nature reviews physics

https://doi.org/10.1038/s42254-025-00839-z

Technical review

Check for updates

Data challenges and prospects of high-resolution spectroscopy of exoplanets

Sergei N. Yurchenko [1] ✉, Jonathan Tennyson [1] & Matteo Brogi [2,3]

Abstract

Understanding the atmospheres of exoplanets is crucial for unravelling their formation, evolution and potential habitability. High-resolution cross-correlation spectroscopy (HRCCS) has emerged as a powerful tool for probing exoplanetary atmospheres, enabling the detection of molecular species and the characterization of atmospheric dynamics. However, the reliability of these detections depends critically on the accuracy of laboratory spectroscopic data, particularly precise line positions and the careful statistical treatment of observational data. This Technical Review explores the interplay between laboratory data and high-resolution exoplanet spectroscopy, emphasizing the growing shift from isolated molecular detections to comprehensive whole-atmosphere characterization. We discuss the specific challenges of producing high-quality laboratory data and outline the needs of the exoplanetary community in this context. Key topics include the reliability of HRCCS detections, typical jargon of HRCCS and the ethical considerations in data attribution. By bridging the perspectives of laboratory spectroscopy, quantum chemistry and observational astronomy, we provide recommendations for advancing the field towards a more robust and self-consistent framework for exoplanetary atmospheric studies.

Sections

[1]Department of Physics and Astronomy, University College London, London, UK. [2]Dipartimento di Fisica, Università degli Studi di Torino, Turin, Italy. [3]Osservatorio Astrofisico di Torino, INAF, Pino Torinese, Italy.
✉e-mail: s.yurchenko@ucl.ac.uk

Key points

- High-resolution cross-correlation spectroscopy (HRCCS) is a powerful tool. It has become a leading method for characterizing exoplanet atmospheres from ground-based observatories.
- Laboratory data are crucial. The accuracy and completeness of molecular line lists and laboratory data directly affect the reliability of HRCCS detections.
- Interdisciplinary collaboration is needed. Stronger communication between astrophysicists, spectroscopists and database providers is essential to ensure that laboratory data meet the needs of exoplanet spectroscopy.
- False positives and non-detections remain a challenge. HRCCS detections can be affected by incomplete or inaccurate spectroscopic data, necessitating robust statistical methods and improved molecular databases.
- Future telescopes will push the field forward. Next-generation ground-based facilities such as the Extremely Large Telescope, the Giant Magellan Telescope and the Thirty Meter Telescope will expand HRCCS applications, requiring even more precise spectroscopic data to maximize their scientific impact.

Introduction

Nearly 6,000 exoplanets have been discovered in the past three decades (see, for example, exoplanetarchive.ipac.caltech.edu). Spectroscopy of their atmospheres across the widest possible spectral range (0.2–24 μm currently) is thought to be the main tool to address fundamental questions about planetary composition, formation, evolution and, ultimately, habitability. Among the various observational techniques, high-resolution cross-correlation spectroscopy (HRCCS) from the ground has recently emerged as an efficient and robust tool to characterize exoplanet atmospheres and to complement space observatories. High-resolution spectroscopy (HRS) promises to greatly expand the frontier of exoplanet science in the next decade, when it will be applied to observations obtained with the next generation of ground-based facilities, such as the European Southern Observatory Extremely Large Telescope, the Giant Magellan Telescope and the Thirty Meter Telescope.

Key to the interpretation of exoplanet spectroscopy, and in particular to HRCCS, is the theoretical knowledge of the spectra of target atomic and molecular species. This Technical Review considers the interplay between the provision of laboratory data, experimental or theoretical, and its application to HRCCS. This is a timely subject, as potential discoveries from HRS can be hindered or biased in cases of insufficient (incomplete and/or inaccurate) knowledge of theoretical spectra. Indeed, the unusual conditions (from the point of view of the Solar System) found on most of the known exoplanets, involving elevated temperatures and high fluxes of stellar radiation, mean that the required data are often not readily measurable in the laboratory or not measured accurately enough for HRS.

This Review mainly targets two communities: spectroscopic (data providers) and astrophysical (data users). The data providers need to appreciate the data needs specific for exoplanet spectroscopy, such as wavelength windows, temperature and pressure ranges, accuracy requirements, and how the data are going to be used. In turn, the astrophysical community (data users) needs to understand the limitations and capabilities of the high-resolution laboratory data, both experimental and theoretical, as well as of the underlying techniques. Additionally, it is important that database builders learn both sides of the story to provide efficient and easily accessible repositories for the high-resolution data.

We outline the main observational techniques and the specifics of HRCCS. We then continue with a review of the sources of laboratory and theoretical data, and we recap the state of molecular detections so far. We conclude with some recommendations for scientists working on these problems.

Observational techniques

Thus far, atmospheric characterization has primarily been achieved through spectroscopy of transiting exoplanets. These are planets that, when viewed from Earth, pass in front of their parent star (primary transit) and are later occulted by it (secondary eclipse). Although the star and planet are never spatially resolved, the changing orbital configuration produces measurable variation in the flux from the system. The fraction of starlight blocked during transit, measured as a function of wavelength, constitutes the transmission spectrum. The exoplanet light blocked by the stellar occultation, also measured as a function of wavelength, is called the dayside spectrum or sometimes the secondary eclipse. Depending on wavelength, it encodes both reflected starlight and the planet's own thermal emission. These two techniques are complementary as they probe different parts of the planetary atmosphere, with the transit penetrating into the atmosphere whereas the reflection occurs from its upper layers.

Spectroscopy of transiting exoplanets is currently the major technique used by the James Web Space Telescope (JWST) and the Hubble Space Telescope (HST) as well as some ground-based telescopes, with results already hinting at important atmospheric processes such as condensation[1] as well as the onset of thermal inversion layers[2]. Although the method indirectly reconstructs the spectrum by measuring the depth of transits and eclipses in narrowband photometry, the data products can be treated as conventional spectra at resolving powers $20 < R < 3{,}000$ (where $R = \lambda/\Delta\lambda$ and λ is the spectroscopic wavelength), which in the context of this Review we define as low resolution. Even at relatively low R, transmission and dayside spectroscopy encodes the full physics of an exoplanetary atmosphere, such as abundances, temperature and so forth. Transit spectroscopy is most efficient with space observations, where the photometric stability is higher. Transit spectral features are usually more pronounced for hot, gaseous giant exoplanets than for cooler, rocky exoplanets[3], where the contrast between planet and star is higher and the atmospheres are more expansive. We note that statistically only 1–2% of all planets should transit their star when viewed from Earth, although, mainly owing to the highly successful Kepler and TESS (Transiting Exoplanet Survey Satellite) missions, a very much higher proportion of the exoplanets that have thus far been detected do actually transit.

Planets on much larger orbits, when they are young enough still to glow from their internal heat following formation (typically less than about 20 Myr old), can be studied via direct imaging, which achieves the daunting task of separating the faint light from the exoplanet from that of the parent star by blocking the latter using coronographs[4]. The technique favours large angular separations between the planet and the star, thus nearby systems. In fact, there are only a few dozen directly imaged exoplanets discovered so far, and most of them linger

at the mass boundary between a planet and a brown dwarf[5]. Once spatial separation is achieved through direct imaging, more conventional spectroscopic techniques can be applied that involve directly measuring the wavelength-dependent flux of the exoplanet, either from the ground or from space. Consequently, the resolving power can in principle be very high (R = 100,000 or higher), but current instrumentation has so far been limited to low-resolution spectroscopic capabilities. This situation is expected to change in the future as new instruments and missions are planned, such as the Habitable Worlds Observatory and European Extremely Large Telescope, and the importance of direct imaging grows.

