Publications of the Astronomical Society of the Pacific, 128:066001 (43pp), 2016 June doi:10.1088/1538-3873/128/964/066001
© 2016. The Astronomical Society of the Pacific. All rights reserved. Printed in the U.S.A.

State of the Field: Extreme Precision Radial Velocities*

Debra A. Fischer[1], Guillem Anglada-Escude[2,3], Pamela Arriagada[4], Roman V. Baluev[5,6], Jacob L. Bean[7], Francois Bouchy[8], Lars A. Buchhave[9], Thorsten Carroll[10], Abhijit Chakraborty[11], Justin R. Crepp[12], Rebekah I. Dawson[13,14], Scott A. Diddams[15,16], Xavier Dumusque[17], Jason D. Eastman[18], Michael Endl[19], Pedro Figueira[20], Eric B. Ford[13,14], Daniel Foreman-Mackey[21,22], Paul Fournier[23], Gabor Fűrész[24], B. Scott Gaudi[25], Philip C. Gregory[26], Frank Grundahl[27], Artie P. Hatzes[28], Guillaume Hébrard[29,30], Enrique Herrero[31], David W. Hogg[21,32], Andrew W. Howard[33], John A. Johnson[18], Paul Jorden[34], Colby A. Jurgenson[1], David W. Latham[18], Greg Laughlin[35], Thomas J. Loredo[36], Christophe Lovis[17], Suvrath Mahadevan[13,14,37], Tyler M. McCracken[1], Francesco Pepe[17], Mario Perez[38], David F. Phillips[18], Peter P. Plavchan[39], Lisa Prato[40], Andreas Quirrenbach[41], Ansgar Reiners[42], Paul Robertson[13,14], Nuno C. Santos[20,43], David Sawyer[1], Damien Segransan[17], Alessandro Sozzetti[44], Tilo Steinmetz[45], Andrew Szentgyorgyi[18], Stéphane Udry[17], Jeff A. Valenti[46], Sharon X. Wang[13,14], Robert A. Wittenmyer[47,48], and Jason T. Wright[13,14,37]

[1] Department of Astronomy, Yale University, New Haven, CT 06511, USA; debra.fischer@yale.edu
[2] Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
[3] School of Physics and Astronomy, Queen Mary University of London, 327 Mile End Road, London E1 4NS, UK
[4] Department of Terrestrial Magnetism, Carnegie Institute of Washington, Washington, DC 20015, USA
[5] Central Astronomical Observatory at Pulkovo of Russian Academy of Sciences, Pulkovskoje shosse 65, St Petersburg 196140, Russia
[6] Sobolev Astronomical Institute, St Petersburg State University, Universitetskij prospekt 28, Petrodvorets, St Petersburg 198504, Russia
[7] Department of Astronomy and Astrophysics, University of Chicago, 5640 S. Ellis Avenue, Chicago, IL 60637, USA
[8] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, F-13388 Marseille, France
[9] Centre for Star and Planet Formation, Natural History Museum of Denmark & Niels Bohr Institute, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark
[10] Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D-14482 Potsdam, Germany
[11] Astronomy and Astrophysics Division, Physical Research Laboratory (PRL), Ahmedabad 380009, India
[12] Department of Physics, University of Notre Dame, South Bend, IN, USA
[13] Department of Astronomy and Astrophysics, The Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802, USA
[14] Center for Exoplanets and Habitable Worlds, The Pennsylvania State University, University Park, PA 16802, USA
[15] National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA
[16] Department of Physics, University of Colorado, 2000 Colorado Avenue, Boulder, CO, 80309, USA
[17] Observatoire Astronomique de l'Université de Genéve, 51 chemin des Maillettes, CH-1290 Versoix, Switzerland
[18] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[19] The University of Texas at Austin and Department of Astronomy and McDonald Observatory, 2515 Speedway, C1400, Austin, TX 78712, USA
[20] Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal
[21] Center for Cosmology and Particle Physics, Department of Physics, New York University, NY, USA
[22] Astronomy Department, University of Washington, Box 951580, Seattle, WA 98195, USA
[23] Fibertech Optica Inc., Kitchener, ON N2M 5C6, Canada
[24] MIT Kavli Institute for Astrophysics and Space Research, 77 Mass Avenue, 37-515, Cambridge, MA 02139, USA
[25] Department of Astronomy, Ohio State University, 140 W. 18th Avenue, Columbus, OH 43210, USA
[26] Physics and Astronomy, University of British Columbia, 6244 Agricultural Road, British Columbia, V6T 1Z1, Canada
[27] Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark
[28] Thüringer Landessternwarte, Sternwarte 5, Tautenburg 5, D-07778 Tautenburg, Germany
[29] Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre & Marie Curie, 98bis boulevard Arago, F-75014 Paris, France
[30] Observatoire de Haute-Provence, Université d'Aix-Marseille & CNRS, F-04870 Saint Michel lObservatoire, France
[31] Institut de Ciències de lEspai (CSICIEEC), Campus UAB, Facultat de Ciències, Torre C5 parell, 2a pl, E-08193 Bellaterra, Spain
[32] Center for Data Science, New York University, NY, USA
[33] Institute for Astronomy, University of Hawaii, Honolulu, HI 96822, USA
[34] e2v Technologies, 106 Waterhouse Lane, Chelmsford, Essex, UK
[35] UCO Lick Observatory, Department of Astronomy and Astrophysics, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
[36] Center for Radiophysics and Space Research, Space Sciences Building, Cornell University Ithaca, NY 14853-6801, USA
[37] Penn State Astrobiology Research Center, The Pennsylvania State University, University Park, PA 16802, USA
[38] NASA Headquarters, Washington, DC, USA
[39] Department of Physics, Astronomy and Material Science, 901 S National Avenue, Missouri State University, Springfield, MO 65897, USA
[40] Lowell Observatory, 1400 West Mars Hill, Road, Flagstaff, AZ 86001, USA
[41] Landessternwarte, Zentrum für Astronomie der Universität Heidelberg, Königstuhl 12, D-69117 Heidelberg, Germany
[42] Institut für Astrophysik, Georg-August-Universitä, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany
[43] Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal
[44] INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, I-10025 Pino Torinese, Italy
[45] Menlo Systems GmbH, Am Klopferspitz 19a, D-82152 Martinsried, Germany
[46] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
[47] School of Physics and Australian Centre for Astrobiology, University of New South Wales, Sydney, NSW 2052, Australia

* The content of this publication emerged from presentations and discussion by the 150 participants in the Second Workshop on Extreme Precision Radial Velocities, held at Yale University. The talks and posters from this meeting are archived at http://eprv.astro.yale.edu.

[48] Computational Engineering and Science Research Centre, University of Southern Queensland, Toowoomba, QLD 4350, Australia
Received 2016 January 21; accepted 2016 February 29; published 2016 May 17

Abstract

The Second Workshop on Extreme Precision Radial Velocities defined circa 2015 the state of the art Doppler precision and identified the critical path challenges for reaching $10\,\mathrm{cm\,s^{-1}}$ measurement precision. The presentations and discussion of key issues for instrumentation and data analysis and the workshop recommendations for achieving this bold precision are summarized here. Beginning with the High Accuracy Radial Velocity Planet Searcher spectrograph, technological advances for precision radial velocity (RV) measurements have focused on building extremely stable instruments. To reach still higher precision, future spectrometers will need to improve upon the state of the art, producing even higher fidelity spectra. This should be possible with improved environmental control, greater stability in the illumination of the spectrometer optics, better detectors, more precise wavelength calibration, and broader bandwidth spectra. Key data analysis challenges for the precision RV community include distinguishing center of mass (COM) Keplerian motion from photospheric velocities (time correlated noise) and the proper treatment of telluric contamination. Success here is coupled to the instrument design, but also requires the implementation of robust statistical and modeling techniques. COM velocities produce Doppler shifts that affect every line identically, while photospheric velocities produce line profile asymmetries with wavelength and temporal dependencies that are different from Keplerian signals. Exoplanets are an important subfield of astronomy and there has been an impressive rate of discovery over the past two decades. However, higher precision RV measurements are required to serve as a discovery technique for potentially habitable worlds, to confirm and characterize detections from transit missions, and to provide mass measurements for other space-based missions. The future of exoplanet science has very different trajectories depending on the precision that can ultimately be achieved with Doppler measurements.

Key words: instrumentation: spectrographs – methods: observational – methods: statistical – techniques: radial velocities – techniques: spectroscopic

Online material: color figures

1. Introduction

The past two decades have been a golden era for the discovery of exoplanets—thousands of exoplanets have been detected using Doppler measurements, transit photometry, microlensing, and direct imaging. Pioneering technology has driven remarkable acceleration in the rate of detections—of special note: the clever use of an iodine reference cell for wavelength calibration and modeling of the instrumental profile to achieve radial velocity (RV) precisions of a few meters per second using general purpose spectrometers (Marcy & Butler 1992; Butler et al. 1996); the stabilized High Accuracy Radial Velocity Planet Searcher (HARPS) spectrometer, which set a new standard in RV precision (Pepe et al. 2002; Mayor et al. 2003); ground-based transit surveys that evolved rapidly from demonstrating the existence of transiting extrasolar planets (Charbonneau et al. 2000; Henry et al. 2000) to achieving 1 mmag photometric precision (Johnson et al. 2009); and the *Kepler* space mission with a dramatic improvement over the precision of ground-based photometry. The *Kepler* mission confirmed earlier suggestions (Mayor & Udry 2008) that most of the stars in our galaxy have planetary systems and that small planets are ubiquitous (Howard et al. 2012; Fressin et al. 2013; Buchhave et al. 2014). However, while ever smaller and lower mass planets have been detected in the past few years, all current techniques fall just short of detecting so-called exo-Earths (planets that are nearly the mass of the Earth, orbiting at habitable zone distances) around nearby stars.

A generation of students has grown up thinking of exoplanet science as a booming field where the rate of discovery and characterization has a steep and positive trajectory. However, the key to growth in this field has been the sequential improvements in technology and measurement precision and this is also a fundamental requirement for the future of the field. If we keep making the same observations with the same instruments, we will improve the signal-to-noise ratio (S/N) and population size, but we will basically have the same results and the field will stagnate; advances in strategy, analysis and instrumentation are required to reach higher RV precision and to push into new parameter space.

New space missions including the *Transiting Exoplanet Survey Satellite* (*TESS*; Ricker et al. 2014), the *CHaracterizing ExOPlanet Satellite* (*CHEOPS*; Fortier et al. 2014) and *PLAnetary Transits and Oscillation* of stars (*PLATO*; Rauer et al. 2014) will have the precision to detect transiting planets with small radii and will observe bright nearby stars. Doppler observations will provide critical support for these missions and will remain a key asset in the search for exo-Earths (Mayor

Table 1
Current Doppler Planet Search Programs

Spectrograph	Slit or Fiber	Temp Control	Spectral Resolution	Wavelength Range (nm)	Wavelength Calibrator	SMP ($m\,s^{-1}$) S/N = 200	Number of Stars	Duration of Program
HARPS	f	Y	115,000	380–690	ThAr	0.8	2000	2003
HARPS-N	f	Y	115,000	380–690	ThAr	0.8	500	2012
PARAS	f	Y	67,000	380–690	ThAr	1.0	27	2012
CHIRON	f	Y	90,000	440–650	Iodine	1.0	35	2011
SOPHIE	f	Y	75,000	387–694	ThAr	1.1	190	2011
PFS	s	Y	76,000	390–670	Iodine	1.2	530	2010
HIRES	s	Y	55,000	364–800	Iodine	1.5	4000	1996
Levy (LCPS)	s	Y	110,000	376–970	Iodine	1.5	100	2013
Levy (CPS)	s	Y	100,000	376–940	Iodine	2.0	300	2013
SONG	s	N	90,000	440–690	Iodine	2.0	12	2014
HRS	s	Y	60,000	408–784	Iodine	3.0	100	2001–2013
Hamilton	s	N	50,000	390–800	Iodine	3.0	350	1987–2011
UCLES	s	N	45,000	478–871	Iodine	3.0	240	1998
Tull	s	N	60,000	345–980	Iodine	5.0	200	1998

et al. 2014). If we can improve our current instrumental precision and if we can distinguish stellar photospheric velocities from Keplerian motion, then we have a chance of detecting, confirming and characterizing a bounty of exoplanets in new mass and period regimes around nearby stars.

In 2015 July, we held the Second Workshop for Extreme Precision Radial Velocities (EPRVs) at Yale University with three goals: to examine the current state of the Doppler precision, to discuss recent advances, and to chart a course for eliminating the remaining obstacles on the road toward $10\,cm\,s^{-1}$. We discussed the future game-changers and the current bottlenecks. This paper summarizes the presentations and discussions from this workshop. Section 2 presents the state of the art for fourteen different RV surveys highlighted at the workshop and compares some of the key attributes correlated with high Doppler precision. In Section 3, we discuss instrumentation challenges and solutions. In Section 4 we discuss the challenges to high precision that are not instrumental in nature, such as stellar atmospheric velocities and telluric lines in the Earth's atmosphere and capture the discussion about statistical analysis techniques that show promise in distinguishing stellar photospheric signals from Keplerian motion.

2. State of the Art

It is important to first clarify what we mean by Doppler precision. RV measurements are currently carried out by one of two methods: the iodine technique or the cross correlation technique (Lovis & Fischer 2010). The estimated single measurement precision (SMP; or averaged nightly velocity precision) is a good indication of the photon-limited "on sky" precision. The long-term velocity rms (several days to years) is yet another indicator, exposing both systematic instrumental errors and the limitations in analysis techniques for treating variability from the stellar photosphere (so-called stellar jitter).

Participants at the workshop discussed the Doppler precision for current RV planet search programs. In standardized formats, presenters were asked to show the dependence of SMP on the S/N for the equivalent of a $3\,km\,s^{-1}$ pixel bin at 550 nm and to provide the RV rms of their program stars without removing trends or Keplerian signals. This is the first side-by-side comparison that has ever been made for Doppler planet search programs and we owe a great debt of gratitude to the teams that contributed data for this comparison.[49] Their generosity in sharing accumulated data allows the community to consider the possible relative importance of environmental stability, spectral resolution, wavelength coverage, calibration techniques and observing cadence. No comparison metric is perfect and there will still be discrepancies that arise because of differences in the stellar samples and the scientific goals of the programs.

Table 1 lists the key parameters for key Doppler surveys. The programs are sorted by the demonstrated SMP. We summarize some of the fundamental parameters for the programs, including whether the coupling of light is accomplished with a slit or fiber, whether there is any environmental control (however, the extent of environmental control varies widely), the spectral resolution, wavelength range, method of wavelength calibration, SMP for S/N of 200 (at 550 nm), the number of stars on the program and the time baseline of the program. Fiber fed, stable instruments with high spectral resolution represent the state of the art circa 2015. Future instruments will require even more stringent specifications in order to gain another order of magnitude in RV precision. High resolution and broad bandwidth may end up being critical

[49] Additional current or planned PRV spectrographs can be found in Tables 2 and 3 of Plavchan et al. (2015).

parameters for fitting out perturbing signals from stellar photospheres and telluric contamination.

The Doppler programs listed in Table 1 are discussed in more detail below in order of the date that each planet survey began.

2.1. 3 m Lick Hamilton, 1987–2011

The Hamilton spectrograph (Vogt 1987) at Lick Observatory[50] was used to search for exoplanets between 1987 and 2011 (Butler & Marcy 1997; Fischer et al. 2014). The program effectively ended in 2011 when insulating material on the cell burned and changed the iodine spectrum of the cell. The Hamilton echelle spectrograph is a general purpose instrument that can be fed with either the 0.6 m Coudé Auxiliary Telescope or the 3 m Shane telescope. The wavelength range can be selected to include wavelengths from 390 to 900 nm in a single observation. The instrumental resolution depends on the selected slit width, varying from 30,000 to 115,000, but the typical resolution used for Doppler planet hunting was $R \sim 50{,}000$. The instrument is located in a coudé room and experiences diurnal and seasonal temperature swings of several degrees. One pioneering aspect of this program was the use of an iodine reference cell for the wavelength calibration and modeling the line spread function (Marcy & Butler 1992; Butler et al. 1996) to reach a measurement precision of 3 m s^{-1}. The use of a reference cell is ideal for general-purpose instruments because the forward modeling of iodine accounts for variations in the SLSF and instrumental drifts. The iodine absorption lines only span a wavelength range from 510 to 620 nm; this limits the Doppler analysis to about 110 nm of the spectrum and high S/N is required in order to model the multi-parameter SLSF.

The search for exoplanets at Lick Observatory began with a sample of 109 stars and the program was allocated roughly 25 nights per year; that allocation was augmented with as many as 50 additional nights per year on the smaller Coudé Auxiliary Telescope. The first assessment of the occurrence rate of exoplanets (Cumming et al. 1999) and the first assessment of the impact of the impact of photospheric magnetic activity on RV precision (Saar et al. 1998) was made using data from the Lick planet search program. The sample size was increased (Fischer et al. 1999) to include an additional ~300 stars, with a focus on a subset of metal-rich stars. Johnson et al. (2007) carried out a search for planets orbiting subgiants. The Hamilton spectrograph was also used to study the impact of optical scrambling on RV measurements (Spronck et al. 2010, 2012a, 2012b, 2013).

With the iodine modeling technique for the Lick Hamilton and CTIO CHIRON programs, the spectrum is divided into many small chunks, typically 2 or 3 Å wide, and the wavelength solution, the spectral line spread function, and the Doppler shift are modeled for every chunk. Each spectral chunk provides an independent estimate of the relative Doppler shift that is good to about 30–50 m s^{-1} and 3σ outliers are rejected. Then, a weighted uncertainty is calculated for each chunk as follows. First, the difference, d_i, between the velocity for each ith chunk and the median velocity of all chunks in a given observation is calculated:

$$d_i = \mathrm{vel}_i - \mathrm{median}(\mathrm{vel}). \quad (1)$$

The uncertainty in the measurement for each chunk, σ_i, reflects the ability of that chunk to consistently report a Doppler shift and it is determined over several observations as the standard deviation of d_i for n observations:

$$\sigma_i = \mathrm{stddev}(d_i[1], ..., d_i[n_{\mathrm{obs}}]). \quad (2)$$

Thus, d_i for a particular chunk may be large (or small) with many observations, but as long as it is consistently large (or small), σ_i for the chunk will be small. This allows for offsets between chunks (e.g., that might occur because of stitching errors in the charge-coupled device (CCD) or other instrumental issues). Chunks with values of d_i that are erratic from one observation to the next will have a large uncertainty, σ_i. A subtle corollary is that the sigma for a given chunk is better determined as more observations are accumulated, i.e., the uncertainty for a specific observation can change slightly as future observations are analyzed.

The SMP at Lick was limited by instrumental errors, rather than photon statistics. The Hamilton spectrograph was not a stabilized instrument and this set the floor for the SMP of Doppler measurements to about 3 m s^{-1} (Butler et al. 1996) with rms velocities greater than 5 m s^{-1}. Figure 1 (left) shows the dependence of estimated internal errors on the S/N of the observations. The different families of errors (depicted with different plot symbols) correspond to eras where the CCD detectors were changed on the Hamilton spectrograph. In particular, the worst RV precision for the Lick program occurred when a thinned CCD (dewar 6) was used; this device suffered from charge diffusion that broadened the SLSF and yielded a significantly degraded velocity precision. Figure 1 (right) is a histogram of the rms of the velocities. No trends or Keplerian signals were fitted out and most stars have rms velocities that were greater than the 10 m s^{-1} boundary of the plot.

The Lick planet search program demonstrates the importance of using a stable, special-purpose instrument. Several different detectors were used over the course of the 25 year program and each of these changes to the instrument introduced systematic offsets in the time series RV data. Improving the instrument stability was difficult in the large coudé room where large seasonal temperature were observed. The Lick planet search program was retired as the new Automated Planet Finder (APF) was being brought online.

[50] Presentation by Debra Fischer.

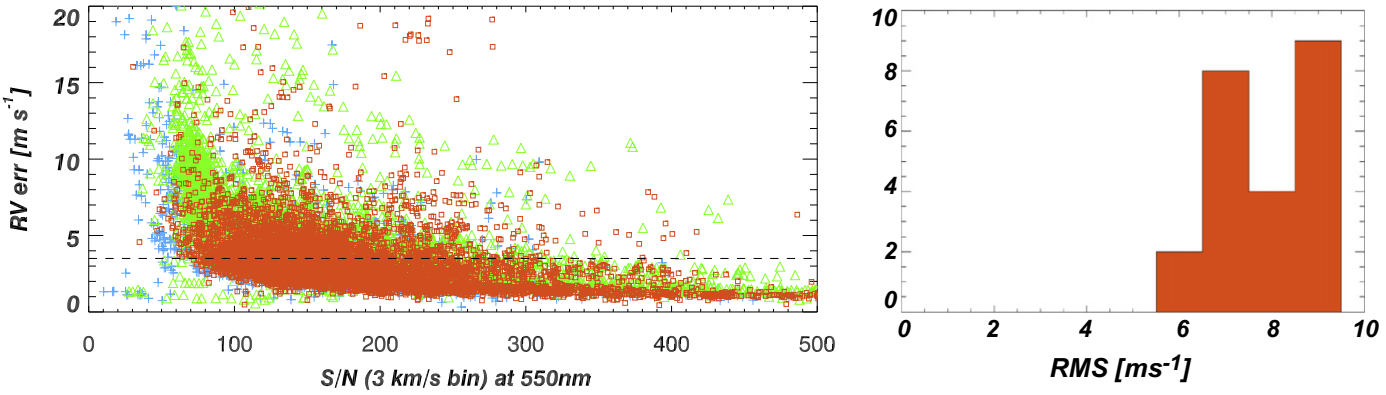


Figure 1. Single measurement precision as a function of S/N (left) for dewar 13 (blue triangles) used from 1994 to 1997, dewar 6 (green diamonds) from 1998 to 2000, and dewar 8 (red circles) from 2001 to 2011. The radial velocity rms scatter (right) is plotted for data from the Hamilton spectrograph at Lick Observatory; most stars have rms scatter greater than the 10 m s^{-1} cutoff. (Figures provided by Debra Fischer.)
(A color version of this figure is available in the online journal.)

2.2. 10 m Keck I with HIRES Spectrograph, 1996

In the early days of RV planet searches, HIRES (Vogt et al. 1994) on the Keck 10 m telescope[51] was one of the instruments that made high impact contributions to exoplanet detections. The program began in 1996 and has continued operations to this day. The spectrometer was designed as a slit-fed instrument and originally configured for a Doppler planet search with a wavelength range of 389–618 nm and a 0.″8 slit that provides a moderate resolution of about 55,000 (higher resolution can be achieve by narrowing the slit with an accompanying loss of light). In 2004, the detector was upgraded to a mosaic of three CCDs, providing wavelength coverage of 364–479 nm, 498–642 nm, and 655–797 nm for each detector respectively. The wavelength calibration is achieved with an iodine reference cell and the Doppler analysis is therefore restricted to wavelengths between 510 and 620 nm. Although HIRES is not in a vacuum enclosure, the temperature is a fairly stable (1° ± 1°C) in the instrument room.

The primary science goals are planet detection and characterization for a broad range of stellar populations (Vogt et al. 2010a, 2010b). A hallmark of the Keck program is the collaborative integration of projects from several principal investigators, using a time-sharing scheme that pools 40–60 nights per year. This strategy improves the observing cadence for several programs, including the California Planet Search (CPS; Howard et al. 2010), the N2K ("next 2000") search for gas giant planets around metal rich stars (Fischer et al. 2005), a search for planets orbiting subgiants (Johnson et al. 2007) and the "Eta-Earth" program to systematically search for low mass planets (3–30 $M_{\oplus}$) orbiting in the habitable zone of the nearest 230 GKM stars (Howard et al. 2009). Keck HIRES programs also follow up transiting planets (e.g., *Kepler*, *K*2, HAT-NET) to determine masses (Marcy et al. 2014a, 2014b; Weiss & Marcy 2014) and to measure the Rossiter–McLaughlin effect (Winn et al. 2010, 2011a, 2011b; Sanchis-Ojeda et al. 2015). Keck data have been used to derive planet occurrence rates (Cumming et al. 2008; Howard et al. 2012) and spectra have been used for characterization of host star metallicity and activity (Fischer & Valenti 2005; Valenti & Fischer 2005; Isaacson & Fischer 2010). Collectively, more than 4000 stars have been observed and analyzed from Keck-HIRES.

Before the HIRES detector upgrade, a number of detector-related noise sources produced flux dependent errors. After correcting for this effect, the SMP was 3 m s^{-1}. When the detector was upgraded in 2004 August, the SMP improved to about 1.5 m s^{-1} for an S/N of 200 and reached a floor of about 1 m s^{-1} for S/Ns approaching 500. The RV rms is typically greater than 2 m s^{-1} because of instrumental errors, limitations of the iodine technique (Isaacson & Fischer 2010; Spronck et al. 2015), and the impact of stellar photospheric velocities, or stellar "jitter." Figure 2 (left) shows the dependence of estimated internal errors for CPS observations as a function of the S/N of the observations. Figure 2 (right) is a histogram of the rms of the velocities (without attempting to remove any trends or Keplerian signals).

Although the iodine technique enables the use of general purpose instruments for precise Doppler measurements, this method requires high S/N (about 200 per pixel) for the forward modeling process. The variable SLSF contributes to increased scatter in the RV measurements and limits the ability to take advantage of the exceptional signal to noise that is possible with a 10 m telescope. The CPS team is now designing SHREK, a new stabilized instrument for the Keck II telescope.

[51] Presentation by Andrew Howard.

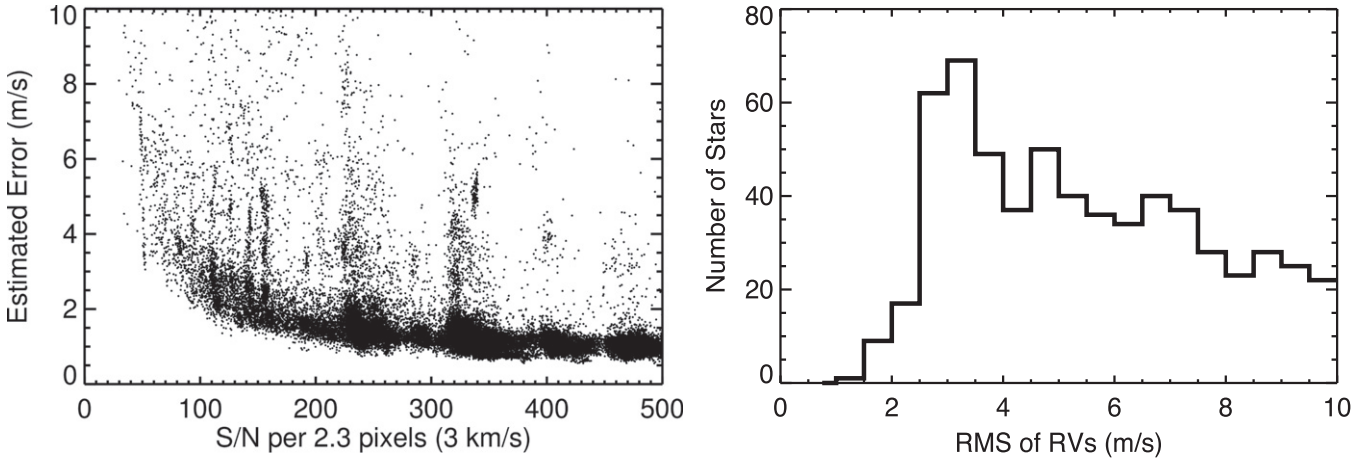


Figure 2. Radial velocity precision (left) and rms velocity scatter for Doppler measurements by the California Planet Search team using the HIRES at Keck. (Courtesy of Andrew Howard.)

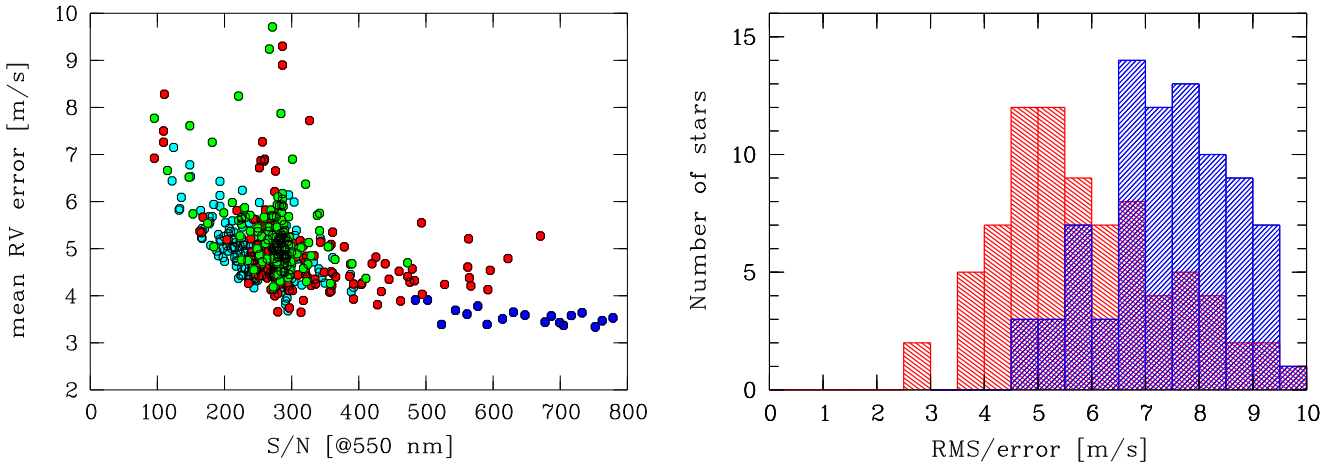


Figure 3. (Left) Estimated single measurement precision as a function of S/N using the Tull spectrograph on the 2.7 m Harlan Smith Telescope at McDonald Observatory. Different color plot symbols are for different stars. (Right) Radial velocity scatter (right) for the stars observed with this same spectrograph at McDonald Observatory. (Courtesy of Michael Endl.)

