean, isolated lines that can be fit with EW methods over blended lines, which require spectral synthesis techniques. By not fitting the star’s entire spectrum simultaneously, some of the spectrum’s chemical information goes un-utilized. By calculating the CRLBs for several high-resolution spectrographs, we illustrate the chemical information that can be accessed through full-spectrum-fitting techniques.

In the context of extragalactic studies, two commonly used single-slit echelle spectrographs are Magellan/MIKE and Keck/HIRES. Both instruments provide high-resolution spectra across the entire optical regime and have been used extensively for abundance measurements in MW globular clusters (e.g., Boesgaard et al. 2000, 2005; Venn et al. 2001; Koch & Côté 2010) and in nearby dwarf galaxies (e.g., Shetrone et al. 1998; Frebel et al. 2014, 2016; Koch & Rich 2014; Ji et al. 2016a, 2016b, 2016c, 2020).

We also consider two spectrographs on the VLT: UVES and X-SHOOTER. UVES is a high-resolution spectrograph with a more limited wavelength coverage (only 4800–6800 Å) but is capable of observing up to eight stars at a time when connected with the Fibre Large Array Multi Element Spectrograph (FLAMES) fiber feed.³¹ It has been used to observe RGB stars

in MW globular clusters (e.g., Alves-Brito et al. 2006) and in nearby dwarf galaxies (e.g., Shetrone et al. 2003; Letarte et al. 2006; Hill et al. 2019; Lucchesi et al. 2020). X-SHOOTER has also been used to measure abundances of bright stars in dwarf galaxies (Starkenburger et al. 2013; Spite et al. 2018) and provides slightly lower resolution than MIKE, HIRES, and UVES but significantly higher throughput and broader wavelength coverage.³²

As discussed in Section 3.1.3, a 1 hr exposure of a $m_V = 19.5$ RGB star is typically insufficient for high-resolution spectrographs to overcome the read-noise-limited regime of faint object spectroscopy. Instead we consider a more realistic 6 hr (~ 1 night) of integration, which yields $S/N > 15$ (10) pixel^{-1} at 4500 Å and $S/N > 20$ (20) pixel^{-1} at 7500 Å for HIRES (MIKE) when adopting the ETC configurations in Table 4.

Figure 10 shows the CRLBs for HIRES, MIKE, FLAMES-UVES, and X-SHOOTER. As expected, high-resolution spectra provide very precise detailed chemical abundance patterns. HIRES, MIKE, and X-SHOOTER are forecasted to measure a dozen elements to nearly 0.01 dex and over 30 elements to better than 0.3 dex. UVES, with its smaller wavelength coverage and lower S/N (5–10 pixel^{-1}), is still forecasted to recover over 20 elements. This high precision is predicted despite the low S/N (< 20 redward of 4500 Å) of these observations, demonstrating the potential power of full spectrum fitting applied to high-resolution spectra. While at low S/N any given absorption feature might be only weakly informative, the ensemble of all spectral features still provides strong constraints on the chemical abundances of a star.

The chemical information for many of the elements in Figure 10 can be traced to the same large numbers of features below ~ 5000 Å as previously discussed in Section 4.2. While these absorption features are still subject to blending, the higher resolution of these instruments increases the rms depth of the absorption feature and alleviates degeneracy between elements. This results in increased abundance precision over low-resolution instruments at fixed wavelength coverage. We can see this effect when comparing the CRLBs of the two HIRES settings, which have the same wavelength coverage but different resolving powers—the CRLBs scale with resolving

³¹ In this way, it straddles the boundary of the single-slit, multiorder spectrographs discussed in this section and the highly multiplexed, single-order spectrographs discussed in Section 4.3.2.

³² X-SHOOTER’s NIR arm extends wavelength coverage to 2.48 μm , but due to the limitations of our line list we only consider wavelengths shorter than 1.8 μm .

power $\sigma_{\text{CRLB}} \propto R^{-1/2}$ as expected for instruments with the same wavelength range.³³

In addition to elements previously discussed in Sections 4.1 and 4.2, HIRES can recover the abundances of neutron-capture elements Sm, Er, Tb, and Os to better than 0.3 dex. At $R \sim 50,000$ there are nearly 100 Sm lines with gradients $>5\%$ dex⁻¹ and over 30 lines with gradients of 10%–30% dex⁻¹ in the HIRES wavelength range—all of which are below 4500 Å. The same spectrum has ~ 15 (5) absorption lines with gradients of $>5\%$ dex⁻¹ (10% dex⁻¹) absorption lines for Er (Tb) blueward of 5000 Å. Os can be recovered to ~ 0.3 dex from no more than five absorption lines with $>5\%$ dex⁻¹ gradients.

MIKE’s bluer wavelength coverage is largely offset by its lower resolving power ($R \sim 28,000$) and very low S/N (<5 pixel⁻¹) below 5000 Å. Nevertheless, MIKE achieves slightly better precision for Tb and Er, which have two to three times more lines between 3500 and 3900 Å than they do at wavelengths longer than 3900 Å. MIKE’s recovery of N is aided by strong molecular absorption bands at $\lambda 3550$ and $\lambda 3800$ and another in the red at $\lambda 9150$. Its higher precision for Al and S compared to HIRES is the result of additional atomic absorption lines beyond 8500 Å and its higher S/N in the red.

X-SHOOTER, despite its lower resolution ($R \sim 10,000$), recovers most elements as precisely as, if not better than, MIKE and HIRES. For C, N, and O, X-SHOOTER can achieve precisions two to three times better than MIKE and HIRES as a result of its larger wavelength coverage. It is sensitive to both the CNO molecular bands in the blue optical and the NIR molecular features beyond 1 μm . Si, Mg, Na, Al, K, and S also have a handful of absorption features in the NIR, enabling one to two times higher precision with X-Shooter than MIKE and HIRES. Furthermore, because the NIR is generally less dense with absorption features, the gradients for these elements are less degenerate with other stellar labels and can thus be more precisely recovered.

The comparatively lower precision of FLAMES-UVES can be attributed to its shorter (and redder) wavelength coverage, which does not include nearly as much of the high-information density spectral regions as the other spectrographs considered here. Furthermore, the S/N is roughly two to three times lower than that of MIKE or HIRES. Depending on the desired science, however, the multiplexing capabilities of UVES may more than make up for its lower throughput and wavelength coverage.

At low S/N (e.g., 5 pixel⁻¹), there may be a concern that the assumptions of Gaussianity, which underlies the CRLB, may not be valid. However, we show in Appendix E that the CRLBs are robust to the level of ~ 0.01 dex down to S/N ~ 5 pixel⁻¹. Thus, we believe non-Gaussianity to have a minimal impact on the CRLBs, especially compared to other practical limitations (e.g., model fidelity) that make it difficult to fully realize the precision forecasted by the CRLBs.

UVES and the UVB arm of X-SHOOTER oversample their spectra with 5 pixels/FWHM. If the pixels in these spectra are not completely independent as we assume, the CRLBs may not be as precise as we present here (see Section 6.4.3).

³³ A factor of R^{-1} from the scaling of the absorption feature rms depth and a factor of $R^{1/2}$ from the scaling of S/N with dispersion. For these two HIRES settings, $R^{-1/2} \sim 0.85$.

4.3.2. High Resolution, Single Order

Another approach to high-resolution spectroscopy involves using order-blocking filters that block all but one order of the echelle spectrum. Doing so allows for improved multiplexing, but limits the observed wavelength to a small window of 50–300 Å. Historically, the primary application of these instruments for extragalactic archaeology has been the efficient measurement of precise radial velocities in dwarf galaxies (e.g., Walker et al. 2007, 2009b), but these spectra clearly contain chemical information as well.

We consider three such high-resolution, single-order, fiber-fed MOS: VLT/FLAMES-GIRAFFE, MMT/Hectochelle, and Magellan/M2FS. Due to the nature of order blocking in these instruments, there is great flexibility in deciding what small portion of spectrum to observe. In this work, we will only look at spectral regions targeted by existing observations and save a detailed analysis of the optimal wavelength windows for a future paper. For M2FS, this includes a “HiRes” and a “MedRes” setting around the Mg I b triplet ($\lambda\lambda 5183, 5172, 5167$), which have been used for membership determination and [Fe/H] measurement in several MW satellites (e.g., Walker et al. 2007, 2009b, 2015a, 2016). The RV31 order-blocking filter was used on Hectochelle for similar purposes (e.g., Walker et al. 2009b, 2015b; Spencer et al. 2017) and is also utilized by the H3 MW halo survey (Conroy et al. 2019a, 2019b). On FLAMES-GIRAFFE, five settings have been used by the DART (Dwarf Abundances and Radial Velocities Team) program to measure various abundances and radial velocities in Local Group dwarf galaxies: LR8, HR10, HR13, HR14A, and HR15 (e.g., Hill et al. 2019; Theler et al. 2019). Details for all of these instruments and settings can be found in Table 2.

Just as with the previous high-resolution CRLBs, we consider 6 hr of integration on our log $Z = -1.5$ RGB star at 100 kpc and the ETC configurations in Table 4.

Figure 11 shows the forecasted precision for these single-order echelle spectrographs. As expected, the limited wavelength coverage of these setups severely reduces their chemical abundance recovery compared to the full-optical, high-resolution spectrographs presented in Figure 10. Even most low-resolution spectrographs can achieve comparable or better abundance recovery in a fraction of the time as presented in Figures 8 and 9. This is because the information content scales proportionally with the square root of the number of absorption features. A smaller wavelength range means fewer lines for a given element and worse precision.

Nevertheless, given the narrow wavelength range covered by these orders and the low S/N (~ 15 – 30 pixel⁻¹), it is promising that more than a handful of elements beyond Fe can be recovered to better than 0.3 dex. We first consider the abundance precision for M2FS and Hectochelle (Figure 11; top), which cover 5100–5300 Å. This narrow region of the spectrum contains numerous absorption lines of Fe, and to a lesser extent also of Ni, Ti, Co, Cr, and Nd, which enable their recovery. All three filters were designed to include the Mg I b triplet and as a result Mg can also be measured. There are also a few (<5) strong ($\sim 10\%$ – 30% dex⁻¹ at $R \sim 32,000$) lines each for Ca, Sc, Y, and Cu in this wavelength range that enable the M2FS MedRes configuration and Hectochelle to recover these elements. Hectochelle achieves slightly higher precision due to its higher resolving power. Because M2FS’s HiRes filter has a more limited wavelength range, it misses a

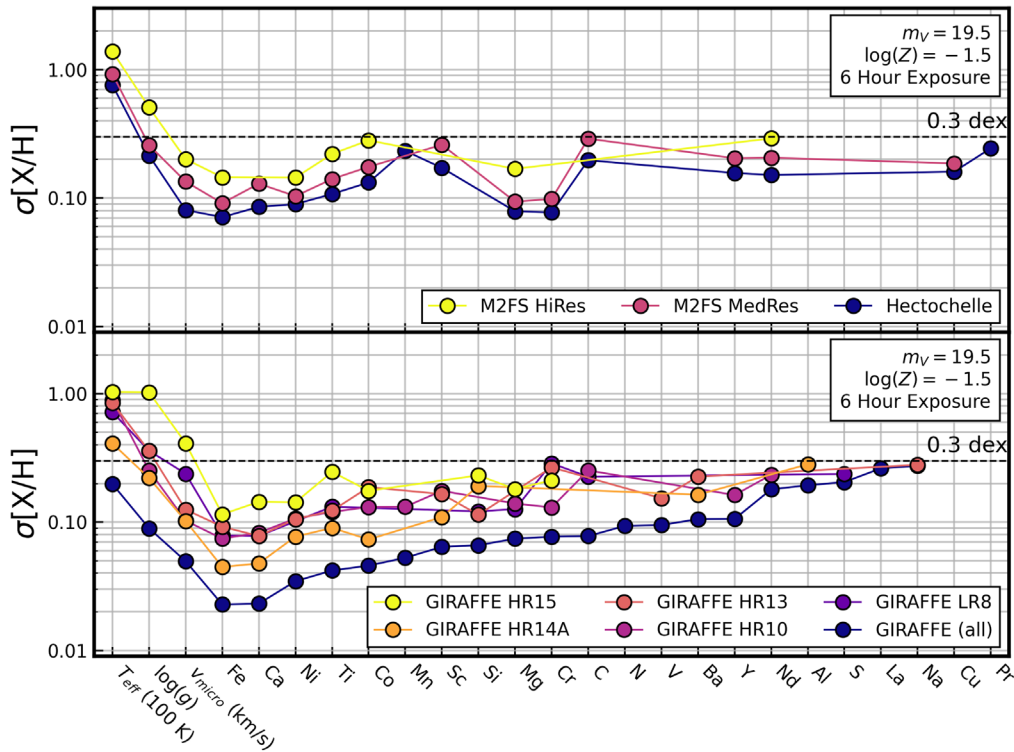


Figure 11. Same as Figure 10 but for multiplexed, single-order echelle spectrographs. CRLBs for Magellan/M2FS and MMT/Hectochelle are included in the top panel, and CRLBs for various VLT/FLAMES-GIRAFFE orders are included in the bottom panel. Elements are ordered by the precision forecasted for a combined analysis of all five GIRAFFE orders shown. The CRLBs suggest that even very small regions of spectrum, when well chosen, may contain nonnegligible chemical information.

considerable fraction of these lines and thus cannot measure these abundances as precisely.

Next we consider GIRAFFE, which has several orders that span the entire optical spectrum. Fe, Ca, Ni, Ti, and Co all have numerous strong lines ($>10\% \text{ dex}^{-1}$ at $R \gtrsim 20,000$) below 7000 \AA , enabling their recovery by all high-resolution order-blocking filters. Mn, however, has the majority of its strongest lines between 5300 and 5600 \AA and is thus only recovered by HR10. The same is approximately true for Y and Nd. Ba has two moderate absorption features ($>10\% \text{ dex}^{-1}$ at $R \gtrsim 20,000$) at $\lambda 6143$ and $\lambda 6499$ in the HR13 and HR14A filters, respectively, but is better recovered in HR14A because of the filter’s higher S/N and resolving power. The combination of throughput and resolution enables HR14A to achieve higher precision for its recoverable elements than the other individual filters, though its redder wavelength coverage precludes it from measuring elements whose lines reside primarily at wavelengths bluer than 6000 \AA .

For reference, we also include in the bottom of Figure 11 the CRLB for the combined analysis of all five GIRAFFE orders as was done in Hill et al. (2019). It is clear that by combining the many information-carrying absorption features across all orders provides a significant improvement in the possible stellar label precision and enables the measurement of elements that no individual filter alone could recover (e.g., N and La). However, to achieve the S/N and abundance precision found here, would require 6 hr of integration on each of the five GIRAFFE orders for a total of 30 hr of integration. Still, it is useful to compare this precision to that

of low-resolution MOS and high-resolution single-slit echelle spectrographs. While low-resolution, blue-optical spectroscopy can achieve similar precision abundance determinations for a similar number of stars in a small fraction of the time, the kinematic information in these observations is limited—at $R \sim 2000$, the precision of radial velocity measurements is only $\sigma_{RV} \sim 150 \text{ km s}^{-1}$, which is good enough for membership determination, but not for detailed kinematic studies. In contrast, $R \sim 20,000$ spectra yield $\sigma_{RV} \sim 5 \text{ km s}^{-1}$, which are precise enough for stellar multiplicity determinations, orbit reconstruction, and dark matter mass measurements. Furthermore, these high-resolution observations will be less prone to systematics incurred by model imperfections in blended lines.

A drawback to high-resolution, single-slit echelle spectrographs is the amount of time required to build up large samples of stars. In 30 hr of integration time, assuming 6 hr per pointing and ignoring overheads, HIRES, MIKE, and X-SHOOTER could observe five stars, while five echelle orders (6 hr each) could be acquired by GIRAFFE for ~ 100 stars. Ultimately, the choice of instrument and observing strategy is highly dependent on the science case and whether higher abundance precision or a larger sample size is most valuable and whether precise radial velocities are needed. However, in the specific case of chemodynamical studies of dwarf galaxies, where both chemical and kinematic information are desired for a large number of stars, it may be worth trading in full optical coverage for specific wavelength regions and higher multiplexing.