The HRCCS method is based on the cross correlation of the observed high-resolution spectrum with a pre-calculated model spectrum of the exoplanetary atmosphere. The method uses an accurate radiative transfer model to amplify the faint signature of the exoplanet. HRCCS takes advantage of the Doppler shift caused by the planet's orbital motion to distinguish exoplanetary spectroscopic features from the telluric spectra. HRCCS involves high-resolution spectra with resolving powers greater than $R \approx 25{,}000$, often in the range R = 50,000–100,000, which is achieved with ground-based telescopes, typically in the spectral range between 0.4 and 2.5 µm. A list of ground-based telescopes equipped with high-resolution instruments is shown in Table 1. For more details on these and other spectroscopic techniques used to study atmospheres of exoplanets, see previous reviews[5–8].

In contrast to the transit technique, HRCCS does not reconstruct a planet's spectrum from light curves, but rather it looks for a differential Doppler shift between the signal of an exoplanet and other unwanted contaminants. Across a few hours of observations, which correspond to a sizeable fraction of an orbit for close-in exoplanets, such a Doppler shift changes by tens to a hundred kilometres per second. Practically speaking, the data analysis proceeds by filtering out telluric and stellar lines, and then continues by cross correlating a Doppler-shifted model spectrum (with line positions coming from laboratory determinations) with the filtered observations. Cross correlation has been widely used to measure stellar radial velocity and discover exoplanets, including the first planet around a main-sequence star[9]. When measuring stellar radial velocity, an important advantage is that even though stars wobble by only tens of metres per second, their signal is at least thousands of times stronger than that of an exoplanet. Thus the data must be cleaned and prepared in different ways for stellar and planetary applications. The high-resolution Doppler dispersion technique is well established in stellar spectroscopic studies as a powerful complementary tool to the direct spectral observations[10].

Not being limited to planetary orbits seen edge-on, HRCCS can target both transiting and non-transiting planets[11,12], and even directly imaged planets[13], offering a strong complement to space-based spectroscopy. However, only relative measurements (line depth relative to the continuum and to each other) can be used to reconstruct the physical conditions of an atmosphere, owing to the loss of the absolute continuum in the process of removing telluric and stellar lines. Furthermore, unlike conventional spectroscopy, HRCCS does not produce a discernible spectrum, as the signal-to-noise ratio per line is insufficient. It produces instead a level of correlation between model and data, which has to be translated into a detection significance by means of signal-to-noise theory or statistical tests on the correlation coefficients. This step means that it is hard to assess whether a model is a good fit to the data, but it is still possible to rank which model (among a family) best fits the data[14], even by means of Bayesian retrievals[15].

Table 1 | Main characteristics of ground-based high-resolution spectrographs currently in operation

Instrument	Telescope	Site	R/1,000	Range
CARMENES	CAHA 3.5 m	Calar Alto	80	0.96–1.71 µm
			96	0.52–0.86 µm
SPIRou	CFHT	Mauna Kea	75	0.95–2.35 µm
IGRINS-2	Gemini-N	Cerro Pachon	45	1.45–2.6 µm
NIRSPEC	Keck	Mauna Kea	50	1.0–5.3 µm (*)
MAROON-X	Gemini-N	Mauna Kea	80	0.500–0.920 µm
NIRPS	ESO 3.6 m	La Silla	75/82	0.971–1.854 µm
HDS	Subaru	Mauna Kea	80/110/165	0.35–1.0 µm
IRD	Subaru	Mauna Kea	70	0.97–1.75 µm
GIANO	TNG	La Palma	50	0.95–2.45 µm
HARPS-N	TNG	La Palma	112	0.378–0.691 µm
HARPS	ESO 3.6 m	La Silla	112	0.378–0.691 µm
CRIRES+	VLT	Paranal	50/100	1.0–5.3 µm (*)
ESPRESSO	VLT	Paranal	70/140/190	0.380–0.788 µm
UVES	VLT	Paranal	80/110	0.30–1.10 µm (*)

Instruments marked with (*) cannot observe the entire spectral range simultaneously: resolving power (R) and/or spectral range covered depend on the grating or settings selected. CAHA, Calar Alto Observatory; CFHT, Canada–France–Hawaii Telescope; ESO, European Southern Observatory; TNG, Galileo National Telescope; VLT, Very Large Telescope.

Broadly speaking, HRCCS is used to claim detections (or non-detections) of individual atomic or molecular species, thus reconstructing a qualitative view of the chemical inventory. Increasingly, HRCCS is being used to derive quantitative constraints on temperature, metallicity (that is, the elemental abundance with respect to solar values) and ratio of elemental abundances such as the carbon-to-oxygen ratio (C/O), considered to be important tracers of planet formation and early evolution[16,17]. In the latter case, instead of the cross-correlation function, specific likelihood functions sensitive to line intensity and shape have been designed[15,18]. Owing to the nature of correlation and the very high resolving powers, HRCCS is particularly well suited to unambiguously identifying thermal inversions[19,20] and directly inferring planet rotation and atmospheric winds[21,22].

For low-resolution spectroscopic studies of transiting exoplanets, the presence of distinct and well-known molecular spectral features that are recognizable even before detailed analysis can significantly enhance the plausibility of retrievals. For example, the water band centred at 1.4 µm has been often reported in HST spectra[1,23,24]; owing to its distinct spectral shape and the higher-resolution capabilities of JWST, water can be clearer and less ambiguous seen in JWST spectra[25]. The recent JWST spectra of WASP-39 b[26] and WASP-107 b[27] have a statistically significant presence of CO_2 in the infrared. Even though the spectral shape can be misleading, and detections require a detailed and comprehensive retrieval and analysis, these features not only serve as immediate indicators of atmospheric composition but also help to constrain the parameter space, reducing the complexity of retrieval models.

Conversely, those using HRCCS often have to operate 'in the dark' with no or very limited clues on the molecular compositions and therefore must assume a much more extensive set of molecules. As a consequence, analysis of HRCCS has to rely completely on the quality of

Box 1 | Sources of laboratory data

The main source of infrared spectroscopic data up to 2011 was HITRAN[135] and its high-temperature offshoot HITEMP[55]. HITRAN is a collection of largely empirical, high-quality molecular line lists with the main emphasis on terrestrial applications — that is, of temperature (T) about 296 K and mainly composed of molecules important for the Earth atmosphere. In terms of accuracy, both of the line positions and intensities, it is highly suitable for high-resolution cross-correlation spectroscopy (HRCCS) applications. In terms of the temperature coverage as well as the molecular variety, it is not. HITRAN becomes increasingly incomplete for T hotter than room temperature. HITEMP provides a collection of molecular line lists for a much more limited number of molecules (currently eight), with the main emphasis on completeness, and, at least until more recently[94,136], it is not suitable for higher-resolution applications.

GEISA[137] is another HITRAN-like database with, from an exoplanet perspective, only marginal differences in terms of the species and spectral ranges provided.

The Kurucz[138] and VALD[139] databases provide data on mostly diatomic molecules. TheoReTS[140], which contains nine molecular line lists of polyatomic molecules (most notably of CH_4), and NASA Ames[141], with line lists for six polyatomic molecules, use similar methods to ExoMol; MoLLIST[132] is based on experimental measurement whereas the CaSDa (Calculated Spectroscopic Databases) database contains synthetic line lists resulting from fits of the effective Hamiltonian and dipole moment parameters[142]. CDMS[143] and JPL[144] provide molecular lists limited to microwave with the emphasis on studies of the interstellar medium. As discussed below, microwave data are currently not applicable for HRCCS applications.

ExoMol[39], established in 2011 with European Research Council funding, has become the largest provider of molecular spectroscopic data for exoplanetary, stellar and other hot atmospheric studies as well as being used in industrial and other terrestrial applications. It contains extensive line lists for almost 100 molecules (nearly 300 if one counts isotopically substituted species known as isotopologues) and more than a trillion transitions. Some of the individual line lists for polyatomic species alone contain hundreds of billions of transitions.

the model spectrum – that is, on the quality and quantity of the input laboratory data – even to detect a signature, as an incorrect spectral model would produce no (or negligible) correlation. This reliance of the HRCCS studies on laboratory data leads to tangible challenges that we describe in the following section.