(A color version of this figure is available in the online journal.)

2.3. 2.7 McDonald and Tull Coudé Spectrograph, 1998

The Tull spectrograph (Tull et al. 1995) has been used at the 2.7 m Harlan J. Smith telescope at McDonald Observatory[52] for exoplanet searches since 1998 and more than 10,000 RV measurements have been made over the past 15 years. The observing time is scheduled roughly once per month. The spectral resolution is 60,000 and the wavelength range is 345–1000 nm. The spectrometer is not environmentally stabilized and it is slit-fed; wavelength calibration is carried out with an iodine reference cell, restricting Doppler analysis to the wavelength range of 510–620 nm.

As shown in Figure 3 there are about 200 stars on the exoplanet survey with the Tull spectrograph; 82 of these stars have an extensive baseline of observations. The primary science goal of this program is the detection of Jupiter analogs at 4–5 au.

The internal errors represent the formal uncertainty of the weighted mean of the individual RVs that are modeled for small spectral chunks. These chunks are typically 2–3 Å wide and the internal error depends on the overall width of the distribution of the accepted chunks (extreme deviating chunk RVs are rejected) and their total number. A detailed description

[52] Presentation by Michael Endl.

can be found in Endl et al. (2000). The mean SMP errors peak at about 5 m s^{-1} and the velocity rms scatter peaks at about 6–7 m s^{-1}. When observing the sun through a solar port where there is a more homogeneous illumination of the slit, the RV data routinely improves to a precision of 2–3 m s^{-1} demonstrating the importance of uniform illumination of the spectrometer optics.

The most important weaknesses of this program are that this is a large multi-setup, multi-user instrument with complex, large optics, relatively poor telescope image stability and guiding and a 2 pixel sampling of the instrumental profile. The detector is a Tektronix CCD (circa 1990). The project could likely be improved with a new 15 micron CCD, octagonal fiber link, wavefront sensor for focus, and fast tip/tilt for image stabilization.

2.4. 3.9 m AAT and University College London Echelle Spectrograph (UCLES) Coudé Spectrograph, 1998

The Anglo-Australian Planet Search (AAPS) has been in continuous operation since 1998, using the UCLES at the 3.9 m Anglo-Australian Telescope.[53] The 1″ slit is used for the AAPS and corresponds to a resolution of 45,000. The instrument wavelength range extends from 4780 to 8415 Å and the instrument employs an iodine reference cell for wavelength calibration. For flexible scheduling, the program uses an automated queue-based observing system, which controls instrument configuration, removes operator error, and implements parallel detector readout and telescope slewing.

UCLES on the AAT has concentrated on the same sample of 240 stars from 1998 to 2012, detecting more than 45 new planets orbiting these stars (Butler et al. 2001; Carter et al. 2003; Bailey et al. 2009; O'Toole et al. 2009; Tinney et al. 2011; Wittenmyer et al. 2014). The AAPS has entered its final phase of operations and is concentrating on the key science question: What is the occurrence rate of Jupiter-analogs, gas giant planets still orbiting beyond the frost line with no inner gas giant planets? Owing to the legacy of the 17 year continuously observed sample, the AAPS remains a leader in the detection of long-period planets.

Although UCLES is a non-stabilized spectrograph, the AAPS has achieved a velocity precision of ~3 m s^{-1} per epoch at a signal-to-noise of 200 per pixel and a resolution of 45,000. The quoted velocity uncertainties from the AAPS include the effects of photon-counting uncertainties, residual errors in the spectrograph point-spread function (PSF) model, and variation between the underlying iodine-free template spectrum and spectra observed through the iodine cell. The spectrum is broken up into several hundred 2 Å chunks and velocities are derived from each chunk (Butler et al. 1996). The final RV measurement is determined as the mean of the chunks, and the quoted internal uncertainties are estimated from the scatter of the individual chunks (Butler et al. 1996). Figure 4 is a histogram of the rms of the velocities (without attempting to remove any trends or Keplerian signals).

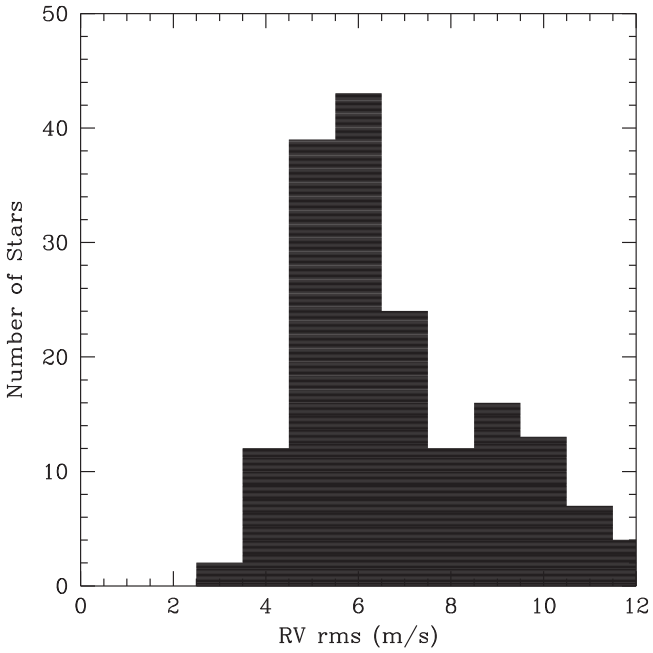


Figure 4. Radial velocity scatter (right) for UCLES. (Courtesy of Robert Wittenmyer.)

Over the lifetime of the AAPS, the team has worked to improve the stability and precision of the RV measurements. In 2005 the program began using longer integrations, which averaged over stellar oscillation noise (O'Toole et al. 2008) and delivered higher S/N (typically 400). As a result, the systematic and stellar uncertainties were reduced below 2 m s^{-1} as demonstrated for the bright triple-planet-hosting star 61 Vir (Vogt et al. 2010b). This long-term stability is reflected in the number of long-period systems detected by AAPS; the majority of AAPS planets ($62 \pm 13\%$) have periods longer than one year.

To advance the precision at the AAT, a next-generation spectrograph, *Veloce*, is being built as a replacement for UCLES, with commissioning planned in late 2016. *Veloce* is an asymmetric white-pupil echelle spectrograph that employs the innovations used by other PRV instruments (e.g., G-CLEF, SALT-HRS, and KiwiSpec) to maximize velocity stability: dual-shell thermal enclosure, vacuum-enclosed echelle grating, octagonal fibers, an innovative ball-lens double scrambler, and wavelength calibration with an ultra-stablized pulsed-laser frequency comb (Murphy et al. 2007). This calibration source will be injected into the simultaneous calibration fiber feed to achieve a wavelength calibration better than 20 cm s^{-1}.

[53] Information provided by Rob Wittenmyer.

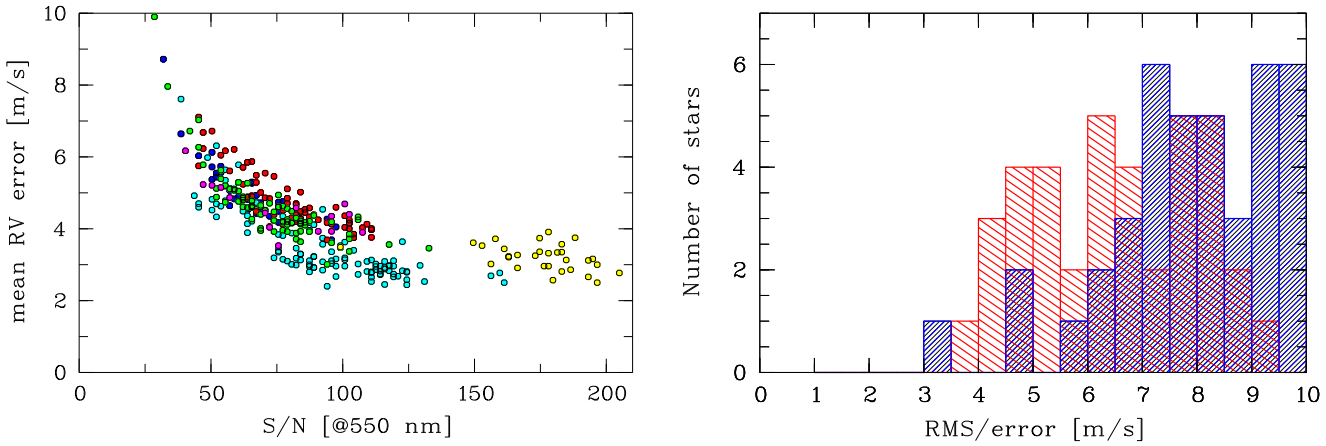


Figure 5. Radial velocity precision as a function of S/N (left) and radial velocity scatter (right) for HRS. (Courtesy of Michael Endl.)
(A color version of this figure is available in the online journal.)

2.5. 9.2 m Hobby–Eberly Telescope (HET) and High Resolution Spectrograph (HRS) Spectrometer, 2001

The HRS (Tull 1998) was commissioned at the 9.2 m HET at McDonald Observatory[54] in 2001 and carried out a search for exoplanets through 2013. The instrument and telescope are currently closed for the HET upgrade project, but will be recommissioned in 2016. The wavelength range of HRS is 408–784 nm and the spectral resolution is $R = 60{,}000$. An iodine reference cell is used for wavelength calibration. The instrument is not stabilized for temperature or pressure. Observations are carried out in queue mode scheduling and stars are observed between 1 and 5 times per month. Over the past 10 years, more than 1600 RV measurements have been obtained.

The primary targets on the HRS program are 41 faint stars; these are typically M dwarfs with apparent magnitudes of $10 < V < 12$. There is also a program to carry out follow-up of *Kepler* transiting planet candidates for mass determination.

There are about 65% slit losses with the HRS. This occurs because the slit must be narrowed to reach a resolution of 60,000. As a result, the HRS observations have an S/N that is typically around 100; this reduces the Doppler precision when modeling with the iodine technique and yields estimated SMP errors that range from 4 to 8 m s^{-1} with a long term velocity rms that is greater than 6 m s^{-1} (Figure 5).

The instrument does not have an exposure meter; this is particularly critical for HET because the telescope design produces variable illumination at the telescope pupil. The HRS-upgrade project is now underway with plans to provide image slicers for an increase in resolution to 70,000, a new efficient cross disperser, octagonal fiber feeds to provide some scrambling, and an exposure meter. The improved efficiency should allow for observations of stars that are about two magnitudes fainter with the same S/N. The team notes that the last twelve RV measurements of sigma Draconis show a low rms of 1 m s^{-1}. These observations were taken when the outer ring of the HET mirror segments were already taken off, pointing to a possible problem with the optical quality of the camera that was improved by stopping down the pupil to stabilize the PSF.

2.6. 3.6 m European Southern Observatories (ESO) Telescope and HARPS, 2003

The HARPS spectrometer was commissioned in 2003 (Pepe et al. 2002; Mayor et al. 2003) on the 3.6 m telescope at La Silla[55] in Chile and was the first spectrograph to be specifically designed for a Doppler exoplanet survey. The instrument operates in a vacuum enclosure with a stable temperature of $17° \pm 0.°01$ C and the operating pressure is kept stable to better than 0.01 mbar since pressure changes will introduce drifts in the spectrum at the level of about 100 m s^{-1} per mbar. The fiber-fed instrument has a resolution of 115,000 and the calibration of the entire wavelength range from 380 to 690 nm is achieved with simultaneous reference from a Thorium–Argon (ThAr) lamp with a second fiber to track instrumental drifts during the night. Doppler measurements are obtained by cross-correlating each spectrum with a template mask. HARPS uses a fiber double scrambler (Avila & Singh 2008) to invert the near and far fields producing a scrambling gain of about 10,000 for consistent illumination of the optics. HARPS ushered in a new era for RV measurements, and was the first instrument to deliver better than 1 m s^{-1} RV measurement precision.

[54] Presentation by Michael Endl.

[55] Presentation by Christophe Lovis.

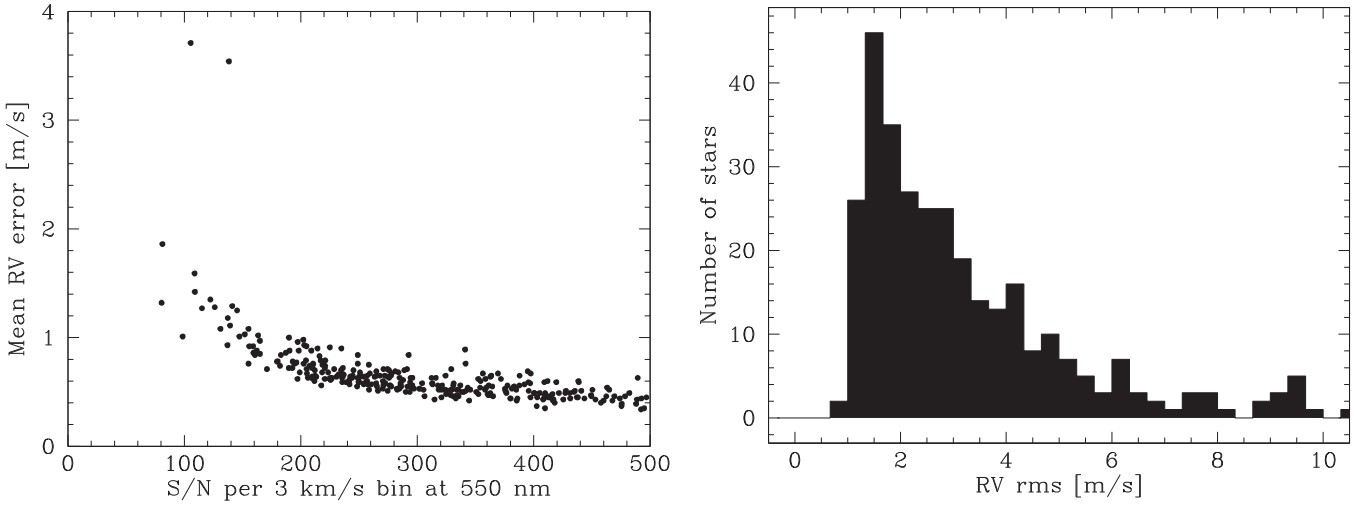


Figure 6. Single measurement precision as a function of S/N (left) and radial velocity scatter (right) at HARPS for the target sample of FGK dwarfs to search for super Earth and Neptune mass exoplanets with HARPS. (Courtesy of Christophe Lovis.)

Several exoplanet programs are carried out with HARPS, including a search of metal-rich and metal-poor stars (Santos et al. 2014), an M dwarf exoplanet search (Bonfils et al. 2013), and a search for super-Earth and Neptune mass planets around FGK stars (Pepe et al. 2011). The particular HARPS program that was highlighted at the EPRV workshop was the high precision program that focuses on detection of the lowest mass exoplanets to reveal population statistics around FGK type stars.

HARPS and HARPS-N both use the cross correlation technique (Baranne et al. 1996; Pepe et al. 2002). The internal uncertainties are calculated as the quadratic sum of three different contributions: (1) photon and readout noise, (2) wavelength calibration error, and (3) instrumental drift error. The RV uncertainty caused by photon and readout noise is obtained by propagating error bars from the spectrum to the weighted CCF (Pepe et al. 2002), and computing the fundamental RV uncertainty on the CCF using the formula in Bouchy et al. (2001). A global uncertainty on the wavelength calibration is estimated from the rms dispersion of ThAr lines around the wavelength solutions, as well as the total number of fitted ThAr lines. Finally, a drift uncertainty is also computed as the fundamental RV uncertainty on the simultaneous reference fiber. Combined together, these error sources provide a lower bound to the true error bars on the RVs. Additional contributions from stellar photospheric velocities, background contamination, and other instrumental effects should also be added in quadrature to the internal errors (Santerne et al. 2015; Díaz et al. 2016).

It is worth noting that HARPS was a significant financial investment for the ESO, reflecting a strong vision and political commitment to the improvement of spectrometers for exoplanet detection. As Figure 6 (left) shows, a velocity precision of $0.8\,\mathrm{m\,s^{-1}}$ is achieved with S/N of 200. Higher S/N improves the SMP to better than $0.5\,\mathrm{m\,s^{-1}}$. The long term velocity rms scatter shown in Figure 6 (right) is also significantly lower than the velocity scatter with the iodine technique; many stars show an rms that is less than $2\,\mathrm{m\,s^{-1}}$. HARPS pushed the instrumental errors lower by a factor of two relative to the general-purpose spectrometers with iodine reference cells for wavelength calibration.

Importantly, the team sees evidence for photospheric activity that is imprinted in the stellar spectrum by using several indicators: line bisector shapes and variations in the full width half max for the CCF, and emission in the cores of Ca II H&K lines (Queloz et al. 2001; Gomes da Silva et al. 2012; Figueira et al. 2013) The precision on HARPS is affected by guiding errors and variations in the focus, both of which lead to variable illumination of the spectrometer optics and instabilities in the SLSF Last summer, the team implemented new octagonal fibers to produce a higher scrambling gain for the light injected into the spectrograph. The instrument is also now equipped with a laser frequency comb (LFC) for a better wavelength calibration. The commissioning tests with the LFC demonstrate a dramatic improvement in precision and also identified stitching errors in the pixel format of the CCD (discussed in Section 3.1).

An important lesson learned from HARPS is that higher S/N, higher resolution, and higher fidelity spectra are needed to make significant progress on disentangling stellar activity from the center of mass (COM) velocities for chromospherically quiet stars. In response to this, the team is building ESPRESSO on the Very Large Telescope (VLT) at Paranal in Chile.

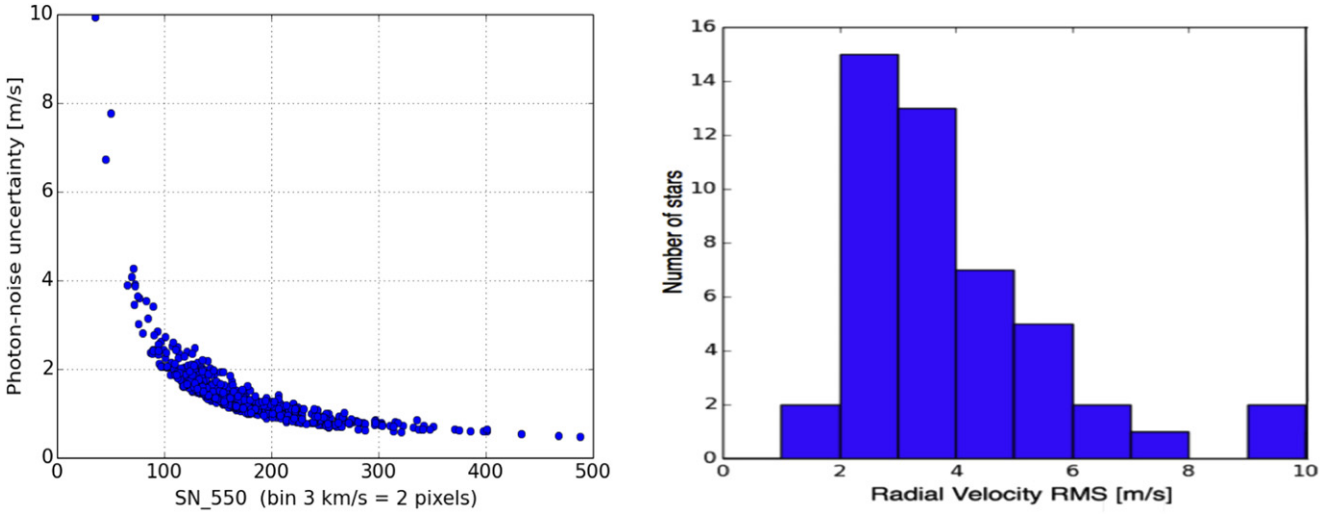


Figure 7. Theoretical uncertainty based on S/N and radial velocity scatter for the SOPHIE spectrograph. (Courtesy of Francois Bouchy.)
(A color version of this figure is available in the online journal.)

2.7. Haute Provence 1.93 m Telescope and SOPHIE Spectrometer, 2006

The ELODIE spectrograph (Baranne et al. 1996) that was used to discover 51 Peg b (Mayor & Queloz 1995) and other exoplanets with the 1.93 m telescope at Observatoire de Haute-Provence was replaced by the SOPHIE[56] spectrograph in 2006 (Perruchot et al. 2008). That instrument has a resolution of 75,000 with spectral sampling of two pixels full width at half maximum (FWHM). The wavelength range is 387–694 nm and wavelength calibration is carried out with thorium argon injected in a second fiber for simultaneous wavelength calibration. In 2011, a significant upgrade was made to the spectrometer, and an octagonal fiber was added to provide better scrambling and to improve the precision (Bouchy et al. 2013).

Among the different SOPHIE planet surveys, the high-precision program focuses on 190 G & K dwarfs using ∼16% of the telescope time distributed in fractional nightly allocations. The primary science goal is to search for Super-Earth and Neptune mass planets. There are also follow up programs for transit candidates and plans for future *TESS* follow-up.

As illustrated in Figure 7 (left) the estimated SMP errors are about 1 m s^{-1} for S/N of 200. This S/N is reached in about 15 minutes on $V = 7.5$ stars. This exposure time is adequate to average over *p*-mode oscillations in solar type stars. There are 49 stars with more than 20 measurements and Figure 7 (right) shows that the velocity rms peaks at ∼3 m s^{-1}.

One challenge with SOPHIE is the thermal coupling between the spectrograph and the telescope pillar, which introduces annual uncontrolled temperature variations. There are plans to improve the thermal control. The ThAr wavelength calibration is one of the current limiting factors in the RV precision and there are plans to implement a Fabry–Pérot (FP) etalon for simultaneous drift and wavelength calibration. A new data reduction package is also being installed with all of the improvements from the most recent HARPS pipeline.

One lesson from SOPHIE is that the charge transfer inefficiency (CTI) of the CCD is a critical parameter. In addition, scrambling of both the near and far field of the fiber output beam is important; the octagonal fiber was critical for helping to keep the RV precision of SOPHIE competitive. Finally, all available spectroscopic indicators (e.g., line bisectors, FWHM of the cross correlation function (CCF), Ca II H&K line core emission) should be used to help interpret any signals in the RV measurements.

2.8. 6 m Magellan Telescope with the Planet Finder Spectrograph (PFS) Spectrometer, 2010

The PFS (Crane et al. 2006) was commissioned at the 6.5 m Magellan telescope[57] in 2010 and continues to be used to search for exoplanets. About 30 nights are allocated in 10 day runs each semester. The echelle grating is in a vacuum enclosure and the instrument has active thermal control. This slit-fed spectrograph has a resolution of 76,000 with a 0.″5 slit.

[56] Presentation by Francois Bouchy.

[57] Presentation by Pamela Arriagada.

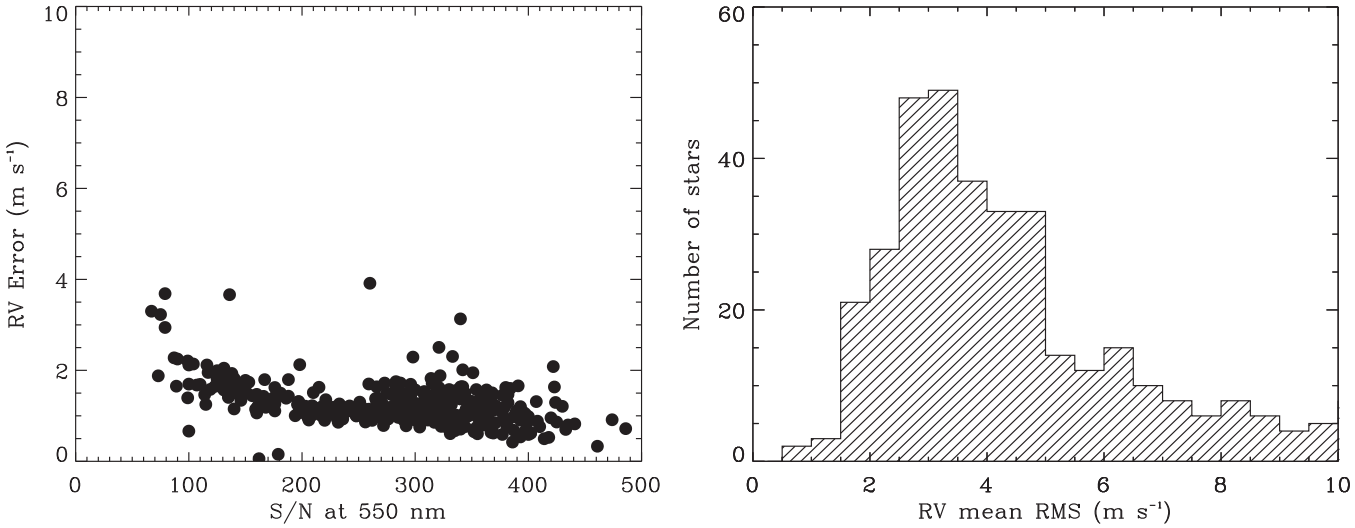


Figure 8. Estimated single measurement precision as a function of S/N (left) and radial velocity scatter for PFS. (Courtesy of Pamela Arriagada.)

The wavelength range extends from 390 to 670 nm and an iodine cell is used to calibrate the spectrum between 510 and 620 nm for Doppler measurements.

There are about 530 stars on the Magellan PFS Doppler survey. The primary goal of this program is the detection of low mass companions around stars that are closer than 50 parsecs. There is also a collaborative program to obtain Rossiter–McLaughlin measurements to determine the coplanarity of transiting exoplanet orbits and follow-up programs for the Hungarian Automated Telescope-South (HAT-S) and the *Kepler* "*K*2" mission.

Figure 8 (left) shows that the internal SMP for PFS is about 1.2 m s^{-1} at an S/N level of 200, with an rms floor (Figure 8, right) that typically ranges between 2 and 4 m s^{-1}. The typical S/N is about 300 for most Magellan PFS observations.

There are plans to refine the optical alignment of the spectrograph to achieve better image quality. Additional upgrades are being implemented this year, including replacing the CCD system with a device that has smaller pixels, better electronics and faster readout. A pupil slicer is planned that will feed six optical fibers to increase the resolving power and to introduce some scrambling.

The team is also improving the analysis to reduce the number of free parameters in the forward modeling of the iodine spectra. The higher resolution template spectra provide higher precision especially for late-type stars. High cadence observations are an advantage for detecting short-period planets during the 10 day observing runs. However, the gaps in these data sets also cause statistical sampling problems for detecting lower amplitude systems where the prospective Keplerian velocity amplitudes are comparable to the photospheric velocity jitter.

2.9. 1.5 m CTIO Telescope with the CHIRON Spectrometer, 2011

The CHIRON spectrometer (Tokovinin et al. 2013) was commissioned at the SMARTS 1.5 m CTIO telescope[58] in 2011 and is still being used for exoplanet searches. The instrument was upgraded and recommissioned in 2012 with a new echelle grating in a vacuum enclosure, two stages of thermal control (the spectrograph enclosure and an outer room maintain temperature stability to ±1 C), an exposure meter for calculating the observation midpoints for barycentric velocity corrections, an octagonal fiber feed for better scrambling, and a new CCD controller for faster readouts. The wavelength range of CHIRON extends from 440 to 650 nm with partial orders from 620 to 870 nm. Wavelength calibration is accomplished with an iodine cell, limiting the Doppler information to the wavelength subinterval from 510 to 620 nm. The bare fiber produces a spectral resolution of 28,000. An image slicer was built to increase the resolution to 90,000 and maximize throughput; however, the Doppler precision with the slicer was limited to a few meters per second, probably because of unstable illumination of the slicer and imperfect flat-fielding through the fiber slicer. Instead, a slit mask in front of the fiber is used for the high precision mode. This increases the resolution to either $R = 90{,}000$ or $R = 140{,}000$, albeit at the cost of lost light.

The primary goal of the CHIRON program was the search for planets around alpha Centauri A and B and the program also included 35 bright stars. The low throughput associated with a HRS on a small telescope required two hour exposures

[58] Presentation by Debra Fischer.