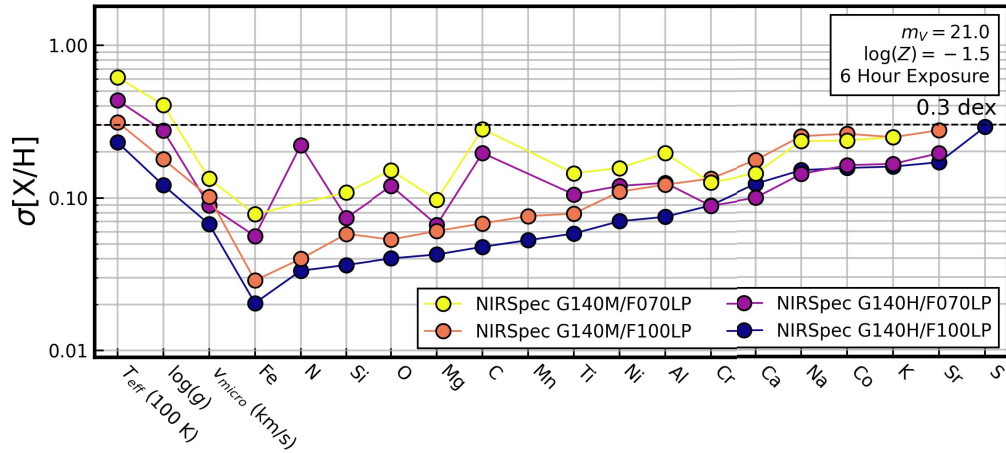


Figure 12. CRLBs for four gratings on JWST/NIRSpec assuming a 6 hr exposure of a $\log Z = -1.5$, $m_V = 21$ TRGB star. The elements are ordered by decreasing precision as forecasted up to 0.3 dex. These CRLBs represent the abundance precision that can be measured for RGB stars in M31 or in dwarf galaxies at the edge of the Local Group.

5. Forecasted Precision of Future Instruments

In this section, we forecast the precision achievable by instruments currently in their construction or design stages. Our lengthy, but incomplete, list includes JWST/NIRSpec, 30 meter class ELTs, and several planned survey facilities (e.g., MSE, FOBOS). Because many of these instruments are still undergoing conceptual and practical revisions, the specifications we adopt in this section are estimates based on the best currently available information.

5.1. JWST/NIRSpec

The unprecedented angular resolution of the Near-Infrared Spectrograph (NIRSpec) on JWST opens up a new domain of crowded-field extragalactic stellar spectroscopy that is currently at or beyond the limits of the most powerful ground-based telescopes (e.g., faint stars in the disk of M31 or beyond the Local Group).

In this analysis, we consider four of the nine NIRSpec MOS disperser–filter combinations whose details can be found in Table 2. We consider 6 hr of integration and a $\log Z = -1.5$ TRGB star at a magnitude of $m_V = 21$, which is similar to observing such a star in M31 or at the edge of the Local Group.

Figure 12 shows the CRLBs for JWST/NIRSpec. We predict that NIRSpec can recover between 13 and 17 individual elemental abundances to better than 0.3 dex despite low resolution of these spectra ($R < 3000$) and the faintness of the target star. This is quite promising for the future of extragalactic stellar spectroscopy as the field moves toward more distant and crowded extragalactic systems. For comparison, ground-based observations are presently limited to measuring only $[\text{Fe}/\text{H}]$, bulk α -element enhancements, and a few other elements in the M31’s halo and satellites (e.g., Collins et al. 2013; Vargas et al. 2014b; Escala et al. 2019b; Gilbert et al. 2019; Kirby et al. 2020).

Figure 12 also shows that for the same filter (i.e., wavelength coverage) the slightly higher resolution of the G140H grating provides an advantage in precision over the G140M grating despite the reduced S/N (100 pixel^{-1} versus 160 pixel^{-1} at $1.2 \mu\text{m}$). Just as in Section 4.3, this is consistent with the CRLBs scaling with $R^{-1/2}$ at fixed wavelength coverage.

Further, we see that the redder F100LP filter provides better abundance precision than the blue F070LP filter. This is due to a combination of factors including the F100LP’s larger wavelength coverage and marginally higher S/N. Though it is true that blue-optical wavelengths are rich in information, the situation changes in the red, where molecular bands in the NIR are more information rich than the red optical.

In fact, the abundance precision benefits greatly from information contained at wavelengths longer than $1.4 \mu\text{m}$ provided by the F100LP filter. These redder wavelengths include numerous molecular features like the strong H_2O absorption lines that extend to $1.8 \mu\text{m}$. Also included are bands of CN ($\lambda 1.1 \mu\text{m}$), OH ($\lambda 1.4 \mu\text{m}$), and CO ($\lambda 1.5 \mu\text{m}$), features, which enable precise determinations of C, N, and O. In addition to Fe, Si, and Mg, which have absorption features somewhat uniformly distributed from 7000 \AA to $1.8 \mu\text{m}$, the F100LP filter also enables precise recovery of Mn, which has ~ 10 lines between 1.2 and $1.4 \mu\text{m}$ with strengths greater than $1\% \text{ dex}^{-1}$ (at $R = 2700$).

The redder wavelength coverage of the F100LP filter also allow for more precise recovery of T_{eff} and $\log g$. This is the result of both Paschen lines at $\lambda\lambda 1.05, 1.09$, and $1.28 \mu\text{m}$ and Brackett lines redward of $1.46 \mu\text{m}$. These lines are all sensitive to atmospheric parameters and thus provide strong constraints on T_{eff} and $\log g$ (and to a lesser extent Fe, Si, Mg, and Al).

The bluer wavelength coverage of the F070LP filter does provide better recovery for Ti, Ca, Na, and Cr. Constraints on Ti abundance come from several TiO bands blueward of $1 \mu\text{m}$ and constraints on Cr come from roughly a dozen weak ($< 2\% \text{ dex}^{-1}$ at $R = 2700$) lines blueward of $1.2 \mu\text{m}$. The precision of Ca and Na is a result of the Ca I triplet at $\lambda\lambda 8498, 8542, 8662$ and Na I doublet at $\lambda\lambda 8185, 8197$ as discussed previously in Section 4.2.1.

We conclude by noting potential challenges in achieving the NIRSpec CRLBs. NIRSpec’s elemental precision is strongly contingent on the information content of complicated molecular features. As a result, the abundances measured by NIRSpec may be quite sensitive to assumptions of the model atmosphere, molecular network, and line lists employed. Achieving the reported CRLBs and avoiding large systematics at $R < 3000$ will require careful treatment of this portion of the spectrum.

In addition, due to the rigid nature of NIRSpec’s mechanical slit mask, it will frequently be the case that stars will lie slightly off the center of their slit. In addition to a small cut in S/N to lost light, this introduces deviations to the expected LSF of the spectrum. Accounting for this effect will be important for abundance recovery to approach the forecasted precision and avoid systematics caused by variations in the LSF. Efforts to calibrate NIRSpec early in the lifetime of JWST should help to mitigate this issue.

5.2. Extremely Large Telescopes

The advent of extremely large telescopes (ELTs) with apertures in excess of 30 meters have the potential to revolutionize extragalactic archaeology. Their higher angular resolution and increased light-collecting power will enable the spectroscopic observation of resolved stars in some of the most distant and compact systems in and around the Local Group. The Thirty Meter Telescope (TMT; 30 meter aperture), the European-Extremely Large Telescope (E-ELT; 39 meter aperture), and the Giant Magellan Telescope (GMT; 24.5 meter aperture) all have plans for a highly multiplexed spectrographs—TMT/WFOS, E-ELT/MOSAIC, GMT/GMACS, and GMT/G-CLEF.

5.2.1. Low-resolution ELT MOS

We first consider the three low-resolution spectrographs, WFOS, MOSAIC, and GMACS, which all enable observations of 100+ stars across the full optical spectrum at resolving powers between $R \sim 1000$ and $R \sim 5000$. The configurations we consider are listed in Table 2. As in Section 4.2, we assume a 1 hr observation of our fiducial $\log Z = -1.5$ RGB star with $m_V = 19.5$ and the ETC configurations in Table 4.

Figure 13 presents the CRLBs for these ELT spectrographs. We predict that all three optical ELT spectrographs are capable of measuring 30–40 elemental abundances to better than 0.3 dex. In addition to all Fe-peak elements and most α elements, this includes 22 neutron-capture elements spanning all three r- and s-process peaks. Of these, 12, 9, and 8 can be recovered to better than 0.1 dex by GMACS (G3), MOSAIC (HMM-VIS), and WFOS (B2479/R1392), respectively.

Many of these elements have only weak features below 4000 Å, which necessitate high S/N in the blue optical and near-UV for their recovery. Tb and Tm, for example, have ~ 20 absorption lines with 1%–3% dex^{-1} gradients at $R \sim 3500$, but nearly all are found at wavelengths shorter than 4000 Å. Similarly, Pd, Os, and Hf have fewer than 10 absorption lines of similar strengths, which are also predominantly located blueward of 4000 Å. The strongest line of Th is at $\lambda 4019$ with a gradient of $\sim 1.5\% \text{ dex}^{-1}$, while ~ 20 weaker (0.5%–1.0% dex^{-1}) features exist between 3100 and 4000 Å. Despite the limited chemical information, spectrographs on ELTs are capable of measuring these elements because their large-aperture telescopes and blue wavelength coverage can achieve $S/N \sim 100$ at 4000 Å.

The informative power of blue-optical spectroscopy can be further seen in the comparatively poorer abundance recovery of MOSAIC’s HMM-Vis and HMM-NIR settings. Because the optical arm only extends to 4500 Å, it cannot capitalize on the information-rich, near-UV stellar spectrum. The NIR is expected to recover even fewer abundances than the optical arm due to the lower information density beyond 8000 Å. Nevertheless there are some elements (e.g., Ca, Si, Sr, O, Al,

and S) whose absorption features are better observed in the NIR. CN absorption in the red and NIR also allow for recovery of C and N to a similar degree to what can be done with spectra down to 4500 Å. We note, however, that because the JWST NIRSpec ETC was repurposed to provide S/N in the NIR for MOSAIC, the S/N used here does not include the effects of troublesome NIR telluric features. As a result, we expect the abundance precision of MOSAIC’s HMM-NIR spectra to be noticeably worse in practice.

Figure 13 (top) illustrates the trade-offs in S/N, wavelength coverage, and resolution at a fixed number of detector pixels for three different GMACS gratings. As predicted by Ting et al. (2017a), the abundance precision of a detector with fixed pixel real estate under the assumption of uniform distribution of chemical information is relatively invariant of the resolving power. Of course, there are slight differences in the expected precision of the gratings. For many elements, G2 ($R = 1000$) performs more poorly than the higher resolution gratings, which is likely due to strongly blended lines at $R = 1000$ and the resulting increased covariance between elements. It is also apparent that the chemical information is not uniformly distributed; there are several abundances (e.g., Cr, C, Ba, Al, Dy, Gd, and K) which the G4 grating recovers noticeably worse if not at all because the absorption features of these elements lie outside of its reduced wavelength coverage. These elements are predominantly those with few strong features that lie below 4200 Å. Similar conclusions can be drawn from a comparison of the three WFOS grating combinations.

5.2.2. High-resolution ELT MOS

Here, we consider G-CLEF, a GMT first-light, fiber-fed echelle spectrograph. While it is primarily optimized for very high-resolution ($R \sim 100,000$) single-slit spectroscopy across the optical, it will also feature a MOS mode that will combine modest multiplexing, Keck/HIRES-like spectra, and a 24.5 meter aperture telescope that will dramatically increase the feasibility of high-resolution spectroscopy of stars beyond the immediate vicinity of the Local Group (see Tables 2 and 3 for details). We calculate the S/N using the G-CLEF ETC given the same observational conditions used for the forecasting of existing high-resolution instruments (see Table 4).

Figure 14 shows the CRLBs of G-CLEF with the HIRES 1 $''$ 0 CRLBs for comparison. We forecast that G-CLEF observations will recover 30 elements to better than 0.1 dex (and nearly 40–0.3 dex) similar to HIRES and the other single-slit high-resolution spectrographs analyzed previously in Section 4.3.1. In addition to achieving HIRES-like abundance recovery, G-CLEF’s multiplexing enables the simultaneous observation of up to 40 stars at a time. This dramatically increasing the feasibility of high-resolution studies of substantial numbers of stars in extragalactic systems (for both chemistry and kinematics).

The reason G-CLEF does not achieve substantially better abundance precision than its 10 meter class analogs appears to be largely a consequence of G-CLEF’s lower predicted throughput. Despite having a much larger light-collecting power, G-CLEF acquires roughly the same S/N as Keck/HIRES at wavelengths shorter than 6000 Å where most of the chemical information resides. G-CLEF achieves higher S/N ($\sim 35 \text{ pixel}^{-1}$ compared to $\sim 20 \text{ pixel}^{-1}$) at longer wavelengths, but this only yields small improvements in abundance precision. Furthermore, G-CLEF’s

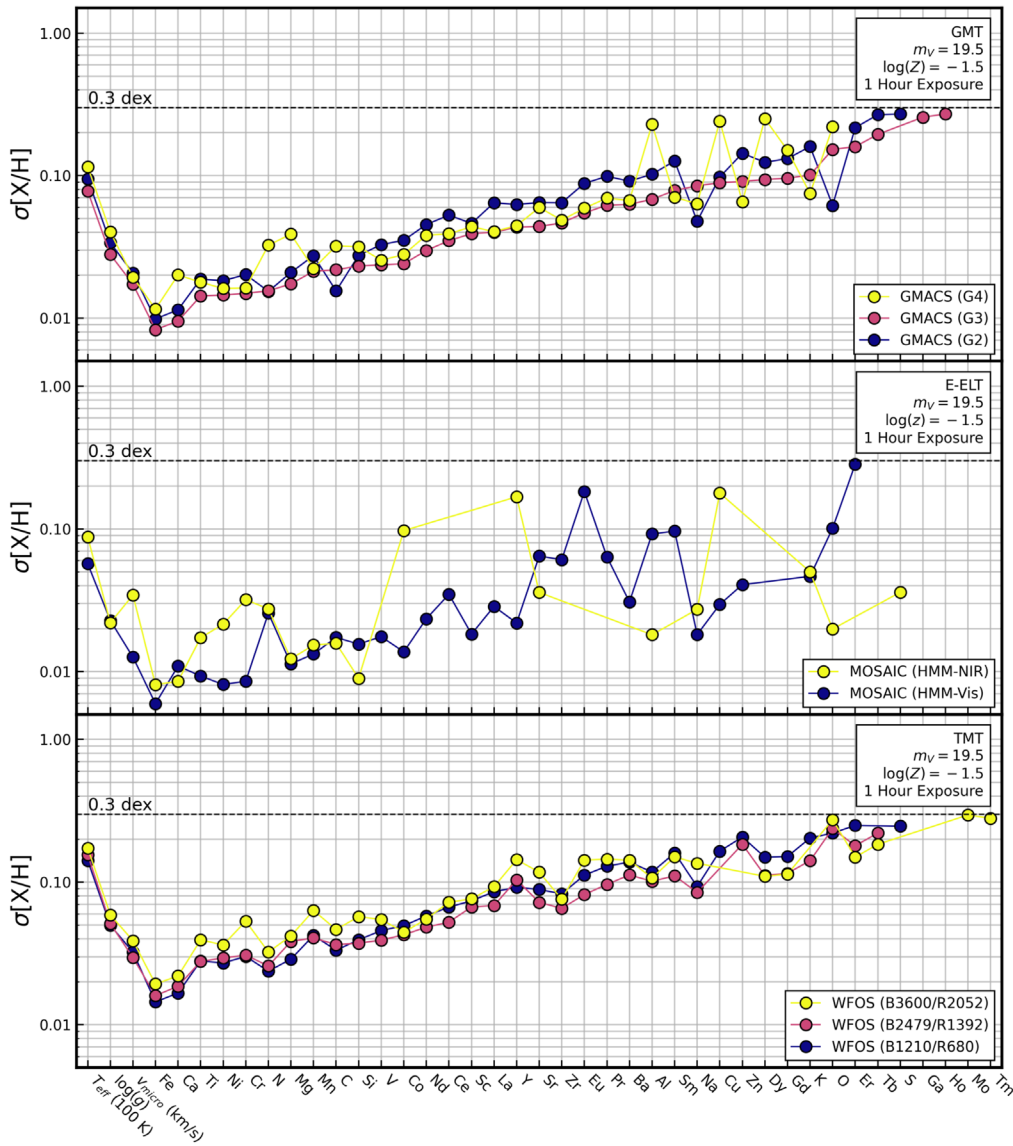


Figure 13. Same as Figure 8 but for the low-resolution ELT spectrographs GMT/GMACS, E-ELT/MOSAIC, and TMT/WFOS.

bluer wavelength coverage is at $S/N < 5 \text{ pixel}^{-1}$ and thus provides little additional information.

5.3. Spectroscopic Surveys

Galactic archaeology in the MW has been revolutionized by several large-scale spectroscopic surveys (e.g., RAVE, Steinmetz et al. 2006; SEGUE, Yanny et al. 2009; LAMOST, Luo et al. 2015; GALAH, de Silva et al. 2015; APOGEE, Majewski et al. 2017; and DESI,³⁴ DESI Collaboration et al. 2016b). These surveys have collected millions of stellar spectra from which detailed abundance patterns have been measured. The success of these surveys in the realm of stellar abundance measurements is in part due to the high quality and homogeneity of the spectra collected. This has allowed for rigorous, self-consistent analyses, the implementation of data-driven approaches, and the refining of stellar models. However, similarly ambitious

observing campaigns outside the MW are in their early stages, primarily because it requires a dedicated survey instrument on a 10 meter class telescope.

The next decades is poised to bring the field of extragalactic stellar spectroscopy its first large sample of homogeneously collected stellar spectra. For example, the Prime Focus Spectrograph (PFS) on Subaru will begin science observations in early 2020. PFS will dedicate ~ 100 nights to surveying M31’s disk and halo, making it the largest extragalactic stellar spectroscopic survey to date (Tamura et al. 2018).