Completeness versus accuracy of laboratory data

There are several mainstream sources of molecular line lists applicable for spectroscopic studies of exoplanetary atmospheres: HITRAN, HITEMP, ExoMol, GEISA, CDMS, JPL, VALD and Kurucz's database (detailed in Box 1).

The ExoMol database is the biggest source of molecular spectroscopic data for exoplanetary and other applications. Before about 2016, the main focus of ExoMol was the completeness of the molecular data, in terms of temperature and wavelength coverage, which was shown to be crucial for correctly reproducing observed spectra of astronomical objects with $T \gg 300$ K (ref. 28). Note that hot Jupiter exoplanets generally have temperatures in the range 1,000–2,000 K, and ultra-hot Jupiters (UHJs) can be as hot as stellar photospheres, with the hottest objects (such as Kelt-9 b) having a measured dayside temperature well over 4,000 K (ref. 29). Furthermore, some rocky exoplanets ('lava planets') can also be hot, with dayside temperatures exceeding 3,000 K (ref. 30).

The current state-of-the-art methodology used by ExoMol, TheoReTS and NASA Ames is based on the numerical solution of the quantum mechanical Schrödinger equation, separately, for both electronic and nuclear motion[31]. The calculations start from first principles or ab initio to obtain a 'spectroscopic model', consisting of potential energy surfaces, dipole moment surface and other properties describing the behaviour of the molecule and its interaction with light. Pure ab initio spectroscopic models are usually not adequate for any qualitative spectroscopic applications, with the typical accuracy limited to $R \approx 5{,}000$, and are therefore always refined by fitting to experimental data. A typical spectroscopic model consists of an empirically refined potential energy surface and an ab initio dipole moment surface (see, for example, ref. 32). In fact, empirical adjustment is a crucial part of state-of-the-art line list production techniques capable of achieving the accuracy of theoretical line positions of $R \approx 10{,}000$, but not much higher. As a result, not all theoretical data obtained using this technique are suitable for HRCCS applications (see, for example, refs. 33,34). Thus the quality of the molecular line lists is essentially limited by the existing laboratory data.

In turn, in HRCCS studies focused on detection of species, completeness is not the most critical factor, but accuracy is. Pure cross correlation is highly sensitive to line position, less so to their relative intensity, and insensitive to global scaling factors. As completeness mostly affects the level of the continuum, it does not play a crucial role in aiding or hindering a detection, and small sets of high-quality molecular lines, as few as 50–100 for diatomics such as CO or a few thousands for more complex species, can be sufficient for a good detection.

In response to this increasing need for high-resolution spectroscopic data of $R > 50{,}000$, ExoMol has established the new functionality ExoMolHD[35] based on the direct use of the laboratory measurements in the line list. At the core of this functionality is the measured active rotational–vibrational energy levels (MARVEL) approach[36]. The MARVEL procedure is applied to high-resolution laboratory spectra to extract high-accuracy empirical energy levels, which can then be fed back into ExoMol line lists. By construction, MARVEL energies retain the experimental accuracy ($R = 50{,}000$–$100{,}000$), thus providing accurate predicted line positions not only for those (usually strongest) transitions used to derive them. Importantly, this methodology leads to an increase in the number of accurately predicted transitions compared with the original input data, typically by a factor of about 20 (ref. 37) or almost 750 in the extreme example of VO (ref. 38). The active nature of MARVEL means that as new high-resolution laboratory measurements of spectra become available, it is straightforward to update the MARVEL results and hence the associated line list.

For polyatomic molecules, however, a standard of $R > 50{,}000$ is still not possible for the billions of lines. Even for our best example, the water line list POKAZATEL, out of 5.7 billion transition lines, only ~1.5 million are of sufficient quality. A quality indicator of the ExoMol database structure is therefore necessary, as a guide for selecting lines suitable for HRS applications. The solution suggested by ExoMolHD is to generate and provide uncertainties for every line position in the database as well as the associated functionality to select lines of the required quality[39,40]. The uncertainties of the accurate (MARVELized) line positions are essentially experimental (typically 0.1 cm^{-1} or less),

while the uncertainties of the calculated lines are estimated based on the expected accuracy of the calculations.

ExoMol, as well as other databases, aims for coverage throughout the infrared and visible range. It should be noted, however, that the current HRCCS applications do not use the mid-infrared. That is despite the ground-based instrumentation, such as CRIRES+ at the Very Large Telescope (VLT) or NIRSPEC at Keck, that can observe at wavelengths corresponding to the L, M and N bands (3.0–5.0 μm) and also despite the availability of the high-quality laboratory data in this region. This exclusion is because at these wavelengths, especially longer than 4.1–4.2 μm, the environment surrounding the telescope (and the telescope itself) begins to thermally radiate a significant photon flux; in other words, observations are essentially impossible for all but the brightest stars in the sky, at least for signal-to-noise ratios sufficient for detecting an exoplanet atmosphere. Therefore, despite the availability of very high-quality molecular data in these spectral regions (and beyond, to the mid-infrared and far infrared), sadly these are not currently exploitable in HRCCS studies of exoplanets. The good news, however, is that these limitations will be strongly reduced with the bigger aperture of the Extremely Large Telescope[41].

State-of-the-art of HRCCS detections

In this section, we mostly review detections of molecules in exoplanet atmospheres that are relevant to the provision of line lists, highlighting when the associated data quality (or its lack) was responsible for detections (or non-detections). We then briefly review the current attempts at measuring key atmospheric parameters by means of Bayesian retrievals, this practice being much more recent in the community.

At the time of writing, the vast majority of atomic and molecular species discovered in exoplanets have been claimed using HRCCS. These include CO, H_2O, TiO, HCN, CH_4, NH_3, C_2H_2, OH, VO and CrH as well as a long list of atoms and atomic ions. A notable absence in the list is CO_2, whose 4.3-μm band is instead featured regularly in recent JWST spectroscopy[26,42–44] as its opacity peaks in a region of the spectrum dominated by thermal background in ground-based observations as explained above. Although the detection of CO_2 in the near infrared presents considerable challenges due to telluric contaminations[45], a high-resolution search for CO_2 in exoplanetary atmospheres is under way[46]. Another molecule of interest, SO_2, which has recently been detected in transit by JWST in WASP-39 b[26], is expected at lower abundance and therefore more difficult to detect. Its signatures are also scarce at the wavelengths most commonly targeted by HRCCS, 1.35–2.45 μm, which is expected to change in the future by extending observations to the L (3.0–3.9 μm) and possibly M (4.0–4.9 μm) bands.

The detection significance σ in HRCCS is typically evaluated using statistical measures that quantify how strongly the observed signal matches the expected molecular template[11,12]. It is often expressed as the signal/noise (S/N) of the cross-correlation function (CCF) peak. For example, a 5σ detection is referred as $S/N = 5$ (see Supplementary Information for definitions of the detection significance).

The detection of molecules in exoplanetary atmospheres using HRCCS is illustrated in Table 2, where the first, typical or recent cases are presented. Figure 1 illustrates the corresponding spectra of these molecules in the form of cross sections at $T = 1{,}000$ K.