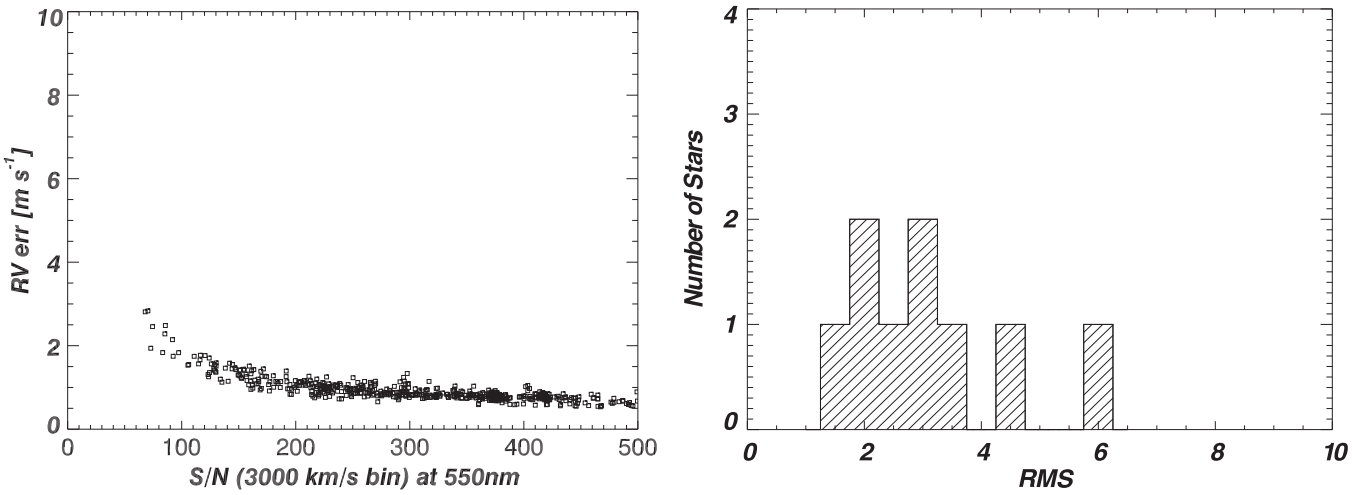


Figure 9. Single measurement precision (left) and radial velocity scatter for 10 bright stars observed with CHIRON at CTIO. (Courtesy of Debra Fischer.)

on stars fainter than $V = 5$. This ultimately lead to a program re-scope to concentrate on the brightest stars with the goal of understanding the limitations of RV precision for detection of lower mass planets. The program uses queue scheduling (Brewer et al. 2014) and this enables observations of a few bright stars on virtually every clear night.

The radial velocities and uncertainties are calculated as described in Section 2.1. At an S/N of 200, the SMP errors for CHIRON are $1\,\mathrm{m\,s^{-1}}$ and the SMP improves to $0.5\,\mathrm{m\,s^{-1}}$ as the S/N approaches 500. Figure 9 (left) shows the dependence of estimated internal errors on the S/N of the observations. Figure 9 (right) is a histogram of the rms of the velocities (no trends or Keplerian signals were removed) with a minimum rms of $2\,\mathrm{m\,s^{-1}}$. The fairly narrow range in rms velocities occurs because the program concentrated on chromospherically quiet stars.

The alpha Cen survey suffered from increasing contamination as the separation of the two binary stars decreased from 6″ in 2011 to 4″ in 2013 (when the project was put on hold). The contamination was significant with 2.″7 fiber, selected to accommodate the frequent poor seeing at the observatory. A technique was developed to fit for scaled flux contamination when modeling the RV of each star, however this only provided a precision of $3\,\mathrm{m\,s^{-1}}$.

CHIRON demonstrated the advantage of a fiber-fed instrument and very high cadence Doppler observations. The biggest challenge with CHIRON is the struggle to obtain the high S/N needed for $1\,\mathrm{m\,s^{-1}}$ precision with the iodine technique; this is difficult on a small aperture telescope and there were additional challenges with poor guiding and telescope control system failures. As a result, the program goals shifted to understanding the limitations of RV precision by focusing on a small subset of the very brightest stars.

Perhaps the most important lesson from CHIRON is that even with a more stable instrument, parameter degeneracies in forward modeling were identified that ultimately limit the RV precision (in addition to stellar jitter) with the iodine technique (Spronck et al. 2015). While a more precise wavelength calibrator that spans a broader wavelength range on CHIRON is desirable, the instrument is not in a vacuum enclosure and CHIRON was not designed to be stable enough for simultaneous thorium argon or LFCs. The lesson here is that a change in wavelength calibration is not always an easy retrofit; it is something that needs to be designed into the instrument from the start. The team is now building on lessons learned with CHIRON in the design of a new spectrometer, the EXtreme PREcision Spectrograph (EXPRES) for the 4.3 m Discovery Channel Telescope (commissioning is planned for summer 2017). This vacuum-enclosed white pupil design spectrometer will have a resolution of $R = 150{,}000$, a broad wavelength range, and calibration with a LFC.

2.10. Paras

PARAS (PRL Advanced Radial-velocity Abu-sky Search) is currently India's only exoplanet search and characterization program[59] and started science observations in 2012 (Chakraborty et al. 2010, 2014). PARAS obtains a single-shot spectral coverage of 3800–9500 Å at a resolution of 67,000 with 4 pixel sampling. RV measurements are calculated using simultaneous wavelength calibration with a ThAr hollow cathode lamp. The spectrograph is maintained under stable temperature conditions and enclosed in a vacuum vessel. The enclosure has active temperature control to correct for thermal conduction through the concrete pier under the spectrograph; temperature is monitored both inside the vacuum chamber and in the room,

[59] Presentation by Abhijit Chakraborty.

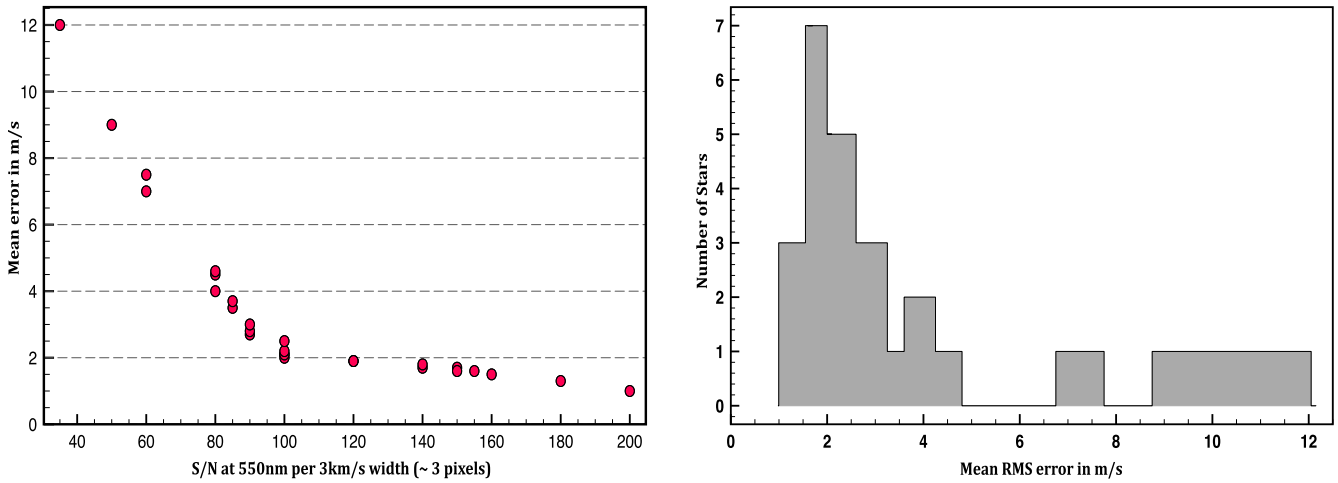


Figure 10. Precision and radial velocity scatter for PARAS. (Courtesy of Abhijit Chakraborty.) The plot on the left shows the nightly RV scatter as a function of S/N in 3 km s^{-1} bins and the plot on the right is a histogram of the long term rms scatter for 27 bright stars.
(A color version of this figure is available in the online journal.)

and corrections for the heat loss or gain are applied by changing the heater power. This provides an rms thermal stability of 25 ± 0.009C during the night. The typical vacuum stability is about 0.05 millibar during a night of observations. An e2v deep depletion 4k × 4k CCD is used for imaging the spectra along with the ARC Inc., Leach Controller for CCD electronics and interface with Computer. Simultaneous ThAr exposure along with a ThAr exposure immediately after the observation is used to correct for instrument drifts with an rms error of 90 cm s^{-1} or better (Chakraborty et al. 2014).

The blaze peak efficiency of the spectrograph between 5000 and 6500 Å, including the detector, is about 30%; and about 25% including the fiber transmission. The total efficiency, including spectrograph, fiber transmission, focal ratio degradation (FRD), and telescope (with 81% reflectivity considering primary and secondary mirrors) is ∼7% in this same wavelength region.

The PARAS data analysis pipeline (Chakraborty et al. 2014) is custom-designed and based on the REDUCE routines of Piskunov & Valenti (2002). The pipeline performs the routine tasks of cosmic ray correction, dark subtraction, order tracing, and order extraction. A thorium line list is used to create a weighted mask which calculates the overall instrument drift, based on the simultaneous ThAr exposures.

RVs are derived by cross-correlating the target spectra with a suitable numerical stellar template mask. The stellar mask is created from a synthetic spectrum of the star, and contains the majority of deep photospheric absorption lines (Pepe et al. 2002, and references therein). RV measurement errors include photon noise errors (Bouchy et al. 2001) and other errors like those associated with fitting the CCF. Typically, two to three RV points are used to calculate the nightly mean velocity. Chakraborty et al. (2014) demonstrate a mean nightly rms scatter of 1.2 m s^{-1} for the star Sigma Draconis over a period of seven months and 1 m s^{-1} for HD 9407 over ∼30 days. Additional velocities now show 1.4 m s^{-1} over two years for Sigma Draconis and 1 m s^{-1} for tau Ceti over a period of 40 days.

Figure 10 shows the estimated measurement precision of PARAS for 27 bright stars since the beginning of observations in 2012. These sources have at least 10 epochs of observations spanning more than 30 days. On a clear photometric night with 90% reflectivity on the telescope primary and secondary mirrors, an S/N of 200 per 3km s^{-1} RV bin is reached in about 15 minutes on a 6.5 mag star.

The small telescope aperture of 1.2 m with its large tracking errors (of the order of 1″–1.″5) is one of the greatest short comings of the project. A combination of octagonal and circular fibers (without a double scrambler) helps to stabilize illumination of the spectrograph optics. Instead of a pinhole at the telescope focal plane, a focal reducer is used to focus the F/4.5 beam on the octagonal fiber. The image of the star slightly overfills the fiber which is about 2″ on the sky. This helps to take care of telescope tracking and atmospheric scintillation issues at the cost of some light loss.

The team plans to move PARAS to a new 2.5 m telescope in late 2019 or early 2020 with a fast tip-tilt image stabilizer and an atmospheric dispersion compensator (ADC) at the Cassegrain focus for proper injection of light into the fiber. The new telescope will have rms tracking and jitter errors of less than 0.″1. The temperature control and spectrograph room will be redesigned to alleviate the present deficiencies like the heat conduction issues from the pier and to achieve thermal stability of 25 ± 0.001C. This will eliminate the need for taking large

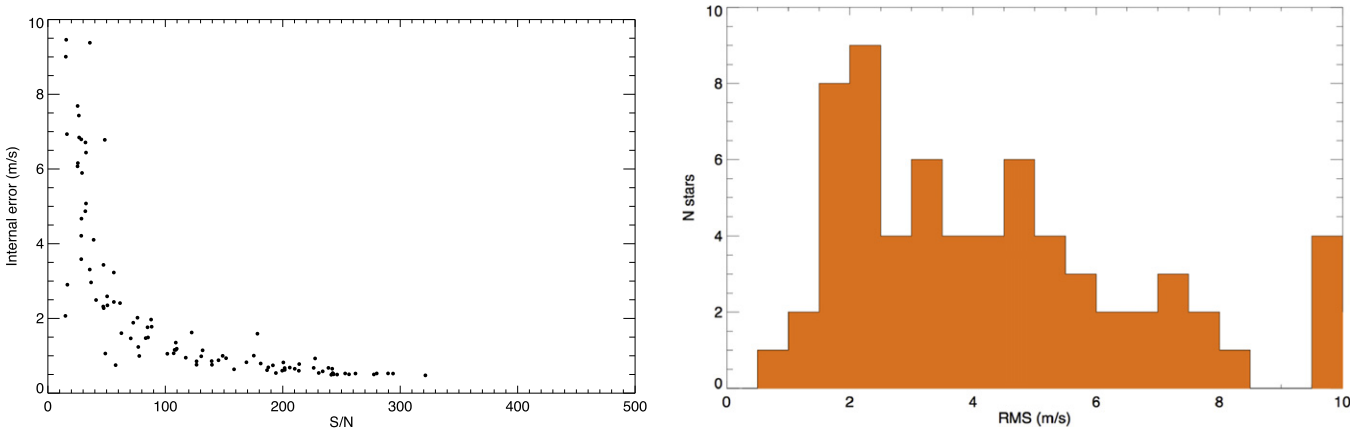


Figure 11. Histograms of the estimated SMP and radial velocity scatter for stars with more than 10 observations at HARPS-N. (Courtesy of Lars Buchhave.) (A color version of this figure is available in the online journal.)

number of ThAr calibrations during the night, providing additional time for more stellar observations.

2.11. 3.5 m Telescopio Nazionale Galileo (TNG) Telescope with the HARPS-N Spectrometer, 2012

In 2012, a near twin of HARPS was commissioned at the TNG observatory.[60] The vacuum enclosed instrument, HARPS-N, has temperature and pressure stabilization, a resolution of 115,000 and wavelength range from 383 to 693 nm. HARPS-N was initially commissioned with ThAr simultaneous calibration, with a planned upgrade to an optical frequency comb. The instrument is fed by a combination of octagonal fiber and double scrambler, which inverts the near and far fields. A failure of the CCD delayed the debut of this instrument, but regular operations are now underway.

HARPS-N is being used for several collaborative exoplanet detection programs (Cosentino et al. 2014), including the Global Architecture of Planetary Systems (GAPS; Covino et al. 2013) program. In summer 2015, a novel program was started to carry out solar observations with the goal of reaching sufficient measurement precision to detect Venus (Dumusque et al. 2015a). The solar observations are one example of a project that is not photon starved, and uses high cadence (5 minutes) for about 4 hr each day, using a compact solar telescope to feed the HARPS-N spectrograph. When using the HARPS-N LFC, the error from the wavelength calibration is only 6 cm s^{-1} and the estimated precision is a few meters per second over the course of a week. The project uses a photometric $F - F'$ correction (Aigrain et al. 2012; Dumusque et al. 2015a) from the *Solar Dynamo Observatory* images to reduce this to 1 m s^{-1}.

The velocities and uncertainties are calculated as described in Section 2.6. Figure 11 shows that the measurement precision at HARPS-N is similar to HARPS. The estimated SMP for S/N of 200 is between 0.5 and 1 m s^{-1}. The rms velocity scatter peaks at about 2 m s^{-1}.

RePack is the name of a Doppler pipeline for HARPS-N, written by Lars Buchhave. Here, the internal uncertainties are calculated by cross correlating either a high S/N observed spectrum or a co-added combined spectrum (i.e., a template) against the observed spectra. Typically, a delta function template is used. This template consists of weighted delta functions based on the measured line centers of known absorption lines, so that each line can contribute to the CCF with some weight. Individual lines have a relatively poor wavelength determination, often with several hundreds of meters per second uncertainties. The final RV of each order can thus be several hundreds of meters per second offset compared to the other orders because of the (combined) uncertainties in the position (wavelengths) of the delta function lines. The offset is the average offset between the weighted delta function lines compared to the actual stellar spectrum. This can create problems when the spectrum is shifted around by the barycentric correction, because lines will move from higher to lower and lower to higher S/N regions on the detector (because of the blaze function). In other words, the combined RV offset of the order can change because e.g., a strong line might move toward lower S/N if it moves toward the edge of the order.

In contrast, an observed template will yield RVs for the individual orders that are more realistically scattered around the RV shift between template and observation. With RePack, an RV measurement is determined for each individual order using a CCF and the internal uncertainty is calculated as the photon weighted rms of the RVs for each order, divided by the square root of the total number of orders used. The final RV is measured by co-adding the CCFs for each order and then determining the position of the co-added CCF peak.

[60] Presentations by David Phillips and Lars Buchhave.

The precision of HARPS-N allows for the detection of additional errors that are normally masked by poorer precision data. A 3 minute thermal cycle on the cold plate in the CCD cryostat at HARPS-N was identified and the thermal cycling is believed to produce an expansion-contraction cycle that slightly shifts the position of the CCD. The temperature variations from the cold plate probe correlate directly with wavelength (velocity) shifts in the thorium argon calibration lines for short 20 s exposures. The stability of the wavelength solution was restored by taking 5 minute thorium-argon observations to average over the cold-plate jitter cycle. Two different data analysis pipelines have been tested at HARPS-N; this was a useful exercise that helped to identify new areas for improving the RV precision.

2.12. 2.4 m APF with the Levy Spectrometer, 2013

The 2.4 m APF (Radovan et al. 2010; Vogt et al. 2014) is located at Lick Observatory[61] and the Levy spectrograph was commissioned at the APF in 2013. The instrument uses a narrow slit to reach resolutions up to 150,000. Like Keck-HIRES, the wavelength calibration is carried out with an iodine cell and therefore the wavelength range used to derive Doppler velocities is limited to 510–620 nm. However, the spectra span from 374 to 950 nm, enabling the use of chromospheric diagnostics like the Ca II H&K lines.

The telescope time on the APF is divided between two groups; the Lick-Carnegie Planet Search (LCPS) team at UC Santa Cruz observes about 100 stars and the CPS team at UC Berkeley and the University of Hawaii observes a few hundred stars. The science goals of both teams include detection and characterization of low mass exoplanet systems and follow-up observations to better resolve components of multi-planet systems. In the future, the APF will provide support for the NASA *TESS* mission. A big advantage of the APF is that it is a robotic instrument and the observing scripts can be launched from remote sites (Burt et al. 2015). The unique combination of large telescope time allocation and small user group allows the facility to obtain very high cadence measurements.

The CPS team aims for a fixed target S/N of 200 per observation; this yields estimated internal errors of about 2 m s^{-1}. Typically many observations are taken for each star during a given night and the velocity is averaged to give better precision. Because the CPS team is observing stars with low chromospheric activity and a known legacy of Doppler velocities from Keck HIRES, the rms of the velocities is a tighter distribution than Keck HIRES, centered on 3–4 m s^{-1}. Figure 12 (left) shows the dependence of estimated internal errors on the S/N of the observations. Figure 12 (right) is a histogram of the rms of the velocities (without attempting to remove any trends or Keplerian signals).

The Lick Carnegie team has estimated errors of about 1.5 m s^{-1} per observation at an S/N of 200 and 1 m s^{-1} at S/N of 400 and the velocity rms is typically greater than 2.5 m s^{-1}. Figure 13 (left) shows the dependence of estimated internal errors on the S/N of the observations. Figure 13 (right) is a histogram of the rms of the velocities (no trends or Keplerian signals were removed).

The precision of the Levy spectrometer is limited to about 1 m s^{-1} by the iodine reference cell. Although the flexible scheduling has been critical for this program, only a relatively small number of observations can be taken each night because of the modest 2.4 m telescope aperture.

2.13. Hertzprung 1 m Telescope with the SONG Spectrometer, 2014

The SONG spectrometer (Grundahl et al. 2011) was commissioned in mid-2014 at the Mount Teide Observatory in Tenerife.[62] The instrument has a series of slits and is operated with a spectral resolution of 90,000 and an optical wavelength range of 440–690 nm. The instrument is not in a stabilized vacuum enclosure so an iodine reference cell is used to measure Doppler shifts between 510 and 620 nm. The detector is a 2k by 2k Andor device with 2 s readout, suitable for measuring the oscillation periods of stars. There is a fast tip/tilt system to help stabilize the slit illumination.

The science program focuses on a relatively small sample of twelve stars with the goal of obtaining asteroseismology data, reaching a better understanding of stellar physics and detecting exoplanets. Except for time lost to weather, this is a dedicated facility that obtains very high cadence observations every night. The telescope will be part of a network that spans the globe for nearly continuous observations of stars. The estimated SMP error (Figure 14) is about 2–4 m s^{-1} with an rms of about 4 m s^{-1}. The best precision is obtained for K dwarfs, but there is still large scatter in the RV errors.

SONG is optimal for bright targets. However, the program is new, so the data analysis pipelines are still being developed. The iodine cell is in a collimated beam and demonstrated fringing; this cell has just been replaced with a cell that has wedged optical windows that are counter-rotated. The small 1 m telescope aperture means that the stars must be brighter than V of 6.5 for reasonable exposure times and high S/N.

2.14. Future RV Programs

In addition to these ongoing Doppler planet searches, there are several funded new spectrometers that will come online in the few years, including the first node of Network of Robotic Echelle Spectrographs (NRES) on the LCOGT telescopes in 2015 with a target RV precision of 3 m s^{-1} (Eastman et al.

[61] Presentations by Greg Laughlin and Andrew Howard.

[62] Presentation by Frank Grundahl.

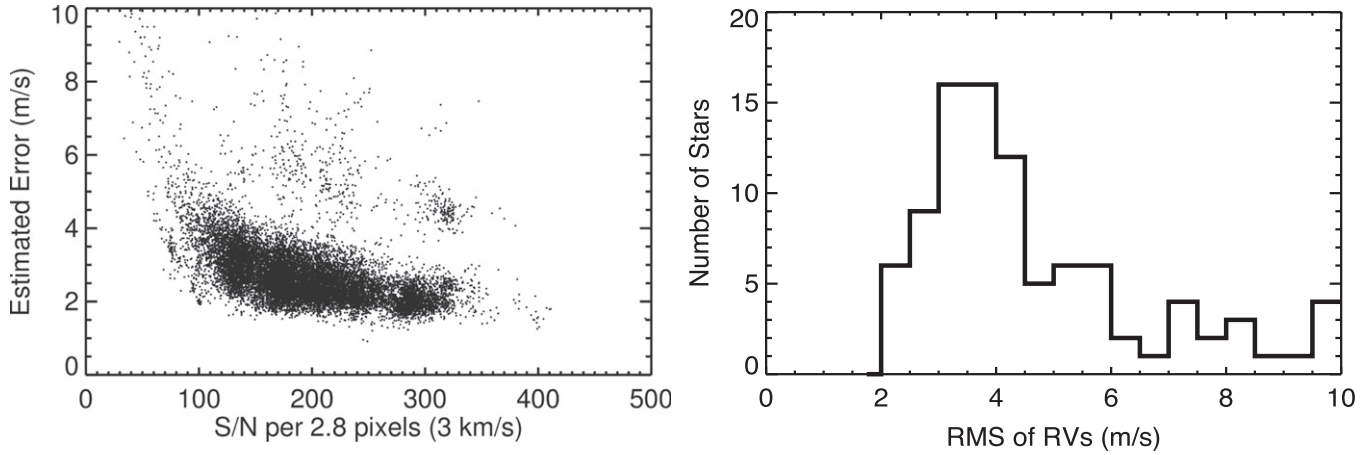


Figure 12. Estimated precision (left) and radial velocity scatter (right) for Doppler measurements obtained by the California Planet Search team using the Levy Spectrograph at the Automated Planet Finder. (Courtesy of Andrew Howard.)

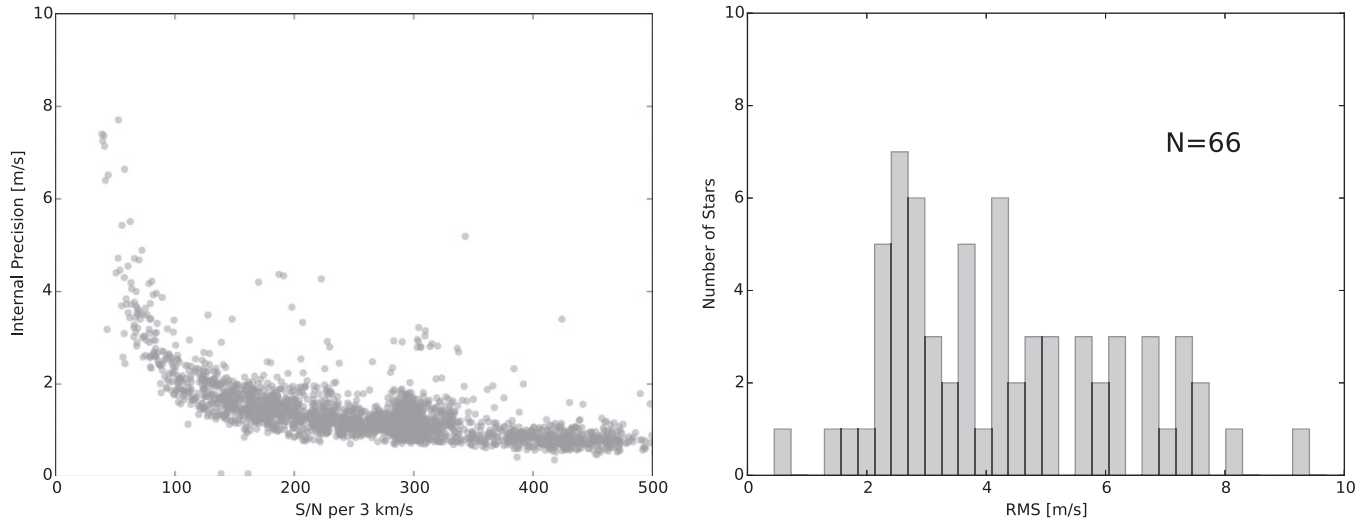


Figure 13. Radial velocity precision as a function of S/N (left) and velocity rms (right) for the Lick Planet Search team with the Levy Spectrograph at the Automated Planet Finder. (Courtesy of Greg Laughlin.)

2014), CARMENES in 2015 at the 3.5 m Calar Alto Observatory (Quirrenbach et al. 2014), HPF in 2016 on the 9 m HET (Mahadevan et al. 2012), MINERVA 1 m telescope array in 2016 at Mt. Hopkins (Swift et al. 2015), *Veloce* in 2016 at the AAT; ESPRESSO in 2017 at the 8 m VLT (Pepe et al. 2013), EXPRES in 2017 on the 4.3 m Discovery Channel Telescope at Lowell Observatory, SPIRou in 2017 at the 3.6 m Canada–France–Hawaii Telescope (Artigau et al. 2014b), the NNEXPLORE spectrograph at the 3.5 m WIYN telescope in 2018, iLOCATOR at the 8.4 m Large Binocular Telescope in 2018 (Bechter et al. 2015) and G-CLEF in 2020 at the GMT (Szentgyorgyi et al. 2012). For a more comprehensive list of future instruments, see Tables 2 and 3 in Plavchan et al. (2015).

3. Instrumentation Challenges

Efforts to identify confusing sources of velocity scatter from stellar photospheres will be more effective if the instrumental precision is securely below the target RV precision for the stars.[63] These requirements set the top level specifications for observing cadence and spectral resolution, which in turn dictates the wavelength bandpass, type of wavelength calibration, S/N, precision for the barycentric correction, and required stability for the spectral line spread function (Connes 1985; Bouchy et al. 2001; Podgorski et al. 2014).

Asymmetric white pupil spectrograph designs offer several advantages for PRV spectrometers. The overall size of the

[63] Presentations by Andrew Szentgyorgyi and Suvrath Mahadevan.