The MSE will replace the Canada–France–Hawaii Telescope with an 11.25 meter dedicated survey telescope, while FOBOS is a next-generation instrument proposed for the Keck telescopes with time dedicated for a stellar (extra)galactic archaeology survey. Both MSE and FOBOS are much earlier in their conceptual design and plan to be on sky by ~ 2030 (Bundy et al. 2019; MSE Science Team et al. 2019).

The details for these spectrographs can be found in Tables 2 and 3. For all three survey instruments we consider our standard 1 hr of integration time of our fiducial $\log Z = -1.5$, $m_V = 19.5$ RGB star and the ETC configurations in Table 4.

³⁴ In Appendices D and F, we forecast the precision of the ongoing LAMOST MW survey and the recently begun DESI survey of MW halo stars. For forecasted precision of other MW surveys we refer the reader to Ting et al. (2017a).

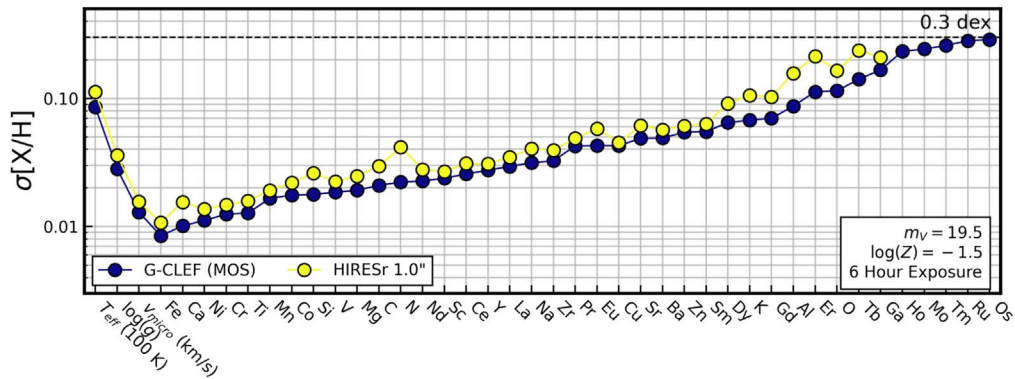


Figure 14. Same as Figure 10 but for high-resolution ELT spectrograph GMT/G-CLEF.

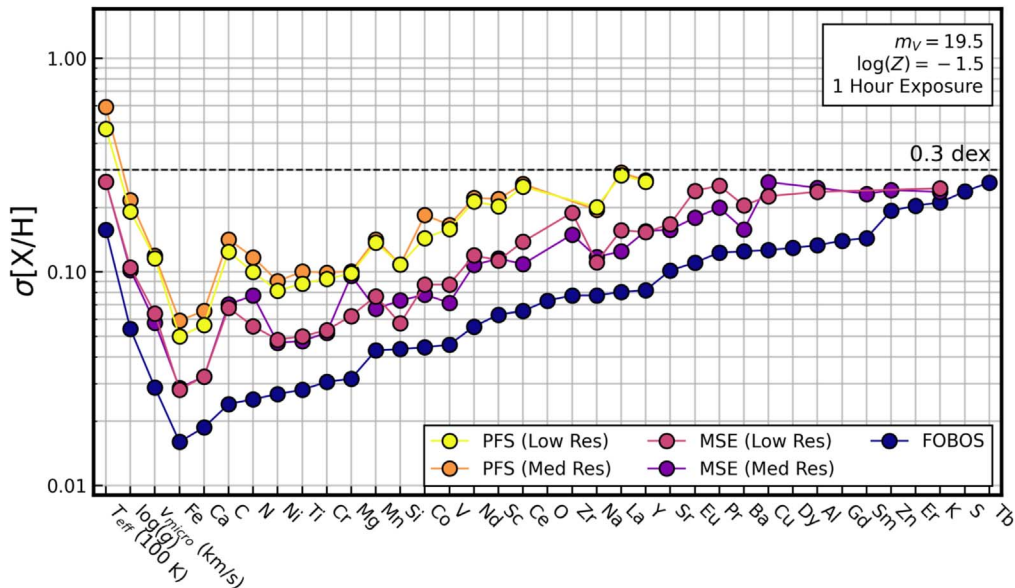


Figure 15. Same as Figure 8 but for the survey instruments PFS, MSE, and FOBOS.

We present the abundance precisions of PFS, MSE, and FOBOS for this observing scenario in Figure 15. All three spectrographs are capable of similar chemical abundance precision to blue-optimized spectrographs considered in Sections 4.2.1 and 4.2.2 (e.g., DEIMOS 1200B, LRIS, and MODS), recovering >20 elements to better than 0.3 dex. As seen in previous analyses, there are only minor differences between the low- and medium-resolution setting on PFS and MSE. The increase in resolution is roughly canceled out by decreases in S/N and wavelength coverage. In this comparison, the additional wavelength coverage beyond $1 \mu\text{m}$ by the NIR and red arms of PFS and MSE (low-res) provide improved precision of Si and Al, but not C, N, and O, which would require even redder spectra that extend past $1.4 \mu\text{m}$.

Despite the relatively similar specifications of these three survey spectrographs, there is a considerable spread in their forecasted abundance precision. This can be attributed to two predominant factors. The first and most important factor is the S/N of the observations. Throughout most of the optical, PFS achieves an S/N only one-half to three-quarters that of FOBOS and MSE. In addition, FOBOS's blue sensitivity enables an $S/N > 10 \text{ pixel}^{-1}$ down to 3500 \AA for these observations, while the S/N of MSE and PFS drops below an S/N of 10 pixel^{-1} at $\sim 4000 \text{ \AA}$.

The second factor contribution to the higher precision predicted for FOBOS is its higher wavelength sampling (6 pixels/FWHM), which is nearly twice that of MSE and PFS. Even holding all other instrument specifications constant (e.g., wavelength coverage, resolving power, S/N), the higher sampling alone leads to a $\sqrt{2}$ improvement in the forecasted precision. Of course, oversampling the spectrum by this degree in practice would likely lead to increased correlations between adjacent pixels, resulting in a smaller improvement than our naïve scaling with $n^{-1/2}$ predicts (see Appendix C).

6. Discussion

6.1. Information-rich Blue Spectra

In the context of extragalactic spectroscopy (i.e., at medium and low resolution), a key result of this paper is the importance of the blue-optical spectrum for measuring abundances. Spectral regions bluer than $\sim 4500 \text{ \AA}$ are rich in absorption features of α elements and r- and s-process elements and overall enable the recovery of more than double the number of elements than red-optical-only wavelengths. This finding echoes the power of low-resolution, blue-optical spectra highlighted in Ting et al. (2017a) and demonstrated by Xiang et al. (2019) with LAMOST spectra.

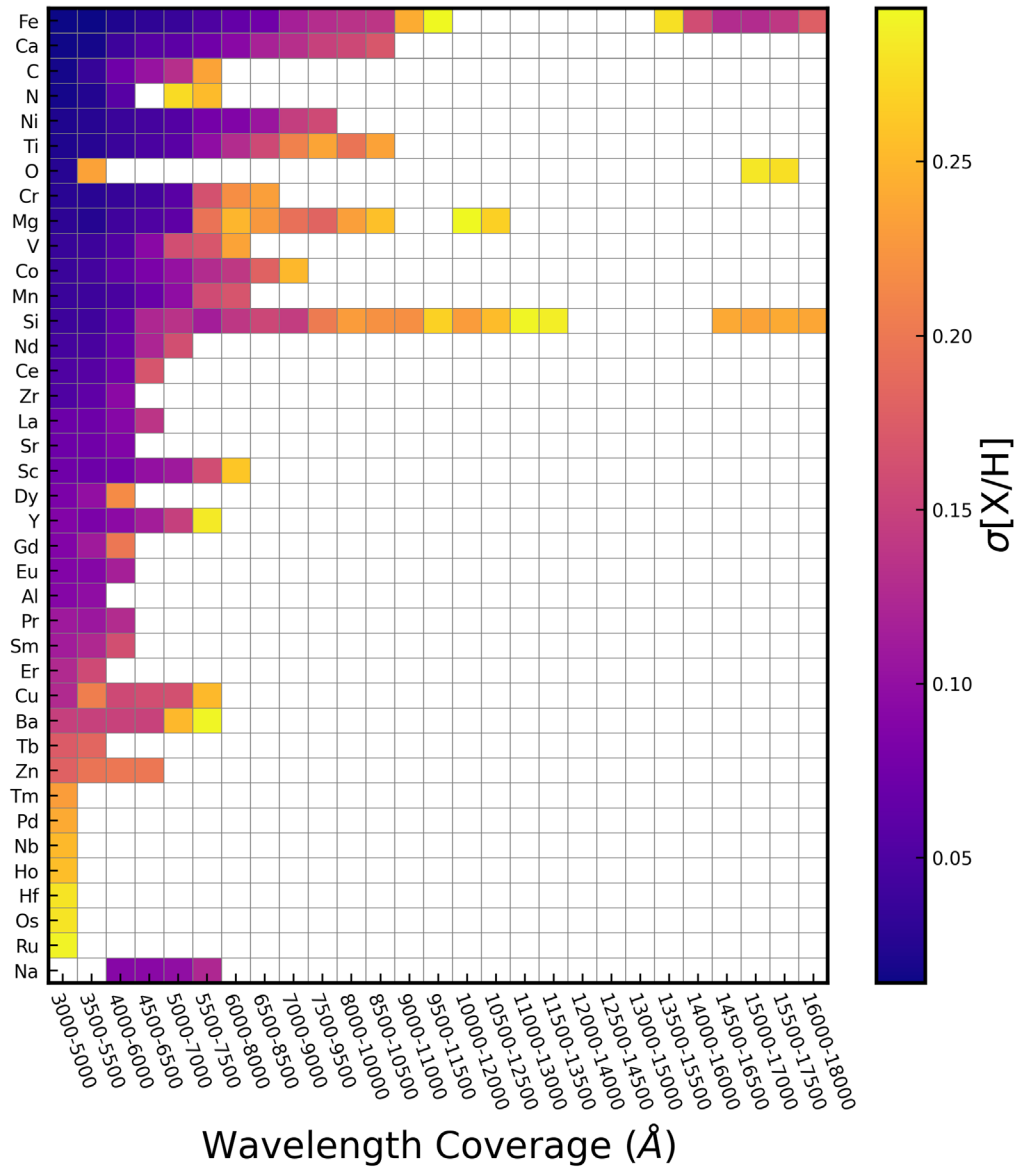


Figure 16. CRLBs for a $\log Z = -1.5$ RGB star observed in 2000 Å wavelength regions from 3000 Å to 1.8 μm , assuming $R = 2000$, $R_{\text{samp}} = 3$, constant throughput, a K2V stellar SED, and $S/N = 100 \text{ pixel}^{-1}$ at 6000 Å. This figure demonstrates the high density of chemical information found at wavelengths shorter than 4500 Å, especially for many neutron-capture elements.

Figures 16 and 17 summarize the power of blue-optical spectroscopy for abundance recovery. To generate these figures, we have simulated a spectra with $R \sim 2000$ and 5000, respectively, and a spectral sampling of 3 pixels per resolution element for a $\log Z = -1.5$ RGB star. We then computed the CRLB for each element for the 2000 Å wavelength regions shown on the x -axis. We assume a K2V SED, constant throughput with wavelength, and an S/N of 100 pixel^{-1} at 6000 Å ($\sim 40 \text{ pixel}^{-1}$ at 3000 Å; $\sim 55 \text{ pixel}^{-1}$ at 1.5 μm). Each cell is color-coded by the CRLB precision.³⁵

Figures 16 and 17 show that the largest number of elements can be recovered in the spectrum spanning 3000–5000 Å. In this range, 38 (49) elements are recovered to a precision of < 0.3 dex for $R = 2000$ (5000). The number of elements available drops to

28 (34) in the 2000 Å range between 4000 and 6000 Å, indicating the rich information available below 4000 Å.

In the 5000–7000 Å range, 18 (22) elements can be recovered. As the wavelength coverage shifts redder, fewer elements are precisely measurable. At $R = 2000$, no elements, including Fe, can be measured from 2000 Å regions between 1.2 and 1.5 μm . This is because there are few absorption features for any elements—Fe with only ~ 20 lines with gradients larger than $1\% \text{ dex}^{-1}$ has the strongest of any element in this portion of the spectrum. The paucity of lines means there is little information to break the degeneracy between the poorly constrained T_{eff} and $\log g$ ($\sigma_{T_{\text{eff}}} > 300 \text{ K}$ and $\sigma_{\log g} > 1.5 \text{ dex}$) and the elemental abundances. Applying the same priors as in Section 4.1.4 enables the recovery of Fe, Si, and Mn to better than 0.3 dex. As the wavelength coverage moves farther into the near-IR (1.5–1.8 μm), the number of elements that can be recovered increases as a result of molecular features (e.g., H_2O and CO) and larger numbers of Fe, Si, Mg, and Al lines (see

³⁵ To first order, the precision of a given element from a combination of two or more wavelength windows can be found by taking the inverse square sum of the abundance’s precision in the relevant wavelength ranges.

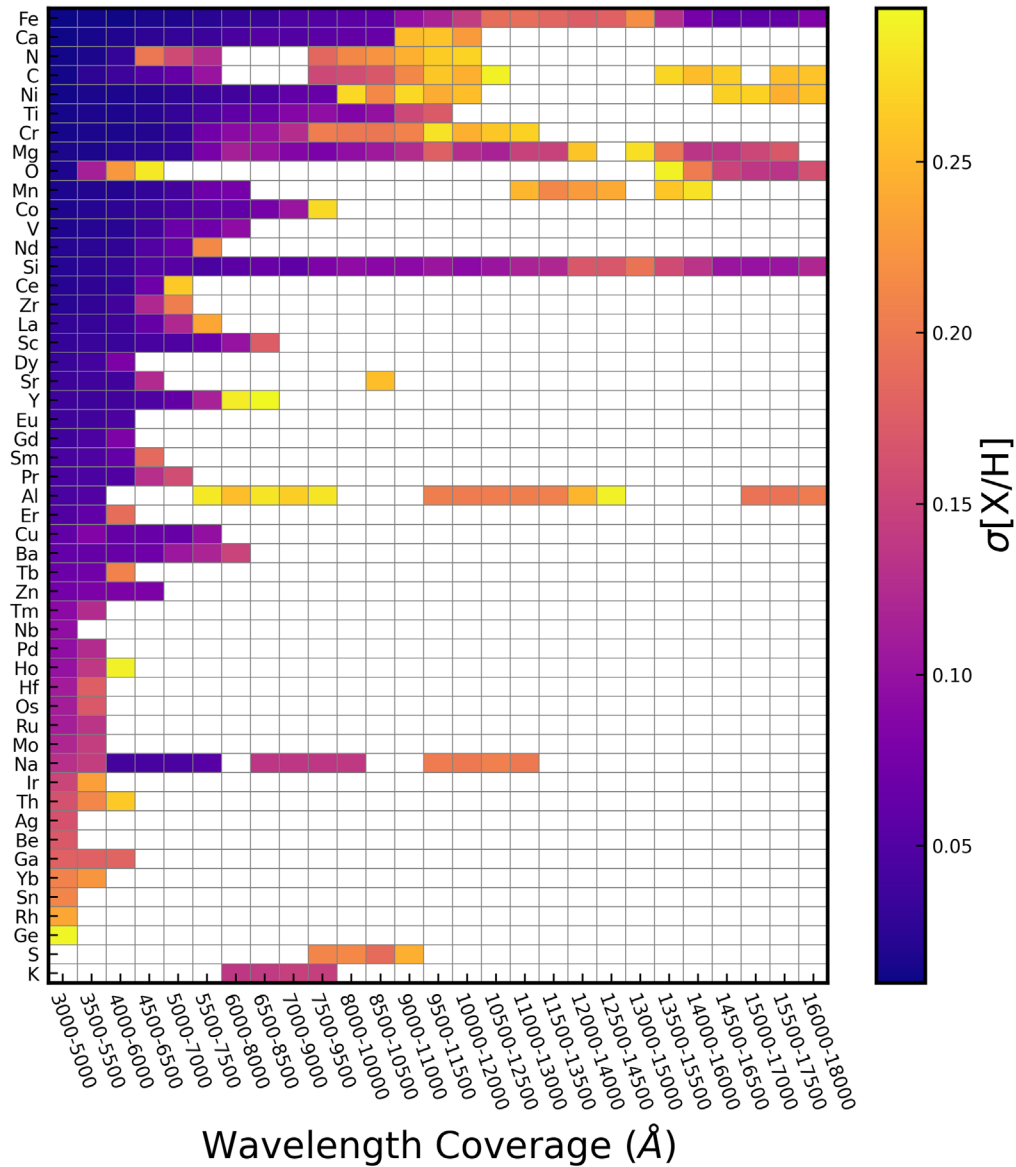


Figure 17. Same as Figure 16, except for $R = 5000$.