CO

Carbon monoxide was the first molecule to be detected with HRCCS in an exoplanetary atmosphere — on the exoplanet HD 209458 b. CO was long known to be ubiquitous in stellar atmospheres and predicted to be the dominant carbon-bearing molecule for planets at $T > 1{,}200$ K. Still, detection of CO from space has only recently been convincingly claimed[48] and is still potentially confused within the neighbouring

Table 2 | Detection of molecular species using high-resolution cross-correlation spectroscopy at the time of writing

Molecule	Object	Instrument	*S/N*	σ	λ (μm)	*R*	Ref.	Line list
CO (first)	HD 209458 b	CRIRES		5.6	2.291–2.349	100,000	11	HITRAN[54]
CO (best σ)	GQ Lupi b	CRIRES	11.6		2.302–2.331	100,000	122	HITRAN[54]
H_2O (first)	51 Peg b	CRIRES		5.6	3.1805–3.2659	100,000	52	HITRAN 2008[54]
H_2O (best)	HD 209458 b	CARMENES	6.4	8.1	0.96–1.06, 1.06–1.26	100,000	123	HITEMP 2010[55]
CH_4	HD 209458 b	GIANO		5.6	0.95–2.45	50,000	124	HITEMP 2020[94]
HCN	HD 209458 b	CRIRES		4.7	3.18–3.27	100,000	125	ExoMol (Harris)[126]
HCN	HD 189733 b	CRIRES		5	3.18–3.27	100,000	127	ExoMol (Harris)[126]
HCN	HD 209458 b	GIANO		10.1		50,000	14	ExoMol (Harris)[126]
NH_3	HD 209458 b	GIANO		5.3		50,000	14	ExoMol (CoYuTe)[128]
CO_2	HD 209458 b	GIANO		3		50,000	14	HITEMP 2010[55]
C_2H_2	HD 209458 b	GIANO		6.1		50,000	14	ExoMol (aCeTY)[129]
TiO	WASP-33 b	Subaru		4.8	6.164–7.396, 7.685–8.810	165,000	66	Schwenke (1998)[67]
TiO	WASP-189 b	HARPS		5.6	0.460–0.690	112,000	70	ExoMol (ToTo)[68]
VO	WASP-76 b	Gemini North		6	0.490–0.920	85,000	75	ExoMol (VOMYT)[76]
CrH	WASP-31 b	Gemini North		5.6	0.860–0.890	66,000	130	Burrows et al.[131]
OH	WASP-33 b	Subaru	5.4	5.5	0.97–1.75	70,000	84	MoLLIst[132]
OH	WASP-76 b	CARMENES	6.1	5.5	0.97–1.75	80,400	133	MoLLIst[132]

λ, spectroscopic wavelength; *R*, resolving power; σ, detection significance; *S/N*, signal-to-noise ratio.

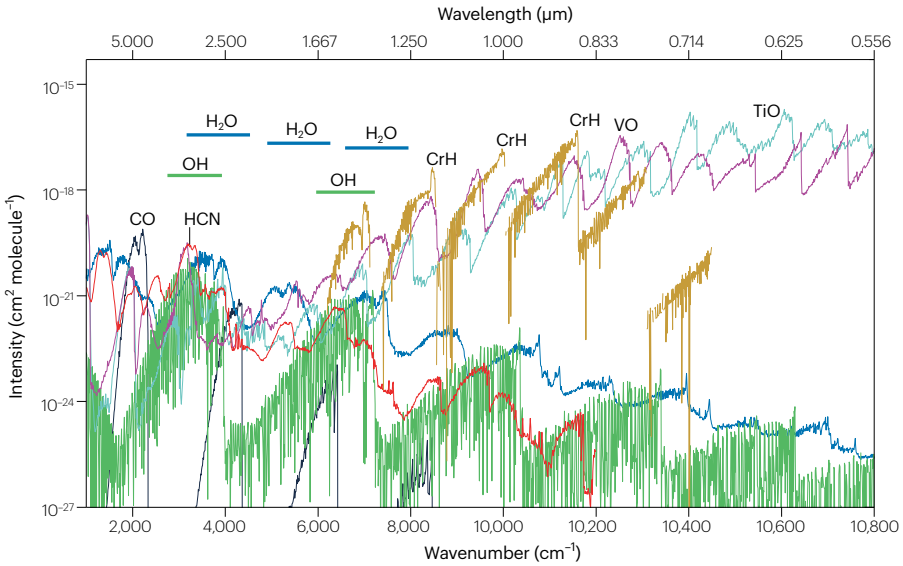

Fig. 1 | Spectra of seven molecules detected using high-resolution cross-correlation spectroscopy, based on line lists from ExoMol and HITRAN. The spectra correspond to CO (black), HCN (red), OH (green), H_2O (blue), CrH (olive green), VO (magenta) and TiO (cyan). Spectra are shown for $T = 2{,}000$ K computed with the Doppler line profile. The blue and green solid bars indicate regions where H_2O and OH have strong absorption. Data from refs. 54,56,85,126,131.

signature of CO_2 at low resolution[49]. Since then, CO has been detected in nearly every exoplanet observed with HRCCS so far, including directly imaged planets targeted via the technique. Even the first detection of an isotopically substituted molecule in an exoplanet was made using HRCCS for ^{13}CO (ref. 50). This is not surprising from the spectroscopic data point of view. It has a simple, very regular spectrum containing relatively few lines, which makes it an ideal prototype for HRCCS studies. In the first measurement, even the relatively small simultaneous coverage of VLT/CRIRES ($R \approx 100{,}000$, 0.07 μm centred at 2.3 μm), which covered 30–40 strong CO lines, was sufficient to achieve a detection at 5.6σ. The model used in the analysis was calculated with molecular data from HITEMP. In the context of CO, the line list is extensively validated and complete to high temperatures[51], and is available from both HITRAN and ExoMol.

H_2O

Despite water being predicted to be a dominant species in many exoplanet atmospheres and observed in many transit studies, the successful search for it using HRCCS took slightly longer. The first confident detection was published in 2013[52] (Fig. 2), which is the same year as the first robust detection from space with Hubble/WFC3[53]. One hindrance to its detection is that the opacity of water is relatively weak in the spectral window centred at 2.3 μm which was usually targeted when looking for CO. In fact, the detection by Birkby et al.[52] was obtained with observations around 3.2 μm. Furthermore, laboratory data of H_2O at that time were below the required quality. The existing line lists were either accurate and incomplete (HITRAN 2008[54]) or complete but not sufficiently accurate (HITEMP 2010[55]). Nevertheless, the latter HITEMP list was successfully used to detect water up to the turn of the decade, when the more accurate and complete water line list POKAZATEL[56] became available.

A comparison between the BT2 line list[57], which is used in HITEMP 2010, and the more accurate and complete POKAZATEL[56] is shown in Fig. 3a around 3.2 μm at $R \approx 70{,}000$. Although there are a number of strong and accurate features in both line lists, manifested in coinciding lines, the number of differences is still large enough to affect the cross correlation when using these data. Interestingly, a laboratory study of hot (1,723 K) H_2O emissions in the near infrared pronounced POKAZATEL to be six times more accurate than BT2[58]. Using incomplete or inaccurate line lists can affect the HRCCS detection, resulting in biased retrieved parameters[15], although in the case of water the effects seem to be marginal[59].

A more concerning difference arises when comparing the earlier line list by Partridge and Schwenke[60] to the two above, where little to no match is observed at very high R across wide spectral ranges. In fact, it has been shown that in this case detection of water can be hindered or otherwise made impossible via HRCCS[47,61]. The Partridge and Schwenke[60] line list is based on high-quality theoretical calculations, but the line positions are all calculated, rather than empirically obtained by using MARVEL energy levels as in the more recent line lists. In fact, many molecular line lists are still based on calculated levels, meaning that a subsequent MARVELization procedure is pivotal for line lists aimed for HRCCS applications.

TiO and VO

Titanium oxide and vanadium oxide have long been considered key molecules for exoplanetary studies owing to their strong optical and infrared opacity, expected high abundance in hot Jupiters and stratospheric temperature inversions[62,63].