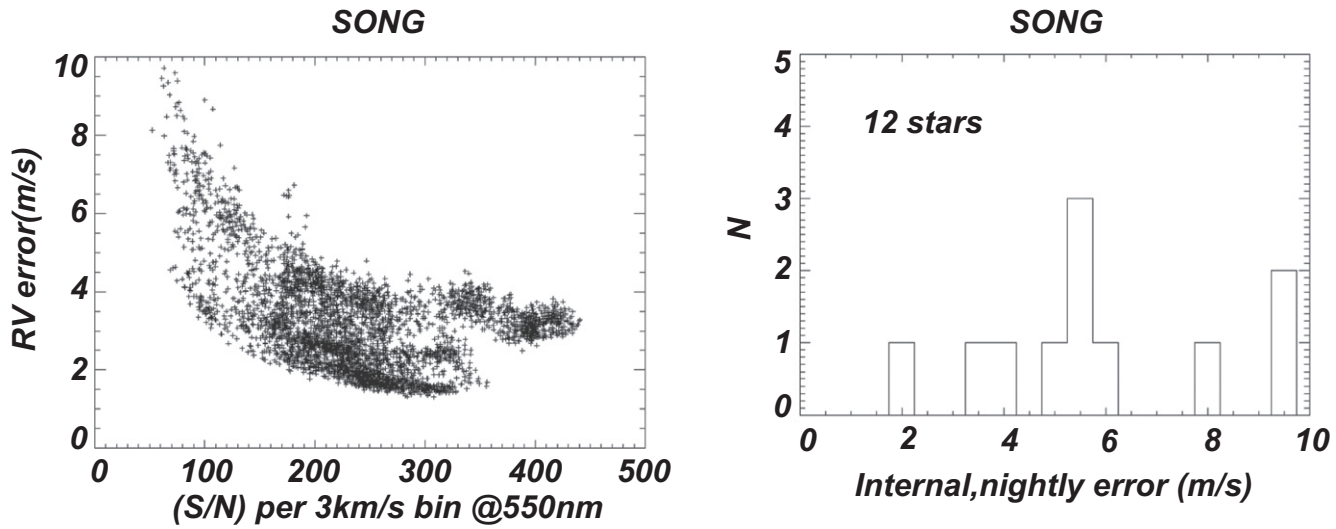


Figure 14. Radial velocity precision as a function of S/N for SONG and the histogram of nightly errors for 12 stars. One star has internal nightly errors that are greater than 10 m s^{-1}. (Courtesy of Frank Grundahl.)

optics and the instrument is reduced, and vignetting and aberrations are reduced. The white pupil design offers higher efficiency and better image quality than conventional designs with the same resolution. White pupil designs have intrinsic cylindrical field curvature that is traditionally mitigated with a toroidal (cylindrical) field flattener that must be placed very close to the detector. A curved CCD would be an ideal solution, but these are not commercially available. Instead a Mangin mirror can be used in place of the folding transfer mirror to compensate for cylindrical field errors. However, care must be taken since the Mangin mirror can produce significant ghosts if it is not wedged.

High velocity precision requires high S/N and broad wavelength bandpass, so a telescope of at least a modest aperture is required for all except the brightest stars. The size of spectrometer optics increases with the size of the telescope unless clever designs are implemented (e.g., the anamorphic design of ESPRESSO or single mode fibers (SMFs) behind adaptive optics systems). The need for both high resolution and broad bandwidth often implies that two (or more) cameras will be required, each with its own challenges. Often, the efficiency for the blue arm of the instrument is challenging because of the lower transmission of optical glasses below 420 nm and the fibers must be kept short to minimize attenuation of light. Spectra obtained from the red arm of the spectrometer have telluric contamination that shift across the rest frame of the stellar spectra because of the barycentric velocity of the Earth.

Camera designs for spectrometers fall into one of two categories. Standard designs have 7–9 lenses that are independently testable. Fabrication and alignment of these (mostly spherical) optics is straightforward; however, there can be a significant hit in terms of efficiency, ghosting and scattered light. New high-dispersion glasses are now available (Ohara and Nikon i-Line) that can reduce the number of optical elements needed for standard camera designs. The second type of camera is a pupil-folded camera that uses aspheric components. These are not independently testable and carry some risk in manufacturing and alignment; however, the optics are very efficient with few ghosting and scattered light problems.

The need to precisely centroid spectral lines and to track line profile variations from stellar photospheres (granulation, spots, plage, variations in long-term magnetic fields) sets top-level requirements for environmental stability. The solution from HARPS was to put the instrument in a vacuum chamber with vibration isolation. Temperature and pressure in vacuum enclosures are routinely controlled at the level of 10 mK and 10^{-2} Torr respectively; at this level, refractive index variations are no longer a problem. However, temperature stability below 1 mK and pressure stability below 10^{-7} Torr is desirable because this will stop molecular transfer, providing a system that is almost completely radiatively coupled. This is the state of the art for infrared instruments with liquid nitrogen tanks and charcoal or zeolite getters for cryo-pumping.

The ease of environmentally stabilizing a spectrometer is inversely proportional to the size of the vacuum enclosure. Most enclosures are high grade steel, which is easy to weld, or aluminum, which is more difficult to weld. Optical benches have been made out of aluminum (NRES, KiwiSpec, HPF), mild steel (HARPS, ESPRESSO), invar and composite material. All have virtues and drawbacks. Aluminum is light weight, low cost, and low technical risk, but may pose a high

performance risk for RV precision because of the thermal properties that must be carefully controlled. Steel is heavy, but easy to process, with high thermal inertia and low cost. Invar is heavy with high thermal inertia; it is difficult to manufacture, posing a moderate technical and RV performance risk. The design trade study for G-CLEF found that carbon fiber epoxy met the systems engineering requirement with excellent conductivity and low thermal inertia. There is some technical risk since this is a new material and it is known to outgas water. However, in a vacuum, outgassing will decrease over time to acceptable levels (may need to pump for about a year).

As the spectrometers become more complex, a higher level of systems engineering is required with well-considered budgets for throughput, RV precision, sequencing, timing, and delivery (e.g., Podgorski et al. 2014). New software programs are being developed to integrate these error budgets and to track compliance; this makes it easier to understand the design impact if, for example, a vendor delivers a component that does not meet the specification (many components are still contracted to vendors with a best effort commitment). The error budget should distinguish between error sources that can be calibrated out and those that cannot. In their error budget for G-CLEF, Podgorski et al. (2014) describe error sources that affect the calibration and observations equally and therefore can be calibrated out with an extremely stable instrument. These include:

1. Thermal stability: thermo-elastic distortions of the spectrograph can shift the spectra and affect focus. For G-CLEF temperature loads of 0.001 C were applied in the FEA model (using the Smithsonian Astrophysical Observatory Bisens software) and a lateral motion of 110 (55 cm s^{-1}) was identified; the solution was to adopt lower CTE materials and improve thermal control.
2. Mechanical stability: the spectrometer is affected by changes in atmospheric pressure outside the enclosure that result in deformation of the spectrograph causing lateral shifts and focus errors; vibrations can also cause lateral shifts, focus errors, and PSF broadening; instability of materials can occur; for example from moisture desorption of the composite bench with G-CLEF.
3. Pressure changes: because the wavelength calibrator follows the same path and column density as the program observations this can be calibrated out if the drift is slow compared to the observation times.
4. CCD stitching errors: the lithographic process for making CCDs steps the mask in discrete steps, leaving positional errors after blocks of 512 pixels. As Doppler shifts or barycentric motion causes the spectral line to cross one of these regions, a spurious wavelength shift will be measured (Dumusque et al. 2015b). According to Paul Jorden, the stitching errors have been reduced by a factor of three with the current e2v devices.

Other error sources that cannot be calibrated out (Podgorski et al. 2014), include:

1. Barycentric errors: current state of the art algorithms, when combined with precise exposure meters, provide corrections good to 2 cm s^{-1}.
2. Software fitting errors: affect both the calibrator and extraction of the science spectra.
3. Micro-vibration: it is possible to obtain ~1 mg vibration control, but smaller vibrations can be coupled to the detector and would broaden the lines.
4. Detector errors: including controller readout errors, intra-pixel quantum efficiency (QE) variations, non-identical pixel sizes, detector heating on readout, dependence of CTE on amount of carried charge.
5. Stray light: to minimize this, use baffling and ghost or scattered light analysis in the instrument design and keep the echelle orders well-separated.
6. Tracking: this includes both telescope and on-instrument tracking. For negligible contribution to the RV error, displacements should be less than a quarter of the fiber diameter with a scrambling gain of 10,000. This is possible with octagonal fibers and a double scrambler to invert the near and far field.
7. Telescope focus errors and variable seeing conditions will lead to variable instrument illumination, though translating this variation int an RV error is difficult.
8. Imperfect ADC: will vary the spatial illumination of the optics in a wavelength dependent manner.

Imperfect atmospheric dispersion compensation can potentially introduce systematic effects in the measurements. For example, for fiber-fed spectrometers, the photocenter of the image might not fall at the center of the fiber at all the wavelengths. This could change the measured spectral energy distribution, introducing time-correlated changes in the slopes of the echelle orders. Data reduction and extraction should account for this, and algorithms must be tested against the range of changing conditions during commissioning. A second effect is that chromatic dependence in the fiber illumination may couple with scrambling efficiency and produce variations in the spatial illumination of the optics in a wavelength dependent manner.

3.1. Detector Technology

CCDs are the detector of choice for optical spectrographs due to their high QE, low noise, large dynamic ranges, excellent linearity, and large pixel format substrates.[64] CCD detector technology continues to provide a steady improvement in performance along with new and innovative features, which help to make extreme precision RV measurements possible.

[64] Presentations by Andy Szentgyorgyi and Suvrath Mahadevan.

This section summarizes the current state of the art for CCD technology and highlights some of their limitations in optical spectrograph performance.

CCDs are sensitive over a broad wavelength range from 350 to 950 nm, but require anti-reflection (AR) coatings to achieve high QE. Simple coatings must be optimized for specific wavelength ranges and thus are not ideal for the broad wavelength ranges of some spectrographs. More sophisticated graded coatings are available, which have spatial gradients in the AR coating to match the spectrograph bandpass. The spatial pattern of the AR coating can be customized for the spectrograph design—optimized spectral response is ideal for fixed format spectrographs. In addition to improving the QE, the AR coatings will minimize reflections and thus minimize ghosts and fringes.

Inherent to CCD fabrication there are slight variations in the pixel spacing and geometry that can introduce errors in extreme precision RV measurements. Construction of the CCDs manufactured by e2V[65] starts with a metal mask with a pattern of amplifiers, registers and image region that defines the structure of the device. The image area of the CCD is fabricated with a photolithography process by projecting a block of 512×512 pixels onto the silicon image region. The number of pixels in a block is limited by the field flatness of the camera. To produce CCDs larger than 512×512 pixels requires that the block mask be stepped and reimaged (stitched) multiple times to form larger area devices. A stepper motor moves the camera to position for projecting the adjacent block images and each block has a random alignment error relative to the stepper reference frame as illustrated in Figure 15.

Two machines are used by e2V for the stitching process. The Ultratech stepper will lay down up to two blocks in x and multiple blocks in y with an alignment precision of 250 nm. The stitching errors for these devices can be seen by eye where the adjacent blocks overlap. The newer Nikon stepper will lay down multiple blocks in x and y with a stepper alignment precision of 90 nm. While each stitched block is the same by design, there will be alignment differences between the blocks. Even with perfect alignment at the edges of the blocks, there can be a measurable boundary effect because of overlapping of the masks. Both pixel size variations and stitching errors can be calibrated with a LFC or laser metrology (Shao et al. 2013).

Pixels on either side of the boundary can have slightly different dimensions. There are magnification and distortion errors in the photolithography process that can influence alignment. The pixel sizes have random uncertainties of ± 150 nm (or 1% for a 15 micron pixel) from the tolerances of the lithographic process.

CCDs and complementary metal-oxide-semiconductor (CMOS) chips exhibit pattern noise caused by QE variations, multiple output amplifiers, surface defects, and process variations. The pattern noise is small in modern detectors and fixed structure can be removed with calibration images. CCD output amplifiers can be manufactured with low read noise (<2 e^- rms) and consistent electrical gain so that multiple outputs do not contribute significantly to pattern noise. Process variations in detector fabrication, such as pixel size, thin film thickness, doping, metallization, and impurities cause a pixel response non-uniformity (PRNU). The pixel response uniformity in back-thinned CCDs is best at 650 nm, with the non-uniformities increasing at UV wavelengths from laser annealing process and at red wavelengths due to fringing. The PRNU can be improved with a good AR coating and thick silicon devices.

To improve the QE response in the red and near IR, deep depletion CCDs are fabricated with high-resistivity silicon material (high ρ) and have less silicon removed during the thinning process to produce thick (50–200 micron) devices. This results in the pixel sky scraper effect where the height of the pixel is very large compared to its width: for example, the LSST chips are 100 microns thick with 10 micron pixels. The high ρ silicon allows a greater depletion layer within the tall pixel that minimizes the field-free region near the surface and allows electrons produced there to be more readily captured and collected within the pixel. This minimizes charge diffusion effects—electrons wandering into neighboring pixels—and improves the modulation transfer function (MTF), which is a measure of the resolution capabilities of the CCD. Conversely, the high ρ silicon also introduces electrostatic fields in the CCD that could adversely affect charge collection: when one pixel

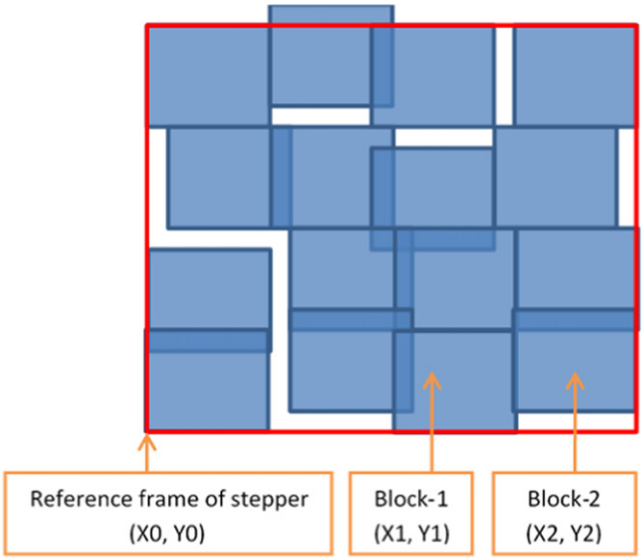


Figure 15. Example of the pixel image blocks required to create a 2k × 2k pixel CCD with exaggerated random position errors relative to the reference frame (image credit to e2v technologies).
(A color version of this figure is available in the online journal.)

[65] Presentation by Paul Jorden.

fills up with signal charge it can influence the collection of charge in neighboring pixels because of these electrostatic effects (Weatherill et al. 2014). In other words, the high ρ silicon improves the MTF, but subtle variations in the MTF are possible as pixels fill up with charge and change the electric field lines in surrounding pixels. Deep depletion devices are also available and can exhibit tree ring patterns from resistivity variations in silicon boule (Plazas et al. 2014).

CTI is a measure of charge transfer across each pixel and is typically between 0.0010% and 0.0001%. CTI losses occur in traps that collect and hold some number of electrons before releasing them at a later time. For a 10,000 e^- signal, a CTI of 0.001% corresponds to a loss of 100 electrons per 1000 pixel transfers. This generally results from 100 single electron traps within a 1000 pixel row or column. CTI is an average of charge trapping events over the whole detector array. CTI is a function of temperature and transfer frequency (clocking) as well as the signal level per pixel (Bouchy et al. 2009). Higher clock voltages can improve CTI but also introduce the risk of clock-induced charge (CIC) trapping, although tri-level clocking will reduce CIC. On the larger format devices that are required for high-resolution spectrometers, the process of transferring charge can dissipate power in the silicon chip and cause warping that can affect the measured RV. The ESPRESSO team estimates that 2 nm rms stability is required for 10 cm s^{-1} RV precision.

Charge traps occur in discrete locations and are stable so that it is possible to identify and characterize them using the pocket-pumping technique (Wood et al. 2014). Because radiation in space produces ongoing deterioration of the detector by creating traps, EUCLID plans to use pocket-pumping to identify the charge traps and then apply post processing corrections. Detectors used for ground-based facilities will not suffer from ongoing degradation and can probably be characterized in the lab prior to installation in the instrument; the measurement of charge traps would allow a correction to be applied to correct for CTI.

Silicon has a thermal coefficient of expansion of 2.6×10^{-6}/C and thus contracts on cooling; a 4k 15 micron thick sensor will shrink by 12 microns when cooled. For large format CCDs it is very important that the silicon is bonded to a package with a similar expansion rate to prevent stresses from building up as the devices are cooled. Silicon carbide (SiC) is an ideal material for CCD packages because it has a high thermal conductivity and low density with high strength, and a coefficient of expansion (4.0×10^{-6}/C) that is well matched to the silicon CCD. Both e2V and STA offer SiC packages for their large format CCDs.

The detector technology described in this summary was focused on specific detectors currently in use in precision optical RV spectrographs. It is by no means an exhaustive analysis of all detector options and features available today. This summary instead highlights some of the important types of considerations that need to be made when using CCD detectors in extreme precision RV spectrographs. Through a combination of advancements in detector technology and better calibration techniques, the detector performance required to achieve 10 cm s^{-1} precision is within reach.

3.2. Wavelength Calibration

Historically, wavelength calibration for astronomical observations has been carried out with emission lamps or absorption cells. These wavelength calibrators have a price performance that is hard to beat.[66] A crisis has been precipitated by a decision by all suppliers, globally, of hollow cathode lamps to cease the manufacture of ThAr hollow cathode lamps with metallic cathodes. All ThAr lamps are now made with thorium oxide, which introduces impurities into the cathode. These impurities produce undesirable spectral features—sometime referred to as "grass"—which quantifiably compromise wavelength calibration of cross dispersed echelle spectrographs.[67] These traditional calibrators have inherent limitations that preclude a Doppler precision of better than about 1 m s^{-1}. The ThAr spectrum changes as the lamp ages and the emission lines are broad, saturated and irregularly spaced. If the ThAr spectrum passes through a fiber that is adjacent to the science fiber, the calibration is made for pixels that are offset from the science spectrum and an ultra-stable spectrometer is required for a reliable interpolation. Furthermore, there is no conventional lamp that provides adequate calibration in the near infrared (NIR); charge from argon lines diffuses into neighboring pixels and the thorium lines do not have adequate density.

The iodine reference cell technique is ideal for an unstabilized or general-purpose spectrograph, because the forward modeling technique accounts for variations in the spectral line spread function and modeling of the wavelength and dispersion with every observation. However, the iodine technique has a precision that is limited to ~1 m s^{-1} for several reasons. First, the wavelength region is limited to 510–620 nm, where the molecular iodine lines form, and there are light losses of about 25% from the iodine cell. The superimposed absorption spectrum also masks line profile variations that might be diagnostic of stellar activity. In addition, the intrinsic iodine spectrum is required for the forward modeling process; at the level required for sub-meter per second precision, the PSF of the high-resolution, high-S/N Fourier transform spectrometer (FTS) scans cannot be precisely deconvolved. Furthermore, there are degeneracies between parameters used in the forward modeling (Spronck et al. 2015) to measure Doppler shifts. Even if the spectrometer SLSF is stabilized, the Levenberg–Marquardt algorithm used to model shifts in the

[66] Presentation by Suvrath Mahadevan.

[67] Abhijit Chakraborty has a program to develop ThAr lamps for astronomers in collaboration with Indias atomic energy program.

stellar lines is driven by chi-squared minimization and will fit out weak telluric contamination or some of the photospheric noise; this adds scatter to the measurement of the COM velocity of the star.

Both the absolute wavelength of the calibrator and the spectrometer wavelength dispersion must be known with incredible precision if instruments are to achieve the precision necessary to detect Earth analogs. The reflex velocity induced by the Earth on the Sun is 9 cm s^{-1}, or equivalently, a 100 kHz shift in frequency over one year. The ideal wavelength calibrator for the detection of Earth analogs must have a SMP of about 1 cm s^{-1} and must be stable for the expected duration of the observations (years).

Ideally, light from the calibrator would follow the same light path through the spectrometer as the starlight. However, because a superimposed wavelength calibrator will obscure the stellar spectrum and make it more difficult to extract astrophysical Doppler signals, the calibration source must be spatially offset or temporally interleaved with the stellar observations. The ideal calibrator would have spectral features that are uniformly spaced (to maximize information content), unresolved (making it useful for caliibrating instrumental errors and the instrumental profile), well separated (ensuring features can be cleanly resolved by a high-resolution spectrometer), and robust to ensure long-term accuracy is maintained.

3.2.1. High Precision Wavelength Calibrators

To improve the wavelength calibration precision from lamps or reference cells, several photonic sources, including LFCs, stabilized etalons, line-referenced electro–optical combs, tunable FP, and chip-scale microcombs are being studied for precision RV spectrometers.[68] These methods are summarized in Figure 16.

LFCs can meet the requirements of an ideal wavelength calibrator. LFCs use femtosecond mode-locked lasers to produce extremely narrow emission lines (Murphy et al. 2007; Osterman et al. 2007; Li et al. 2008; Steinmetz et al. 2008). The LFC has perfectly uniform spacing of its individual emission line modes; this has been tested to 20 digits of precision (Ma et al. 2004). Moreover, LFCs with broad octave-spanning spectra can employ self-referencing to measure the carrier envelope offset, which provides a straightforward means to stabilize the absolute optical frequency of every comb element, in addition to the mode spacing. This creates an optical synthesizer where the frequency of every comb tooth is known with the same precision as the microwave standard (such as cesium clocks or GPS signals) that provide the reference signal for the LFC. And because such microwave standards can be traced to the internationally defined SI second, the absolute frequencies of the comb are stable for decades. For example, when referenced to the free and wide-spread signals from the GPS (global positioning system) clocks, the long-term absolute frequency of the comb could be stable to better than 10^{-11} (i.e., <0.3 cm s^{-1}) (Lombardi 2008; Quinlan et al. 2010), meeting or exceeding the precision and stability required for RV planet searches. The advantage of the self-referenced LFC is significant, in that it directly ties the comb frequencies to international standards (not artifacts or secondary references) and leverages the tremendous technological infrastructure of GPS and other global navigation satellite systems.

Among the challenges for LFCs is the innately close mode spacing of typically 0.1–1 GHz, which requires spectral filtering of ~98% of the modes with one or more FP cavities to produce a calibration grid with 10–40 GHz spacing (commensurate with the resolving power of a high-resolution spectrometer). Although filtered, the unwanted modes do not completely go away (Sizer 1989; Osterman et al. 2007; Braje et al. 2008; Quinlan et al. 2010; Ycas et al. 2012). More important than the fact that residual side modes remain, is that the suppressed remnants can be asymmetricand thus potentially shift the line center as seen by the spectrograph. Careful attention to this matter is required in order to provide a comb with sufficient side mode suppression and symmetry such that it does not affect the calibration at the level of desired RV performance (Probst et al. 2014).

Another challenge with LFCs is that spectral broadening with nonlinear optics is required to generate a comb that spans a broad bandwidth or operates at optical wavelengths. This is straightforward at the native mode spacing of 0.1–1 GHz, but due to a corresponding reduction in pulse peak power, significant broadening is more challenging at 10+ GHz mode spacing. A complication of this is that the side modes that one would like to suppress can be parametrically amplified during nonlinear spectral broadening and can even overtake the principal comb line (Chang et al. 2010). In some systems, this problem has been overcome by using up to 3 concatenated filter cavities in order to provide sufficient suppression of unwanted side modes (Probst et al. 2014). Another approach to dealing with parametric sidemode reamplification is to first spectrally broaden at the native repetition rate, and then spectrally filter the broadened comb with a purely linear FP filter cavity. This approach has been undertaken by the teams providing LFC calibrators for HARPS-N (Langellier et al. 2014) and the Penn State Habitable zone Planet Finder (HPF). For example, with the HPF instrument the original 250 MHz laser at 1550 nm is broadened into the Y and J band with a highly nonlinear fiber at the native repetition rate to avoid parametric amplification of the side modes; the signal is then filtered using cavities with low dispersion mirrors. While sidemodes may still exist, their impact on the spectral calibration can be calculated from simple linear optics and measured in situ (Li et al. 2008).

[68] Presentation by Scott Diddams.

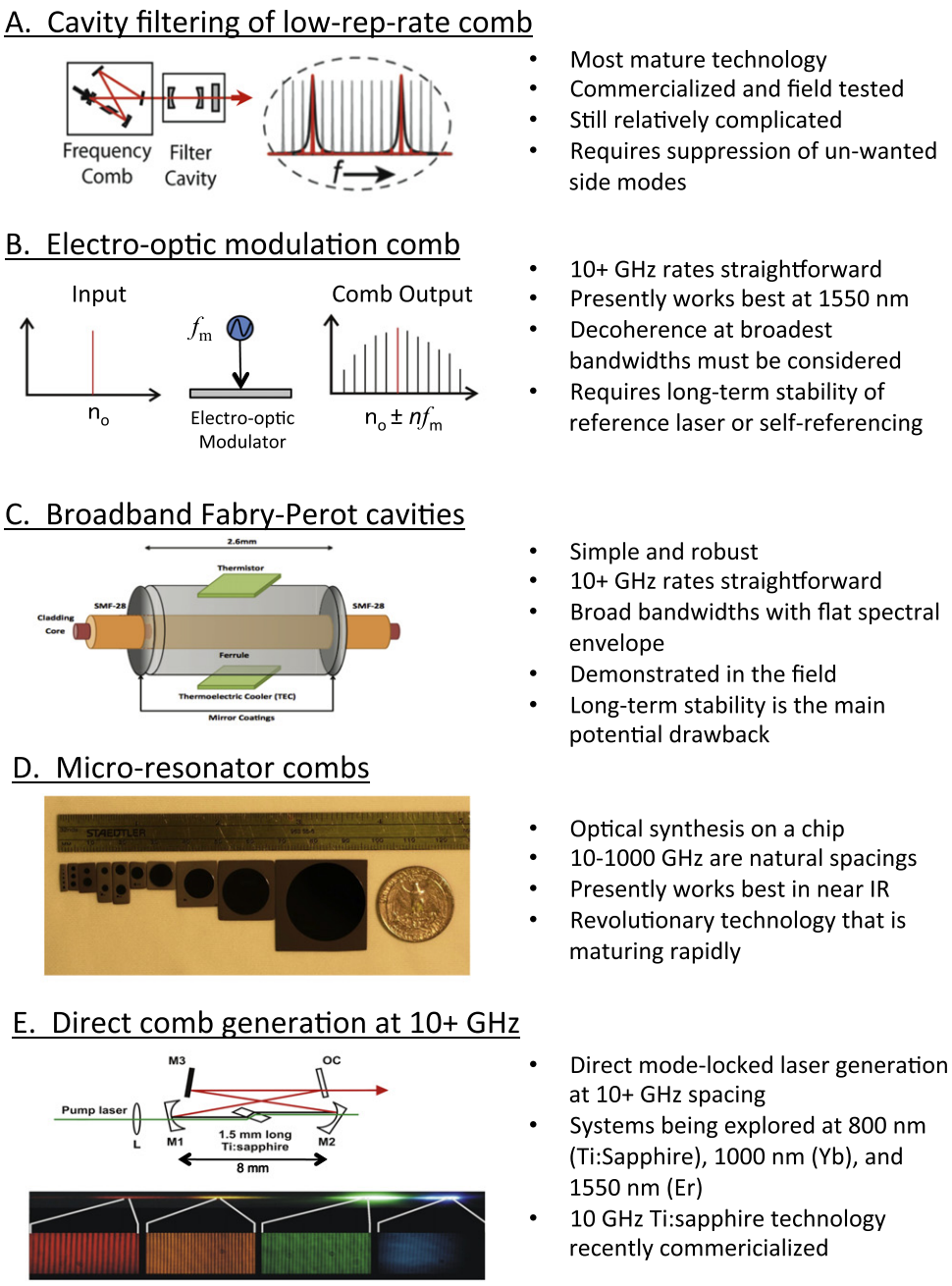


Figure 16. Overview of various photonic sources that have been employed, or are being studied, as calibration sources for precision radial velocity spectroscopy. Frames (A), (C), (D), and (E) are adapted with permission from Osterman et al. (2007), Halverson et al. (2014), Bartels et al. (2009), Li et al. (2012), respectively. (A color version of this figure is available in the online journal.)

As an example of how far the LFC technology has matured in the past few years, Menlo Systems[69] now offers a commercial GPS-locked LFC that is designed to reach an accuracy of $4\,\mathrm{mm\,s^{-1}}$ at optical wavelengths (Probst et al. 2014). The system uses an state of the art IR fiber laser as a robust seed source. The low repetition rate (corresponds to a small mode spacing) of the fundamental femptosecond-laser is increased by mode filtering through three FP cavities with finesse of about 3000. Any mode spacing between 1 GHz and 25 GHz—in multiple integers of 250 MHz—can thus be choosen to best fit the spectrographs resolving power. The filter narrowband IR spectrum is then amplified in an fiber

[69] Presentation by Tilo Steinmetz.

amplifier and subsequently broadened in a specially designed photonic crystal fiber.

The system uses an IR fiber laser with mode filtering through three FP combs, signal amplification and spectral broadening. The use of FP cavities with a high finesse of 3000 significantly reduces the sidemodes and thereby minimizes their reamplification during spectral broadening. The spectrum is flattened to better than a few dB by using a liquid crystal display with active feedback loop, and the line intensities have low fluctuations over a broad optical wavelength band of 400–700 nm with 15–25 GHz mode spacing.

Two Menlo system LFCs (with mode spacings of 18 and 25 GHz) were tested during a campaign at HARPS from 2015 April 8–18. The comb illuminated 48 echelle orders with 12,000 lines per channel. The HARPS pipeline produces 1-d extracted spectra for every order and a Gaussian was used to fit the LFC line centers. Modal noise in the fibers changes the intensity profile of the comb line recorded on the detector. Fiber agitation is required to ensure each comb mode has a smooth, stable SLSF that is not dominated by the speckle interference pattern associated with multimode fibers illuminated by coherent sources. The wavelength solution was determined with an accuracy of $5\,\mathrm{cm\,s^{-1}}$ after binning 5 exposures, and improved to better than $1\,\mathrm{cm\,s^{-1}}$ by binning 20 exposures.