APOGEE results: Ness et al. 2015; García Pérez et al. 2016; Ting et al. 2019).

Beyond increasing the number of elements that can be recovered, the blue optical is rich in the absorption lines of neutron-capture elements. For this reason, the blue-optical portion of the spectrum has long been targeted by high-resolution spectroscopy (e.g., Sneden & Parthasarathy 1983; Cowan et al. 2002; Sneden et al. 2003; Hansen et al. 2015).

However, as shown in Figure 18, these elements have strong gradients even at low resolution ($R \sim 2000$). Sr and Eu, for example, have a handful of absorption lines between 3500 and 4500 Å with gradients of 4%–8% dex⁻¹. Other elements, like Zr, Ce, and Nd, have a forest of weaker ($\sim 2\%$ dex⁻¹) absorption lines that extend blueward of 4500 Å. The results of Figures 16–18 together indicate that full spectral fitting methods have the potential to recover neutron-capture elements outside the immediate vicinity of the MW.

The high information density of the blue optical also introduces challenges to abundance recovery. For example,

the large number of lines makes it challenging to define a continuum. Most spectral fitting routines operate on normalized spectra and the lack of a clearly defined continuum introduces additional sources of uncertainties into the fitting process.

A second challenge is the blending of absorption lines. The blending of spectral features is not inherently a problem for full spectral fitting, provided that all stellar labels are fit simultaneously to account for degeneracies. However, doing so requires a high degree of trust in the stellar atmosphere models, radiative transfer treatment, and line lists. When lines are resolved, individual lines that are imperfectly modeled (e.g., from non-LTE or 3D effects) can be isolated and ignored. But when lines are severely blended as they are in the blue optical, identifying and masking (or calibrating) problematic lines become a nontrivial, but crucial, endeavor.

Finally, blue-optical spectra will typically have lower S/N than redder observations of the cool RGB stars we are considering—their flux peaks at ~ 6100 Å. To achieve the same S/N at 3000 Å as at 6100 Å requires at least 50% longer

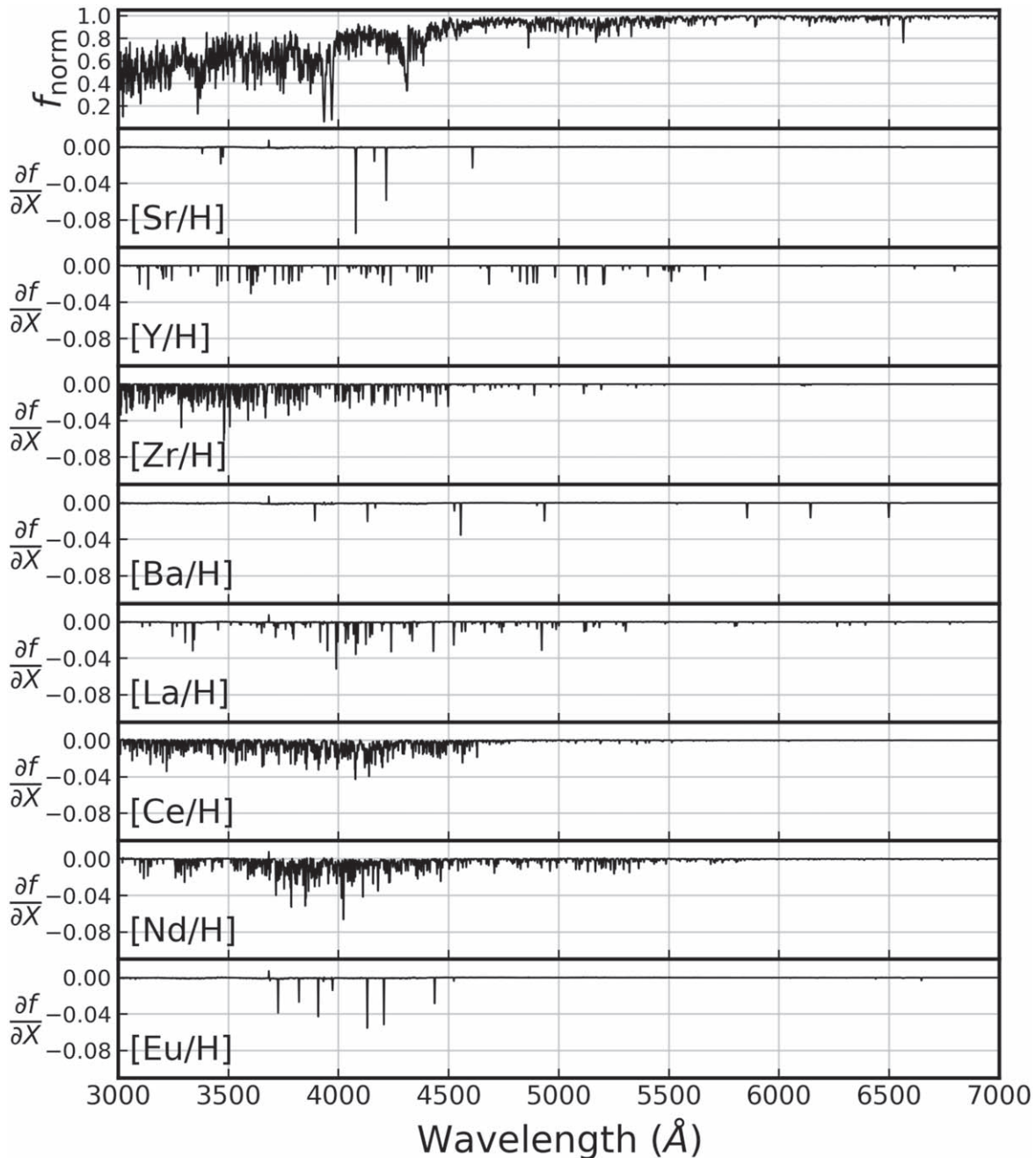


Figure 18. (Top) Spectrum of a $\log Z = -1.5$ RGB star convolved down to $R = 2000$. (Below) Gradients of the spectrum with respect to r-/s-process elements recoverable by LBT/MODS given the setup in Section 4.2.2. Most of the information for these elements is at wavelengths shorter than 4500 \AA . Not shown in this figure are three modest Sr lines with gradients of $1\% \text{ dex}^{-1}$ between 1.0 and $1.1 \mu\text{m}$ and a handful of weak Y lines (all with gradients of $<0.5\% \text{ dex}^{-1}$) that lie redward of 7000 \AA .

integration times in the blue.³⁶ We have attempted to take this into account by using ETCs with SEDs of cool stars to determine realistic S/Ns of our the observing scenarios.

Taken together, the challenges of dealing with line blending and lower S/N has meant that medium- and low-resolution blue-optical spectroscopy has seldom been used for extragalactic stellar chemical abundance measurements.

These difficulties, however, do not invalidate the enormity of the information content contained in the near-UV and blue portions of a star's spectrum. Given the current designs of

upcoming instruments and surveys, we will soon be awash in low-resolution blue stellar spectroscopy and the potential for major advances in abundance determinations. Fully taking advantage of this data set will not be trivial and will take significant investments in stellar models, instrumental calibrations, and spectral fitting techniques, but we believe that it will be well worth the investment.

6.2. Stellar Chemistry beyond 1 Mpc

At present, a full night (~ 6 hr) of observing time on a 10 meter telescope is necessary to measure $[\text{Fe}/\text{H}]$, $[\alpha/\text{Fe}]$, and a few individual elemental abundances in stars as faint as

³⁶ Assuming a constant throughput and a K2V stellar SED.

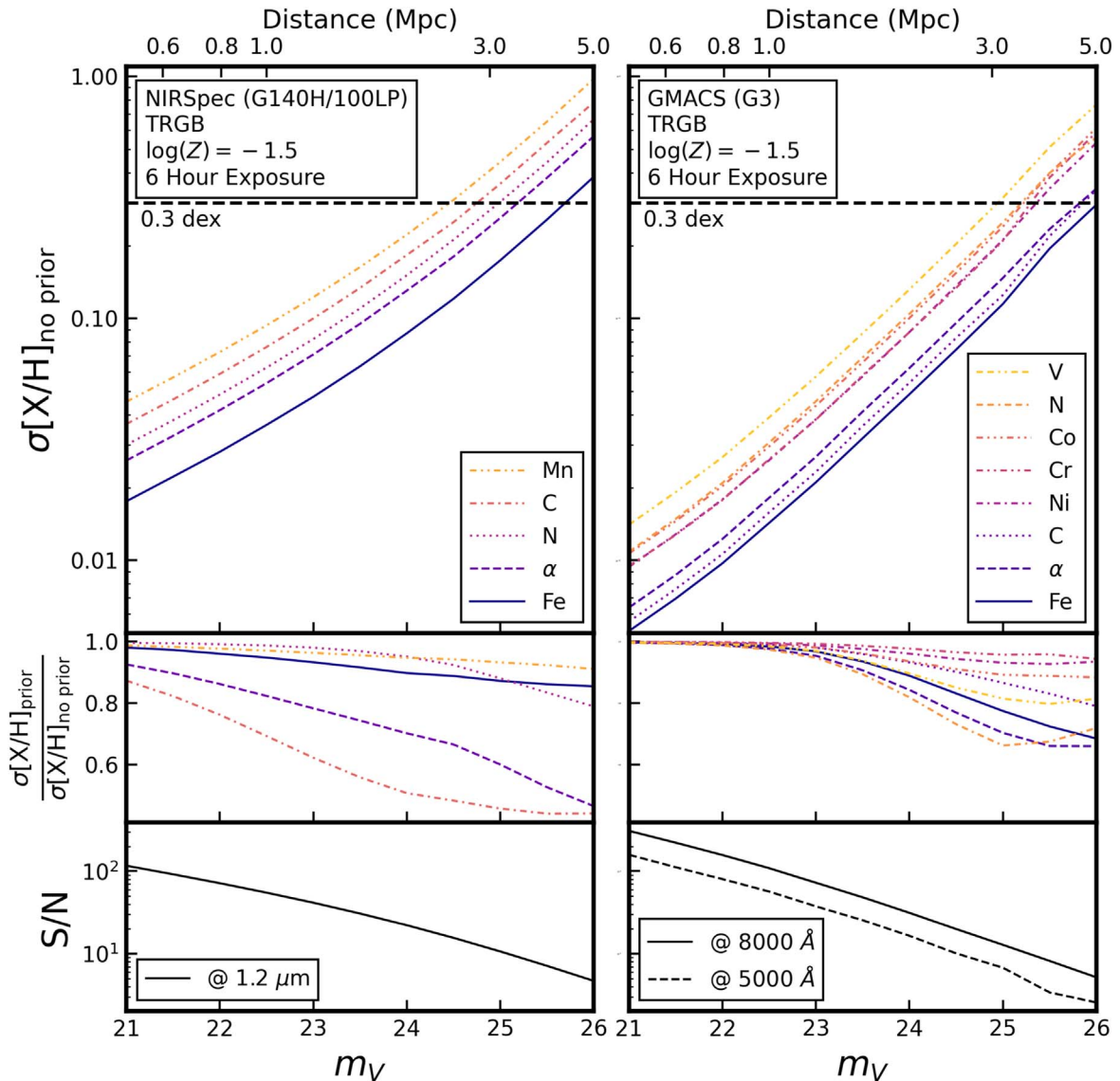


Figure 19. CRLBs for the JWST/NIRSpec G140H/100LP (left) and the GMT/GMACS G3 (right) setups given a 6 hr observation of a $\log Z = -1.5$ TRGB star as a function of apparent magnitude and distance. The middle panels show how the CRLBs improve when assuming Gaussian priors of $\sigma_{T_{\text{eff,prior}}} = 100$ K, $\sigma_{\log g, \text{prior}} = 0.15$ dex, and $\sigma_{v_{\text{micro,prior}}} = 0.25$ km s $^{-1}$. The S/N at a characteristic wavelength is plotted in the bottom panels for each instrument. Small wiggles in the G3 S/N at 5000 Å (and CRLBs) are due to interpolation errors in the extraction of data from the GMACS ETC at low S/N. JWST and ELTs will enable the recovery of Fe and α to better than 0.3 dex beyond 4 Mpc and out to ~ 3 Mpc for a handful of other elements.

$m_V \sim 23$ (e.g., Vargas et al. 2014a, 2014b; Escala et al. 2019a, 2019b; Gilbert et al. 2019; Kirby et al. 2020). While this enables the measurement of stellar metallicities in the halo of M31 with current facilities, measuring elemental abundances in systems at greater distances and stellar densities is currently out of reach, due to long integration times, read-noise limitations, and crowding. Outside the Local Group, stellar spectroscopy is not possible for resolved stars.

However, both JWST/NIRSpec and the ELT spectrographs will excel in the observation of faint stars in crowded systems. They provide Hubble-like angular resolution ($\lesssim 0''.2$) for spectroscopy, can achieve reasonable S/N for faint stars in modest integration times, and are sensitive to the spectral features of many elements (see Sections 5.1 and 5.2.1).

Figure 19 illustrates the potential of JWST and the ELTs for resolved star spectroscopy in and beyond the Local Group. Here we plot the CRLB for several elements as a function of

distance for two telescope configurations: JWST/NIRSpec (G140H/100LP) and GMT/GMACS (G3; see Table 2). For these calculations, we assume 6 hr observations of a $\log Z = -1.5$ TRGB star (see Table 1) and replace the CRLBs of individual α elements (O, Ne, Mg, Si, S, Ar, Ca, and Ti) with a CRLB for $[\alpha/H]$.³⁷ The CRLBs indicate that JWST and GMACS will be able to measure the Fe abundance to 0.3 dex in individual stars out to 4.4 and 5.0 Mpc, respectively.³⁸ GMACS is capable of recovering α abundances, primarily through Ca features and to a lesser extent from Ti, Mg, and Si features, out to 4.5 Mpc. For NIRSpec, α is recovered through a combination of Si, O, and Mg features (in order of decreasing importance) out to 3.5 Mpc.

³⁷ The gradients for α were calculated as in Section 3.2 except that offsets were applied to all α -element abundances in lockstep instead of individually.

³⁸ We note that the S/N for both instruments is quite low beyond 4 Mpc: < 10 pixel $^{-1}$ for NIRSpec and < 5 (< 10) pixel $^{-1}$ for GMACS at 5000 (8000) Å.

The small wiggles in the G3 S/N at 5000 Å (and CRLBs) seen beyond 25 Mpc are the result of interpolation errors in the extraction of data from the GMACS ETC at low S/N.

We also calculate the Bayesian CRLB using the same Gaussian priors as in Section 4.1.4 ($\sigma_{T_{\text{eff,prior}}} = 100$ K, $\sigma_{\log g, \text{prior}} = 0.15$ dex, and $\sigma_{v_{\text{micro,prior}}} = 0.25$ km s⁻¹). The middle panel of Figure 19 illustrates that these priors can improve the precision of C and α (N, Fe, and α) by up to a factor of 2 (1.5) for JWST (GMACS) observations of faint stars.

In addition to Fe and α , NIRSpec and GMACS are capable of recovering a handful of other individual abundances at a distance of ~ 3 Mpc—N, C, and Mn for NIRSpec and C, Ni, Cr, Co, N, and V for GMACS. These elements can all be measured to better than 0.2 dex at 2 Mpc and 0.1 dex at 1 Mpc. Other elements not shown that can also be recovered to 0.3 dex out to 1 Mpc include Mn, Nd, Sc, Ce, La, Zr, Y, Pr, Sm, Ba, Na, K, Al, Sr, Eu, Cu, Gd, Zn, and Dy for GMACS and Ni, Al, and Cr for JWST. This would not only enable precise chemical abundance measurements of stars in M31 and its satellites, but also enable detailed chemical enrichment studies of galaxies at the periphery of the Local Group and beyond, including potential new faint galaxy discoveries by LSST.

Though we did not explicitly compute the CRLBs as a function of distance for TMT/WFOS and E-ELT/MOSAIC, we expect that each of these powerful facilities have similar abundance recovery potentials for stars outside the Local Group.

6.3. Planning Observations

For stellar abundance work, selecting the appropriate spectrograph, setup, and exposure time for a specific science case can be daunting given the large number of facilities and instrumental configurations. This can often lead to inefficiencies in observational strategies.

As illustrated in Sections 4 and 5, the CRLB provides a useful and quantitative way to evaluate abundance recovery for a given spectroscopic setup. As an example, consider the comparison of Keck spectrographs and gratings in Figure 8, which displays the numerous trade-offs of each setup on an element-by-element basis. LRIS generally provides the most chemically informative spectra, but if high multiplexing is a priority, the 1200B grating on DEIMOS is likely the better choice. However, if a specific element is of interest (e.g., Ca), one of the lower-resolution DEIMOS gratings might be more valuable than the 1200B grating.