TiO is commonly found in the atmospheres of M-type stars and brown dwarfs[64]; therefore, the search for this molecule in exoplanet atmospheres started relatively early. The first claimed detections, through low-resolution transmission spectroscopy in WASP-33 b[65] and HRCCS[66], both came in 2017. An earlier non-detection via HRCCS of HD 209458 b[33] was attributed to the insufficient quality of the line list used[67], which correlated poorly even with M-dwarf spectra. The 2017 detection (4.8σ) was obtained by dayside optical spectroscopy of WASP-33 b with the Subaru telescope (0.62–0.74 μm, 0.77–0.88 μm).

A new line list for TiO (ToTo) was generated by ExoMol[68] in response to the pressing data needs from exoplanetary studies. Interestingly, with this new line list a reanalysis of the original 2017 data yielded a weaker signal[69], which in fact led the authors to doubt their original claim. Subsequently, ToTo was used to detect TiO in WASP-189 b (5.6σ)

in HARPS observations (0.46–0.69 μm)[70]. The authors removed some spectroscopic sub-windows (0.507–0.522 μm, 0.569–0.581 μm, 0.591–0.615 μm and 0.621–0.628 μm) where they suspected the line list to be less accurate. In Fig. 3b, we use the ExoMolHD functionality to show the accuracy of the ToTo line list for TiO (ref. 68) by plotting cross sections from the accurate lines only (with ExoMolHD uncertainties better than 0.1 cm^{-1}) as well as the total cross sections at $T = 1{,}700$ K and overlaying them with the regions excluded in ref. 70. One can see that although generally the selection in ref. 70 was sensible, there are still regions where this could be done more efficiently. This example illustrates the capability of the new format of the ExoMol line lists, ExoMolHD; to provide information on the uncertainties of the spectral lines in the database, see ref. 39. In fact, an updated version of the ExoMol ToTo line list for TiO with the ExoMolHD functionality and improved line positions from recent experimental studies[71–73] has been constructed[74]. High-resolution ExoMol lines can be accessed directly from a new ExoMolHR database[40].

VO has proven to be an even more elusive molecule than TiO. Despite many attempts, it has only been detected in 2023 via HRCCS of WASP-76 b with the MAROON-X instrument at Gemini North[75]. The detection ($S/N \approx 6$) was aided by the ExoMol line list VOMYT[76] used between 0.49 and 0.92 μm at $R \approx 85{,}000$. In spite of this detection, the VOMYT line list lacked sufficient quality for HRCCS studies, which is a cited reason for the many non-detections[34]. Although VOMYT is sufficiently complete for the low-accuracy transit spectroscopy, line positions are certainly not good enough for high-resolution applications for a number of reasons. The spectroscopic model used to build VOMYT[76] did not benefit from the (now) standard MARVELization procedure (see, for example, ref. 39) to improve the accuracy of the calculated energies (and thus line positions). Furthermore, VOMYT does not include the spectroscopic effects caused by hyperfine splitting of the lines (transitions between states affected by the nuclear spin effects), which is especially pronounced for VO owing to the large nuclear spin ($I_V = 7/2$) and nuclear quadrupole moment of V. The hyperfine effects are usually small and therefore can be ignored in spectroscopic applications, but VO is a special case.

A new line list for VO, called HyVO[38], has recently been released. HyVO is based on a MARVEL study that required the assignment of previously only partially analysed high-resolution laboratory spectra[77], experimental line positions from 14 experimental sources (see ref. 77 for details), extension of ExoMol diatomic nuclear motion code Duo to explicitly consider hyperfine effects[78], and extensive theoretical studies to develop an accurate spectroscopic model for VO (refs. 79–81). HyVO considers the same rovibronic transitions as VOMYT but uses a greatly improved spectroscopic model combined with some improvements of the VO transition dipole moment curves. The VO line lists are illustrated in Fig. 3c,d. The light blue boxes in Fig. 3c indicate spectral regions where HyVO is more accurate, as supported by the experimental data. Although the general shapes of the bands agree reasonably well, including the band heads, the higher-resolution comparison shows a significant shift in the line positions in the spectrum of the improved VO line list HyVO. Improvements represented by HyVO have successfully led to a detection of VO in the atmospheres of TOI-1518 b[82] and the confirmation of the presence of VO in WASP-76 b[83] with increased detection significance.

OH

The radical OH is a product of thermal dissociation of water, and this species is predicted to be particularly abundant at the very high temperatures (≥2,200 K) of the UHJs. As this class of planet is a relatively new discovery, detections of OH are also relatively new. The first was announced in 2021 based on HRCCS of WASP-33 b in emission (5.5σ)[84] using SUBARU/IRD (0.97–1.75 μm, $R \approx 70{,}000$), where an empirical line list[85] was used. The OH line list has proven to be sufficiently accurate and complete in infrared, at least for local thermal equilibrium[86]; however, the quality is expected to degrade in the shorter-wavelength region populated with electronically excited states and with associated photodissociation effects[87]. OH is an important species for deriving the oxygen content of UHJs[88], and thus its detection is

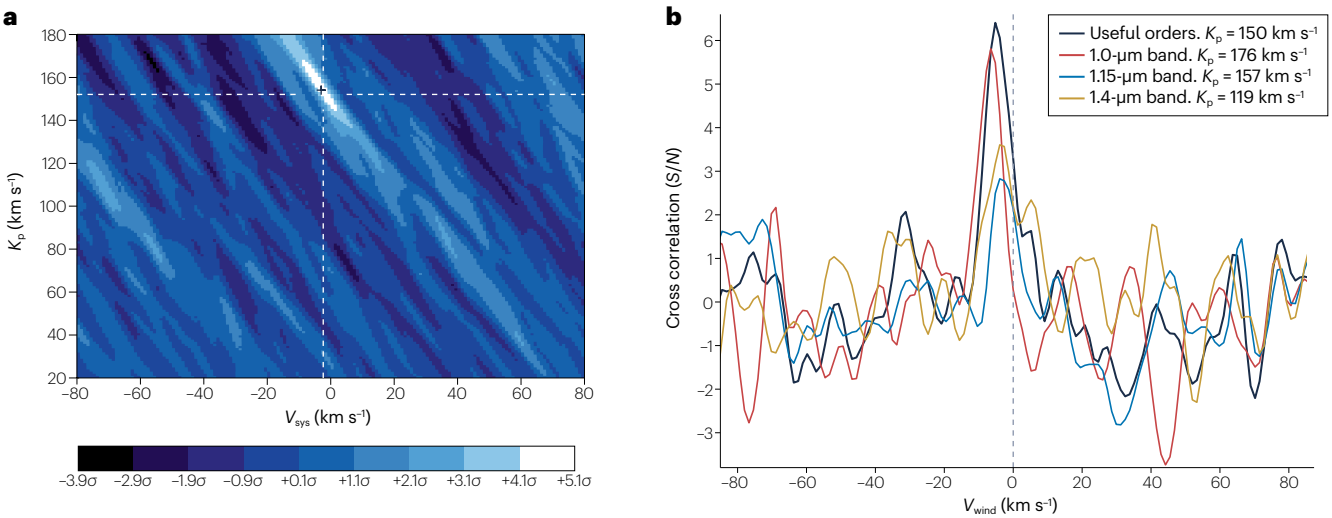

Fig. 2 | Examples of high-resolution cross-correlation spectroscopy detections. a, Detection of H_2O via high-resolution cross-correlation spectroscopy from Very Large Telescope (VLT)/CRIRES emission spectroscopy of HD 189733 b[52]. **b**, Calar Alto Observatory (CAHA)/CARMENES transmission spectroscopy of HD 209458 b[123]. The *x* axes in both figures represent the velocity of the cross-correlated signal in the frame of the planet (in km s^{-1}). The *y* axis in panel **a** represents the semi-amplitude of the planet's radial velocity (K_p, km s^{-1}), whereas in panel **b** it represents the cross-correlation signal-to-noise ratio, with colours indicating different portions of the near-infrared spectrum used in the analyses. For details, see ref. 123.