The Menlo LFC systems have been developed for long term fail-safe turnkey operation. The system takes 10 minutes to change from the off state (laser amplifiers off) to standby mode with the laser on, locks enabled, and high power amplifier off. From the standby mode, it only takes 10 s to switch to full operation. There are some expected maintenance costs: the broadening fiber needs to be changed once per year and the pump diodes can fail every year or two. These components are built in as modules that can be easily replaced at a cost of a few thousand dollars.

3.2.2. Line-referenced Electro–Optical Frequency Combs (LR-EOFC)

In contrast to the mode-locked laser at the heart of the LFC, the LR-EOFC provides a spectrum of lines generated by electro-optic modulation. A pump laser is locked to an atomic or molecular transition and is then phase modulated and amplified to produce tunable sidebands with a spacing set by the modulation frequency (Ishizawa et al. 2011). Yi et al. (2015) have demonstrated a stability of <200 kHz ($\sim 0.3\,\mathrm{m\,s^{-1}}$) with a LR-EOFC at H-band wavelengths. The advantage of this technique is that there are no side modes; the system only generates the carrier with frequency sidebands that can be referenced to a microwave standard. Comb generation makes use of commercial off the shelf telecommunication components, but has currently been operated in wavelength bands where optical amplifiers and phase modulators are available (1–2 μm) from the telecom industry.

There is still technology development needed to create EOFC broadband combs with constant power per comb mode, but the same techniques employed by Probst et al. (2014) should be adaptable to such combs. Another challenge is that although the line spacing of the LR-EOFC is fixed to a microwave reference, the seed laser for the comb needs to be locked to prevent wavelength drifts. In present systems, this has been provided by a laser that has its frequency locked to transitions in hydrogen cyanide or acetylene (Yi et al. 2015). While this is straightforward, robust and commercially available, the long-term stability (e.g., months to years) of such molecularly stabilized lasers is not known. Alternative atomic references for such LR-EOFCs could include narrower saturated absorption transitions in acetylene (Edwards et al. 2004) or the well-studied rubidium D1 and D2 lines. The spectral broadening is also more difficult at the high mode spacing of the EOFC due to the commensurate reduction in pulse energy. However, very recently a 10 GHz EOFC centered at 1550 nm has been broadened to octave span and self-referenced such that the frequency of its seed laser can also be referenced to a microwave standard, and thereby inherit the stability advantages similar to the LFC described above (Beha et al. 2015).

3.2.3. Micro-combs

One of the most exciting developments in LFC technology is a new approach to generate broad bandwidth combs in chip-scale devices. Such microcombs use the nonlinear Kerr effect in a dielectric micro-ring to produce a comb with line spacing that can be tuned between 10 and 800 GHz without the need for filtering (Del'Haye et al. 2007, 2008; Kippenberg et al. 2011; Jung et al. 2014). While still at a relatively early stage of technical development, such microcombs hold the promise of frequency comb stability and accuracy in a very robust, compact, and potentially inexpensive silicon chip package. At present, the main challenges associated with microcombs are related to deterministic low-noise operation and increased spectral span, as required for self-referencing. Additionally, the broadest bandwidth microcombs presently operate with mode spacing of 200 GHz, which may too high for many high-resolution spectrographs (Brasch et al. 2014), although microcomb technologies exist down to mode spacings of a few GHz (Li et al. 2012). Finally, microcombs generally have native wavelengths in the near IR (1550 nm) but doubled or tripled frequencies can shift the comb to red and green wavelengths and self-referencing to microwave standards can be used for 2/3 or full octave-spanning combs (Del'Haye et al. 2015; Jost et al. 2015). The development of microcombs is progressing rapidly, and it is likely that in the near future this

new integrated technology could offer an attractive alternative for precision RV applications.

3.2.4. FP Interferometers

FP interferometers provide a dense grid of lines with a technological approach that is simpler and a price tag that is lower than LFCs.[70] However, the systems need to be referenced to LFCs, lamps or atomic features (Halverson et al. 2014; McCracken et al. 2014; Reiners et al. 2014; Bauer et al. 2015; Schwab et al. 2015). Precise vacuum and temperature controls are generally required to stabilize the devices and it is important to track and correct for drifts in the dispersion.

Etalons are fixed length FP interferometers that produce broadband optical combs when illuminated with a broad bandwidth white light source. The HARPS etalon has Zerodur spacers to control the index of refraction, scrambled uniform illumination, and stringent temperature and pressure control (Wildi et al. 2012). These broadband sources have adjustable mode spacing, simplicity and good short-term stability. The night-to-night drift of the etalon lines is about $10\,\mathrm{cm\,s^{-1}}$ and the stability over 60 days is about $1\,\mathrm{m\,s^{-1}}$ (Wildi et al. 2012). Near-IR fiber etalons have achieved better than $1\,\mathrm{m\,s^{-1}}$ stability over 12 hr timescales. In most cases, the frequency drift is not directly measured or corrected. Reiners et al. (2014) have suggested including a gas cell (e.g., Rb) stabilized laser to track frequency drifts and to derive wavelength offset corrections.

Tunable FPIs that are white-light illuminated also produce a broadband comb of lines like the etalon, however they employ feedback to stabilize the wavelength zero point of the comb with locking to optical atomic standards like the rubidium D2 line at 780 nm. In order to provide sufficient precision, the FP cavity finesse needs to be high (at least 50–100) at the wavelength of the laser lock, while the finesse can be decreased via engineering of the coatings to provide broad bandwidth coverage at other wavelengths.

Reaching $1\,\mathrm{cm\,s^{-1}}$ wavelength calibration precision is a grand challenge. By way of comparison, the Ramsey fringe of the cesium clock has a line width of about 1 Hz and a microwave oscillator can be locked to the central fringe with a precision of 10^{-13} in one second; and with averaging one can achieve a fractional measurement uncertainty of about 10^{-16} (Heavner et al. 2014). In terms of splitting the line, this is the most precise spectroscopy that has ever been done and allows the line centroid to be determined to one part in 1,000,000. However, this measurement is carried out in the lab with state-of-the-art microwave sources and laser-cooled cesium atoms. By comparison, astronomers who aim to reach $1\,\mathrm{cm\,s^{-1}}$ precision (equivalent to 15 kHz resolution in the visible) are using stellar spectral features that are many GHz wide. If successful, this grand challenge would rival the line-splitting precision of work done by the AMO physicists; however, with the additional challenges of the accurate collection, dispersion and recording (at the photon counting level) of the relatively broad and faint spectral absorption lines.

[70] Presentation by Ansgar Reiners.

3.3. Fibers and Scrambling

The spectrograph projects the image of the entrance slit onto the detector as a function of wavelength. Images of the telescope pupil (far field) also illuminate the echelle and other optics; alt-az telescopes that employ image derotators can end up with rotating spider mounts on the echelle. Any variability in the illumination of the spectrograph optics (say, because of guiding errors or changes in seeing) manifests as variability in the spectrum. This is a severe limitation for precise RV measurements. Optical fibers provide an important benefit; not only do they offer a flexible way to couple light from the telescope into a spectrograph,[71] they also offer the additional advantage of providing a more stable input illumination (Heacox 1986).

However, light will exhibit modal noise when traveling through a multi-mode fiber; this is particularly important for coherent light sources like lasers or LFCs. Modal "noise" is a manifestation of mode coupling that produces an irregular intensity distribution emerging from a fiber (Baudrand & Walker 2001). When a laser is coupled into a multi-mode fiber, the light travels in many modes along the fiber; these modes are out of phase with each other, resulting in the complicated speckle interference pattern that illuminates the spectrograph optics. If nothing moves, the spatial speckle pattern is time invariant. However, stresses or motion of the fiber cause temporal shifts in the speckle pattern. When multi-mode fibers are used, some effort is required to scramble, or mix the modes. This can be done by shaking or agitating the fiber (Lemke et al. 2011; McCoy et al. 2012; Plavchan et al. 2013b).

When multi-mode fibers are used, some effort is required to scramble, or mix the modes. Many experiments have been carried out showing that when light is injected into a multi-mode fiber with a circular cross-section, the output intensity distribution is not completely independent of the input illumination. Circular fibers have reasonably good azimuthal scrambling, but poor radial scrambling. This results in variable illumination of the slit and spectrograph optics, and these undesired perturbations are imprinted in the spectrum, adding noise to the data.

The ideal optical fiber has high throughput and a well-scrambled output that is independent of the way that light enters the fiber. To achieve high transmission, the fiber length should be minimized; AR coatings can be applied to the fiber

[71] Presentations by Paul Fournier and Suvrath Mahadevan and breakout session led by Gabor Fűrész and Paul Fournier.

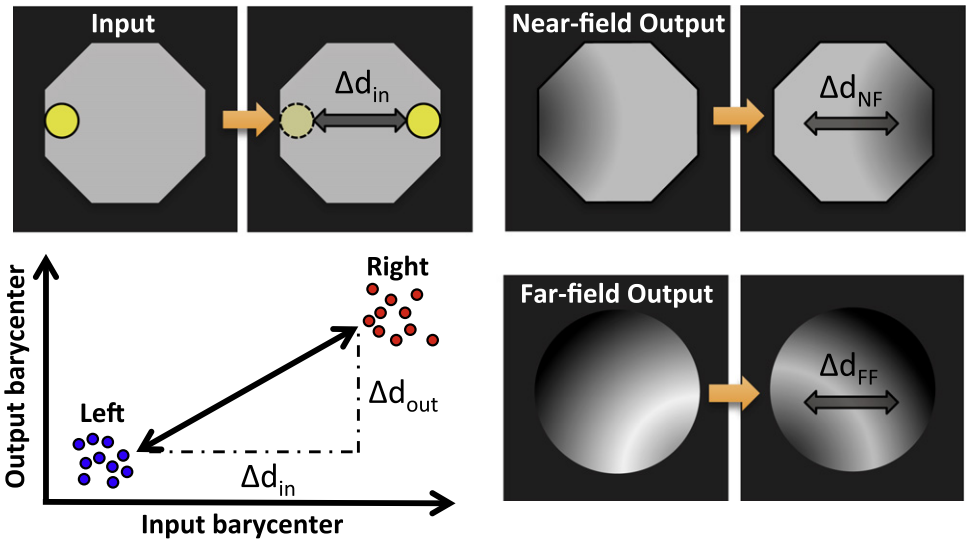


Figure 17. Cartoon illustrating measurement of scrambling gain by translating the test fiber across a fixed illumination spot and recording the output at each position. Measurements are made for the fiber near field and far field. (Courtesy of Sam Halverson.)
(A color version of this figure is available in the online journal.)

faces, and the FRD can be minimized by controlling the light injection cone to match the numerical aperture or acceptance angle of the fiber. Some FRD will still occur during the fiber fabrication and mounting process, from micro-roughness and micro-bending, subsurface damage and other induced stress in the fiber (Carrasco & Parry 1994; Avila 1998; Ramsey 1988). The use of connectors can also be a significant source of FRD and soft, low shrinkage glue must be used to avoid strains. Fiber manufacturers do not typically measure FRD.

The use of fibers with non-circular cross-sections (octagonal, square or rectangular) break the symmetry of the wave guide and provide an easy way to scramble modes (Chazelas et al. 2010; Avila 2012; Spronck et al. 2012a). Scrambling performance is improved by exchanging the in-focus near field (the image) and the out of focus far field (pupil illumination) between two fibers (Hunter & Ramsey 1992). Variations in the fiber near field directly lead to fluctuations in recorded spectral features while variations in the far field change the illumination of spectrometer optics, leading to distortions in the PSF. Both effects lead to spurious velocity shifts and are detrimental to extreme precisoin measurements.

As shown in Figure 17, the scrambling gain of a system—for the near field—can be measured by taking the ratio of the displacement of the light source at the entrance of the fiber to the residual shift in the COM intensity distribution of the fiber output (Avila & Singh 2008). Scrambling gains of up to a few hundred are achieved for pure circular fibers; octagonal core geometry yields a scrambling gain of up to about 1000, while octagonal fibers with double scramblers exhibit ~10–20 k scrambling gain (Feger et al. 2012). Effects of inadequate scrambling in the far field can only be evaluated, by modeling how the observed pupil illumination irregularity propagates through an optical system (Stürmer et al. 2014). This can be done by ray-tracing through the instrument optical model, which leads to an estimate of the RV measurement sensitivity to far field variations. It has to be noted, though, that predicting the exact behavior and effects of the far field for an as-built instrument is almost impossible. It is because the as-built aberrations can significantly effect how the far field propagates through an optical system. Eliminating sensitivity of the as-built spectrographs, or variations in the far field of fiber systems, is key to further improve PRV performance.

SMFs offer an interesting alternative to multi-mode fibers (Crepp 2014). True to their name, SMFs propagate a single mode (actually, two polarization modes are possible) of light; as a result, the fibers deliver a uniform intensity distribution to the spectrometer—the SLSF has unparalleled stability—and there is no modal noise from the fiber. The challenge is coupling light into the small 5 μm diameters of the SMF, however, extreme adaptive optics systems with high strehl ratios like the Subaru Coronagraphic Extreme Adaptive Optics (SCExAO) have reached fiber coupling efficiencies of 68% with an 8 μm fiber core (Jovanovic et al. 2014).

4. Extracting Keplerian Signals

Once the Doppler shift has been measured in a spectrum, the time series velocities are analyzed for Keplerian signals. At this point, there are systematic error sources that add to the formal errors to obscure the detection of low amplitude systems. Telluric lines (from the Earth's atmosphere), stellar oscillations,

sunspots, plage, granulation and meridional flows, all modify the shape of spectral lines as a function of time, and produce obscuring velocity signals.

In many cases, astronomers have not been able to physically distinguish between stellar jitter and COM velocity with activity diagnostics; instead, some magnitude of the jitter is added in quadrature when using chi-squared minimization for fitting for Keplerian signals in RV time series data. Regardless of whether the magnitude is assumed (e.g., based on spectral type) or inferred from the data, this approach is clearly insufficient; it implicitly assumes that jitter can be represented as a white noise source and does not leverage information about the time-coherence in the stellar activity signal.

Another approach for mitigating stellar activity is to first measure apparent RV shifts (neglecting activity) and then apply a bandpass filter in the Fourier (frequency) domain with the aim of removing signals due to stellar activity. This approach has been used with some success (Howard et al. 2011, 2013; Fulton et al. 2015). However, the Fourier components are fixed; since the underlying noise sources change in time, the extracted signal will be distorted. In the best case, this introduces small systematic errors in the orbital parameters. In the worst case, true planetary signals are attributed to stellar activity or the analysis process introduces spurious coherent velocity variations that are mistaken for planets.

The following cases offer a representative sample of putative planetary systems that have been debated, and the different approaches applied by the community:

1. GJ 581, with emphasis in the case of GJ 581d planet candidate initially described Mayor et al. (2009). Subsequent arguments relating the signal to stellar rotation, activity and the statistical treatment of the problem can be found in Robertson et al. (2014) and Hatzes (2016).
2. α Cen Bb, which corresponds of a signal of $\sim$0.5 m s^{-1} on a time-series strongly dominated by noise with structure. The initial claim was made in Dumusque et al. (2012), and some doubts on the series whitening approach and caveats about statistical significance can be found in Hatzes (2013) and Rajpaul et al. (2016).
3. HD 41248b & c. These two Doppler signals were initially attributed to a pair of resonant planets in Jenkins et al. (2013). Arguments in favor and against of activity induced variability were given in Santos et al. (2014) and Jenkins & Tuomi (2014).

4.1. Statistical Analysis Techniques

The interpretation of RV data increasingly relies upon statistical analysis techniques.[72] In medical fields, there has been a near crisis in the ability to replicate the statistical analysis of scientific results (Ioannidis 2005). This is in part because scientists often know what results they expect and this can influence their interpretation, but it is also because of the misuse of null hypothesis tests, such as false alarm probability (FAP) and Kolmogorov–Smirnoff analysis. The repercussions can include financial cost, wasted time re-assessing false discoveries, and public mistrust of science. Our community seems to appreciate the importance of reaching out to statisticians and developing more sophisticated approaches to data analysis.

The FAP test is a common tool for testing the reality of a Keplerian signal in RV data (Marcy et al. 2005). This statistical test is usually carried out with a bootstrap Monte Carlo simulation. A Keplerian fit to the initial prospective signal is made and the χ^2_ν value is saved. The velocity data are then scrambled, while retaining the times of the observations and the errors; for each realization, a blind search is made for the best Keplerian value and the resulting χ^2_ν is recorded. If a signal was really contained in the original velocities, then the subsequent scrambled data are likely to have poorer χ^2_ν values. This approach has the advantage of characterizing the impact of non-Gaussian systematic errors in a data set. The reported FAP is the fraction of Monte Carlo realizations that yielded χ^2_ν values lower than the original set. As noted by Marcy et al. (2005), the FAP test specifies the probability that incoherent noise could have yielded a superior Keplerian fit by chance. It does not necessarily show that the planet interpretation is correct—the signal could arise from window functions or stellar activity or systematic instrumental errors.

Statisticians prefer to call this a *p*-value because the language "FAP" seems to carry an implication about the reality of the signal or the planet interpretation. The *p*-value tells us about the probability of obtaining a particular signal given the null hypothesis (i.e., given the assumption that the data are merely incoherent noise).

4.1.1. Bayesian Analysis

The field of RV exoplanet detection has undergone a dramatic shift over the past two decades.[73] For many early years, "discovering a planet" meant that astronomers had: (1) identified a possible model to explain a series of Doppler observations as due to a planet and measurement noise, and (2) rejected the null hypothesis that a simpler model without the planet could reasonably explain the data. This approach worked well for planet surveys that focused on well-characterized and quiet FGK dwarf stars with RV amplitudes that were much larger than individual Doppler errors and plausible stellar activity signals. Astronomers did not want to risk incorrect detections near the threshold of their errors. When in doubt, the solution was to collect more observations.

[72] Presentation by Tom Loredo.

[73] Presentation by Eric Ford.

More recently, astronomers have begun adopting a Bayesian framework for characterizing the masses and orbits of exoplanets, as well as performing rigorous model comparison to quantify the evidence for planet detections. In this framework, the criteria for detecting a planet is that the model including a planet has a higher posterior probability than the model or models without a planet.

The paradigm shift to adopt Bayesian inference has been fueled by the appreciation for greater statistical rigor, more accurate estimates of parameter uncertainties, and the greater sensitivity possible with a rigorous statistical analysis. Perhaps most importantly, a Bayesian approach allows us to quantitatively answer questions that we actually want to ask. For example, we can quantify the evidence for there actually being N planets, rather than just rejecting the null hypothesis. Many statisticians like to think of Bayesian model comparison as a quantitative "Occam's razor" for comparing competing models. Others prefer to make principle decisions using a utility function after after marginalizing over uncertainty in models and parameters.

Of course, simply switching to a Bayesian framework does not solve all our problems. One barrier to widespread adoption of Bayesian method is the necessity of computing multi-dimensional integrals. This can be computationally expensive, particularly for models with many parameters and when comparing the evidence for multiple competing models. This barrier has been largely overcome for parameter estimation applied to Doppler surveys thanks to increased computational power and improved sampling algorithms. A second barrier is the intellectual investment required to understand the underlying algorithms. For example, the choice of the proposal distribution in Metropolis–Hastings Markov chain Monte Carlo algorithms often makes the difference between fast and extremely slow convergence. Therefore, one should always check multiple convergence diagnostics to identify any signs of non-convergence. Bayesian model comparison is particularly computationally expensive, so it is tempting to use short-cuts, such as the Bayesian Information Criterion (BIC). The BIC was designed to identify the most probable model in abundant data settings, but its approximations are not sound for quantifying model *uncertainty* in small or modest data settings, as arise in Doppler planet hunting. Fortunately, the exoplanet community is gradually gaining experience with successful application of Bayesian methods to interpret observations.

A third barrier is the misperception that the need to explicitly choose priors is an inherent weakness of the Bayesian approach. If anything, it is a feature that the Bayesian framework clarifies the assumptions before entering into an analysis and provides a straightforward path for testing the sensitivity of conclusions to those assumptions. Of course, it is useful to choose priors wisely. Often physical intuition, such as symmetries, or knowledge of previously identified planetary systems, can inform the choice of priors. A general principle is to err on the side of overly broad priors. When in doubt, one should perform a sensitivity analysis to quantify whether key results would change for different reasonable choices of priors (Tuomi & Anglada-Escudé 2013).

For complex models it is also wise to verify that one's model works well on simulated data to ensure that the chosen priors are appropriate for a given application. For model checking refer to Gelman et al. (2013). Too often people using Bayesian methods ignore model checking, because it doesn't have a neat and tidy formal expression in the Bayesian approach. But it is no less necessary to do goodness-of-fit type checks for a Bayesian analysis than it is for a frequentist analysis.

A more fundamental challenge is that one must identify an appropriate model for interpreting observations, regardless of whether using a frequentist or Bayesian approach. Analysis is straightforward when the physical model is known, but inference is much more challenging when the model is physically incomplete. Early searches for giant planets appropriately considered only Keplers laws, but modern searches for low mass planets are significantly affected by stellar activity, especially when surveys include active stars. Unfortunately, we do not currently have a practical model of stellar activity, nor even a clear path to such a model. Activity indicators sometimes correlate with velocities, but this is not always the case.

The problem of model incompleteness has been one cause for apparent contradictory results concerning signals that seem to be highly significant but are close to the measurement uncertainty level. For example, a discontinuity of a few $\mathrm{m\,s^{-1}}$ (e.g., from a change of the gas cell or Th Ar lamp) will typically be much better fit by a sinusoid than a model with white noise only. If the model could account for real discontinuities, it would be clear that the addition of a sinusoid yielded a poorer match to the data. In a Bayesian context, this leads to an overestimation of the evidence because the sum over the models is only done for a subset of them. Thus, evidence ratios should not be interpreted as definitive proof for detections. This issue affects the model probabilities even if the multidimensional integrals are executed exactly. It is also related to the discussion about prior choices, in the sense that zero probability is implicitly assigned to a large number of alternative models. One way to mitigate this confusion is to secure minimal evidence ratio thresholds based on simulated data that is as close as possible to the real data. A more informative, but more resource intensive approach, is to confirm or refute signals for the same star using different instruments or techniques. For example, highly significant candidate signals were spotted in UVES data, but could not be confirmed with HARPS measurements (Tuomi et al. 2014). When applied to a large enough sample of objects, it is possible to construct more complete models or design a properly calibrated probability threshold system to identify reliable exoplanet signals.

It is useful to keep in mind that "All models are wrong; some models are useful" (attributed to George Box 1979). Even after one verifies that a model and algorithm perform well when analyzing simulated data, one must still verify that they are useful for actual astronomical observations. This will require extensive astronomical observations, extracting additional information from each spectrum to probe stellar activity, new models for joint RV and activity data, and improved sampling algorithms. Therefore, future Doppler detections of low-mass planets will require very large amounts telescope time. In an effort to make efficient use of precious astronomical resources, it is natural to seek out the most powerful statistical methods to increase the sensitivity to low-mass planets and scientific impact of Doppler surveys. Our community must identify realistic science goals, design observing strategies optimized to match those goals, and persuade time allocation committees to allocate sufficient time to characterize long-term and short-term stellar variability and systems of multiple planets.

We will likely be forced to learn to live with uncertainty. Given our upcoming requests for large investments of astronomical observatories, we will need to publish observations of both strong and marginal detections and learn to responsibly convey the limitations of our studies to other scientists and the public. This greater cultural shift toward emphasizing reproducible research is essential to the credibility of planet detections, astronomy and science in general.

4.1.2. Gaussian Processes

A Gaussian process regression is a powerful statistical analysis technique (beyond simply minimizing chi-squared in our models) that can be applied to the detection of weak signals in the presence of red noise.[74] Both the systematic and correlated noise sources can be included in the model and should be fit simultaneously with the signal (not sequentially) and marginalized out to correctly remove or minimize the parameters that are not of interest and to thereby avoid distorting the underlying signals.

Chi-squared is defined as the sum of the squared differences between the data points, y_i, and the model (m_i are theoretical points for a Keplerian signal), divided by the independent uncertainties ${\sigma_i}^2$, as shown in Equation (3). It is also possible to switch from summation to vector notation and to relate chi-squared to the natural log of a Gaussian likelihood function, L, by writing the difference of the data[75] and model as a vector difference with the noise represented as a diagonal co-variance (V) matrix (Equation (4)).

$$\chi^2 \equiv \sum_i \frac{[y_i - m_i]^2}{{\sigma_i}^2} \tag{3}$$

$$\begin{aligned} \chi^2 &= [y - m]^T \cdot V^{-1} \cdot [y - m] \\ V &= C \\ C_{ij} &= {\sigma_i}^2 \delta_{ij} \\ -2 \ln L &= [y - m]^T \cdot V^{-1} \cdot [y - m] + \ln\det V + N. \end{aligned} \tag{4}$$

If there are multiple components to the (Gaussian) noise, their covariances add in the noise covariance tensor, V. For example, if a second noise source is added through the covariance tensor, Q, the variance tensor becomes $V = C + Q$ as shown in Equation (5). This approach is straight forward and allows for the analysis of very high dimensional spaces and high fidelity signal recovery.

$$\begin{aligned} -2 \ln L &= [y - m]^T \cdot V^{-1} \cdot [y - m] + \ln\det V + N \\ V &= C + Q \\ C_{ij} &= {\sigma_i}^2 \delta_{ij}. \end{aligned} \tag{5}$$

Systematic sources of error can be modeled as the sum of M basis vectors with off-diagonal terms. If a Gaussian prior can be placed on the linear amplitudes, the systematic noise can be represented by a rank-M covariance matrix and marginalized to obtain a best fit. For example, in the case of *Kepler* photometric time-series data, the systematic noise sources (e.g., from the spacecraft jitter) can be modeled as the sum of M basis vectors; in this case, the basis vectors might contain information about the behavior of other stars.

$$\begin{aligned} -2 \ln L &= [y - m]^T \cdot V^{-1} \cdot [y - m] + \ln\det V + N \\ V &= C + B \cdot \Lambda \cdot B^T \\ B &= \text{block of } M \text{ basis vectors} \\ \Lambda &= M \times M \text{ prior variance}. \end{aligned} \tag{6}$$

The noise tensor, V, encodes beliefs about the contaminating Gaussian noise and it can be designed to include a component for complex systematic noise. The models may be enormous, but it is important to fit for all of the nuisance parameters. For expressions like V in Equation (6), it is numerically safer to factorize the matrix than to compute inverse matrices or determinants directly.[76]

Baluev (2013) treated time-correlated red noise sources as a Gaussian random process with an exponentially decaying correlation function and found that some of the announced planets around GJ 581 were illusions of red noise.[77] Good statistical methods for handling red noise sources from stellar photospheres and instrumental errors are needed before we can determine the ultimate limitation to Doppler precision from stellar noise (Baluev 2015). In recent years, Bayesian inference has become a central tool for the study of planetary or stellar signals in any RV data set (e.g., Faria et al. 2016).

[74] Presentation by David Hogg.
[75] Also see the presentation by Guillam Anglada-Escude.
[76] In the python numpy package, never use `inv`, always use `solve`; never use `det`, always use `slogdet`.
[77] Presentation by Roman Baluev.