Given the simplicity in its computation, we suggest that CRLBs should be standardized as part of observational planning for resolved star spectroscopic abundance measurements as a logical extension of the standard ETC usage. An ETC determines the S/N of a spectrum based on the integration time and observing conditions, and the CRLB in turn relates that S/N into an expected abundance precision. Figure 3 provides a clear example of how calculating CRLBs for an instrument can inform an observing strategy. If the intended science goals necessitate simply measuring Fe and an α element out beyond 100 kpc, an hour-long exposure with the D1200G grating will likely suffice, allowing for a handful of fields to be observed in a night. However, if the science requires measuring specifically the α element, magnesium, an integration time of three or more hours is necessary per field and a different observing strategy is required.

6.4. Caveats and Assumptions

In this section we discuss in more detail the assumptions adopted in our calculation of CRLBs, namely that (1) the model spectra perfectly reproduce real stellar spectra, (2) the likelihood and noise properties are Gaussian, and (3) that adjacent pixels are uncorrelated. We save a more technical discussion of the CRLB for a biased estimator for Appendix A.

6.4.1. Model Fidelity

Model fidelity is a fundamental assumption inherent in all problems of parameter estimation. The CRLB of stellar spectra is no exception to this as the gradient spectra used in the above calculations are strongly dependent on the physical assumptions and spectral line lists that underpin any spectral synthesis model. It is important to keep in mind that the CRLB makes no claims about the accuracy of stellar label measurements, merely the possible precision. Nevertheless, incomplete or incorrect line lists will leave out or misplace spectral information, while models that assume 1D atmospheres in local thermodynamic equilibrium (LTE) may incorrectly predict the spectral response to varying stellar labels for non-LTE lines. It is thus important to strive for consistency and consider the CRLBs calculated using the models relevant to the spectral fitting that will be conducted. While comparing CRLBs of different models is a valuable exercise to evaluate systematics in the predicted CRLBs, this should not be done to pass judgment on model quality.

A common practice in full-spectrum fitting is the masking of spectral regions that are known to be poorly fit by the spectral model to avoid introducing potential systematics into the analysis. Often the poor fit is due to non-LTE effects, but may also be the result of 3D effects, poorly calibrated oscillator strengths, or an incomplete (or incorrect) line list (see Nissen & Gustafsson 2018 and references therein). When these regions are masked, so too is the information that it holds. In such a case the appropriate CRLB should be calculated with gradient spectra masked in the same regions (as we do in Section 4.1.1), resulting in a higher uncertainty for the stellar labels. We note, however, that because information adds in quadrature, masking 90% of the lines only worsens the CRLB by a factor of ~ 3 . For a more thorough analysis of the CRLB's dependence on masked regions, see Ting et al. (2017a).

Another underlying challenge for our CRLBs is the assumption that the continuum can be perfectly determined. In the red-optical and near-infrared region of the spectrum, lines are sufficiently sparse that even at $R \sim 2000$ identifying the continuum and dividing it out is routine. Unfortunately, the many absorption features in the blue optical and UV, make it challenging to define a stellar continuum. Instead, a pseudo-continuum is defined using a polynomial function (or some smoothing kernel) and divided out, potentially introducing systematics or additional uncertainty in the normalized flux that will worsen the precision. By similarly normalizing the model spectra (instead of using the true continuum), any systematics introduced through imperfect normalization can be minimized.

Knowledge of the instrumental LSF is necessary to fit observed spectra with model spectra at the same resolving power. In this work, we have assumed a constant LSF. However, in practice, the LSF is not always known to great precision and can vary from object to object depending on where in the field of view the star lies. Use of the wrong LSF is

thus another means by which systematics may be introduced into the fitting of stellar labels. Ting et al. (2017a) showed that at least at moderate resolution ($R \sim 6000$) and high S/N (>200), mismatched LSFs only bias stellar label recovery for differences in broadening greater than 10 km s^{-1} and is unlikely to affect the measurement precision. Spectral fitting at lower resolving powers should be even less sensitive to mismatches in LSF.

In addition, when using rest-frame synthetic spectra, it is necessary to properly determine and correct for the radial velocity of stars. As with the continuum normalization and LSF, we have not quantified the uncertainty in stellar labels that is introduced when the radial velocity is fit simultaneously with other stellar labels. We expect any changes in the CRLBs to be small given that radial velocity is unlikely to correlate with other stellar labels. We will pursue this analysis in a future study.

Even with perfect spectral models, continuum normalization, and instrument characterization, fully extracting the chemical information content of a spectrum requires fitting the full wavelength range (as opposed to measuring EWs) for all stellar labels simultaneously. This is particularly important at low and moderate resolution to account for the degeneracies between labels introduced by blended spectral features. In practice, this can be computationally challenging owing to the high dimensionality of the stellar label space and the large runtimes needed to generate even 1D LTE stellar atmospheres.

Despite these challenges, the future of extragalactic stellar spectroscopy looks bright as steady progress is being made in all of the aforementioned areas. Attempts to incorporate non-LTE and 3D effects into stellar atmosphere and radiative transfer models have been undertaken by a number of groups (e.g., Caffau et al. 2011; Bergemann et al. 2012; Amarsi et al. 2016). Several groups have committed to further refining line lists through the identification of unknown (or misplaced) lines in stellar spectra (e.g., Shetrone et al. 2015; Andreassen et al. 2016) and the improved calibration of transition oscillator strengths (e.g., Pickering et al. 2001; Aldenius et al. 2007; Pehlivan Rhodin et al. 2017; Laverick et al. 2018). Lastly, full spectrum-fitting techniques have made major strides with spectral “emulators” trained through data-driven (e.g., the Cannon; Ness et al. 2015), ab initio (e.g., the Payne; Ting et al. 2019), or combined (e.g., The DD-Payne; Xiang et al. 2019) methods, which bypass the computationally expensive stellar atmosphere and radiative transfer calculations.

The above challenges to achieving the precision predicted by the CRLBs should not dissuade from the use of CRLBs. Instead, the precision forecasted by the CRLBs provide strong motivation for the continued efforts toward understanding stars, their atmospheres, and their spectra.

6.4.2. Assumptions of Gaussian Posteriors

Implicit in the derivations of Equations (1), (2), and (5) was that of Gaussian likelihoods and uncertainties. When these conditions are not met, the CRLB will inaccurately predict measurement errors and degeneracies between stellar labels. In such situations, a more accurate estimate of the achievable precision can be found using Bayesian sampling techniques. A comparison of the CRLB and the precision predicted by HMC sampling in the low-S/N limit is performed in Appendix E, and we find it robust down to an S/N of 5 in the case of D1200G (assuming a constant S/N with wavelength).

6.4.3. Pixel-to-pixel Correlation

Throughout this study we simplify our analysis by setting the correlation between adjacent pixels to zero when calculating the CRLBs.³⁹ In practice, however, most spectrographs are designed to oversample their spectra such that the number of pixels per resolution element is larger than the Nyquist sampling (~ 2 pixels/FWHM).⁴⁰ As a result, adjacent pixels will show some correlation and not be truly independent as we have assumed.

While this is unlikely to make a large difference for most spectrographs, which only slightly oversample their spectra (3–4 pixels/FWHM), the pixel-to-pixel correlation of spectrographs that more highly oversample (e.g., Hectospec, FLAMES-UVES, FOBOS, and some DEIMOS and LRIS gratings) may be nonnegligible in practice. If instead we believe that only 2 pixels per resolution element are informative then the CRLBs should be a factor of $\sqrt{2}$ ($\sqrt{3}$) larger than presented for spectrographs with a sampling of 4 (6) pixels/FWHM as the CRLBs scale as $n^{-1/2}$. More realistically, additional sampling beyond the Nyquist limit will yield pixels that are still informative, just less so than wholly independent pixels. Thus, we expect the increase in the CRLB to be considerably less than a factor of $\sqrt{2}$ ($\sqrt{3}$) when the correlation of adjacent pixels are taken into account. In Appendix C, we present an illustrative example of the impact of wavelength sampling and pixel-to-pixel correlation on the CRLBs.

7. Chem-I-Calc

Forecasting stellar label recovery for spectroscopic observations is crucial to planning realistic observational campaigns and for validating the reported precision of spectral fitting analyses. However, there are far more combinations of instruments, observational conditions, and stellar targets than can be presented in a single paper. To make the calculation of stellar CRLBs convenient to the astronomical community, we have developed the open-source Python package, Chem-I-Calc—the Chemical Information Calculator (Sandford 2020).⁴¹

The Chem-I-Calc Python package provides all the tools necessary to perform all of the computational work presented in this paper, excluding the generation of high-resolution spectra. All of this paper’s calculations are included in a Jupyter Notebook on the Chem-I-Calc Github repository along with several other helpful tutorials and instructions for downloading the synthetic spectra described in Section 3. The code base is designed to be easy to modify for users that need more flexibility in their CRLB calculations (e.g., for incorporating wavelength-dependent resolution, alternative stellar models, or masking of specific wavelength regions).

While Chem-I-Calc is ready to be used in its current state, it is still under active development. Over time we expect to add additional commonly used spectrographs as presets and include a larger range of stellar types and metallicities as reference stars. We gratefully welcome community feedback and contributions to the Python package.

³⁹ A similar simplification is employed nearly ubiquitously in the measurement of chemical abundances from stellar spectroscopy.

⁴⁰ For most instrumental LSFs, the Nyquist sampling is somewhat larger than 2 pixels/FWHM (see Robertson 2017).

⁴¹ <https://chem-i-calc.readthedocs.io/>

8. Summary and Conclusions

Current and future generations of powerful, highly multiplexed spectrographs on large-aperture telescopes make accessible an enormous wealth of chemical information in the spectra of stars outside the MW. Already these instruments have observed the spectra of tens of thousands of individual stars in extragalactic systems, enabling the measurement of their abundance patterns (e.g., Suda et al. 2017 and references therein). With the advent of large-scale extragalactic spectroscopic surveys and ELTs, the number of stars outside the MW with observed spectra will increase by at least an order of magnitude (Takada et al. 2014; Bundy et al. 2019; MSE Science Team et al. 2019).

The majority of these spectra will be acquired at low and moderate resolution ($R < 10,000$) and feature heavy blending of spectral lines, necessitating the entire spectrum be fit for all stellar labels simultaneously. Recently, novel full-spectral fitting techniques (e.g., The Cannon; Ness et al. 2015, The Payne; Ting et al. 2019, and The DD-Payne; Xiang et al. 2019) applied to stellar spectra from MW surveys have proven capable of measuring dozens of elemental abundances from low-resolution spectra.

With the field of extragalactic stellar spectroscopy poised for substantial growth, it is imperative that we understand the chemical information content of the spectra we collect and the precision to which it enables the recovery of elemental abundances. To that end, we have employed CRLBs to quantify the information content of extragalactic stellar spectra and forecast chemical abundance precision for 41 existing, future, and proposed spectrograph configurations on 14 telescopes. Here we summarize our findings.

1. The CRLB is an efficient method for computing the expected precision of stellar labels determined via full spectral fitting. We find that the precision of literature abundances for the commonly used DEIMOS 1200G grating and the LAMOST MW survey are within a factor of 2 of our CRLBs.
2. Low- and moderate-resolution spectroscopy at blue-optical wavelengths ($\lambda \lesssim 4500 \text{ \AA}$) are incredibly information rich, enabling the recovery of two to four times as many elemental abundances as red-optical spectroscopy ($5000 \lesssim \lambda \lesssim 10000 \text{ \AA}$) at similar resolutions. Further, low-resolution, blue-optical spectroscopy is capable of constraining the abundances of several neutron-capture elements (e.g., Sr, Ba, La, Eu).
3. High-resolution ($R \gtrsim 20,000$) spectra contain substantial chemical information even at low S/N ($\sim 10 \text{ pixel}^{-1}$). Maximizing the precision of abundance recovery from high-resolution spectra benefits from full spectral fitting over EW techniques.
4. Even small ($\sim 100\text{--}500 \text{ \AA}$) windows of low S/N, high-resolution spectra can constrain [Fe/H] and a handful of other elements to better than 0.3 dex.
5. JWST/NIRSpec and ELTs can recover 10–30 elements for red giant stars throughout the Local Group and [Fe/H] and $[\alpha/\text{Fe}]$ for resolved stars in galaxies out to several Mpc with 6 hr (~ 1 night) of integration time.
6. Our analysis strictly concerns the precision, not accuracy, of chemical abundance measurements. In practice, imperfect stellar models, line lists, and data reduction can introduce systematics that can bias abundance measurements and

hinder attainment of near-CRLB precision. Further investment in the development of stellar models and spectral analysis is necessary to maximally use the chemical information content of the spectra collected.

7. CRLBs, like ETCs should be used when planning stellar spectroscopic observations or developing spectroscopic instrumentation. To facilitate the calculation of CRLBs, we present Chem-I-Calc, an open-source Python package for calculating CRLBs of arbitrary spectrograph configurations.

We thank the anonymous referee for constructive comments that helped improve the paper. We thank Bob Kurucz for developing and maintaining programs and databases without which this work would not be possible, as well as all those who have contributed to the design and ongoing utility of the various ETCs used throughout this work. Specifically, we thank Brad Holden, Luke Schmidt, Nicolas Flagey, and Kyle Westfall for providing additional information about the Keck, GMACS, MSE, and FOBOS/WFOS ETCs, respectively, as well as Maosheng Xiang for providing example S/N curves of LAMOST spectra. We would also like to thank Hans-Walter Rix, Kim Venn, Alex Ji, Douglas Finkbeiner, Julianne Delcanton, Josh Speagle, and Gregory Green for insightful discussions.

D.R.W. acknowledges support from an Alfred P. Sloan Foundation Fellowship, an Alexander von Humboldt Fellowship, and a Hellman Faculty Fellowship. D.R.W. and N.R.S. are grateful to the Max Planck Institute for Astronomy for their hospitality during the writing of this paper. Support for this work was provided by NASA through grants HST-GO-15901, HST-GO-15902, and JWST-DD-ERS-1334 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555. Y.S.T. is grateful to be supported by the NASA Hubble Fellowship grant HST-HF2-51425.001 awarded by the Space Telescope Science Institute. The computations in this paper were partially run on the Savio computational cluster resource provided by the Berkeley Research Computing Program at the University of California, Berkeley.

Software: iPython (Pérez & Granger 2007), Matplotlib (Hunter 2007), pandas (McKinney 2010; Pandas Development Team 2020), NumPy (Walt et al. 2011), Astropy (Astropy Collaboration et al. 2013, 2018), PyMC3 (Salvatier et al. 2016), PyTorch (Paszke et al. 2019), SciPy (Virtanen et al. 2019), Chem-I-Calc (Sandford 2020).

Appendix A Biased CRLB

A fundamental assumption adopted in this work is that of perfect models that accurately reproduce observed stellar spectra. However, as in most of astrophysics and as we discussed in Section 6.4, this is not the case in practice. Many spectral features are poorly modeled due to 3D and non-LTE effects, miscalibrated oscillator strengths and transition wavelengths, and imperfect reductions. While these systematic errors primarily affect the accuracy of abundance measurements, they also invalidate our assumption that the MLE, $\hat{\theta}$, is an unbiased estimator of the true stellar labels and may also change the expected precision of the abundance measurements.

If the bias of a particular spectral model is known, this can be included in the prediction of stellar label precision using the

“biased” or “misspecified” CRLB:

$$\sigma_{\text{biased},\alpha} = \sqrt{([I + D]F^{-1}[I + D]^T)_{\alpha\alpha}}, \quad (\text{A1})$$

where F is the FIM as defined in Equation (6), I is the identity matrix, and D is the bias gradient matrix:

$$D = \left[\frac{\partial b}{\partial \theta_\alpha} \right], \quad (\text{A2})$$

where b is the bias of your labels given by

$$b(\hat{\theta}) = E(\theta) - \hat{\theta}. \quad (\text{A3})$$

Because evaluating the bias is both model and instrument dependent, it is beyond the scope of this paper. However, we note that in the simple case of a uniform bias (i.e., measuring the surface temperature of all stars to be 100 K too hot), the normal and biased CRLB are the same. In the more complicated (and realistic) case where the bias is dependent on the stellar labels (i.e., the surface temperature is measured to be 100 K too hot in giant stars but 100 K too cold in dwarf stars) the biased CRLB will differ from normal CRLB. Depending on the direction and amplitude of the bias, this may result in either better or worse precision than in the unbiased case.

The main challenge in practice is not that the CRLBs cannot be used in the presence of bias, but that the bias needs to be known a priori for the CRLB—or any forecast of precision—to be computed accurately.

Appendix B CRLB Calculation

For instruments whose observations span noncontiguous wavelength ranges, the gradient spectra (and 1D S/N arrays) for each of the wavelength ranges are concatenated together. This technique can also be used to combine observations from potentially complementary instruments or observing campaigns, though we do not consider any here. All combinations of wavelength ranges examined in this work are forced to be non-overlapping to avoid a more complicated treatment of the

spectral covariance matrix. This is done even though it means ignoring the additional information that an overlapping region of spectrum might provide.