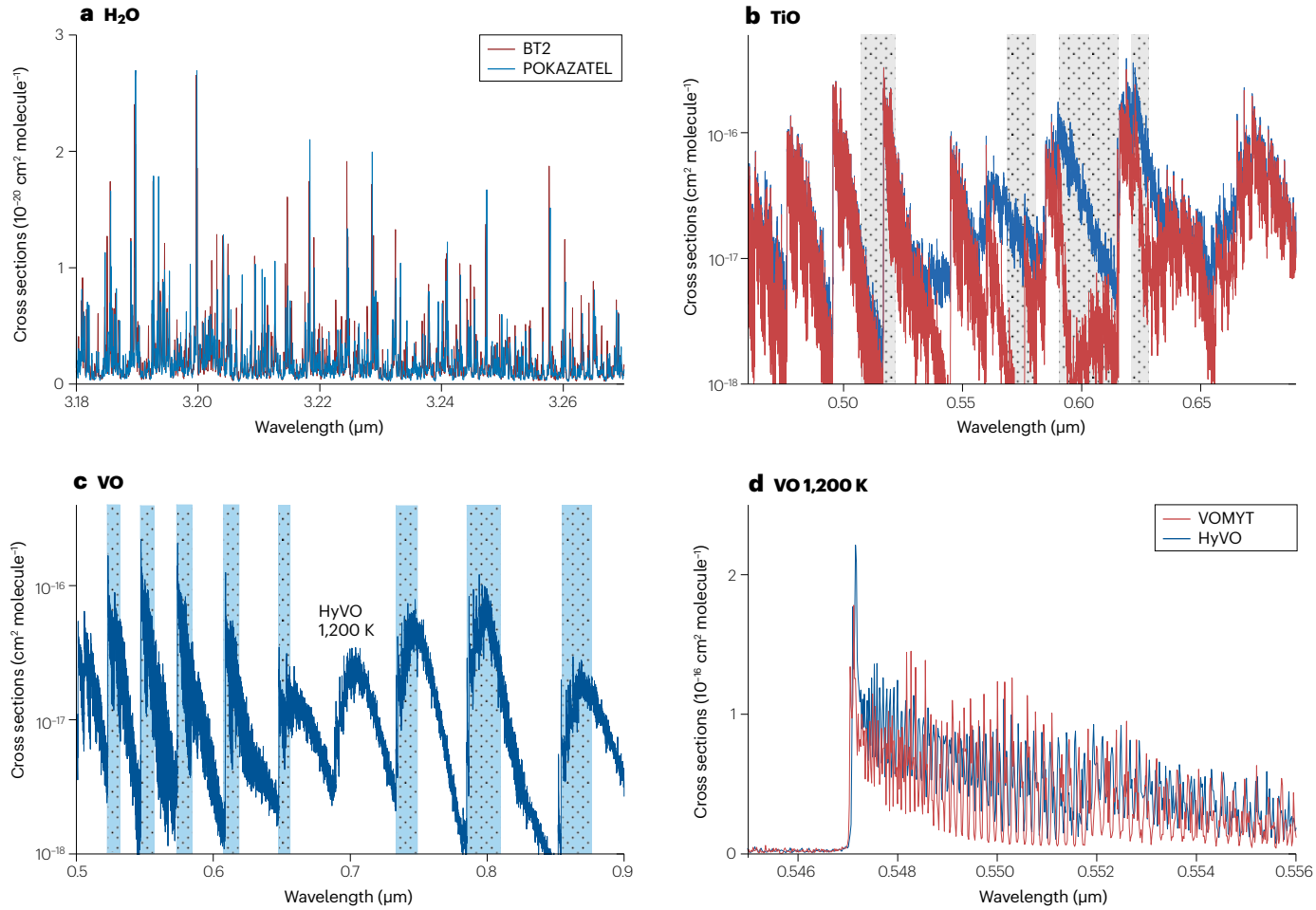

Fig. 3 | Synthetic spectra of molecules. a, Absorption cross sections of H_2O at $T = 1{,}950$ K computed using the BT2[57] and POKAZATEL[56] line lists at $T = 1{,}950$ K, with a Lorentzian line profile. **b**, Absorption cross sections of TiO computed using the ToTo line list[68] (blue line); cross-section contribution from accurate lines with line positions better than 0.1 cm^{-1} as defined by the ExoMolHD uncertainties (red lines). Grey rectangular areas indicate regions deemed inaccurate in Prinoth et al.[70] and excluded from their cross correlations of the spectra of WASP-189 b[134]. **c**, Cross sections of VO at $T = 1{,}200$ K computed using the new HyVO[38]. The light blue rectangular areas indicate regions where HyVO is especially accurate thanks to the extensive MARVELization procedure. **d**, Comparison of the cross sections computed using the new line list HyVO and the old VOMYT line lists[76].

desirable for any planet of this class. The strength of the data was also demonstrated in the first band-by-band HRCCS analysis of OH in WASP-33 b[86]. A new MARVELized rovibronic line list for OH with ExoMolHD uncertainties has been recently released as part of the ExoMol project[89].

CH_4

Methane has long represented a puzzle for the chemistry of exoplanet atmospheres. At their typical pressures, and in equilibrium, atmospheres should transition from CO-dominated to CH_4-dominated around 1,200 K (ref. 49). However, multiple attempts to look for the signatures of CH_4 from space using HST/WFC3 were hindered by the simultaneous presence in the same spectral range of a strong water vapour band. JWST has already provided several detections of CH_4, including in WASP-80 b[90], K2-18b[42] and TOI-270 d[44,91]. Through HRCCS, the search started in 2001[92] and continued systematically but largely unsuccessfully during the 2010s, also partly owing to the hotter temperatures targeted. The first evidence (4.1σ) of methane came from the atmosphere of HD 102195 b with GIANO ($R = 50{,}000$, 0.95–2.45 μm) using the HITRAN 2016 line list[93]. This measurement is puzzling, as it was obtained with a line list that is potentially incomplete at high temperatures (1,200 K), especially in near infrared, and in fact it was not confirmed when reanalysed[59] using the more extensive HITEMP 2020 line list[94]. The latter line list was instead successfully used to detect methane in HD 209458 b (5.6σ)[14], WASP-69 b (4.9σ)[95] and WASP-80 b (4.2σ)[45]. For the latter two cases the claim is not strong; however, the detection of methane in WASP-80 b has been fully confirmed from space by JWST[90].

The HITEMP 2020 dataset[94] is based on the empirically computed CH_4 line list by the TheoReTS group[96,97]. Another popular extensive spectroscopic dataset of CH_4 is the ExoMol line list 10to10[98], which was produced in response to the pressing needs for modelling (low-resolution) transit observations of hot Jupiters[28]. Acknowledging the importance of having a line list suitable for HRCCS studies, ExoMol has produced

a new CH_4 list (MM)[99], built using the ExoMolHD procedure — that is, using an accurate spectroscopic model (potential energy surface) and MARVELization procedure with a new set of experimentally derived energies of methane[100]. Although the MM line list supersedes the quality and completeness of 10to10, MM, HITEMP 2020 and TheoReTS provide essentially similar capability both for the low-resolution and high-resolution exoplanetary applications.

Whereas HITRAN provides highly accurate line lists that are not necessarily very complete at higher temperatures, synthetic line lists such as HITEMP 2020, TheoReTs and MM are examples of extensive molecular data sets constructed to provide the completeness, which is especially important for low-resolution application. It is also useful to retain the accuracy of the underlying laboratory data. They are typically computed using the best empirically improved spectroscopic models for a large range of excitations, containing (or based on) tens of billions of transitions, with some of these transitions (~300,000 for CH_4) replaced with experimentally derived, HITRAN-like values.