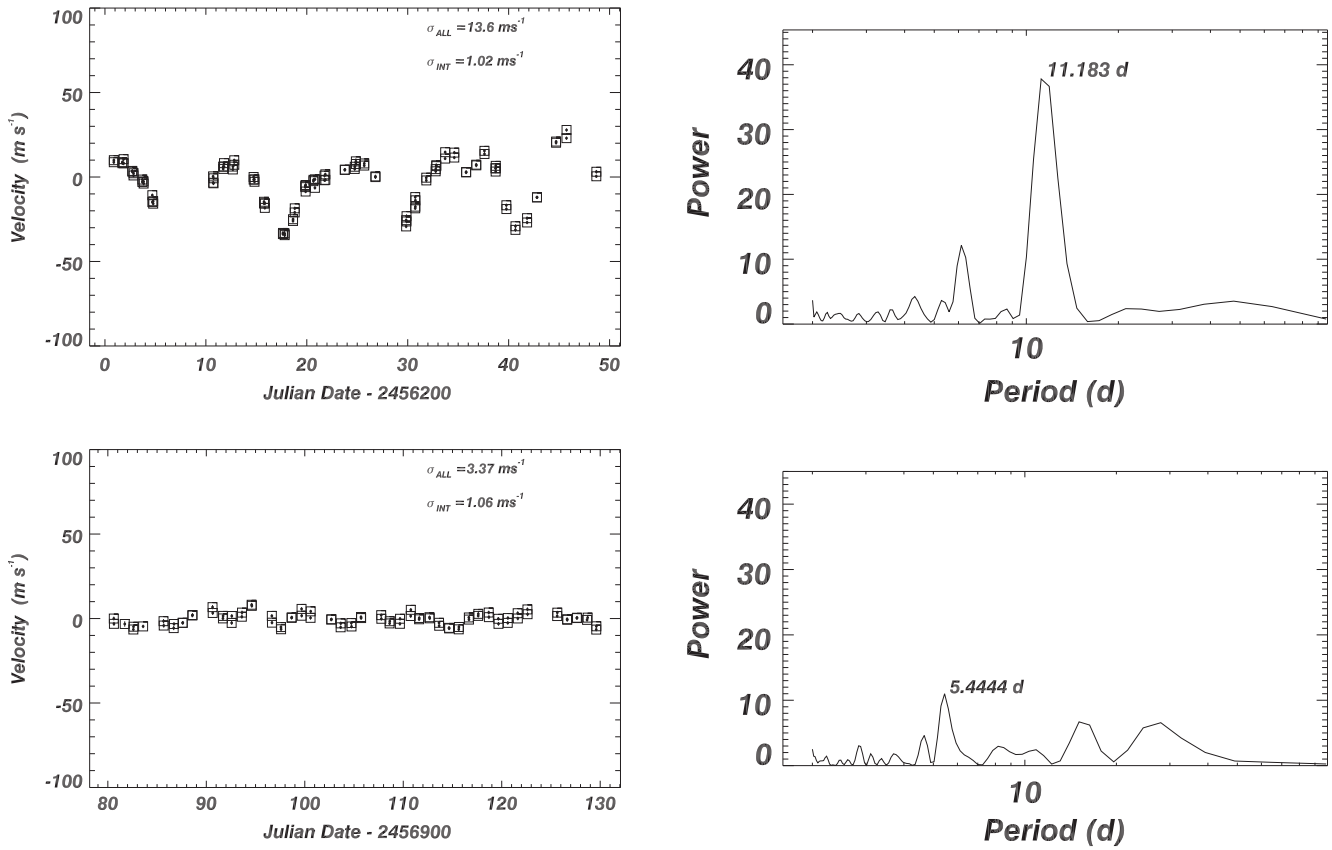


Figure 18. Radial velocity measurements of Epsilon Eridani using the CHIRON spectrometer at CTIO have internal errors of 1 m s^{-1}. In 2012 (top panel) a clear 11.4 day periodicity is seen, which matches the rotation of the star. The rms variability is 13.6 m s^{-1}. This signal attenuates over time and in 2014 (bottom panel) was not detectable. (Courtesy of Debra Fischer.)

4.2. Photospheric Velocities: Stellar Jitter

The contributions to RV measurements that arise from photospheric motions are generally termed "stellar jitter" and are currently an obfuscating source of time-correlated noise for RV measurements that aim to detect weak Keplerian signals. The typical magnitude of stellar jitter is 1–3 m s^{-1} for chromospherically quiet stars (Dumusque et al. 2011). Stellar jitter is caused by a variety of physical processes.[78] Unfortunately, there is no general analytical model for photospheric signals. Activity signals manifest as both incoherent and quasi-periodic variability.

Stars that are quiet can rather suddenly become more active (Santos et al. 2014). For example Figure 18 shows RV measurements of the moderately active star, Epsilon Eridani. Velocities were obtained with the CHIRON spectrograph and are shown for two different epochs. In 2012, a 50 day time series of radial velocities (from JD 2456200 to JD 2456250) has an rms of 13.6 m s^{-1} and shows periodogram power that peaks at 11.18 days, consistent with the 11.35 and 11.55 day spot rotational period observed in 2005 using the *Microvariability and Oscillations of STars* (*MOST*) satellite (Croll et al. 2006). The 11 day RV signal in Epsilon Eridani has been shown to correlate with photometric variability using *MOST* satellite and concurrent RV measurements with the CHIRON spectrograph in 2014 (Giguere et al. 2016). This signal weakens over time with the rms dropping to a 3.37 m s^{-1} and is not seen in the periodogram of velocities obtained in 2014 (from JD 2456980 to JD 2457030).

Cool stars have convective envelopes that support acoustic modes (*p*-modes) with velocity variations at the level of a meter per second on timescales of several minutes (Kjeldsen & Bedding 1995). Granulation in the photosphere is a manifestation of thousands of rising warm gas cells surrounded by a network of descending cool gas (Del Moro 2004); kilometer-per-second velocities average out to some extent, leading to a net blueshift of a few meters per second in Sun-like stars (Gray 2009) and operate on timescales from hours to a few days. The granulation

[78] Presentation by Damien Ségransan.

blueshift depends on stellar properties and for a given star varies as photospheric magnetic fields evolve (Dumusque et al. 2011).

Active regions on stars arise from magnetic fields that thread the photosphere. Magnetic fields coalesce into flux tubes, producing faculae (plage is the chromospheric counterpart) that are bright when small and dark when large (spots). The typical active structure that is observed on the Sun is composed of a dark spot in the center with strong magnetic fields and large flux tubes, surrounded by faculae. On average the faculae region is ten times larger than the dark spot region (Chapman et al. 2001). Magnetic fields act locally to inhibit convection, which suppresses the net convective blueshift induced by convection (Dravins 1982). A dark spot and a faculae are therefore redshifted regions relative to the quiet photosphere (Cavallini et al. 1985).

Magnetic flux tubes form and decay on timescales comparable to the stellar rotation period, typically several days to a few months. Dark cool spots manifest as missing flux at the position of the spots. This breaks the symmetry of a rotationally broadened line, producing an east-west line asymmetry as a function of the rotational phase. This produces a net velocity shift and also changes the FWHM of all the spectral lines, and therefore the FWHM of the spectral CCF profile. As dark spots and bright plage evolve and rotate across the visible hemisphere, they alter the weighting of projected velocities (Saar & Donahue 1997; Dumusque et al. 2014).

Long term magnetic cycles change the convection patterns (e.g., inhibiting convection), thus changing line-bisectors and line-shifts (Dravins 1982). These cycles have timescales from a few hundred days to several years (Santos et al. 2010; Meunier & Lagrange 2013), comparable to the orbital periods of real Jupiter like planets.

It is easier to obtain high precision on very short timescales where the changes in the photospheric velocities that imprint variability in the spectral line profile are minimized. Bourrier & Hébrard (2014) devised an empirical correction for an observed wavelength dependent velocity trend and obtained residual velocity rms of 30 cm s^{-1} when fitting the Rossiter-McLaughin effect for 55 Cancri e.[79] Long term precision is more difficult to maintain because the many sources of stellar jitter add spurious temporally correlated scatter to the COM Doppler velocities. These noise signals are a function of spectral type (lowest jitter for K dwarfs and larger for warmer stars) and evolutionary state of the stars (higher for subgiants or more evolved stars). Fortunately, stellar jitter has two important properties that can be exploited:

1. it is not a persistent Keplerian signal; it waxes and wanes, it is not perfectly coherent, and it varies on timescales that are different from COM radial velocities
2. the underlying physical phenomena that spawn jitter have detailed spectroscopic, photometric, wavelength dependent, and polarization signatures that are in principle distinguishable from simple wavelength shifts due to Keplerian Doppler shifts.

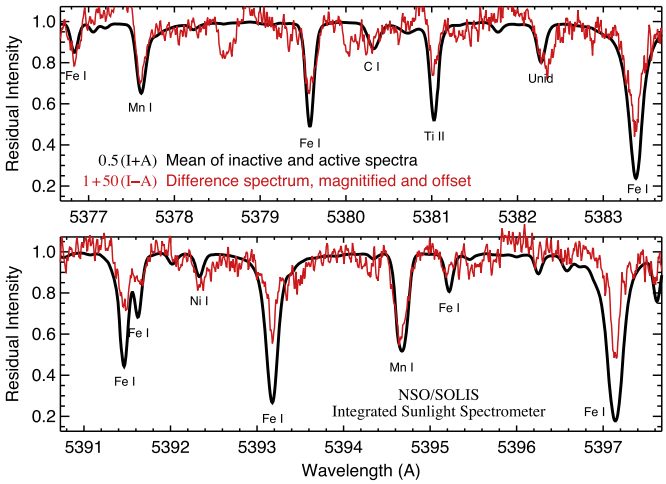


Figure 19. Full disk solar spectrum and the magnified difference between active and inactive states. Individual lines respond differently to activity changes, providing a way to help distinguish surface phenomena from Keplerian velocity shifts. (Courtesy of Jeff Valenti.)
(A color version of this figure is available in the online journal.)

Figure 19 compares the full disk solar spectrum at times of low and high activity, overplotting the magnified difference spectrum (red) on the average solar spectrum (black). Individual lines repond differently to changes in photospheric activity and provide a way to distinguish surface phenomena from Keplerian velocities.

4.2.1. Spectral Diagnostics of Jitter

Astronomers have used diagnostics of stellar activity to decorrelate photospheric signals and radial velocities.[80] These diagnostics include measurements of emission in the cores of Ca II H&K spectral lines, changes in the spectral line bisector, or variability in the FWHM of the spectral CCF (Dumusque et al. 2014). Emission in the cores of H α, the Na–D lines, or the sodium doublet at about 819 nm (Schlieder et al. 2012) provides more information about stellar activity for M dwarfs, which do not have much flux at the Ca H&K wavelengths.[81]

The variations in the line bisector and FWHM are induced by active regions and are increasingly pronounced with stellar rotation. However, when looking at slow rotators like the Sun, it is more difficult to resolve the tiny variations in the bisector and FWHM, making these ineffective diagnostics of stellar

[79] Presentation by Guillaume Hébrard.

[80] Presentation by Nuno Santos.

[81] Presentation by Paul Robertson.

activity (Desort et al. 2007; Dumusque et al. 2014; Santos et al. 2014). Observing with higher spectral resolution and with high sampling should help and additional spectroscopic diagnostics must be found. Currently, if significant signals can be detected in activity diagnostics, it is at least a warning that prospective Doppler signals may be spurious (Queloz et al. 2001).

Two important upgrades were made to the HARPS pipeline that improved the precision of spectroscopic activity diagnostics. First, it was found that subtraction of scattered light (especially from the ThAr calibration fiber) near the Ca II H&K lines improved the $\log R'_{\rm HK}$ measurements (Lovis et al. 2011). Second, the blue-to-red flux ratio varies because of weather conditions; imposing a fixed spectral energy distribution to the star before calculating the CCF vastly improved the CCF bisector and FWHM diagnostics[82] (Cosentino et al. 2014). The very stable spectral line spread function has been a decisive advantage for high-fidelity spectrographs like HARPS that allows for recovery of spectroscopic activity indicators, even for chromospherically quiet stars.

A caveat is that the activity indicators can be affected by instrumental effects and should not be used blindly. Over the past decade, there has been a small focus drift in HARPS that might have been mis-interpreted as a drift in FWHM from instrinsic stellar variability. Because this was seen for many stars, it is clearly an instrumental effect and was fitted out. In another example, low frequency modulation in the radial velocities was observed as a function of air mass in the star 18 Sco (and has also been seen in other stars). The source of this variation is ambiguous; is this granulation? seeing? S/N?

The stellar spectrum contains a wealth of information that remains to be explored. The physics of stellar atmospheres should guide us in devising clever spectroscopic diagnostics and different stars may require different approaches. Regions of faculae and spots in the Sun have been modeled and demonstrate unequivocally that these features produce velocity variations of a few meters per second over one activity cycle (Lagrange et al. 2010; Meunier et al. 2010; Borgniet et al. 2015). Similarly, Haywood et al. (2016) shows correlations for simultaneous Solar Dynamic Observatory data and HARPS velocity measurements. Other useful approaches might include Zeeman modeling of the magnetic field strength (Reiners et al. 2013) and magneto-hydrodynamic simulations to better understand the effect of granulation on line formation (Cegla et al. 2015). The HARPS-N solar telescope will provide excellent data to understand further the effect of stellar activity on RV measurements (Dumusque et al. 2015a). Studies like these associate physical phenomena with RV variations and offer an important means for disentangling photospheric signals from COM velocities.

[82] Presentation by Christophe Lovis.

4.2.2. Identifying Jitter with Doppler Imaging

Stellar astronomers carry out Doppler imaging on rapidly rotating stars by inverting time series spectral line profile information.[83] Conventional wisdom suggests that it is not possible to carry out Doppler imaging on slowly rotating stars, but progress has recently been made by employing principal component analysis (PCA) of the CCF for time series RV data from HARPS.

To motivate this analysis, solar synoptic magnetograms from 2012 June were used as a model for the magnetic field on the surface of the Sun. Magneto-convection simulations were then used to simulate the temperature and velocity structure with one degree spatial resolution using the MURaM code (Vögler et al. 2005) and a radiative transfer code was used to produce composite spectra for 156 absorption lines at each of 40 rotation phases. A singlular value decomposition (analogous to the CCF) of the time series spectra revealed an apparent RV variation (induced by the photospheric surface phenomena) with the same period as the stellar rotation and an amplitude of $\sim$1.5 m s^{-1}. Contributions to the simulated photospheric velocity variations came from the magnetic field (from Zeeman splitting; $<$1% effect), spots ($<$10% effect) and plage ($\sim$90% effect). The 40 time series CCF profiles were then inverted to produce a reasonably realistic reconstruction of the surface features that were initially used to generate the spectra.

In his workshop presentation, Thorsten Carroll reported promising progress using PCA on the CCF of HARPS spectra for the slowly rotating star HD 41248 (rotation period of 25 days). This star is particularly interesting because of confusion about whether radial velocities are Keplerian signals or a red noise source from photospheric signals (Jenkins et al. 2013; Jenkins & Tuomi 2014; Santos et al. 2014). Carroll found that the first PCA eigenvector contains almost the entire RV signal. The spectral lines from HD 41248 were then divided according to low ($<$1.25 eV) and high ($>$2.5 eV) excitation potential (EP). The CCF of both sets of lines showed essentially the same phase-folded periodicity in the RV; however, an offset in the velocity zero point was observed. Velocities derived from the CCF for the high EP lines were blue-shifted by about 150 m s^{-1} relative to the low EP lines because they form deeper in the stellar atmosphere and where the convective velocities are higher. The eigenfunctions of the CCF contained additional interesting information; the first eigenfunction for the high EP lines showed an asymmetry (a broadening effect), while the first eigenfunction for the CCF of low EP lines was symmetrical. Thus, the spectral line shape varied with depth in the stellar atmosphere in a way that is consistent with the RV signature—this is an impressive demonstration that the CCF velocity signature here comes from convection, not from a COM velocity of the star. Thus, radiative transfer with

[83] Presentation by Thorsten Carroll.

magneto-convection simulations may ultimately offer an important foundation for understanding photospheric velocities as we move toward 10 cm s^{-1}.

4.3. NIR Radial Velocities

For sunlike stars with spots, there may be an advantage to observing at NIR wavelengths.[84] The temperature of a spot can be 500–1500 K cooler than the surrounding photosphere. If we adopt a toy model of two blackbodies with temperatures representing the spot and the stellar photosphere, the contrast between the emitted light is much higher at blue wavelengths than at red wavelengths. As a result, the RV modulation for spotted stars has larger amplitudes in the optical, compared to NIR wavelengths, an effect that has been well-studied in young active stars (Prato et al. 2008; Mahmud et al. 2011). Because of the importance and potential of RV measurements at NIR wavelengths, many future instruments, including CARMENES (Quirrenbach et al. 2014), the Infrared Doppler spectrograph for high-precision RV measurements (Tamura et al. 2012), SPIRou (Artigau et al. 2014b), and HPF (Mahadevan et al. 2012), are being designed to operate in the NIR.

Most of the activity indicators that are used today are simply not as precise or as efficient as we need them to be.[85] There is not a single activity indicator that is 100% efficient and it is unrealistic to assume that all signals without activity correlations are correct. One way to confirm the planet interpretation is to measure Doppler velocities in the NIR; the precision for the NIR radial velocities is now at the level of about 2 m s^{-1} (Figueira et al. 2010b; Crockett et al. 2012).

Huélamo et al. (2008) showed that the RV variations using optical spectroscopy of a star in the TW Hydra association disappeared when observed with CRIRES using NIR wavelengths. Infrared CRIRES data were also used to confirm that variations in several K giants were caused by planetary companions (Trifonov et al. 2015). Bean et al. (2010) used an ammonia gas cell for wavelength calibration on CRIRES to achieve a Doppler measurement precision of 5 m s^{-1} on M dwarf stars.[86] They found that the noise floor for NIR precision was set by telluric contamination; there was no NIR wavelength region in the CRIRES bandpass that was free of telluric lines at the level needed to detect habitable planets, even around M dwarfs.

CRIRES is fed with light from an AO system, which helps to stabilize the slit illumination. While the slit fed spectrograph is ideal for background sky subtraction, intensity scrambling is not possible, so the spectrograph has variable illumination of the optics. The spectrograph is cryo-cooled, however, it is not a stabilized instrument. Bean et al. (2010) found that the use of an ammonia gas cell calibration for the K-band (2.3 μm) required many free parameters to forward model the instrumental point spread function and the system was less efficient because the K-band was not at the peak of the spectral energy distribution of their early M dwarf stars. However, the NIR spectral range does contain significant Doppler information for mid to late M dwarfs (Beatty & Gaudi 2015; Figueira et al. 2016) and advantages of working at NIR wavelengths to mitigate activity signals make it worthwhile to continued efforts to improve precision in the NIR.

With faculae or plage, the associated magnetic fields suppress convection and this perturbs the spectral line profile and produces photospheric Doppler shifts. Although there is only a small temperature difference between faculae and the photosphere; Marchwinski et al. (2015) used the FF' technique (Aigrain et al. 2012) on photometry from the SORCE spacecraft to show that the plage-dominated activity in the Sun exhibits smaller RV scatter at NIR wavelengths than optical wavelengths. So, the NIR may be useful as a diagnostic even in stars that have small spots but significant faculae and plage.

Interestingly, Reiners et al. (2013) found that the Zeeman effect, which increases with wavelength, can spuriously increase the RV signal at redder wavelengths. The amplitude of the RV signal caused by the Zeeman effect alone can be comparable to that caused by temperature contrast; a spot magnetic field of about 1000 G can produce a similar RV amplitude as a spot temperature contrast of about 1000 K. For the active M dwarf AD Leo, Reiners et al. (2013) found that the RV signal increases at longer wavelengths, consistent with a strong influence of the Zeeman effect. Therefore, the RV signal depends on the combination of spot temperature and magnetic field.

There are several instrumental challenges for EPRV with infrared spectrometers; these include detector technology, thermal and environmental control, modal noise, wavelength calibration and sky background (Anglada-Escudé et al. 2012; Plavchan et al. 2013a, 2013b). For most of these issues, progress has lagged behind optical instruments. The use of adaptive optics may be the one area where NIR spectrographs have an advantage over optical instruments and the use of SMF feeds to IR spectrographs (e.g., iLocator, Crepp 2014) offers an interesting path forward.

4.3.1. Simulating Stellar Noise

The community has developed simulation tools for understanding and modeling stellar activity, including the SOAP (Boisse et al. 2012) and SOAP 2.0 (Dumusque et al. 2014) code for simulating realistic spectra that captures the effect of spots and faculae. A new code, StarSim (Herrero et al. 2015) produces realistic synthetic time series spectra that accounts for limb darkening, spots, faculae and convection using a surface

[84] Presentation by Lisa Prato.
[85] Presentation by Pedro Figueira.
[86] Presentation by Jacob Bean.

Table 2
Fitting Teams

	Teams	People	Techniques
1	Torino	M. Damasso, A. Sozzetti, R. Haywood A.S. Bonomo, M. Pinamonti, and P. Giacobbe	Bayesian framework with Gaussian Process to account for red noise
2	Oxford	V. Rajpaul and S. Aigrain	Bayesian framework with Gaussian Process to account for red noise
3	M. Tuomi	M. Tuomi and G. Anglada-Escudé	Bayesian framework with Moving Average to account for red noise
4	P. Gregory	P. Gregory	Bayesian framework with Apodised Keplerian to account for red noise
5	Geneva	R. Diaz, D. Ségransan and S. Udry	Bayesian framework with white noise
6	A. Hatzes	A. Hatzes	Pre-whitening
7	Brera	F. Borsa, G. Frustagli, E. Poretti and M. Rainer	Filtering in frequency space
8	IMCCE	N. Hara, F. Dauvergne and G. Boué	Filtering in frequency space

integration and includes photometric information.[87] The inputs to StarSim include stellar parameters (effective temperature, log g, metallicity, rotation, differential rotation, and inclination) information about the active regions (spot positions and sizes, temperature contrast, faculae temperature contrast, faculae to spot area factor, and evolution of the active region) and information about planets (size, ephemeris, spin–orbit angle, impact parameter). The spectroscopic contribution from the photosphere, spots and faculae are synthesized from the Phoenix BT-Settl models) and stored in a spherical grid representing the star. The synthetic spectrum is calculated by integrating the surface elements that contain temperature, limb darkening, RV and geometry (size and projection). Time series observations are produced as a function of the rotational phase angle. By using the HARPS mask template, the CCF models can be produced and analyzed to try to recover the underlying simulated noise.

4.3.2. RV Fitting Challenge

In advance of the workshop, Xavier Dumusque organized a challenge[88] for fitting Keplerian signals in RV data.[89] He provided simulated RV data sets, using the time stamps of HARPS observations for some well-sampled stars. The data contained stellar signals from from oscillations, granulation, spots and plages, magnetic cycles and Keplerian signals. Eight teams participated in this simulation and used different techniques to recover planetary signals, as detailed in Table 2. The results of the RV fitting challenge are fully described in X. Dumusque (2016, in preparation). Here, we summarize the take away messages of this exercise.

For the RV fitting challenge, 14 data sets were provided, including 4 data sets from real HARPS observations and 10 simulated data sets. A total of 48 planetary signals were present in the data, with semi-amplitudes ranging from 0.16 to 5.85 m s^{-1}. In Figure 20, we summarize the results of the different teams. For each signal detected, we assign a different color flag depending on the true signals present in the data. The different possibilities are:

1. Dark green: the group recovered a planetary signal that existed in the simulated data, and the group was confident in the detection, meaning that they would have published this result.
2. Light green: the group recovered a planetary signal that existed in the data, however the group assigned a low confidence to the detection.
3. Yellow: the group recovered a planetary signal that existed in the data and was confident enough for publication; however, the amplitude or period is slightly wrong compared to the true signal, or an alias of the true signal was detected.
4. Gray: the group recovered a planetary signal that existed in the data, but was not confident in the detection. The amplitude or period was slightly wrong compared to the truth, or an alias of the true signal was detected.
5. Orange: the group recovered a prospective planetary signal that did not exist in the data, but assigned a low confidence level to the detection
6. Red: false positive or false negative, i.e., the group recovered a planetary signal that does not exist in the data, but they were confident enough in the detection for publication (false positive), or the group rejected the detection of a true signal (false negative).
7. White: a simulated planetary signal was present in the data that has an amplitude larger than 1 m s^{-1}, but this signal was not detected.

To study the statistics of exoplanetary populations, it is most important that published planets with correct parameters (dark

[87] Presentation by Enrique Herrero.
[88] https://rv-challenge.wikispaces.com
[89] Presentation by Xavier Dumusque.

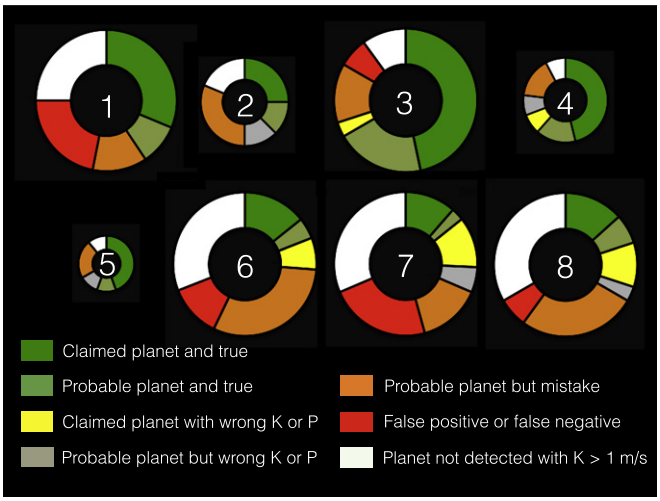


Figure 20. Summary of the result of the RV fitting challenge for teams 1–8. The size of the circles represents the number of systems analyzed: 14 systems for the large size circle, 5 for the medium size and 2 for the smallest size. The meaning of the colors are defined in the legend and in the text. (Courtesy of Xavier Dumusque.)

(A color version of this figure is available in the online journal.)

green) are found. It is also important to understand when false positive or false negative signals (red) are reported and when relatively large signals with amplitudes that exceed error bars are missed (white). Signals that are published with modest errors in their parameters (yellow) will not strongly corrupt the planet population statistics. The RV challenge also helped to illucidate trends regarding the signals (light green, gray, orange) that were either true or false, but that would not be published.

In Figure 20, the most successful groups should have a larger dark green region, and smaller red and white regions. Given this metric for success, we separate the teams in two different groups: teams 1–5 (groups that used a Bayesian framework to compare between models) and teams 6–8 (groups that did not use Bayesian models). Groups 1–4 also used red noise models in addition to Keplerians to account for stellar signals, and group 5 used a white noise model that was free to vary depending on the activity level of the star in addition to Keplerian models. The other groups compared RV signals with signals present in the activity observables (log $R'_{\rm HK}$, FWHM, BIS SPAN) and rejected significant signals it they were also in the RVs or in any of the stellar activity diagnostics.

According to X. Dumusque (2016, in preparation), among the teams that used a Bayesian framework to compare between models and red noise models to account for stellar signals, team 3 was the best at finding true planetary signals; they were confident in the majority of their detections and they made a small number of mistakes. This team detected 85% (17 out of 20) of the true planetary signal present in the data that had a semi-amplitude larger than 1 m s^{-1}. The remaining 3 signals that were not detected all have periods longer than 500 days and their non-detection could be explained either by magnetic cycle RV effects, or simply by the fact that team 3 only fitted up to a second order polonomial to remove any RV trends in the data. This team would have only published 60% of these large amplitude signals, leaving the remaining signals as tentative. When analyzing planetary signals with semi-amplitudes smaller than 1 m s^{-1}, team 3 did also the best job at recovering those signals, with 14% detected (4 out of 20). However the team would have only published 2 of those planets, with amplitudes of 0.48 and 0.69 m s^{-1}. Finally team 3 refrained from announcing non-existing planets with semi-amplitudes larger than 1 m s^{-1}, and would have only announced 2 non-existing systems with smaller semi-amplitudes.

The next two best performing teams were Team 1 and Team 4. While Team 4 ranked third, they actually did a better job than Team 1 on the fitting first five systems. However it was difficult to make a fair comparison because they did not analyze the other systems.

In conclusion, the RV fitting challenge (X. Dumusque 2016, in preparation) showed that teams using a Bayesian framework that allows for model comparison, and that includes a red noise model to account for stellar signals were more efficient at finding true planetary signals, while limiting the number of false claims. Among the different red noise models, it the moving average model used by M. Tuomi and G. Anglada-Escudé (Team 3) was the most efficient. The use of Gaussian Processes was also an efficient approach, however this technique seems to be sensitive to the a priori determination of the rotation period; it seems likely that some improvement can be made here (X. Dumusque 2016, in preparation). The apodized Keplerian technique (Team 4; Gregory 2016) also yields good results. Gregory (2016) finds that it is possible to achieve a factor of about six reduction in the stellar activity noise in the simulated RV sets with the apodized Keplerian technique. Unfortunately, Team 4 only analyzed only the first 5 systems as they were developing their technique, and this ultimately penalized their ranking. Finally, planetary signals with amplitudes above 1 m s^{-1}were detected almost all of the time, while only 14% (4 out of 28) of true smaller signals were discovered. Out of these 4 true small signals, 2 would have been published; however, 2 false planetary signals would have been published as well. Therefore, even with the best models of stellar signals, planetary signals with amplitudes less than 1 m s^{-1} are rarely extracted correctly with current precision and current techniques.

4.4. Aliasing

RV data sets will generally include aliased signals, which are interactions between gaps in data sampling and signals.[90] A sinusoidal signal in time series data is a delta function in the frequency domain. Because observations are taken at specific times, a window function is applied to the data; in the time domain this results in sampling gaps. The convolution of the sinusoidal frequency and the window function produces multiple peaks in the frequency domain.

How do we know what the time sampling is for complex sampling patterns? Solar and sidereal day and annual sampling gaps naturally appear in most RV data sets and these aliases and their harmonics produce peaks at positions that can be calculated:

$$f_{\rm alias} = |f_{\rm true} \pm f_{\rm sample}|. \quad (7)$$

Dawson & Fabrycky (2010) show that aliases imprint a set of predictable peaks; by modeling a noise free data set, it can be possible to confirm Keplerian signals in the presence of aliases. However, low S/N data is more ambiguous when trying to identify aliasing, and chromospheric activity can produce a forest of peaks with associated aliasing signals. Stellar activity does not produce an exact sinusoidal signal; instead, the imprinted perturbations are stochastic and quasi-periodic. The potent combination of variable intensity in the stellar noise signal and poor sampling leads to ambiguity in identifying aliases that can sabotage the identification of weak Keplerian signals.