From this point, the calculation of the CRLBs from the gradient spectra and spectral covariance is simply a matter of matrix multiplication and inversion. However, because the gradient spectrum for some labels is much larger than for others (i.e., Fe compared to Nb), the FIM may be nearly singular and thus unstable to inversion. We take several steps to avoid matrix inversion problems and calculate robust CRLBs:

- (i) We divide the spectral gradient with respect to T_{eff} by 100.
- (ii) If $F_{\alpha\alpha} < 1$ for any label, α , we set $F_{\alpha j} = F_{j\alpha} = 0$ and $F_{\alpha\alpha} = 10^{-6}$.
- (iii) We compute the Moore–Penrose pseudo-inverse of the FIM (Moore 1920; Penrose 1955).

The purpose of (i) is to place df/dT_{eff} on roughly the same scale as $df/d[X/H]$. This keeps the eigenvalue of the FIM with respect to T_{eff} from dwarfing those of the other labels. As a result, the CRLB for T_{eff} is in units of 100 K. Step (ii) avoids zero eigenvalues for labels with very little information in the spectrum. It also removes the covariance of these labels with all other labels, which would otherwise make the matrix nearly singular. This results in a CRLB of $\sim 10^3$ for these labels, which can safely be ignored. Finally, by calculating the pseudo-inverse instead of the true inverse of the FIM in (iii), we avoid numerical instabilities when attempting to invert near-singular matrices.

When including prior information into our CRLB calculations, we add the inverse variance of these priors to the relevant diagonal entries of the FIM as outlined in Equation (12) before inverting the FIM as before. To be rigorously Bayesian, we ought to state that we do this for all labels, including those with uninformative priors with zero inverse variance.

Appendix C Wavelength Sampling and Pixel Correlations

To illustrate the impact of assuming the independence of all pixels in the calculation of the CRLB, we consider the

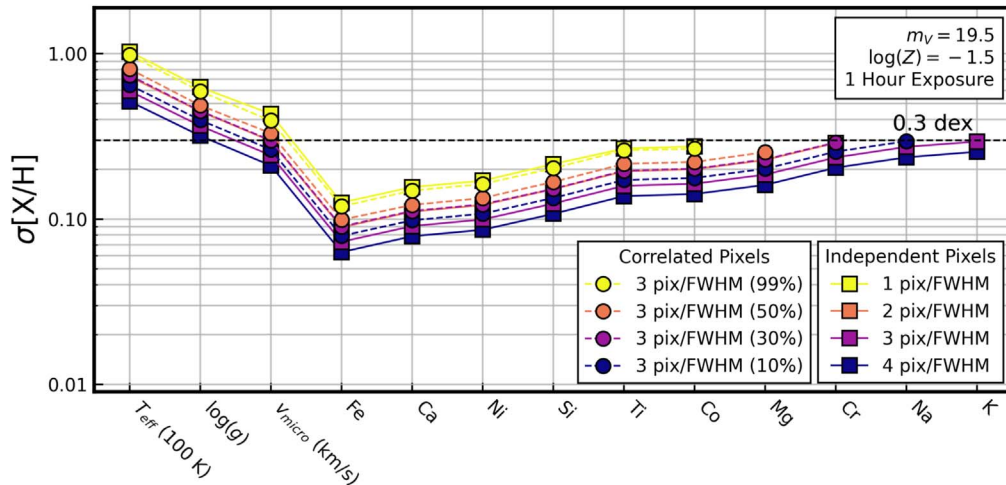


Figure C1. D1200G CRLBs for a 1 hr exposure of a $\log Z = -1.5$, $m_V = 19.5$ RGB star assuming various wavelength samplings and pixel-to-pixel correlations. CRLBs assuming uncorrelated pixels but varying wavelength sampling are represented by squares and solid lines. CRLBs assuming 3 pixels/FWHM but varying degrees of correlation between adjacent pixels are represented by circles and dashed lines. For completely independent pixels, the CRLBs scale proportionally to $n^{-1/2}$, where n is the number of pixels per resolution element.)

simple case where each resolution element is sampled by 3 pixels and all adjacent pixels are correlated by some fraction, c . In such a scenario, the flux covariance is no longer the diagonal matrix presented in Equation (15), but now has diagonal-adjacent terms equal to $c(\sigma)^2$, where $\sigma = (S/N)^{-1}$ at each pixel:

$$\Sigma = \begin{bmatrix} \sigma^2(\lambda_1) & c\sigma^2(\lambda_1) & & & \\ c\sigma^2(\lambda_2) & \sigma^2(\lambda_2) & c\sigma^2(\lambda_2) & & \\ & & \ddots & & \\ & & & c\sigma^2(\lambda_{N-1}) & \sigma^2(\lambda_{N-1}) & c\sigma^2(\lambda_{N-1}) \\ & & & & c\sigma^2(\lambda_N) & \sigma^2(\lambda_N) \end{bmatrix}. \quad (\text{C1})$$

Figure C1 shows the impact of assuming adjacent pixels are 10%, 30%, 50%, and 99% correlated on the CRLB as applied to our fiducial D1200G observation. For comparison, we also include the CRLBs assuming one, two, three, and four completely uncorrelated pixels per resolution element. As expected under the assumption of independent pixels, the CRLBs scale as $n^{-1/2}$, where n is the number of pixels per resolution element.

When adjacent pixels have correlations of 10%, 30%, and 50%, the CRLBs are roughly 8%, 23%, and 35% larger respectively than in the uncorrelated case. These CRLBs are equivalent to calculating the CRLB assuming $n = 2.6$, 2.0, and 1.6 independent pixels per resolution element respectively. In the extreme case that all three pixels are nearly 100% correlated with each other, there is effectively only one independent pixel per resolution element and the CRLB approaches the $n = 1$ pixel/FWHM CRLB or $\sqrt{3}$ times what is found with uncorrelated pixels.

A more realistic treatment of pixel correlation would require adopting a kernel describing the correlation of pixels beyond just the adjacent ones. This, however, requires a deep knowledge of each instrument, which is beyond the scope of this paper.

Appendix D

Comparison with LAMOST DD-Payne Abundances

In Section 4.1.1, we found our CRLBs for D1200G to be in good agreement with the precision reported by Kirby et al. (2018). D1200G observations of metal-poor RGB stars, however, provide only a single point of comparison between our forecasts and what might be expected in practice. Because so few full spectral fitting techniques are currently used in extragalactic contexts, similar comparisons are quite challenging.

Instead, we turn to an example within the Galaxy to provide an additional comparison. Specifically, we compare our CRLBs to the internal precision reported by Xiang et al. (2019) for observations of MW stars by the LAMOST spectrograph (Cui et al. 2012). Xiang et al. (2019) employed the DD-Payne⁴² for full-spectral fitting and used repeat observations to quantify the internal precision of their measurements.

Table D1
Cuts on LAMOST DR5

| |
|---|
| $4600 < T_{\text{eff}} \text{ (K)} < 5000$ |
| $2.3 < \log g < 2.7$ |
| $-0.1 < [\text{Fe}/\text{H}] < 0.1$ |
| $-0.1 < [\alpha/\text{Fe}] < 0.1$ |
| $40 < g\text{-band S/N (pixel}^{-1}\text{)} < 60$ |
| $\chi^2 \text{ Flag} = \text{good}$ |
| $[\text{X}/\text{Fe}] \text{ Flag} = 1$ |

Because LAMOST observed primarily MW stars, we calculate the CRLBs for a typical solar-metallicity K-Giant star ($T_{\text{eff}} = 4800$ K, $\log g = 2.5$, $v_{\text{micro}} = 1.7$ km s⁻¹, $\log(Z) = 0$, and solar abundance patterns). To estimate the S/N of the LAMOST spectra, we use the mean flux variance from several LAMOST spectra of giant stars with a g -band S/N of 50 pixel⁻¹. As in our comparison to Kirby et al. (2018), we make several cuts on the sample in order to fairly compare the reported precision with our CRLBs, which we list in Table D1. These cuts leave the reported precision for approximately 6000 stars.

Because Xiang et al. (2019) report their abundance precision in terms of [X/Fe], we add $\sigma[\text{Fe}/\text{H}]$ in quadrature to $\sigma[\text{X}/\text{Fe}]$ so that the CRLBs are on the same scale. Xiang et al. (2019) do provide estimated systematic uncertainties for their measurements, but because CRLBs are a measure of precision and not accuracy, we do not include them in this comparison.

Figure D1 shows the reported measurement precision of these stars compared to our LAMOST CRLBs. Similar to our comparison with Kirby et al. (2018), we find that most abundances reported by Xiang et al. (2019) are within a factor of ~ 2 of our CRLBs. The largest difference is in the precision of T_{eff} , which is reported to be 27 K, nearly three times larger than our predicted precision (10 K). This is not wholly unreasonable given the subtle and highly model-dependent effects that T_{eff} has on spectral features. The reported precision for Fe (0.029 dex) is also more than a factor of 2 larger than our forecast (0.013 dex)—though the absolute difference is quite small. We suspect this is driven by the larger uncertainties found for T_{eff} and $\log g$ by Xiang et al. (2019) and the substantial correlation these labels have with Fe in giant stars.

Interestingly, we find that the precision reported for Ni, O, and C outperforms the CRLB by a factor of 1.2, 1.7, and 2.1. We suspect that this might be the result of “gradient aliasing” in the DD-Payne, whereby the model picks up spectral gradient features from elements other than the one it attributes them to. This is a common challenge in data-driven methods, and while Xiang et al. (2019) attempted to mitigate it by regularizing the model with ab initio spectral gradients, some gradient aliasing may remain. For the remaining abundances, there are several reasons why slightly poorer precision might be expected in practice, including model fidelity and imperfect calibrations (see Section 6.4 for further discussion).

Together, the comparisons conducted here and in Section 4.1.1 illustrate that the CRLBs are quite reasonable representations of contemporary abundance measurements.

⁴² The DD-Payne is a hybrid spectral model that is trained on high-resolution measurements from GALAH and APOGEE and regularized on ab initio spectral gradients.

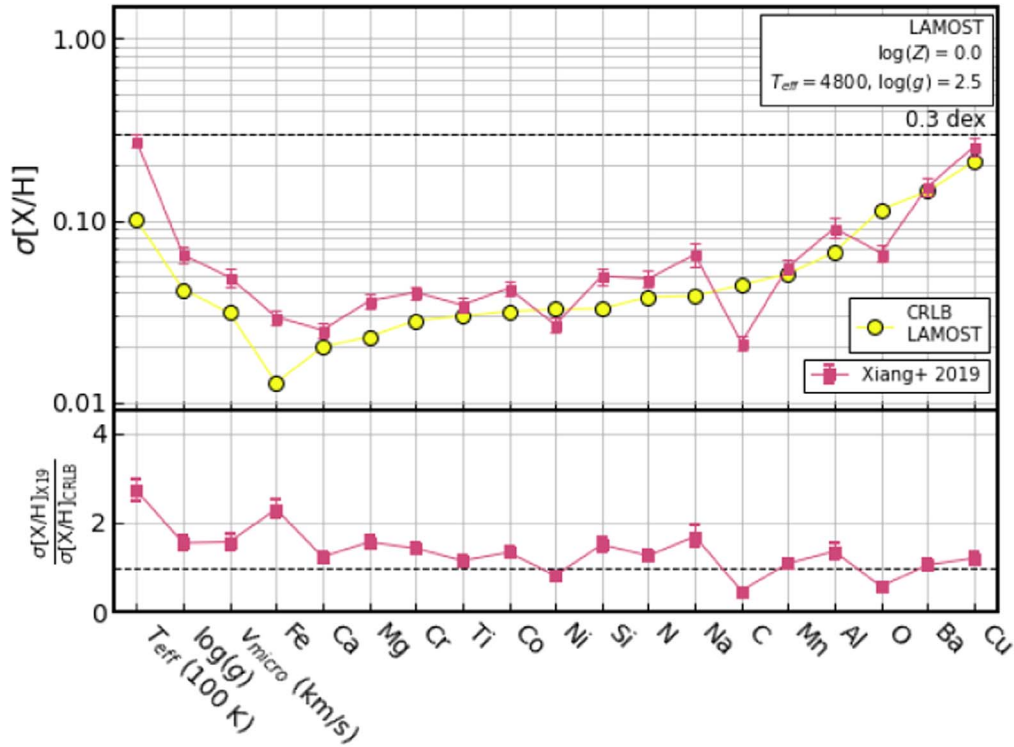


Figure D1. (Top) LAMOST CRLBs for a typical solar-metallicity K-Giant with a g -band S/N of 50 pixel⁻¹ overplotted with the internal precision of ~ 6000 comparable stars report by Xiang et al. (2019). Error bars denote the upper and lower quartiles of the sample’s precision. (Bottom) The ratio of the forecasted LAMOST CRLBs to the reported precision for each stellar label. As found with the comparison to Kirby et al. (2018) in Figure 4, the measurement uncertainties for most elements are generally a factor of $\lesssim 2$ larger than the CRLBs. The reported precision for Ni, C, and O slightly outperform the CRLBs, which may be the result of additional spectral information included by the data-driven model of Xiang et al. (2019) that is not incorporated in our purely ab initio model.

Appendix E Validation of CRLBs

To validate the robustness of the CRLBs, we infer the stellar labels of a mock spectrum at various S/Ns using an ab initio trained spectral model and an HMC sampling method and compare the precision of this inference with the precision forecasted by the CRLBs. We outline the process of training the spectral model in Appendix E.1 and fitting the mock spectrum in Appendix E.2. The results of this comparison are presented in Appendix E.3.

E.1. Training a Spectral Model

Training a spectral model requires a large set of stellar spectra with known labels that span the relevant parameter space. To generate this training set, we randomly drew 10^4 stellar labels from the following uniform distribution:⁴³

$$\begin{aligned} T_{\text{eff}} &\sim \mathcal{U}(4500 \text{ K}, 5000 \text{ K}), \\ \log g &\sim \mathcal{U}(1.5, 2.1), \\ v_{\text{micro}} &\sim \mathcal{U}(1.4 \text{ km s}^{-1}, 2.4 \text{ km s}^{-1}), \text{ and} \\ [X/H] &\sim \mathcal{U}(-0.5, 0.5), \end{aligned}$$

where in this case X refers to a smaller subset of elements: Fe, Ca, Ni, Si, Ti, Mg, and Co. We only considered 7 elements, limiting the model to 10 stellar labels, to simplify the training process. These specific elements were chosen as they are the most precisely recovered elements by the D1200G setup (see

⁴³ $\mathcal{O}(10^3)$ stellar spectra would likely have been sufficient, but opted to generate 10^4 to further reduce emulation errors.

Table E1
Details of The Payne

| | |
|-----------------------|----------------|
| # Training Spectra | 7500 |
| # Validation Spectra | 2500 |
| # Spectra/Batch | 512 |
| # Hidden Dense Layers | 2 |
| # Neurons/Layer | 300 |
| Activation Function | Leaky ReLU |
| # Training Steps | 10^5 |
| Loss Function | L1 Mean |
| Optimizer | Rectified Adam |
| Learning Rate | 10^{-3} |
| Interpolation Errors | <0.1% |

Section 4.1 and Table 2). The bounds of the uniform distributions are chosen to center on the parameters of our fiducial RGB star (Table 1) and span roughly two times the D1200G (S/N = 50) CRLB for each stellar label, assuming the Gaussian priors of $\sigma_{T_{\text{eff}}} = 100 \text{ K}$, $\sigma_{\log g} = 0.15$, and $\sigma_{v_{\text{micro}}} = 0.25 \text{ km s}^{-1}$ used previously in Section 4.1.4. Spectra were generated and convolved to instrumental resolution as previously described in Section 3.2.

Withholding 2500 spectra for validation, we train an updated version of The Payne⁴⁴ (details in Table E1). To aid the training process, the labels are normalized according to

$$\theta_{i,\text{scaled}} = \frac{\theta_i - \theta_{i,\text{min}}}{\theta_{i,\text{max}} - \theta_{i,\text{min}}} - 0.5, \quad (\text{E1})$$

⁴⁴ https://github.com/tingyuansen/The_Payne

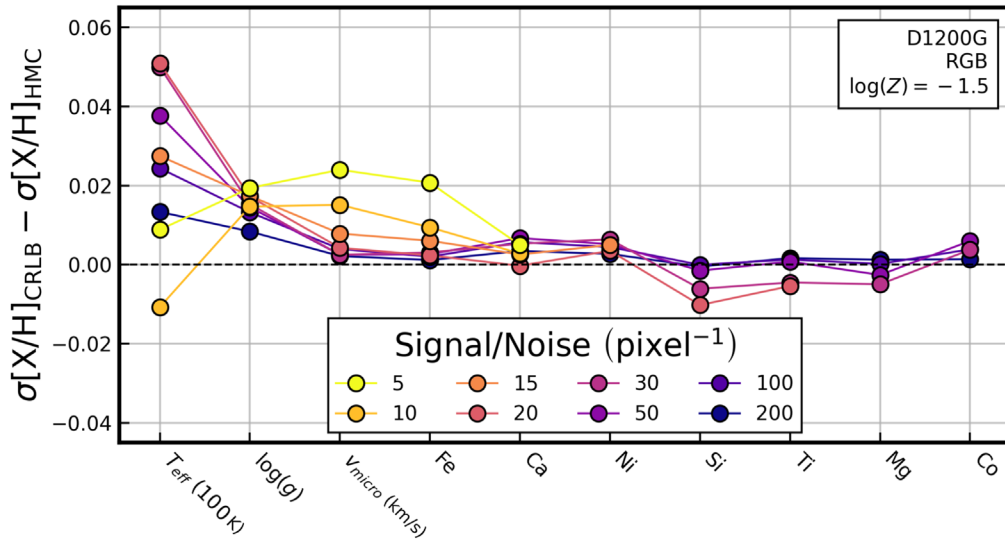


Figure E1. The difference between the CRLB and the stellar label precision found through HMC sampling for a $\log Z = -1.5$ RGB star observed with the D1200G setup. A constant S/N across the wavelength coverage was assumed. Differences are small ($\lesssim 5$ K for $\sigma_{T_{\text{eff}}}$; $\lesssim 0.02$ dex for $\sigma_{\log g}$; $\lesssim 0.02$ km s $^{-1}$ for $\sigma_{v_{\text{micro}}}$; and $\lesssim 0.02$ dex for $\sigma_{[X/H]}$), indicating that the CRLB is a robust predictor of stellar label precision down to at least S/N ~ 15 pixel $^{-1}$.