For other species of interest, see Box 2.

Box 2 | Other species of interest

Nitrogen-bearing species (NH_3, HCN): Nitrogen chemistry is of particular interest for exoplanet spectra, as N is expected to be found primarily as the essentially invisible N_2 in equilibrium; therefore, the by-products ammonia and hydrogen cyanide should be highly sub-abundant[145]. However, this expectation can be modified if disequilibrium chemistry is active, in the form of vertical mixing from the deeper atmospheric layers or photochemistry in the upper layers[146], or the planet metallicity or carbon content is particularly far from solar.

Detection of HCN was indeed claimed in HD 209458 b with the Very Large Telescope (VLT) CRIRES instrument back in 2018[125] and subsequently confirmed at very high significance with GIANO (9.9σ)[14]. But at low-resolution HCN and HCCH are highly degenerate[147], which high-resolution cross-correlation spectroscopy (HRCCS) can sort out. Furthermore, evidence has also been reported for another hot Jupiter, HD 198733 b (5.0σ)[127], via VLT/CRIRES. The ExoMol HCN line list[126] has proved to be robust for these detections; it was the first to undergo a MARVELization-like procedure and was expected to provide sufficient quality for HRCCS studies

The ExoMol CoYuTe line list for NH_3 (ref. 148) is also expected to be sufficiently accurate and complete in the (mid-)infrared, but less so when approaching near infrared, showing pronounced frequency shifts[149]. More laboratory data are needed in the near-infrared spectroscopic region, which however is extremely challenging to analyse even at room temperature[150]. An update of the MARVELization of the ExoMol CoYuTe line list is currently under way to help resolve issues with NH_3 detection. NH_3 has proved more elusive than HCN, and essentially limited to HRCCS with GIANO, again in HD 209458 b (5.3σ)[14], WASP-80 b (5.0σ)[45], WASP-69 b (4.9σ)[95] and HAT-P-11 b (5.0σ)[151]. Although the presence of *N*-bearing species in the cooler (<1,000 K) WASP planets is not unexpected, its detection in the two hot Jupiters is puzzling, because neither of the species is predicted to be abundant enough to show up in the James Web Space Telescope (JWST) transmission spectroscopy, at least when equilibrium chemistry is assumed[152].

Carbon dioxide: CO_2 is a notable species because its abundance in equilibrium chemistry should be negligible for the solar metallicity and C-to-O ratio, but it can grow considerably for higher values of both parameters. Furthermore, the existing line lists for CO_2, UCL-4000 from ExoMol[153] and AI-3000K from NASA Ames[141], should be of the best accuracy applicable for the broad range from mid-infrared to near infrared. Both UCL-4000 and AI-3000K have been extensively validated against accurate experimental spectra of CO_2 at room temperature and contain a large number (over 170,000) of experimental lines or lines of experimental quality (resolving power $R \approx 100{,}000$) from HITRAN 2020[135]. In fact, partially defying expectations, JWST has started to deliver solid detections of CO_2 from space by targeting its main near-infrared band (4.0–4.5 μm)[26]. On the other hand, ground observations and thus HRCCS are confined to the weaker band around 2.0 μm, which is also particularly challenging from a data analysis standpoint as it overlaps with strong CO_2 absorption in the Earth's atmosphere. As such, no claim has been made yet of a CO_2 detection using HRCCS, with the exception of a tentative signature in WASP-80 b (4.4σ)[45].

Other oxides (SiO, SiO_2, MgO and FeO): Silicate atmospheres are predicted to be important absorbers in rock vapour atmospheres of hot super-Earths[111,154,155]. A tentative detection of SiO in the M band for β Pictoris b was recently reported[41] (signal-to-noise $S/N = 4.3$).

Hydrides (FeH, CrH): The search for hydrides in exoplanet atmospheres has not been as widely performed as for other species, but it is nevertheless interesting given the presence of these species in stellar atmospheres. Although FeH has not yet been successfully detected[156], CrH has been recently detected in WASP-31 b (5.6σ)[130] using GRACES/Gemini North and UVES/VLT, in the region of the strongest CrH band in the range of 0.86–0.9 μm with $R \approx 66{,}000$. The CrH line list of ref. 131 was used, which has been validated by high-resolution cross correlation to a low-mass, cold star (Teegarden's star, M8). This is an empirical line list with a varying quality, failing to reproduce many spectral features in stellar spectra[157,158]. A new rovibronic line list for CrH is under construction by ExoMol, which will incorporate new experimental data by the Lyon group[159].

Hydrocarbons (C_2H_2, C_2H_4 and so on): Similarly to CO_2, hydrocarbons are good tracers of a planet's C-to-O ratio and metallicity, but they can also be formed photochemically in the upper atmosphere. However, so far there is just one claim of C_2H_2 observed with GIANO in HD 209458 b (6.1σ)[14]. Given the strength of the detection (above that of NH_3 and CH_4 for the same planet), qualitatively one would expect a non-negligible abundance of this molecule. However, JWST spectroscopy suggests that no feature of C_2H_2 is visible[152]. The ExoMol line list for C_2H_2 (ref. 129) should be sufficiently accurate and complete in the (mid-)infrared, but less so when approaching near infrared; this line list is currently being improved by re-MARVELizing it using new high-resolution laboratory measurements[160–163].

Impact of line list quality on Bayesian retrievals

Although from the perspective of a data provider it is important to know the state of individual molecular detections, HRCCS studies are slowly moving away from claiming individual species and towards using their combined signal to derive the global composition (elemental abundances) and thermal structure of the atmospheres. This further step has only recently been enabled[15,18], and it involves translating cross-correlation functions into likelihood functions. Such functions are then used to explore the full parameter space of an atmosphere through Bayesian retrieval techniques widely used at low resolution[101].

The first two Bayesian retrievals using infrared HRCCS both came in 2021. Pelletier et al.[102] retrieved the CFHT/SPIRou emission spectrum of exoplanet τ Boötis b, which is a non-transiting planet and therefore cannot be studied using the transit technique. They showed that the spectral content in their data was sufficient to constrain the radius of the planet indirectly, and to measure CO log-abundances at a precision (size of confidence intervals) of 0.25–0.3 dex. Remarkably, their measurement found very abundant CO but no H_2O, indicating a planet with high C-to-O ratio and metallicity.

Shortly afterwards, Line et al. used the Gemini-S/IGRINS instrument[103] to retrieve emission spectra of WASP-77Ab, achieving an unprecedented precision of 0.1 dex in the abundances of H_2O and CO. In their work, they used the POKAZATEL line lists for water, but no comparison between HITEMP 2010 and POKAZATEL was provided to assess potential biases at the level of their precision. No other molecular species were found in the mix.

Although the accuracy of the line positions remains the primary driver of good HRCCS detections, when translating the observed spectrum into abundance and temperature constraints, in some cases incomplete line lists can lead to an underestimation of the broadband opacity and thus of the local continuum, especially in the visible, with serious repercussions on the modelled line depths to which HRCCS is sensitive. The importance of the completeness of the lab data for HRCCS studies has not been sufficiently studied and thus requires more understanding.

Furthermore, the line shape also encodes information on the pressure and temperature of the atmospheric layers. Thus it is also important in the long run to estimate the line broadening correctly, particularly pressure broadening in mixtures that are not air, but rather (for the vast majority of cases at the moment) a H_2–He mixture. In the next decade, other heavier mixtures might become relevant, such as dominant-H_2O atmospheres or dominant-CO_2 atmospheres (for example on Venus). Line broadening information for these broadeners remains scarce, particularly at elevated temperatures.