4.5. Barycentric Corrections

Corrections for the velocity of the Earth must be applied in order to recover precise RV measurements of stars for exoplanet detection.[91] Wright & Eastman (2014) outline the magnitude of error contributions to the barycentric velocity correction, including rotational and orbital motion of the Earth, second order corrections to the non-relativistic Doppler formula, secular acceleration, precession, nutation, gravitational redshifts, blueshifts, and the Shapiro delay (for light bending around the Sun). In their publicly available code for barycentric correction,[92] the authors show that their calculations match those from the TEMPO2 (Edwards et al. 2006; Hobbs et al. 2006) code at a rms level of 0.3 mm s^{-1}. The accuracy of TEMPO2 is a gold standard with a precision better than 1 cm s^{-1} that was demonstrated when measuring the Doppler effect of pulse arrival times for the detection of Earth-mass exoplanets orbiting a neutron star (Wolszczan & Frail 1992).

The barycentric correction for RV measurements at the 10 cm s^{-1} level requires a precision of about one second in the flux-weighted exposure time. This is generally accomplished by monitoring the flux of starlight into the spectrograph at regular intervals over the exposures. The barycentric velocity correction should be determined for the time the signals arrive at the solar system barycenter—i.e., at the Barycentric Julian date in Barycentric Dynamical Time (BJD_{TDB}). An error of 1 s in time introduces a semi-amplitude error of about 3 cm s^{-1}, so to reach an error of less than 1 cm s^{-1} the midpoint time must be accurate to about 250 milli-seconds (Eastman et al. 2010; Wright & Eastman 2014).

An effect that was not addressed by Wright & Eastman (2014) is the need to integrate the barycentric correction over the exposure time (Figure 21 is provided from an article, L. A. Buchhave et al. 2016, in preparation). A barycentric correction should be calculated for each sampled flux measurement and then weighted by the photon counts. This updates the older procedure of calculating a single barycentric correction to the photon-weighted midpoint time. The older procedure was deemed as acceptable for most state of the art Doppler surveys because the relatively short exposures incurred an error from the acceleration term in the barycentric velocity that was far below other terms in the error budget. However, this is an error source that must be handled more carefully for higher precision surveys. Even assuming uniform flux during an exposure, the older procedure can introduce a velocity error of ± 25 cm s^{-1} for a 30 minute exposure; this scales as the square of the exposure time.

The optimal flux sampling rate depends, unfortunately, on the apriori unknown rate of variability in the atmosphere. If the atmospheric opacity is constant, then the midpoint time precision is independent of sampling and lower sampling rate is balanced by the benefit of extra photons. Even in an extreme scenario with a 6 s "blackout" at the end of an otherwise uniform 5 minute exposure, integrating flux in 1 minute bins would produce an error in the flux-weighted midpoint time of only 0.6 s. Although more detailed analysis is warranted on the subject, binning flux every ~10 s is reasonable for almost every imaginable weather condition and still provides good sampling for short exposure times of about one minute. For extreme RV precision, it is recommended that the exposure times are limited to about 20 minutes to minimize the error in the barycentric velocity calculation from nonlinear terms discussed above.

For an accuracy in the barycentric velocity of ~1 cm s^{-1} it is also important to sample photon counts as a function of wavelength. For example, if observations are taken at large zenith angle without an ADC the image of the star will be chromatically dispersed; if guiding favors collection of the red light for the first half of the exposure and then collection of blue light for the second half of the exposure, the flux-weighted midpoint would be different for the red and blue photons. However, even with a good ADC, variable wavelength-

[90] Presentation by Rebekah Dawson.
[91] A breakout session on barycentric corrections was led by Jason Eastman and Lars Buchhave.
[92] http://astroutils.astronomy.ohio-state.edu/exofast/

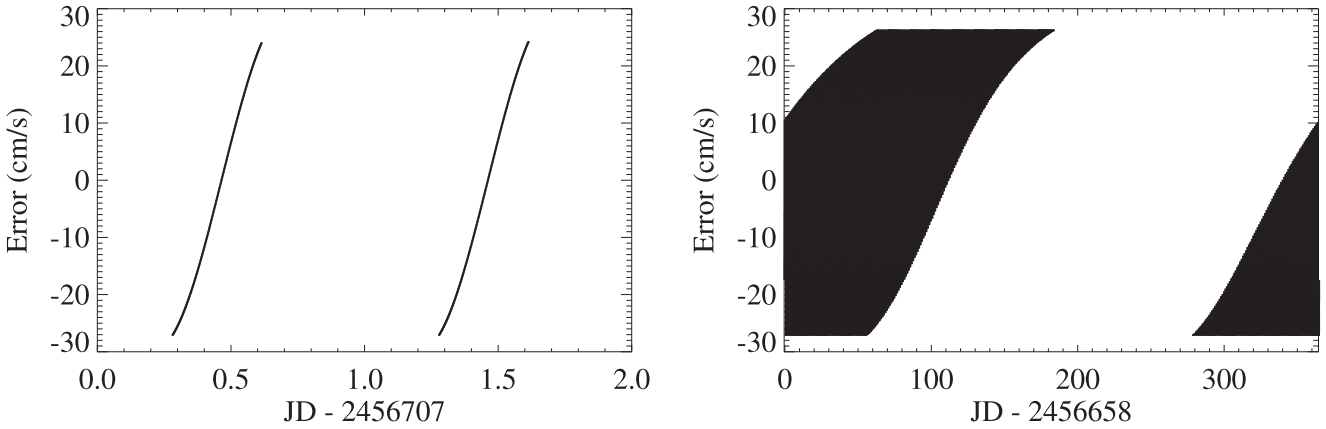


Figure 21. Radial velocity error when acceleration of the barycentric velocity is not properly handled (i.e., when a single barycentric correction is derived for the flux-weighted midpoint time, which is current standard practice in the community). This simulation assumes 30 minute observations of Tau Ceti taken from CTIO with uniform flux during the observation. Velocity errors are only calculated for times when the star would be observable (i.e., airmass less than 2 and the Sun below $-18°$ twilight). (Left) RV errors that could be incurred over a 2 day diurnal cycle. (Right) RV errors over a full year when the same observing constraints are applied. (Figure courtesy of Lars Buchhave, Jason Eastman, Jason Wright.)

dependent extinction can occur as the telescope tracks the star through a range of hour angle, even for short exposures. A basic exposure meter design might use a dichroic to divide the light between two channels; however, a low resolution spectrometer with 8–10 channels is probably preferable for providing a better correction as long as a fast readout can maintain ∼10 s sampling with high S/N. Once the wavelength-dependent barycentric correction has been derived, it should be interpolated (linearly or weighted with an extinction curve) for each wavelength chunk or order and applied when weighting and combining the velocities from different parts of the spectrum. For example, if each order is cross-correlated, then a barycentric correction can be calculated for the central wavelength and applied on an order-by-order basis before calculating the average velocity.

Recommendations for good practices in deriving precise barycentric corrections include sampling the flux during the exposure (at least every minute, but ideally every few seconds), recording all times explicitly as a function of wavelength, ensuring the accuracy of the time stamp in the file headers, and including the time standard (usually GMT or UT).

4.6. Telluric Contamination

Telluric absorption lines from many species (e.g., H_2O, O_2, CH_4, CO_2) are imprinted in a stellar spectrum when starlight passes through the Earth's atmosphere. Some molecular species are well-behaved with only small seasonal variations in column density, while the lines associated with water can vary in depth with changing humidity throughout the night (temporally) or with different positions on the sky (spatially). There is some variability in the wavelengths and line profiles of telluric lines because of atmospheric winds, generally limited to Doppler shifts of about 10 m s^{-1} (Figueira et al. 2010a), corresponding to a fraction of a pixel. More important, the telluric lines raster across the stellar spectrum with time because of the barycentric velocity of the Earth.

Like any attribute that perturbs the shape of the SLSF, telluric contamination must be properly treated for high precision RV measurements.[93] Left untreated, telluric contamination is most important in the NIR where it has an impact of at least a few meters per second (Bean et al. 2010) and where telluric lines can be quite opaque (with $\tau > 0.2$). The impact at optical wavelengths is of order ∼0.2–1 m s^{-1} where (Figure 22) micro-telluric lines have depths of 0.2%–2% (Artigau et al. 2014a; Cunha et al. 2014).

Most groups correct or model telluric lines with physically motivated synthetic models that have a comprehensive line list and use radiative transfer with accurate atmospheric profiles, which is demonstrated to be more accurate than empirical correction using telluric calibration frames (Gullikson et al. 2014; Smette et al. 2015). Most codes are base on the line-by-line radiative transfer model (LBLRTM; Clough et al. 2005) and the High Resolution Transmission (HITRAN) line database (Rothman et al. 2013). Examples of published codes are: TAPAS (Bertaux et al. 2014), TelFit (Gullikson et al. 2014), Molecfit (Smette et al. 2015), and TERRASPEC (Bender et al. 2012). A model of the telluric spectrum from 0.3 to 30 microns is shown in Figure 23. These codes model telluric lines to a precision of 2%–5%; however, for deep lines or lines with large uncertainties in the HITRAN database, the precision

[93] The telluric contamination breakout session was organized by Sharon Wang with panel members Jason Wright, Cullen Blake, Pedro Figueira, Nuno Santos, Peter Plavchan, Andreas Seifahrt, Claire Moutou, Francois Bouchy.

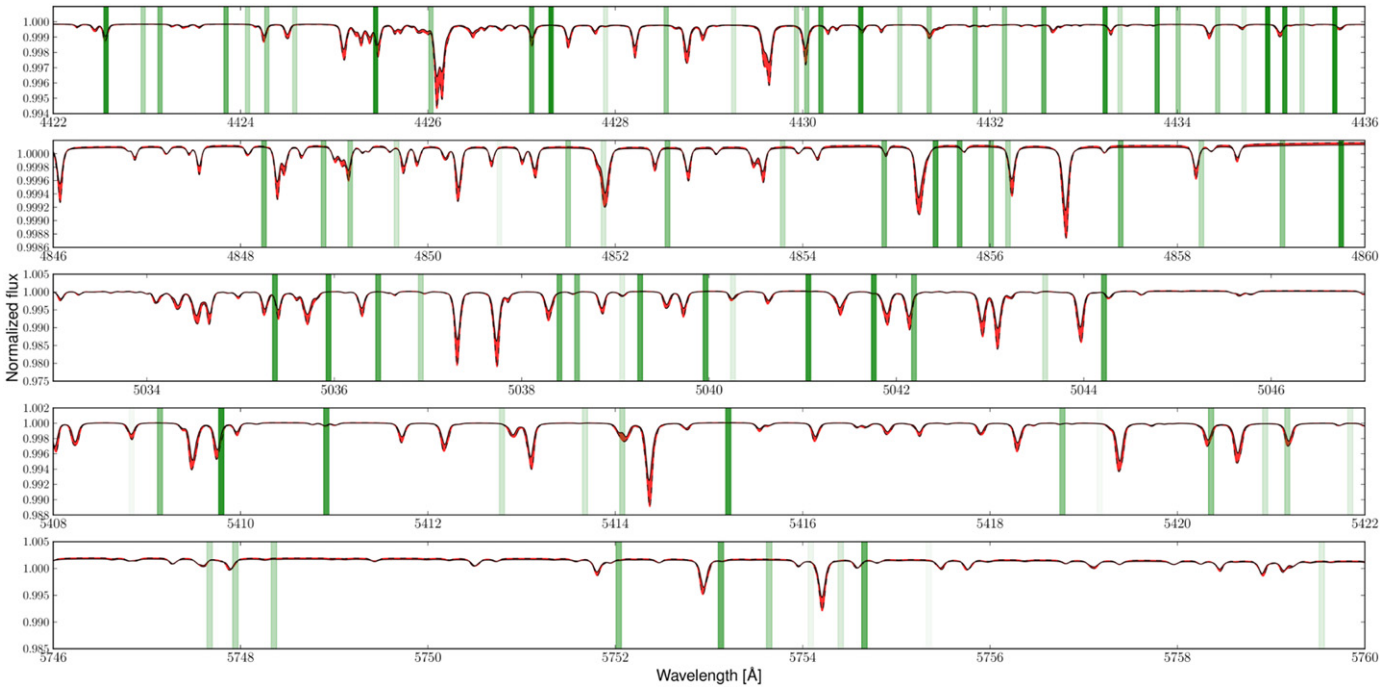


Figure 22. Reproduced Figure 1 from Cunha et al. (2014). Spectrum of micro-telluric lines in the optical spectrum between 442 and 570 nm. (Figure courtesy of Diana Cunha and Nuno Santos.)
(A color version of this figure is available in the online journal.)

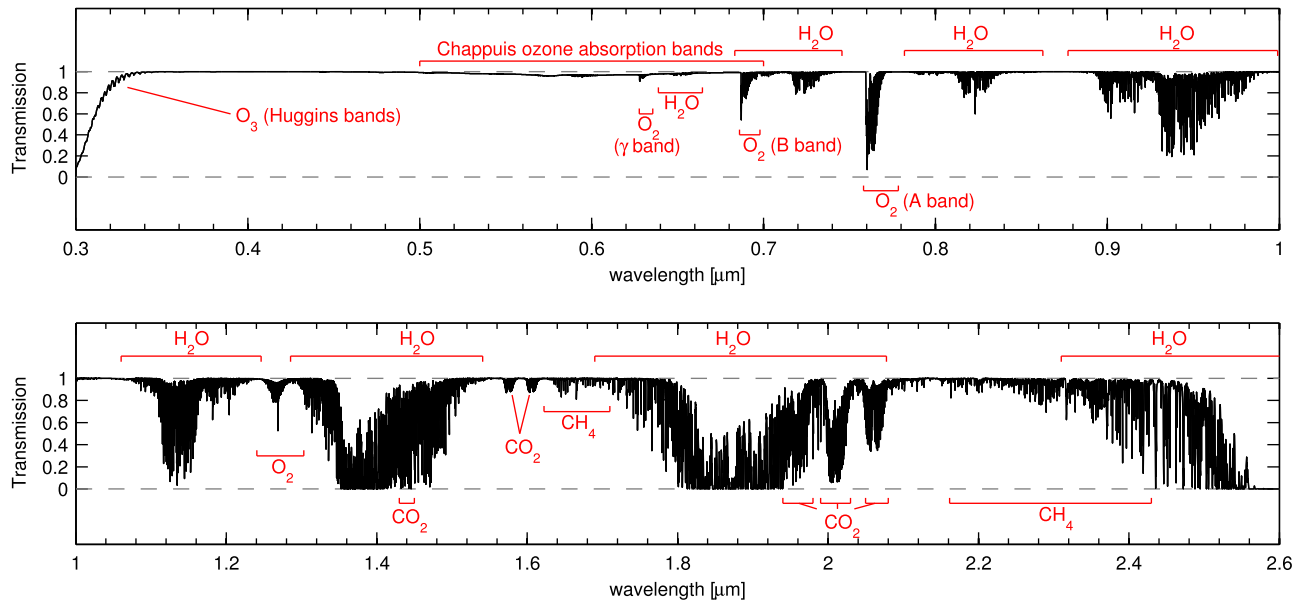


Figure 23. Reproduced Figure 1 from Smette et al. (2015). Spectrum of telluric lines synthesized with a line-by-line radiative transfer code using the annual mean profile for Cerro Paranal at a resolution of about 10,000. (Figure courtesy of Wolfgang Kausch, Alain Smette, Stefan Kimeswenger, Stefan Noll.)
(A color version of this figure is available in the online journal.)

is not as good. Poorly fitted lines are masked out or rejected (Bean et al. 2010; Blake et al. 2010; Seifahrt et al. 2010).

The current telluric modeling precision is limited by: missing lines in the HITRAN database, uncertainties or errors in attributes such as the line position, strength, or shape, limitations in current modeling codes for deriving correct line profiles (i.e., velocity dependence, line mixing effects), insufficient knowledge of real time atmospheric conditions (e.g., water column density variations), and wind-induced line shifts. There are annual meetings on HITRAN where help could be requested for specific lines or species under a range of physical conditions (excitation levels, temperatures). It is also possible that telluric modeling can be improved with empirical corrections using on-sky data or a Fourier Transform Spectrograph (FTS) observation although micro-tellurics will be difficult to observe with high S/N. Perhaps a version of HITRAN can be designed for astronomy applications through compilation and modeling of a large volume of telluric calibration data. This "Astro-HITRAN" would be simpler but more robust for astronomical uses for certain molecular species, and it would allow for scaling of line strengths, for example, for different amount of water vapor content. The variable water species still present a modeling challenge, and Blake & Shaw (2011) have suggested a clever use of global positioning satellite measurements of precipitable water vapor as a prior when scaling these features.

5. Synergy with Other Fields

Precision radial velocities are critical to the success, efficiency and scientific yield of future space missions.[94] There is obvious synergy with transit missions (*TESS*, *CHEOPS*, *PLATO*) in deriving bulk densities for thousands of exoplanets. New connections will be made with Gaia, where of order 10,000 new gas giant planets are expected to be detected. The combination of Gaia astrometry and radial velocities will illuminate several questions about planet formation: the true mass distribution of jovian planets beyond the ice line, the frequency of solar system analogs, whether super Earths are regularly accompanied by giant planets and how gas giant planets affect the presence of terrestrial planets.

RV measurements can also enhance the productivity of direct imaging missions by providing orbital phase information and information about inner companions. The ideal targets for direct imaging are young stars where the gas giant planets will be brighter; these will be amenable to RV follow-up if successful techniques are developed for mitigating photospheric noise in these active stars.

[94] Presentation by Alessandro Sozzetti.

6. Summary

A comparison of several ongoing Doppler surveys shows that most teams have reached about $1\,\mathrm{m\,s^{-1}}$ in single measurement RV precision. The best performing spectrometer is the purpose-built vacuum-enclosed HARPS instrument, with $R = 115{,}000$, a broad wavelength range from 400–700 nm, and a fiber feed system that utlizes octagonal fibers and an optical double scrambler. HARPS reaches a precision of about $0.8\,\mathrm{m\,s^{-1}}$ for observations with S/N of 200 in $3\,\mathrm{km\,s^{-1}}$ bins at 500 nm. The next generation instruments will need to improve upon HARPS with better illumination stability, higher fidelity spectra (more stable spectral line spread function, higher resolution and higher S/N), broader wavelength bandpasses and more stable and precise wavelength calibration, and improved detector performance and characterization efforts.

Tremendous progress has been made over the past five years on technology development for high-resolution optical spectrometers. In particular, the development of frequency combs for the wavelength calibration of spectrometers for Doppler planet searches is now mature technology. Laboratory experiments and initial tests on HARPS indicate that wavelength calibration is likely not the primary limitation in achieving $10\,\mathrm{cm\,s^{-1}}$ Doppler precision. Frequency combs have become powerful tools for characterizing spectrometer drift, estimating the spectral line spread function across the instrument bandpass, and characterizing and calibrating detector imperfections. The remaining work to be done with frequency comb technology is primarily technical implementation, including making systems robust and reliable, and mitigating modal noise when coupling comb light into the spectrograph. A turn-key LFC is now available from Menlo Systems that shows great promise as a long-term calibration device. The Menlo combs have been in operation at the VTT solar telescope in Tenerife (since 2012), at HARPS in La Silla (permanent since 2015), at FOCES in Munich at USM (since 2015) and at Xinglong China (since 2016 February). At USM the LFC has been running continuously for several months. Beyond the more classical astro-comb calibration technologies, there is significant innovation in the fields of microcombs, electro–optical modulation combs, tunable FP, and stabilized etalons as well.

Fiber optic cables are also a well-understood technology. Not only does optical fiber technology allow for spectrometers to be placed in convenient locations relative to telescope focus, but modern fibers can also improve instrument illumination stability significantly while maintaining high efficiency. Care must be taken when connectorizing instrument fiber cables to ensure stress-free mounting (thereby minimizing FRD). Standard fibers used in astronomical instruments do incur modal noise penalties, as only a finite number of modes can be populated within the fiber, but much of this can be mitigated through mechanical agitation. However, modal noise is a

significant issue for coherent sources, such as LFCs, and our community must develop new mitigation techniques if the exquisite precision of next generation wavelength calibration sources is to be fully realized by future spectrometers.

Telluric contamination is another area that the community will need to address for high precision Doppler measurements. The problem is most challenging for infrared observations, but micro-tellurics at optical wavelengths are also a concern at sub meter per second precision. Current approaches involve forward modeling to fit for tellurics or masking out contaminated pixels in the spectrum. Routine monitoring of atmospheric conditions (water vapor distribution and column density, wind etc) is recommended and this information should be stored as meta-data with the spectra.

Barycentric corrections have been demonstrated to be accurate to about $2\,\mathrm{cm\,s^{-1}}$ (Wright & Eastman 2014). The flux-weighted exposure midpoint time should not be used for the correction; instead, barycentric velocity corrections should be calculated throughout an exposure and combined by weighting each point by the exposure meter data. Color-dependent barycentric corrections should also be calculated by collecting the chromatic flux time series as a function of wavelength (in a few or several wavelength bins). Any stellar exposure with sufficient S/N to achieve $<1\,\mathrm{m\,s^{-1}}$ precision should provide enough photons for time series at several wavelength bins. The final Doppler velocity is the combination of measured velocities from a set of lines or wavelength bins; each bin or line would then be corrected according to the BC from the appropriate wavelength region. A high performance atmospheric dispersion corrector (ADC) will reduce the sensitivity to flux-weighted barycentric corrections.

The community is now exploring correct statistical techniques with the goal of improving reproducibility and extracting weaker signals in the presence of time-correlated noise. Frequentist techniques are faster, but Bayesian techniques are more reliable in estimating the true errors and uncovering underlying exoplanet populations from our data. There is common misuse of null hypothesis tests like the *p*-value or Kolmogorov–Smirnoff test. Astronomers are encouraged to team up with statisticians to avoid the pitfalls of applying statistical tests as black boxes. Many of the techniques that seem new to astronomers are well understood techniques for statisticians.

There was extensive discussion about how to handle stellar noise. In the past, it was adequate to identify and decorrelate trends in radial velocities based on calcium H & K emission, or the FWHM or line bisector of the CCF. However, we now appreciate that photospheric velocities, or jitter, is fundamentally imprinted in the spectrum differently from a Keplerian Doppler shift. There are many techniques that show promise in distinguishing jitter from COM velocities, including PCA of spectral lines or the CCF. Stellar jitter is currently the main limitation in the detection of small rocky planets and it merits significant effort by the community. It will not work to avoid stars with jitter (because stellar activity varies over time). We are advancing our understanding of the physics of stellar noise with solar data and photometry from Kepler. Plage is a more important contributor (90% level) to stellar jitter than spots (a 10% problem) for slowly rotating stars.

It is not an adequate strategy to average down hundreds of observations; this wastes precious telescope time and systematic errors don't necessarily average down as the square root of the number of observations. The RV challenge led by Xavier Dumusque showed that the community has not been successful in detecting velocity amplitudes that are smaller than the SMP, even when hundreds of observations are obtained. We must improve SMP as a first step toward detecting low amplitude exoplanetary systems. We must learn how to distinguish stellar noise from Keplerian velocities.

There has been negative advocacy by some in the exoplanet field who claim that a fundamental floor of precision is imposed by photospheric velocities, or stellar jitter at the level of $1\,\mathrm{m\,s^{-1}}$. That statement may be true for many of the current spectrometers and analysis techniques. At a resolution of 50,000 on spectrometers that are not stabilized, stellar noise from chromospherically quiet stars cannot be distinguished from Keplerian velocities. Fortunately, the community at large has continued to press forward on this issue and it seems likely that an instrumental precision of $10\,\mathrm{cm\,s^{-1}}$ will be achieved with the next generation of high fidelity spectrometers. We do not yet know how this will translate into detectability for low amplitude signals in the presences of unavoidable stellar noise; however, it is premature to speculate that we cannot do better. This is an area of active research in the community and there has been promising progress. Research to distinguish stellar activity from COM Doppler shifts must continue to be a high priority, simply because the exoplanet endeavor cannot expand[95] if we do not solve this problem and extract weaker signals in the presence of time-correlated noise.

7. Findings and Conclusions

There have been several reports highlighting the importance of precise radial velocities for supporting NASA missions, including the 2006 Exoplanet Task Force (Lunine et al. 2008), the 2010 Decadal Survey (Blanford 2010), and the 2012 NSF Portfolio Review report (Eisenstein 2012). The cancellation of the $2B Space Interferometry Mission was accompanied by a statement that ground-based RV measurement precision could be pushed to $10\,\mathrm{cm\,s^{-1}}$, enabling at least partial characterization of the architectures of nearby planetary systems. The NASA ExoPAG commissioned a report "RV Prospects Current and Future" (Plavchan et al. 2015) with findings from the precision RV community[96] that complements and foreshadows

[95] Survey input from Rosemary Mardling.
[96] Presentation by Scott Gaudi.

many of the findings from the EPRV workshop. This report highlights a key challenge: EPRVs have transitioned from small PI-based programs to "big" science that requires dedicated resources; however, the NASA and NSF budget models have not kept up with this. While RV precision has been stuck at $1\,\mathrm{m\,s^{-1}}$ for the past several years, there has been progress in technology and analysis techniques that offers promise for moving toward $10\,\mathrm{cm\,s^{-1}}$ precision and this progress needs to be communicated to the astronomy community at large and to the funding agencies.

The "Big Three" science goals for extreme precision radial velocities are:

1. Confirmation and characterization of the many transiting planets discovered by transit surveys (ground-based, *Kepler*, *K2*, *TESS*, *CHEOPS*, *PLATO*). EPRV measurements will uniquely provide mass measurements needed for bulk density measurements.
2. Identification and orbital characterization of planets down to ice giant masses around FGK stars to be imaged by WFIRST-AFTA.
3. The discovery of potentially habitable planets around the nearest and brightest stars.

Resources will be required to meet these ambitious goals. Habitable planets around FGK type stars have velocity semi-amplitudes of order $10\,\mathrm{cm\,s^{-1}}$ and periods of about one year. The discovery of these planets requires high statistical precision and high cadence. Additional work needs to focus on reducing any systematic errors or astrophysical noise to considerably less than $10\,\mathrm{cm\,s^{-1}}$. In additional to controlling systemic errors resulting from instrumental and calibration effects, considerable additional work needs to focus on the critical and difficult task of removing, suppression, or separation of intrinsic stellar noise to considerably better than $10\,\mathrm{cm\,s^{-1}}$. This is in principle a tractable problem because the Doppler variation due to orbiting bodies has a unique signature: all of the lines move by a known amount without changing their shape. Current instruments and detection algorithms are not going to solve this problem; the RV fitting challenge by Xavier Dumusque (Section 4.3.2) showed that no one can reliably detect planets with amplitudes below $1\,\mathrm{m\,s^{-1}}$ in current data sets. We need higher fidelity spectra to make progress and this is a problem that will require the next generation instruments.

7.1. Recommendations

In support of the key role that precision spectroscopy has for space missions, NASA has commissioned the Extreme Precision Doppler Spectrograph for the WIYN 3.5 m telescope.[97] This is a start, but one new instrument is not enough. There is a role for small telescopes with high cadence observations and larger aperture telescope that obtain high resolution, high S/N observations. Participants at this workshop discussed the big ideas that could catapult progress in the field and enable success for the highest priority science goals of precision Doppler surveys.

1. Dedicated moderate to large aperture telescopes. High S/N and very high cadence data are required over long time baselines; this could be achieved for hundreds of stars using dedicated 4–10 m class telescopes. Most of the struggles we have with aliasing and stellar activity is complicated by under-sampled data sets. In the same way that *Kepler* transit detections benefited from high cadence sampling, Doppler measurements would benefit from a paradigm shift with many more observations, perhaps a world-wide effort with coordinated observations at several longitudes for a few stars.
2. Stable Spectrometers. Instruments with exquisite long-term stability are required: spectrometers that are fiber fed with high illumination stability, excellent wavelength calibration, and precise temperature and pressure control represent the immediate future of precision RV measurements. Optical spectrometers contain significant amounts of Doppler information for solar-type stars, but radial velocities from NIR spectrometers have less contamination from stellar activity (spots as well as plages). Based on current technology, solutions to the technical challenges presented at the workshop seem within reach for the spectrometers that are being built today.
3. Proper treatment of stellar noise. It is likely that distinguishing between stellar activity and Keplerian velocities will require very high resolution (perhaps ~150,000), and expanded wavelength coverage (optical plus NIR). The high resolution is not required for measuring line centroids; the line centroid information is resolved at $R \sim 60{,}000$. However, the need to measure higher-order line shape variations, which are a clear signal of stellar activity or changes in the instrumental line spread function rather than a true Doppler shift, likely require much higher resolution and sampling, as well as higher S/N. Simultaneous high-resolution optical and infrared monitoring can mitigate any pathological cases where stellar activity or instrumental effects may cause line centroid shifts without strong line shape variations, since these shifts are neverthless likely to wavelength dependent. In particular, the ionization and line formation depth in the photosphere depends on the wavelength, and thus so do the absorption line shapes through the photospheric velocities.