Table E2
Fixed Stellar Labels at Each S/N

| S/N (pix $^{-1}$) | Fixed Labels |
|--------------------|--|
| 5, 10 | [Ni/H], [Si/H], [Ti/H], [Co/H], [Mg/H] |
| 15 | [Si/H], [Ti/H], [Co/H], [Mg/H] |
| 20 | [Co/H], [Mg/H] |
| 30, 50, 100, 200 | None |

where $\theta_{i,\min}$ and $\theta_{i,\max}$ are the minimum and maximum values represented in the training and validation data sets. After 10^5 training steps, which takes roughly 4 hr on an Nvidia K80 GPU, the model that minimized the L1 mean loss on the validation spectra is chosen as the final model.

We compare ab initio spectra from our validation set to spectra generated with the same labels using `The Payne` and find mean interpolation errors of individual pixels to be less than 0.1%. These errors are much smaller than typical observational uncertainties in the normalized spectra.

E.2. Fitting Mock Spectra with HMC Sampling

The mock spectrum is generated using `The Payne` at the labels of the fiducial $\log Z = -1.5$ RGB star to avoid introducing any bias that may have been introduced in the training of the spectral model—recall that we are interested in precision, not accuracy, here. We assume a constant S/N across the entire spectrum, which manifests as an uncertainty in each pixel of $\sigma = f(\lambda)/(S/N)$, where $f(\lambda)$ is the normalized flux of the model. With the same mock spectrum, we perform the fitting assuming a range in S/N from 5 to 200 pixel $^{-1}$ that is constant across the entire wavelength coverage.

With only 10 stellar labels and likelihoods that we believe to be close to Gaussian, using an MCMC sampling technique would likely be adequate for this scenario. However, because our neural network spectral emulator is differentiable, we opt to use an HMC sampler, making it readily adapted for inference with many more labels where an MCMC sampler might face convergence problems.

We adopt the Gaussian likelihood function in Equation (1) and the following priors:

$$\begin{aligned}
 T_{\text{eff}} &\sim \mathcal{N}^*(4750 \text{ K}, 100 \text{ K}), \\
 \log g &\sim \mathcal{N}^*(1.8, 0.15), \\
 v_{\text{micro}} &\sim \mathcal{N}^*(1.9 \text{ km s}^{-1}, 0.25 \text{ km s}^{-1}), \text{ and} \\
 [X/H] &\sim \mathcal{U}(-0.5, 0.5), \text{ and} \\
 [X^*/H] &\sim \delta(0.0),
 \end{aligned}$$

where $\mathcal{N}^*(\mu, \sigma)$ represents a normal distribution truncated at the limits of the training set so that the model does not extrapolate. Here, X^* refers to elements that the CRLB predict cannot be recovered to better than 0.3 dex at the given S/N. These elements are held fixed at solar value, which is equivalent to applying a delta function prior at $[X/H] = 0.0$. The fixed labels at each S/N are displayed in Table E2.

For each S/N we perform the HMC sampling using 24 parallel chains. Each chain begins with 3000 burn-in samples, which are discarded, followed by another 3000 samples, which constitute our posterior sample.

E.3. Comparison to CRLB

In Figure E1, we plot the difference between the precision predicted by the CRLBs and the standard deviation of the mock fit posteriors for each S/N. In the calculation of the CRLBs, we include the same priors on T_{eff} , $\log g$, and v_{micro} used in the HMC sampling. In addition, for each S/N, we only consider the gradients for the stellar labels that are left free in the sampling (see Table E2), thus holding all other labels fixed at solar values. Instead of calculating spectral gradients from ab initio spectra, we calculate the gradients from our trained spectral model to exclude any systematics introduced by interpolation errors of the model.

In general, we find the CRLBs and the standard deviations of the mock fits to be in agreement at the 0.01 dex level down to an S/N of 10 and at the 0.02 dex level down to an S/N of 5. At very high S/N (200 pixel $^{-1}$), the CRLBs accurately predict the precision of the v_{micro} and all chemical abundances, only very slightly underpredicting the precision of T_{eff} by 1 K and $\log g$

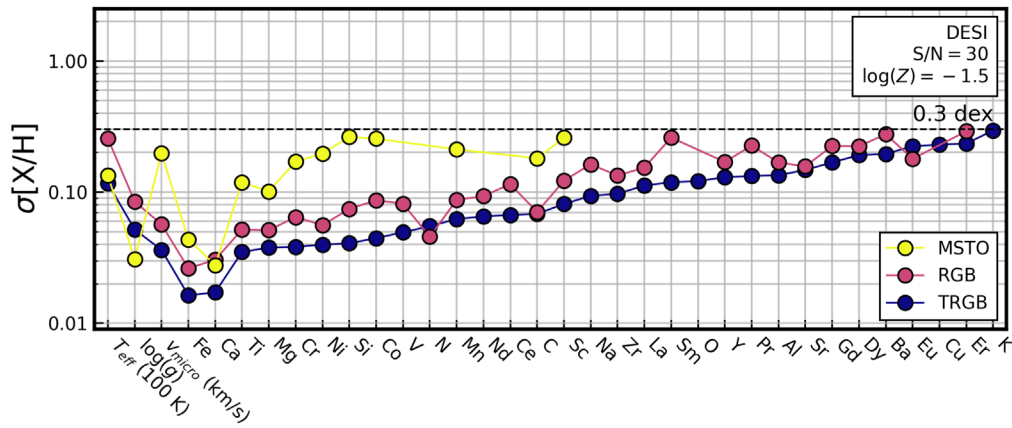


Figure F1. DESI CRLBs of $\log Z = -1.5$ MSTO, RGB, and TRGB stars with a constant S/N of 30 pixel⁻¹. The atmosphere parameters for each star can be found in Table 1. Just as for D1200G, abundance recovery is more precise for cool giants and less precise for hot subgiants.

by 0.01 dex. As the S/N decreases to 20 pixel⁻¹, the difference grows to 5 K and 0.02 dex in T_{eff} and $\log g$, respectively, and the CRLBs slightly overpredict the precision for Si, Ti, and Mg by no more than 0.01 dex. All of these differences remain relatively small compared to the typical precision found for these labels and are the result of the posteriors of these labels being slightly non-Gaussian (negatively skewed).

As the S/N decreases further, the precision of both the mock fit and the CRLB become prior dominated for T_{eff} , $\log g$, and v_{micro} , resulting in a smaller difference in the precision of T_{eff} . This is not the case for the precision of $\log g$ and v_{micro} due to the difference between the Gaussian prior included in the CRLB calculation and the truncated Gaussian included in the HMC sampling. Still, the differences are only ~ 0.02 dex, which is quite minor in relation to the expected precision at S/N < 15. Thus, we find that the CRLB is a robust predictor of stellar label precision down to at least an S/N of 15 pixel⁻¹.

Appendix F DESI CRLBs

DESI is a fiber-fed MOS that covers a wavelength range from 3600 to 9800 Å with a resolving power of 2000–5000. The primary science goal of the DESI survey is not galactic archaeology, nor is the 4 meter Mayall telescope it is mounted on large enough to efficiently observe resolved stars in dwarf galaxies. Nevertheless, it is a particularly interesting spectrograph for stellar chemical abundance measurements. When observing conditions are too poor for faint galaxy work, DESI will target bright galaxies, filling unused fibers with MW stars. This will yield spectra for roughly 10 million MW stars. In addition to many thin- and thick-disk stars, these deep observations are expected to reach MSTO stars in the MW’s halo out to 30 kpc, allowing for a dramatically improved understanding of the stellar halo’s chemical composition. In addition, DESI’s instrumental design has been a major inspiration for current and next-generation survey instruments that will be targeting stars in dwarf galaxies.

Thus, while DESI will not be observing dwarf galaxy stars, we still think it valuable to present the theoretical abundance precision achievable by DESI in the MW halo. For these calculations we assume a uniform S/N of 30 pixel⁻¹, which should be achievable for stars of $m_r = 16.5$ –18 in a short 5–10 minute exposure (DESI Collaboration et al. 2016b). The spectroscopic configuration used is given in Table 2. Because

DESI will be able to observe down to the MSTO in the halo, we calculate the CRLBs for MSTO, RGB, and TRGB stars as done for D1200G in Section 4.1.3.

In Figure F1, we plot the CRLBs for DESI, illustrating its capability to extend the precise chemical abundance measurements of MW-disk surveys out to the MW’s halo. As seen for D1200G in Figure 6, abundance recovery is more precise for cool giants due to stronger absorption features and less precise for hot subgiants, which have weaker absorption features.

ORCID iDs

Nathan R. Sandford <https://orcid.org/0000-0002-7393-3595>
Daniel R. Weisz <https://orcid.org/0000-0002-6442-6030>
Yuan-Sen Ting <https://orcid.org/0000-0001-5082-9536>

References

- Adshad, P., & Easter, R. 2008, *JCAP*, 10, 047
Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:astro-ph/0609591
Aldenius, M., Tanner, J. D., Johansson, S., Lundberg, H., & Ryan, S. G. 2007, *A&A*, 461, 767
Alfaro-Cuello, M., Kacharov, N., Neumayer, N., et al. 2019, *ApJ*, 886, 57
Allende Prieto, C. 2016, *LRSP*, 13, 1
Aller, L. H. 1942, *ApJ*, 96, 321
Aller, L. H. 1946, *ApJ*, 104, 347
Alves-Brito, A., Barbuy, B., Zoccali, M., et al. 2006, *A&A*, 460, 269
Amarsi, A. M., Lind, K., Asplund, M., Barklem, P. S., & Collet, R. 2016, *MNRAS*, 463, 1518
Andreassen, D. T., Sousa, S. G., Delgado Mena, E., et al. 2016, *A&A*, 585, A143
Aoki, W., Arimoto, N., Sadakane, K., et al. 2009, *A&A*, 502, 569
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
Bacon, R., Accardo, M., Adjali, L., et al. 2010, *Proc. SPIE*, 7735, 773508
Bagnasco, G., Kolm, M., Ferruit, P., et al. 2007, *Proc. SPIE*, 6692, 66920M
Bailer-Jones, C. a. L. 2000, *A&A*, 357, 197
Baschek, B. 1959, *ZA*, 48, 95
Bastian, N., & Lardo, C. 2018, *ARA&A*, 56, 83
Battaglia, G., Helmi, A., Tolstoy, E., et al. 2008, *ApJL*, 681, L13
Battaglia, G., Tolstoy, E., Helmi, A., et al. 2006, *A&A*, 459, 423
Battaglia, G., Tolstoy, E., Helmi, A., et al. 2011, *MNRAS*, 411, 1013
Becker, A., Huterer, D., & Kadota, K. 2012, *JCAP*, 12, 034
Bedell, M., Meléndez, J., Bean, J. L., et al. 2014, *ApJ*, 795, 23
Bell, R. A. 1970, *MNRAS*, 148, 25
Bell, R. A., & Branch, D. 1976, *MNRAS*, 175, 25
Bergemann, M., Lind, K., Collet, R., Magic, Z., & Asplund, M. 2012, *MNRAS*, 427, 27

- Bernstein, R., Shtetman, S. A., Zunnels, S. M., Mochnacki, S., & Athey, A. E. 2003, *Proc. SPIE*, 4841, 1694
- Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A22
- Blanco-Cualesma, S. 2019, *MNRAS*, 486, 2075
- Boesgaard, A. M., King, J. R., Cody, A. M., Stephens, A., & Deliyannis, C. P. 2005, *ApJ*, 629, 832
- Boesgaard, A. M., Stephens, A., King, J. R., & Deliyannis, C. P. 2000, *Proc. SPIE*, 4005, 274
- Bundy, K., Westfall, K., MacDonald, N., et al. 2019, *BAAS*, 51, 198
- Caffau, E., Koch, A., Sbordone, L., et al. 2013, *AN*, 334, 197
- Caffau, E., Ludwig, H.-G., Steffen, M., Freytag, B., & Bonifacio, P. 2011, *SoPh*, 268, 255
- Carbon, D. F., Langer, G. E., Butler, D., et al. 1982, *ApJS*, 49, 207
- Carlin, J. L., Grillmair, C. J., Muñoz, R. R., Nidever, D. L., & Majewski, S. R. 2009, *ApJL*, 702, L9
- Carrera, R., Pancino, E., Gallart, C., & del Pino, A. 2013, *MNRAS*, 434, 1681
- Casagrande, L., Schönrich, R., Asplund, M., et al. 2011, *A&A*, 530, A138
- Cenarro, A. J., Cardiel, N., Gorgas, J., et al. 2001a, *MNRAS*, 326, 959
- Cenarro, A. J., Gorgas, J., Cardiel, N., et al. 2001b, *MNRAS*, 326, 981
- Cenarro, A. J., Gorgas, J., Cardiel, N., Vazdekis, A., & Peletier, R. F. 2002, *MNRAS*, 329, 863
- Chapman, S. C., Ibata, R., Lewis, G. F., et al. 2005, *ApJL*, 632, L87
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102
- Cohen, J. G., & Huang, W. 2009, *ApJ*, 701, 1053
- Collins, M. L. M., Chapman, S. C., Rich, R. M., et al. 2013, *ApJ*, 768, 172
- Conroy, C., Bonaca, A., Cargile, P., et al. 2019a, *ApJ*, 883, 107
- Conroy, C., Naidu, R. P., Zaritsky, D., et al. 2019b, *ApJ*, 887, 237
- Cowan, J. J., Sneden, C., Bures, S., et al. 2002, *ApJ*, 572, 861
- Cramer, H. 1946, *Mathematical Methods of Statistics* (Princeton, NJ: Princeton Univ. Press)
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA*, 12, 1197
- Czekala, I., Andrews, S. M., Mandel, K. S., Hogg, D. W., & Green, G. M. 2015, *ApJ*, 812, 128
- Darmois, G. 1945, *Rev. Int. Stat. Inst.*, 13, 9
- Dekker, H., D'Odorico, S., Kaufer, A., Delabre, B., & Kotzlowski, H. 2000, *Proc. SPIE*, 4008, 534
- DePoy, D. L., Allen, R., Barkhouser, R., et al. 2012, *Proc. SPIE*, 8446, 84461N
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016a, arXiv:1611.00037
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016b, arXiv:1611.00036
- de Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, 449, 2604
- Dotter, A. 2016, *ApJS*, 222, 8
- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, *PhLB*, 195, 216
- Duggan, G. E., Kirby, E. N., Andrievsky, S. M., & Korotin, S. A. 2018, *ApJ*, 869, 50
- Echeverría, A., Silva, J. F., Mendez, R. A., & Orchard, M. 2016, *A&A*, 594, A111
- Eriksen, M., & Gaztañaga, E. 2015, *MNRAS*, 452, 2168
- Escala, I., Gilbert, K. M., Kirby, E. N., et al. 2019a, *ApJ*, 878, 42
- Escala, I., Kirby, E. N., Gilbert, K. M., Cunningham, E. C., & Wojno, J. 2019b, *ApJ*, 878, 42
- Evans, C. J., Castro, N., Gonzalez, O. A., et al. 2019, *A&A*, 622, A129
- Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, *Proc. SPIE*, 4841, 1657
- Fabricant, D., Fata, R., Epss, H., et al. 2019, *PASP*, 131, 075004
- Fabricant, D., Fata, R., Roll, J., et al. 2005, *PASP*, 117, 1411
- Feeny, S. M., Wandelt, B. D., & Ness, M. K. 2019, arXiv:1912.09498
- Fischel, D. 1964, *ApJ*, 140, 221
- Font-Ribera, A., McDonald, P., Mostek, N., et al. 2014, *JCAP*, 05, 023
- Fraunhofer, J. 1817, *AnP*, 56, 264
- Frebel, A., Norris, J. E., Gilmore, G., & Wyse, R. F. G. 2016, *ApJ*, 826, 110
- Frebel, A., Simon, J. D., Geha, M., & Willman, B. 2010, *ApJ*, 708, 560
- Frebel, A., Simon, J. D., & Kirby, E. N. 2014, *ApJ*, 786, 74
- Fréchet, M. 1943, *Rev. Int. Stat. Inst.*, 11, 182
- Fulbright, J. P., Rich, R. M., & Castro, S. 2004, *ApJ*, 612, 447
- García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., et al. 2016, *AJ*, 151, 144
- Gilbert, K. M., Kirby, E. N., Escala, I., et al. 2019, *ApJ*, 883, 128
- Gingerich, O. 1969, in *Proc. 3rd Harvard-Smithsonian Conf. Stellar Atmospheres, Comparison of Synthetic Spectra with Real Spectra* (Cambridge, MA: MIT)
- Greenstein, J. L. 1948, *ApJ*, 107, 151
- Hansen, C. J., Ludwig, H.-G., Seifert, W., et al. 2015, *AN*, 336, 665
- Hearnshaw, J. 2010, *JAHH*, 13, 90
- Heiter, U., Jofré, P., Gustafsson, B., et al. 2015, *A&A*, 582, A49
- Hendricks, B., Koch, A., Walker, M., et al. 2014, *A&A*, 572, A82
- Hill, V., Skúladóttir, Á, Tolstoy, E., et al. 2019, *A&A*, 626, A15
- Ho, N., Geha, M., Tollerud, E. J., et al. 2015, *ApJ*, 798, 77
- Holtzman, J. A., Shetrone, M., Johnson, J. A., et al. 2015, *AJ*, 150, 148
- Huggins, W., & Miller, W. A. 1864, *RSPT*, 154, 413
- Hunter, J. D. 2007, *CSE*, 9, 90
- Husser, T.-O., Kamann, S., Dreizler, S., et al. 2016, *A&A*, 588, A148
- Ireland, J. 2005, *ApJ*, 620, 1132
- Jagourel, P., Fitzsimons, E., Hammer, F., et al. 2018, *Proc. SPIE*, 10702, 10702A4
- Ji, A. P., Frebel, A., Ezzeddine, R., & Casey, A. R. 2016a, *ApJL*, 832, L3
- Ji, A. P., Frebel, A., Simon, J. D., & Chiti, A. 2016b, *ApJ*, 830, 93
- Ji, A. P., Frebel, A., Simon, J. D., & Geha, M. 2016c, *ApJ*, 817, 41
- Ji, A. P., Li, T. S., Simon, J. D., et al. 2020, *ApJ*, 889, 27
- Jofré, P., Heiter, U., & Soubiran, C. 2019, *ARA&A*, 57, 571
- Jorgensen, U. G., Carlsson, M., & Johnson, H. R. 1992, *A&A*, 254, 258
- Kalirai, J. S., Beaton, R. L., Geha, M. C., et al. 2010, *ApJ*, 711, 671
- Kamann, S., Husser, T.-O., Dreizler, S., et al. 2018, *MNRAS*, 473, 5591
- Kamann, S., Husser, T.-O., Wendt, M., et al. 2016, *Msngr*, 164, 18
- Kay, S. M. 1993, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory* (1st ed.; Englewood Cliffs, NJ: Prentice-Hall)
- King, A. L., Davis, T. M., Denney, K. D., Vestergaard, M., & Watson, D. 2014, *MNRAS*, 441, 3454
- Kirby, E. N., Cohen, J. G., Simon, J. D., et al. 2017a, *ApJ*, 838, 83
- Kirby, E. N., Cohen, J. G., Simon, J. D., & Guhathakurta, P. 2015a, *ApJL*, 814, L7
- Kirby, E. N., Gilbert, K. M., Escala, I., et al. 2020, *AJ*, 159, 46
- Kirby, E. N., Guhathakurta, P., Bolte, M., Sneden, C., & Geha, M. C. 2009, *ApJ*, 705, 328
- Kirby, E. N., Guhathakurta, P., Simon, J. D., et al. 2010, *ApJS*, 191, 352
- Kirby, E. N., Guhathakurta, P., & Sneden, C. 2008, *ApJ*, 682, 1217
- Kirby, E. N., Guo, M., Zhang, A. J., et al. 2015b, *ApJ*, 801, 125
- Kirby, E. N., Rizzi, L., Held, E. V., et al. 2017b, *ApJ*, 834, 9
- Kirby, E. N., Simon, J. D., & Cohen, J. G. 2015c, *ApJ*, 810, 56
- Kirby, E. N., Xie, J. L., Guo, R., Kovalev, M., & Bergemann, M. 2018, *ApJS*, 237, 18
- Kirchhoff, G. 1860, *AnP*, 185, 275
- Kirchhoff, G. 1863, *AnP*, 194, 94
- Kirchhoff, G., & Bunsen, R. 1860, *AnP*, 186, 161
- Koch, A., & Côté, P. 2010, *A&A*, 517, A59
- Koch, A., Grebel, E. K., Gilmore, G. F., et al. 2008a, *AJ*, 135, 1580
- Koch, A., Grebel, E. K., Kleyna, J. T., et al. 2007a, *AJ*, 133, 270
- Koch, A., McWilliam, A., Grebel, E. K., Zucker, D. B., & Belokurov, V. 2008b, *ApJL*, 688, L13
- Koch, A., & Rich, R. M. 2014, *ApJ*, 794, 89
- Koch, A., Wilkinson, M. I., Kleyna, J. T., et al. 2007b, *ApJ*, 657, 241
- Koch, A., Wilkinson, M. I., Kleyna, J. T., et al. 2009, *ApJ*, 690, 453
- Kurucz, R. L. 1970, *SAOSR*, 309
- Kurucz, R. L. 1993, *SYNTHES* Spectrum Synthesis Programs and Line Data (Cambridge, MA: Smithsonian Astrophysical Observatory)
- Kurucz, R. L. 2005, *MSAIS*, 8, 14
- Kurucz, R. L. 2013, *ATLAS12: Opacity sampling model atmosphere program*, *Astrophysics Source Code Library*, ascl:1303.024
- Kurucz, R. L. 2017, *ATLAS9: Model atmosphere program with opacity distribution functions*, *Astrophysics Source Code Library*, ascl:1710.017
- Kurucz, R. L., & Avrett, E. H. 1981, *SAOSR*, 391
- Lai, D. K., Lee, Y. S., Bolte, M., et al. 2011, *ApJ*, 738, 51
- Latour, M., Husser, T.-O., Giesers, B., et al. 2019, *A&A*, 631, 14
- Laverick, M., Lobel, A., Merle, T., et al. 2018, *A&A*, 612, A60
- Leaman, R., Cole, A. A., Venn, K. A., et al. 2009, *ApJ*, 699, 1
- Leep, E. M., Oke, J. B., & Wallerstein, G. 1987, *AJ*, 93, 338
- Leep, E. M., Wallerstein, G., & Oke, J. B. 1986, *AJ*, 91, 1117
- Lertarte, B., Hill, V., Jablonka, P., et al. 2006, *A&A*, 453, 547
- Li, T. S., Simon, J. D., Drlica-Wagner, A., et al. 2017, *ApJ*, 838, 8
- Longard, N., Martin, N., Starkenburg, E., et al. 2020, *MNRAS*, 491, 356
- Lucchesi, R., Lardo, C., Primas, F., et al. 2020, arXiv:2001.11033
- Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, *RAA*, 15, 1095
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
- Martin, N. F., Geha, M., Ibata, R. A., et al. 2016a, *MNRAS*, 458, L59
- Martin, N. F., Ibata, R. A., Chapman, S. C., Irwin, M., & Lewis, G. F. 2007, *MNRAS*, 380, 281
- Martin, N. F., Ibata, R. A., Collins, M. L. M., et al. 2016b, *ApJ*, 818, 40
- Mashonkina, L., Korn, A. J., & Przybilla, N. 2007, *A&A*, 461, 261
- Mateo, M., Iii, J. I. B., Crane, J., et al. 2012, *Proc SPIE*, 8446, 84464Y

- McKinney, W. 2010, in Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 51
- McLeod, A. F., Kruijssen, J. M. D., & Weisz, D. R. 2020, *ApJ*, 891, 25
- Minnaert, M. 1934, *Obs*, 57, 328
- Moore, E. 1920, *BAMaS*, 26, 394
- MSE Science Team, Babusiaux, C., Bergemann, M., et al. 2019, arXiv:1904.04907
- Muñoz, R. R., Majewski, S. R., Zaggia, S., et al. 2006, *ApJ*, 649, 201
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, 808, 16
- Nissen, P. E., & Gustafsson, B. 2018, *A&ARv*, 26, 6
- Norris, J. E., Gilmore, G., Wyse, R. F. G., et al. 2008, *ApJL*, 689, L113
- Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, *PASP*, 107, 375
- Olszewski, E. W., Schommer, R. A., Suntzeff, N. B., & Harris, H. C. 1991, *AJ*, 101, 515
- Pandas Development Team 2020, pandas-dev/pandas, Zenodo, doi:10.5281/zenodo.3509134
- Pasquini, L., Avila, G., Blecha, A., et al. 2002, *Msngr*, 110, 1
- Paszke, A., Gross, S., Massa, F., et al. 2019, arXiv:1912.01703
- Paxton, B., Bildsten, L., Dotter, A., et al. 2011, *ApJS*, 192, 3
- Paxton, B., Cantiello, M., Arras, P., et al. 2013, *ApJS*, 208, 4
- Paxton, B., Marchant, P., Schwab, J., et al. 2015, *ApJS*, 220, 15
- Payne, C. H. 1925, PhD thesis, Radcliffe College
- Pazder, J. S., Roberts, S., Abraham, R., et al. 2006, *Proc. SPIE*, 6269, 62691X
- Pehlivan Rhodin, A., Hartman, H., Nilsson, H., & Jönsson, P. 2017, *A&A*, 598, A102
- Penrose, R. 1955, *PCPS*, 51, 406
- Pérez, F., & Granger, B. E. 2007, *CSE*, 9, 21
- Pickering, J. C., Thorne, A. P., & Perez, R. 2001, *ApJS*, 132, 403
- Pogge, R. W., Atwood, B., Brewer, D. F., et al. 2010, *Proc. SPIE*, 7735, 77350A
- Pont, F., Zinn, R., Gallart, C., Hardy, E., & Winnick, R. 2004, *AJ*, 127, 840
- Rao, C. R. 1945, *Bull. Calcutta Math. Soc.*, 37, 81
- Robertson, J. G. 2017, *PASA*, 34, e035
- Roederer, I. 2019, *BAAS*, 51, 49
- Roth, M. M., Sandin, C., & Kamann, S. 2018, *A&A*, 618, A3
- Ruchti, G. R., Feltzing, S., Lind, K., et al. 2016, *MNRAS*, 461, 2174
- Russell, H. N. 1929, *ApJ*, 70, 11
- Rutledge, G. A., Hesser, J. E., & Stetson, P. B. 1997, *PASP*, 109, 907
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, *PeerJ Comput. Sci.*, 2, e55
- Sandford, N. 2020, Chem-I-Calc: The Chemical Information Calculator, Zenodo, doi:10.5281/zenodo.3924672
- Shetrone, M., Bizyaev, D., Lawler, J. E., et al. 2015, *ApJS*, 221, 24
- Shetrone, M., Venn, K. A., Tolstoy, E., et al. 2003, *AJ*, 125, 684
- Shetrone, M. D., Bolte, M., & Stetson, P. B. 1998, *AJ*, 115, 1888
- Shetrone, M. D., Côté, P., & Sargent, W. L. W. 2001, *ApJ*, 548, 592
- Shetrone, M. D., Siegel, M. H., Cook, D. O., & Bosler, T. 2009, *AJ*, 137, 62
- Simon, J. D., Drlica-Wagner, A., Li, T. S., et al. 2015, *ApJ*, 808, 95
- Simon, J. D., & Geha, M. 2007, *ApJ*, 670, 313
- Simon, J. D., Li, T. S., Drlica-Wagner, A., et al. 2017, *ApJ*, 838, 11
- Slater, C. T., Bell, E. F., Martin, N. F., Tollerud, E. J., & Ho, N. 2015, *ApJ*, 806, 230
- Snedden, C. 1973, *ApJ*, 184, 839
- Snedden, C. 1974, *ApJ*, 189, 493
- Snedden, C., Cowan, J. J., Lawler, J. E., et al. 2003, *ApJ*, 591, 936
- Snedden, C., & Parthasarathy, M. 1983, *ApJ*, 267, 757
- Spencer, M. E., Mateo, M., Walker, M. G., & Olszewski, E. W. 2017, *ApJ*, 836, 202
- Spite, M., Spite, F., François, P., et al. 2018, *A&A*, 617, A56
- Starkenburger, E., Hill, V., Tolstoy, E., et al. 2010, *A&A*, 513, A34
- Starkenburger, E., Hill, V., Tolstoy, E., et al. 2013, *A&A*, 549, A88
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, *AJ*, 132, 1645
- Strömgren, B. 1940, *Festschrift für Elis Strömgren* (Copenhagen: Einar Munksgaard), 218
- Suda, T., Hidaka, J., Aoki, W., et al. 2017, *PASJ*, 69, 76
- Suntzeff, N. B. 1981, *ApJS*, 47, 1
- Suntzeff, N. B., Mateo, M., Terndrup, D. M., et al. 1993, *ApJ*, 418, 208
- Swan, J., Cole, A. A., Tolstoy, E., & Irwin, M. J. 2016, *MNRAS*, 456, 4315
- Szentgyorgyi, A., Baldwin, D., Barnes, S., et al. 2016, *Proc. SPIE*, 9908, 990822
- Szentgyorgyi, A., Furesz, G., Cheimets, P., et al. 2011, *PASP*, 123, 1188
- Takada, M., Ellis, R. S., Chiba, M., et al. 2014, *PASJ*, 66, R1
- Tamura, N., Takato, N., Shimono, A., et al. 2018, *Proc. SPIE*, 10702, 107021C
- Thaler, R., Jablonka, P., Lardo, C., et al. 2019, arXiv:1911.08627
- Ting, Y.-S., Conroy, C., & Rix, H.-W. 2016, *ApJ*, 826, 83
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Asplund, M. 2018, *ApJ*, 860, 159
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2017a, *ApJ*, 843, 32
- Ting, Y.-S., Conroy, C., Rix, H.-W., & Cargile, P. 2019, *ApJ*, 879, 69
- Ting, Y.-S., Rix, H.-W., Conroy, C., Ho, A. Y. Q., & Lin, J. 2017b, *ApJL*, 849, L9
- Tinsley, B. M. 1980, *FCPh*, 5, 287
- Tolstoy, E., Hill, V., & Tosi, M. 2009, *ARA&A*, 47, 371
- Tolstoy, E., Irwin, M. J., Helmi, A., et al. 2004, *ApJL*, 617, L119
- Tolstoy, E., Venn, K. A., Shetrone, M., et al. 2003, *AJ*, 125, 707
- Unsöld, A. 1938, *Physik der sternatmosphären* (Berlin: Verlag)
- Unsöld, A. 1942, *ZA*, 21, 22
- Vargas, L. C., Geha, M., Kirby, E. N., & Simon, J. D. 2013, *ApJ*, 767, 134
- Vargas, L. C., Geha, M. C., & Tollerud, E. J. 2014a, *ApJ*, 790, 73
- Vargas, L. C., Gilbert, K. M., Geha, M., et al. 2014b, *ApJL*, 797, L2
- Venn, K. A., Irwin, M., Shetrone, M. D., et al. 2004, *AJ*, 128, 1177
- Venn, K. A., Lennon, D. J., Kaufer, A., et al. 2001, *ApJ*, 547, 765
- Venn, K. A., Starkenburg, E., Malo, L., Martin, N., & Laevens, B. P. M. 2017, *MNRAS*, 466, 3741
- Vernet, J., Dekker, H., D'Odorico, S., et al. 2011, *A&A*, 536, A105
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2019, *scipy: SciPy v1.2.1*, Zenodo, doi:10.5281/zenodo.2560881
- Vogel, K., Hilker, M., Baumgardt, H., et al. 2016, *MNRAS*, 460, 3384
- Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, *Proc. SPIE*, 2198, 362
- Walker, M. G., Belokurov, V., Evans, N. W., et al. 2009a, *ApJL*, 694, L144
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2007, *ApJL*, 667, L53
- Walker, M. G., Mateo, M., & Olszewski, E. W. 2009b, *AJ*, 137, 3100
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2015a, *ApJ*, 808, 108
- Walker, M. G., Mateo, M., Olszewski, E. W., et al. 2016, *ApJ*, 819, 53
- Walker, M. G., Olszewski, E. W., & Mateo, M. 2015b, *MNRAS*, 448, 2717
- Wallerstein, G., Leep, E. M., & Oke, J. B. 1987, *AJ*, 93, 1137
- Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, 13, 22
- Wang, Y. 2010, *MPLA*, 25, 3093
- Wright, K. O. 1948, *PDAO*, 8, 1
- Xiang, M., Ting, Y.-S., Rix, H.-W., et al. 2019, *ApJS*, 245, 34
- Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, *AJ*, 137, 4377
- Yoon, J., Whitten, D., Beers, T., Lee, Y., & Placco, V. 2019, *ApJ*, 894, 7