All of this leads to a vision for precision RV "dream machines" requiring large aperture telescopes with dual optical and infrared channels (0.4–1.7 μm), very high optical

[97] Presentation by Mario Perez.

resolution ($R > 150{,}000$), high IR resolution ($R > 50{,}000$), fiber-fed with a very high scrambling gain, environmentally stabilized, and with advanced wavelength calibration (e.g., LFCs). The cost would be of order $20 M per instrument and ideally there would be few of these at various longitudes. In order to carry out the science, these machines would require an assurance of a significant allocation of observing time over several years.

This is a cost that at first glance seems out of reach. However, if precisions approaching 10 cm s^{-1} can be achieved, Doppler surveys may be able to locate the host stars of planets that would be observed by the next generation of flagship direct imaging missions, such as the Habitable Exoplanet Imaging Mission and the Large UV–Optical–IR (LUVOIR) mission, both of which will be studied by NASA over the next few years (see PAG Reports: Boch 2012; Sembach 2015; Gaudi et al. 2016). The cost for such missions is to be determined, but is certainly greater than one billion (by definition for a flagship), and may be up to ten(s) of billions of dollars for the more ambitious architectures that will be considered. The choice of the flagship mission, the technical specifications and the scientific productivity of a flagship mission may well be dependent on whether exoplanets are already known or whether they need to be discovered and it still be helpful to measure the mass of the planet after they are discovered by these direct imaging missions. EPRV is likely to be the only techniques that offers promise for detecting low mass planetary systems around nearby stars before the launch of the next flagship mission. An investment at the $100 M level would be a wise investment that could ultimately save billions of dollars toward the goal of characterizing small rocky planets.

We thank the anonymous referee for providing an exceptionally careful review of this paper and for many suggested changes which improved the quality of the paper. We are grateful to Yale University, to the National Science Foundation grant AST-1458915 and to the NASA Exoplanet Science Institute (NExScI) for providing financial support that enabled this workshop and provided support for participants. D.A.F. thanks NASA NNX12ACG01 C for inspiring the study of extreme precision RV measurements. R.I.D. acknowledges the Miller Institute at the Univ of California, Berkeley for Basic Research in Science. S.A.D. acknowledges support from NIST and the NSF grant AST-1310875. X.D. would like to thank the Society in Science for its support through a Branco Weiss Fellowship. E.B.F. was supported in part by NASA Exoplanet Research Program award NNX15AE21G. Work by B.S.G. was partially supported by NSF CAREER Grant AST-1056524. G.L. acknowledges support from the NASA Astrobiology Institute through a cooperative agreement between NASA Ames Research Center and the University of California at Santa Cruz, and from the NASA *TESS* Mission through a cooperative agreement between M.I.T. and UCSC. A.R. acknowledges support from the European Research Council under the FP7 Starting Grant agreement number 279347 and from DFG grant RE 1664/9-1. N.C.S. and P.F. acknowledge support by Fundação para a Ciência e a Tecnologia (FCT) through the research grants UID/FIS/04434/2013 and PTDC/FIS-AST/1526/2014 as well as through Investigador FCT contracts of reference IF/01037/2013 and IF/00169/2012, and POPH/FSE (EC) by FEDER funding through the program "Programa Operacional de Factores de Competitividade—COMPETE." P.F. further acknowledges support from FCT in the form of an exploratory project reference IF/01037/2013CP1191/CT0001. A.S. thanks the Giant Magellan Telescope Organization for their support for much the work described in his contribution to this paper under Contract No. GMT-INS-CON-00584. S.X.W. is supported by a NASA Earth and Space Science Fellowship (NNX14AN81H). S.X.W. and J.T.W. acknowledge support from NSF grant AST-1211441 for the work on telluric contamination. J.T.W. acknowledges NSF grant AST1109727 and NASA grant NNX12AC01G for work on barycentric corrections. The Center for Exoplanets and Habitable Worlds is supported by the Pennsylvania State University, the Eberly College of Science, and the Pennsylvania Space Grant Consortium. The PARAS program is fully supported and funded by the Department of Space, Govt. of India and Physical Research Laboratory (PRL), Ahmedabed, India. The PARAS team would like to acknowledge the support of the Director, PRL, and the Mt. Abu Observatory staff for running the program and Francesco Pepe (Geneva Observatory) and Larry Ramsey (Pen State University) for many scientific and technical inputs. Two of the PARAS team members, Suvrath Mahadevan and Arpita Roy would like to thank the Centre of Exoplanets and Habitable Worlds at Penn State University for their partial support. The McDonald Observatory planet search is currently supported by the National Science Foundation under Astrophysics grant AST-1313075, and has been supported in the past by various NASA grants.

References

Aigrain, S., Pont, F., & Zucker, S. 2012, MNRAS, 419, 3147
Anglada-Escudé, G., Plavchan, P., Mills, S., et al. 2012, PASP, 124, 586
Artigau, É, Astudillo-Defru, N., Delfosse, X., et al. 2014a, Proc. SPIE, 9149, 914905
Artigau, É, Kouach, D., Donati, J.-F., et al. 2014b, Proc. SPIE, 9147, 914715
Avila, G. 1998, in ASP Conf. Ser. 152, Fiber Optics in Astronomy III, ed. S. Arribas, E. Mediavilla, & F. Watson (San Francisco, CA: ASP), 44
Avila, G. 2012, Proc. SPIE, 8446, 84469L
Avila, G., & Singh, P. 2008, Proc. SPIE, 7018, 70184W
Bailey, J., Butler, R. P., Tinney, C. G., et al. 2009, ApJ, 690, 743
Baluev, R. V. 2013, MNRAS, 429, 2052
Baluev, R. V. 2015, MNRAS, 446, 1493
Baranne, A., Queloz, D., Mayor, M., et al. 1996, A&AS, 119, 373
Bartels, A., Heinecke, D., & Diddams, S. A. 2009, Sci, 326, 681
Baudrand, J., & Walker, G. A. H. 2001, PASP, 113, 851
Bauer, F. F., Zechmeister, M., & Reiners, A. 2015, A&A, 581, A117
Bean, J. L., Seifahrt, A., Hartman, H., et al. 2010, ApJ, 713, 410

Beatty, T. G., & Gaudi, B. S. 2015, PASP, 127, 1240
Bechter, A., Crass, J., Ketterer, R., et al. 2015, Proc. SPIE, 9605, 96051U
Beha, K., Cole, D. C., Del'Haye, P., et al. 2015, arXiv:1507.06344
Bender, C. F., Mahadevan, S., Deshpande, R., et al. 2012, ApJL, 751, L31
Bertaux, J. L., Lallement, R., Ferron, S., Boonne, C., & Bodichon, R. 2014, A&A, 564, A46
Blake, C. H., Charbonneau, D., & White, R. J. 2010, ApJ, 723, 684
Blake, C. H., & Shaw, M. M. 2011, PASP, 123, 1302
Blanford, R. e. a. 2010, New Worlds, New Horizons in Astronomy and Astrophysics (Washington, DC: Natl. Acad. Press)
Boch, J. E. A. 2012, Physics of the Cosmos Program Analysis Group, http://pcos.gsfc.nasa.gov/technology/2012_PCOS_PATR_Final_101612.pdf
Boisse, I., Bonfils, X., & Santos, N. C. 2012, A&A, 545, A109
Bonfils, X., Delfosse, X., Udry, S., et al. 2013, A&A, 549, A109
Borgniet, S., Meunier, N., & Lagrange, A.-M. 2015, A&A, 581, A133
Bouchy, F., Díaz, R. F., Hébrard, G., et al. 2013, A&A, 549, A49
Bouchy, F., Isambert, J., Lovis, C., et al. 2009, in EAS Publ. Ser. 37, Astrophysics Detector Workshop 2008, ed. P. Kern (Les Ulis: EDP Sciences), 247
Bouchy, F., Pepe, F., & Queloz, D. 2001, A&A, 374, 733
Bourrier, V., & Hébrard, G. 2014, A&A, 569, A65
Braje, D. A., Kirchner, M. S., Osterman, S., Fortier, T., & Diddams, S. A. 2008, EPJD, 48, 57
Brasch, V., Chen, Q.-F., Schiller, S., & Kippenberg, T. J. 2014, OExpr, 22, 30786
Brewer, J. M., Giguere, M., & Fischer, D. A. 2014, PASP, 126, 48
Buchhave, L. A., Bizzarro, M., Latham, D. W., et al. 2014, Natur, 509, 593
Burt, J., Holden, B., Hanson, R., et al. 2015, JATIS, 1, 044003
Butler, R. P., & Marcy, G. W. 1997, in IAU Coll. 161: Astronomical and Biochemical Origins and the Search for Life in the Universe, ed. C. Batalli Cosmovici, S. Bowyer, & D. Werthimer (Bologna, Italy), 331
Butler, R. P., Marcy, G. W., Williams, E., et al. 1996, PASP, 108, 500
Butler, R. P., Tinney, C. G., Marcy, G. W., et al. 2001, ApJ, 555, 410
Carrasco, E., & Parry, I. R. 1994, MNRAS, 271, 1
Carter, B. D., Butler, R. P., Tinney, C. G., et al. 2003, ApJL, 593, L43
Cavallini, F., Ceppatelli, G., & Righini, A. 1985, A&A, 143, 116
Cegla, H. M., Watson, C. A., Shelyag, S., & Mathioudakis, M. 2015, Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, 18, ed. G. T. van Belle, & H. C. Harris, 567
Chakraborty, A., Mahadevan, S., Roy, A., et al. 2010, Proc. SPIE, 7735, 77354N
Chakraborty, A., Mahadevan, S., Roy, A., et al. 2014, PASP, 126, 133
Chang, G., Li, C.-H., Phillips, D. F., Walsworth, R. L., & Kärtner, F. X. 2010, OExpr, 18, 12736
Chapman, G. A., Cookson, A. M., Dobias, J. J., & Walton, S. R. 2001, ApJ, 555, 462
Charbonneau, D., Brown, T. M., Latham, D. W., & Mayor, M. 2000, ApJL, 529, L45
Chazelas, B., Pepe, F., Wildi, F., et al. 2010, Proc. SPIE, 7739, 773947
Clough, S. A., Shephard, M. W., Mlawer, E. J., et al. 2005, JQSRT, 91, 233
Connes, P. 1985, Ap&SS, 110, 211
Cosentino, R., Lovis, C., Pepe, F., et al. 2014, Proc. SPIE, 9147, 91478C
Covino, E., Esposito, M., Barbieri, M., et al. 2013, A&A, 554, A28
Crane, J. D., Shectman, S. A., & Butler, R. P. 2006, Proc. SPIE, 6269, 626931
Crepp, J. R. 2014, Sci, 346, 809
Crockett, C. J., Mahmud, N. I., Prato, L., et al. 2012, ApJ, 761, 164
Croll, B., Walker, G. A. H., Kuschnig, R., et al. 2006, ApJ, 648, 607
Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, PASP, 120, 531
Cumming, A., Marcy, G. W., & Butler, R. P. 1999, ApJ, 526, 890
Cunha, D., Santos, N. C., Figueira, P., et al. 2014, A&A, 568, A35
Dawson, R. I., & Fabrycky, D. C. 2010, ApJ, 722, 937
Del Moro, D. 2004, A&A, 428, 1007
Del'Haye, P., Arcizet, O., Schliesser, A., Holzwarth, R., & Kippenberg, T. J. 2008, PhRvL, 101, 053903
Del'Haye, P., Coillet, A., Fortier, T., et al. 2015, arXiv:1511.08103
Del'Haye, P., Schliesser, A., Arcizet, O., et al. 2007, Natur, 450, 1214
Desort, M., Lagrange, A.-M., Galland, F., Udry, S., & Mayor, M. 2007, A&A, 473, 983
Díaz, R. F., Ségransan, D., Udry, S., et al. 2016, A&A, 585, A134
Dravins, D. 1982, ARA&A, 20, 61
Dumusque, X., Boisse, I., & Santos, N. C. 2014, ApJ, 796, 132
Dumusque, X., Glenday, A., Phillips, D. F., et al. 2015a, ApJL, 814, L21
Dumusque, X., Pepe, F., Lovis, C., & Latham, D. W. 2015b, ApJ, 808, 171
Dumusque, X., Pepe, F., Lovis, C., et al. 2012, Natur, 491, 207
Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. J. P. F. G. 2011, A&A, 525, A140
Eastman, J., Siverd, R., & Gaudi, B. S. 2010, PASP, 122, 935
Eastman, J. D., Brown, T. M., Hygelund, J., et al. 2014, Proc. SPIE, 9147, 914716
Edwards, C. S., Margolis, H. S., Barwood, G. P., et al. 2004, OptL, 29, 566
Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, MNRAS, 372, 1549
Eisenstein, D. e. a. 2012, Advancing Astronomy in the Coming Decade: Opportunities and Challenges, http://www.nsf.gov/mps/ast/portfolioreview/reports/ast_portfolio_review_report.pdf
Endl, M., Kürster, M., & Els, S. 2000, A&A, 362, 585
Faria, J. P., Haywood, R. D., Brewer, B. J., et al. 2016, A&A, 588, A31
Feger, T., Brucalassi, A., Grupp, F. U., et al. 2012, Proc. SPIE, 8446, 844692
Figueira, P., Adibekyan, V. Z., Oshagh, M., et al. 2016, A&A, 586, A101
Figueira, P., Pepe, F., Lovis, C., & Mayor, M. 2010a, A&A, 515, A106
Figueira, P., Pepe, F., Melo, C. H. F., et al. 2010b, A&A, 511, A55
Figueira, P., Santos, N. C., Pepe, F., Lovis, C., & Nardetto, N. 2013, A&A, 557, A93
Fischer, D. A., Laughlin, G., Butler, P., et al. 2005, ApJ, 620, 481
Fischer, D. A., Marcy, G. W., Butler, R. P., Vogt, S. S., & Apps, K. 1999, PASP, 111, 50
Fischer, D. A., Marcy, G. W., & Spronck, J. F. P. 2014, ApJS, 210, 5
Fischer, D. A., & Valenti, J. 2005, ApJ, 622, 1102
Fortier, A., Beck, T., Benz, W., et al. 2014, Proc. SPIE, 9143, 91432J
Fressin, F., Torres, G., Charbonneau, D., et al. 2013, ApJ, 766, 81
Fulton, B. J., Weiss, L. M., Sinukoff, E., et al. 2015, ApJ, 805, 175
Gaudi, B. S., Agol, E., Apai, D., et al. 2016, arXiv:1601.00008
Gelman, A., Carlin, J. B., Stern, H. S., et al. 2013, Bayesian Data Analysis (3rd ed.; Boca Raton, FL: CRC Press)
Giguere, M. J., Fischer, D. A., Zhang, C. X. Y., et al. 2016, ApJ, submitted
Gomes da Silva, J., Santos, N. C., Bonfils, X., et al. 2012, A&A, 541, A9
Gray, D. F. 2009, ApJ, 697, 1032
Gregory, P. C. 2016, MNRAS, 458, 2604
Grundahl, F., Christensen-Dalsgaard, J., Gråe Jørgensen, U., et al. 2011, JPhCS, 271, 012083
Gullikson, K., Dodson-Robinson, S., & Kraus, A. 2014, AJ, 148, 53
Halverson, S., Mahadevan, S., Ramsey, L., et al. 2014, PASP, 126, 445
Haywood, R. D., Collier Cameron, A., Unruh, Y. C., et al. 2016, MNRAS, 457, 3637
Hatzes, A. P. 2013, ApJ, 770, 133
Hatzes, A. P. 2016, A&A, 585, A144
Heacox, W. D. 1986, AJ, 92, 219
Heavner, T. P., Donley, E. A., Levi, F., et al. 2014, Metro, 51, 174
Henry, G. W., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2000, ApJL, 529, L41
Herrero, E., Ribas, I., Jordi, C., et al. 2015, in Highlights of Spanish Astrophysics VIII, ed. A. J. Cenarro et al., 494
Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655
Howard, A. W., Johnson, J. A., Marcy, G. W., et al. 2009, ApJ, 696, 75
Howard, A. W., Johnson, J. A., Marcy, G. W., et al. 2010, ApJ, 721, 1467
Howard, A. W., Johnson, J. A., Marcy, G. W., et al. 2011, ApJ, 726, 73
Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, ApJS, 201, 15
Howard, A. W., Sanchis-Ojeda, R., Marcy, G. W., et al. 2013, Natur, 503, 381
Huélamo, N., Figueira, P., Bonfils, X., et al. 2008, A&A, 489, L9
Hunter, T. R., & Ramsey, L. W. 1992, PASP, 104, 1244
Ioannidis, J. 2005, PLoS Med, 2, 8
Isaacson, H., & Fischer, D. 2010, ApJ, 725, 875
Ishizawa, A., Nishikawa, T., Mizutori, A., et al. 2011, OExpr, 19, 22402
Jenkins, J. S., & Tuomi, M. 2014, ApJ, 794, 110
Jenkins, J. S., Tuomi, M., Brasser, R., Ivanyuk, O., & Murgas, F. 2013, ApJ, 771, 41
Johnson, J. A., Fischer, D. A., Marcy, G. W., et al. 2007, ApJ, 665, 785
Johnson, J. A., Winn, J. N., Cabrera, N. E., & Carter, J. A. 2009, ApJL, 692, L100
Jost, J. D., Lucas, E., Herr, T., et al. 2015, arXiv:1508.01398
Jovanovic, N., Guyon, O., Martinache, F., Schwab, C., & Cvetojevic, N. 2014, Proc. SPIE, 9147, 91477P

Jung, H., Stoll, R., Guo, X., Fischer, D., & Tang, H. X. 2014, arXiv:1410.5018
Kippenberg, T. J., Holzwarth, R., & Diddams, S. A. 2011, Sci, 332, 555
Kjeldsen, H., & Bedding, T. R. 1995, A&A, 293, 87
Lagrange, A.-M., Desort, M., & Meunier, N. 2010, A&A, 512, A38
Langellier, N., Li, C. H., Glenday, A. G., et al. 2014, Proc. SPIE, 9147, 91478N
Lemke, U., Corbett, J., Allington-Smith, J., & Murray, G. 2011, MNRAS, 417, 689
Li, C.-H., Benedick, A. J., Fendel, P., et al. 2008, Natur, 452, 610
Li, J., Lee, H., Chen, T., & Vahala, K. J. 2012, PhRvL, 109, 233901
Lombardi, M. 2008, Measure, 3, 56
Lovis, C., & Fischer, D. 2010, in Radial Velocity Techniques for Exoplanets, ed. S. Seager (Tucson, AZ: Univ. Arizona Press), 27
Lovis, C., Ségransan, D., Mayor, M., et al. 2011, A&A, 528, A112
Lunine, J., Fischer, D., Hammel, H., et al. 2008, AsBio, 8, 875
Ma, L.-S., Bi, Z., Bartels, A., et al. 2004, Sci, 303, 1843
Mahadevan, S., Ramsey, L., Bender, C., et al. 2012, Proc. SPIE, 8446, 84461S
Mahmud, N. I., Crockett, C. J., Johns-Krull, C. M., et al. 2011, ApJ, 736, 123
Marchwinski, R. C., Mahadevan, S., Robertson, P., Ramsey, L., & Harder, J. 2015, ApJ, 798, 63
Marcy, G. W., & Butler, R. P. 1992, PASP, 104, 270
Marcy, G. W., Butler, R. P., Vogt, S. S., et al. 2005, ApJ, 619, 570
Marcy, G. W., Isaacson, H., Howard, A. W., et al. 2014a, ApJS, 210, 20
Marcy, G. W., Weiss, L. M., Petigura, E. A., et al. 2014b, PNAS, 111, 12655
Mayor, M., Bonfils, X., Forveille, T., et al. 2009, A&A, 507, 487
Mayor, M., Lovis, C., & Santos, N. C. 2014, Natur, 513, 328
Mayor, M., Pepe, F., Queloz, D., et al. 2003, Msngr, 114, 20
Mayor, M., & Queloz, D. 1995, Natur, 378, 355
Mayor, M., & Udry, S. 2008, PhST, 130, 014010
McCoy, K. S., Ramsey, L., Mahadevan, S., Halverson, S., & Redman, S. L. 2012, Proc. SPIE, 8446, 84468J
McCracken, T. M., Jurgenson, C. A., Fischer, D. A., et al. 2014, Proc. SPIE, 9147, 91473L
Meunier, N., Desort, M., & Lagrange, A.-M. 2010, A&A, 512, A39
Meunier, N., & Lagrange, A.-M. 2013, A&A, 551, A101
Murphy, M. T., Udem, T., Holzwarth, R., et al. 2007, MNRAS, 380, 839
O'Toole, S., Tinney, C. G., Butler, R. P., et al. 2009, ApJ, 697, 1263
O'Toole, S. J., Tinney, C. G., & Jones, H. R. A. 2008, MNRAS, 386, 516
Osterman, S., Diddams, S., Beasley, M., et al. 2007, Proc. SPIE, 6693, 66931G
Pepe, F., Cristiani, S., Rebolo, R., et al. 2013, Msngr, 153, 6
Pepe, F., Lovis, C., Ségransan, D., et al. 2011, A&A, 534, A58
Pepe, F., Mayor, M., Rupprecht, G., et al. 2002, Msngr, 110, 9
Perruchot, S., Kohler, D., Bouchy, F., et al. 2008, Proc. SPIE, 7014, 70140J
Piskunov, N. E., & Valenti, J. A. 2002, A&A, 385, 1095
Plavchan, P., Latham, D., Gaudi, S., et al. 2015, arXiv:1503.01770
Plavchan, P. P., Anglada-Escude, G., White, R., et al. 2013a, Proc. SPIE, 8864, 88641J
Plavchan, P. P., Bottom, M., Gao, P., et al. 2013b, Proc. SPIE, 8864, 88640G
Plazas, A. A., Bernstein, G. M., & Sheldon, E. S. 2014, PASP, 126, 750
Podgorski, W., Bean, J., Bergner, H., et al. 2014, Proc. SPIE, 9147, 91478W
Prato, L., Huerta, M., Johns-Krull, C. M., et al. 2008, ApJL, 687, L103
Probst, R. A., Lo Curto, G., Avila, G., et al. 2014, Proc. SPIE, 9147, 91471C
Queloz, D., Henry, G. W., Sivan, J. P., et al. 2001, A&A, 379, 279
Quinlan, F., Ycas, G., Osterman, S., & Diddams, S. A. 2010, RScI, 81, 063105
Quirrenbach, A., Amado, P. J., Caballero, J. A., et al. 2014, Proc. SPIE, 9147, 91471F
Radovan, M. V., Cabak, G. F., Laiterman, L. H., Lockwood, C. T., & Vogt, S. S. 2010, Proc. SPIE, 7735, 77354K
Rajpaul, V., Aigrain, S., & Roberts, S. 2016, MNRAS, 456, L6
Ramsey, L. W. 1988, in ASP Conf. Ser. 3, Fiber Optics in Astronomy, ed. S. C. Barden (San Francisco, CA: ASP), 26
Rauer, H., Catala, C., Aerts, C., et al. 2014, ExA, 38, 249
Reiners, A., Banyal, R. K., & Ulbrich, R. G. 2014, A&A, 569, A77
Reiners, A., Shulyak, D., Anglada-Escudé, G., et al. 2013, A&A, 552, A103
Ricker, G. R., Winn, J. N., Vanderspek, R., et al. 2014, Proc. SPIE, 9143, 914320
Robertson, P., Mahadevan, S., Endl, M., & Roy, A. 2014, Sci, 345, 440
Rothman, L. S., Gordon, I. E., Babikov, Y., et al. 2013, JQSRT, 130, 4
Saar, S. H., Butler, R. P., & Marcy, G. W. 1998, ApJL, 498, L153
Saar, S. H., & Donahue, R. A. 1997, ApJ, 485, 319
Sanchis-Ojeda, R., Winn, J. N., Dai, F., et al. 2015, ApJL, 812, L11
Santerne, A., Díaz, R. F., Almenara, J.-M., et al. 2015, MNRAS, 451, 2337
Santos, N. C., Gomes da Silva, J., Lovis, C., & Melo, C. 2010, A&A, 511, A54
Santos, N. C., Mortier, A., Faria, J. P., et al. 2014, A&A, 566, A35
Schlieder, J. E., Lépine, S., Rice, E., et al. 2012, AJ, 143, 114
Schwab, C., Stürmer, J., Gurevich, Y. V., et al. 2015, PASP, 127, 880
Seifahrt, A., Käufl, H. U., Zängl, G., et al. 2010, Msngr, 142, 21
Sembach, K. 2015, BAAS, 225, 308.02
Shao, M., Nemati, B., Zhai, C., Carey, S., & Leger, A. 2013, White Paper, http://keplerscience.arc.nasa.gov/K2/docs/WhitePapers/Shao_etal_Kepler_2Wheel.pdf
Sizer, T., II 1989, IJQE, 25, 97
Smette, A., Sana, H., Noll, S., et al. 2015, A&A, 576, A77
Spronck, J. F. P., Fischer, D. A., Kaplan, Z., et al. 2015, PASP, 127, 1027
Spronck, J. F. P., Fischer, D. A., & Kaplan, Z. A. 2012a, in Recent Progress in Optical Fiber Research, ed. M. Yasin, S. W. Harun, & H. Arof (Rijeka: InTech), 353
Spronck, J. F. P., Fischer, D. A., Kaplan, Z. A., Schwab, C., & Szymkowiak, A. 2013, PASP, 125, 511
Spronck, J. F. P., Kaplan, Z. A., Fischer, D. A., Schwab, C., & Szymkowiak, A. E. 2012b, Proc. SPIE, 8446, 84468T
Spronck, J. F. P., Schwab, C., & Fischer, D. A. 2010, Proc. SPIE, 7735, 77350W
Steinmetz, T., Wilken, T., Araujo-Hauck, C., et al. 2008, Sci, 321, 1335
Stürmer, J., Stahl, O., Schwab, C., et al. 2014, Proc. SPIE, 9151, 915152
Swift, J. J., Bottom, M., Johnson, J. A., et al. 2015, JATIS, 1, 027002
Szentgyorgyi, A., Frebel, A., Furesz, G., et al. 2012, Proc. SPIE, 8446, 84461H
Tamura, M., Suto, H., Nishikawa, J., et al. 2012, Proc. SPIE, 8446, 84461T
Tinney, C. G., Wittenmyer, R. A., Butler, R. P., et al. 2011, ApJ, 732, 31
Tokovinin, A., Fischer, D. A., Bonati, M., et al. 2013, PASP, 125, 1336
Trifonov, T., Reffert, S., Zechmeister, M., Reiners, A., & Quirrenbach, A. 2015, A&A, 582, A54
Tull, R. G. 1998, Proc. SPIE, 3355, 387
Tull, R. G., MacQueen, P. J., Sneden, C., & Lambert, D. L. 1995, PASP, 107, 251
Tuomi, M., & Anglada-Escudé, G. 2013, A&A, 556, A111
Tuomi, M., Jones, H. R. A., Barnes, J. R., Anglada-Escudé, G., & Jenkins, J. S. 2014, MNRAS, 441, 1545
Valenti, J. A., & Fischer, D. A. 2005, ApJS, 159, 141
Vögler, A., Shelyag, S., Schüssler, M., et al. 2005, A&A, 429, 335
Vogt, S. S. 1987, PASP, 99, 1214
Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, Proc. SPIE, 2198, 362
Vogt, S. S., Butler, R. P., Rivera, E. J., et al. 2010a, ApJ, 723, 954
Vogt, S. S., Radovan, M., Kibrick, R., et al. 2014, PASP, 126, 359
Vogt, S. S., Wittenmyer, R. A., Butler, R. P., et al. 2010b, ApJ, 708, 1366
Weatherill, D. P., Stefanov, K. D., Greig, T. A., & Holland, A. D. 2014, Proc. SPIE, 9154, 91541A
Weiss, L. M., & Marcy, G. W. 2014, ApJL, 783, L6
Wildi, F., Chazelas, B., & Pepe, F. 2012, Proc. SPIE, 8446, 84468E
Winn, J. N., Albrecht, S., Johnson, J. A., et al. 2011a, ApJL, 741, L1
Winn, J. N., Howard, A. W., Johnson, J. A., et al. 2011b, AJ, 141, 63
Winn, J. N., Johnson, J. A., Howard, A. W., et al. 2010, ApJ, 718, 575
Wittenmyer, R. A., Horner, J., Tinney, C. G., et al. 2014, ApJ, 783, 103
Wolszczan, A., & Frail, D. A. 1992, Natur, 355, 145
Wood, D., Hall, D. J., Murray, N. J., et al. 2014, JInst, 9, C2028
Wright, J. T., & Eastman, J. D. 2014, PASP, 126, 838
Ycas, G. G., Quinlan, F., Diddams, S. A., et al. 2012, OExpr, 20, 6631
Yi, X., Vahala, K., Diddams, S., et al. 2015, arXiv:1501.02509