Another interesting example of using Bayesian retrieval techniques in HRCCS is the recent uniform analysis of six UHJs[104], which determined (log-)abundances of a dozen refractory atomic species (atoms heavier than He and excluding C, N and O) with a precision ranging between 0.2 and 0.5 dex, depending on the species and the data quality. When Fe/H is used as proxy for metallicity and compared with the Solar System trend, the six exoplanets fall below the mass–metallicity relation. This work is one of the very few examples of comparative planetology with the HRCCS method.

Accuracy, parameter space of interest and their associated data requirements

Let us summarize the main points on the physical parameters implicitly covered in the previous sections.

The atmospheric spectral simulations need to be performed over an extended range of atmospheric pressures typically from 10–100 bar (the deepest layers, below the planet's photosphere) to at least 1 μbar (for the strongest molecular absorbers), potentially 10^{-8} bar to model the core of strong atomic lines. This is much broader than the corresponding dynamic range in pressure for low-resolution observations, because line cores — only resolved via HRCCS — absorb at much higher altitude (much lower pressure) than the continuum or line wings.

In terms of temperature, the hottest known exoplanets (UHJs) have equilibrium temperatures reaching 4,000 K. But their upper atmospheric layers can reach 5,000 K or more in the presence of thermal inversions, where temperature increases with altitude instead of decreasing, owing to absorbers (such as TiO and VO) of stellar radiation in the optical[62]. However, these temperatures are unlikely to be relevant for the majority of molecular species, which will be thermally dissociated. Therefore, data provision should be as accurate as possible up to the temperatures corresponding to the thermal dissociation limit at typical pressures of 0.01–1 bar. Depending on the species, this may require measurements and/or calculations extending to 2,500–3,000 K, or even higher.

In terms of accuracy, current spectrographs used for HRCCS are sensitive to velocity shifts between 0.5 km s^{-1} and a few kilometres per second, depending on the signal strength and type of observation (transmission versus emission). Taking 1 km s^{-1} as an order-of-magnitude reference, this means that the relative accuracy in transition frequency must be equal to or better than 3×10^{-6} (that is, $R > 333{,}333$) to avoid smearing of the cross-correlation function, which would affect the detectability of a species and imprint biases in measured velocities and abundances. Across the typical wavelength extent of near-infrared observations, this requirement would translate to ~0.03 cm^{-1} accuracy at 1 μm (10,000 cm^{-1}), or ~0.01 cm^{-1} at 3.3 μm (3,000 cm^{-1}).

Prospects of new HRCCS detections

After the surprise detection of SO_2 in WASP-39 b[26] and more recently in WASP-107 b[27], this species constitutes an obvious candidate for the molecular hunt with HRCCS, together with other sulfur-bearing species, including H_2S, HS, SO and OCS. Very recently, Kirk et al.[105] suggested OCS for the feature at 4.9 μm in WASP-15b using an ExoMol line list[106]. Accurate data for all these molecules are readily available[106–110]. Figure 4 illustrates their spectra coverage from mid-infrared to near infrared. While SO has a strong presence in the near infrared, the other four species have good chances for existing HRCCS detection with infrared instruments.

Additional potential prominent discoveries include other oxides predicted to be abundant on the 'lava planets'. These species include AlO, CaO, NaO, O_2, O_3, FeO, SiO, SiO_2, MgO, KO, TiO, LaO, YO and ZrO, with SiO and SiO_2 being the most prominent identifiers of silicate atmospheres in the infrared[111]. The problem with these lava planet molecules is that, with some exceptions, they have been largely ignored by laboratory studies. For example, there are no line lists for FeO or KO; FeO has a very complex electronic structure[112] that is beyond the reach of current theoretical methods even for a reasonable low-resolution line list, and its complex irregular bands make it difficult to model and analyse experimental spectra[113]. The laboratory data on FeO are limited to microwave[114]. For SiO_2, whose laboratory data on SiO_2 are also limited to microwave, only thanks to the ExoMol machinery that produced a hot theoretical line list for SiO_2 (ref. 115) could the exoplanetary community study and demonstrate its potential importance for the absorption and cooling properties of atmospheres of hot evaporating lava worlds[111]; however, the SiO_2 line list will require high-resolution laboratory data before it is suitable for use in HRCCS studies, and at present there are no useful high-resolution laboratory spectra of this

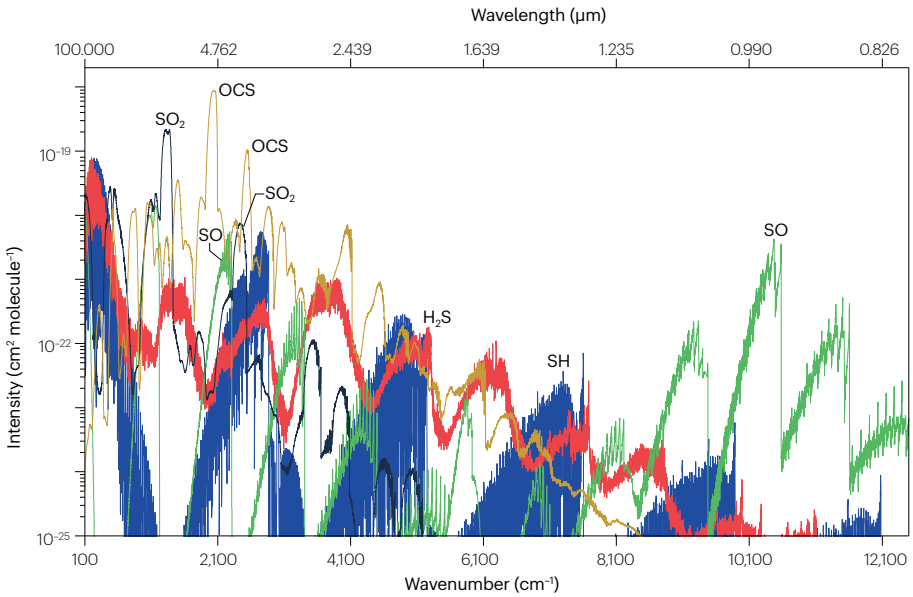

Fig. 4 | Spectra of sulfur-bearing molecules with high-quality ExoMol line lists for high-resolution cross-correlation spectroscopy studies at T = 1,700 K. The colours correspond to SO_2 (black), SO (green), OCS (olive), SH (blue) and H_2S (red). Data from refs. 106,107,109,110 were also used in this plot.

molecule available. Conversely, there are line lists available for SiO, AlO, CaO, MgO, O_3, TiO, LaO and ZrO, which should be accurate enough for use in HRCCS studies.

Discussion

After the detailed, per-species information presented above, we would like to touch upon the question "Are HRCCS detections all equally reliable?" Although we recognize that this is a sensitive topic, some discussion is needed to contextualize molecular detections produced so far by means of HRCCS in order to be useful to both communities of data providers and observers.

First, it has been shown recently that non-robust methods artificially amplify signals or even create false positive detections in HRCCS[116]. This result is not due to a weakness of the method but rather to a statistically flawed analysis. Although improvements in statistical techniques and model validation can help to reduce the likelihood of false positives, understanding the quality of the underpinning laboratory data is equally essential. This is sometimes hard to assess, especially in older work. Indeed, many exoplanetary publications do not habitually fully refer to the source of line list data, often generically referring to previous work or to the general aggregated database (ExoMol or HITEMP) or cross-sections set (HELIOS-K[117] or ExoMolOP[118]) rather than to the actual publication for the selected line list. As the quality (accuracy and completeness) is decisive and varies from source to source, without proper referencing the analysis cannot be fully trusted or reproduced. In addition, there are ethical aspects to consider with respect to acknowledging the appropriate contribution of data providers.

One should also consider and test for possible spurious detections by chance correlations of model spectra with random noise fluctuations or even lines from different species. Although statistically this is hard to achieve with species containing thousands of strong lines across the observed spectral range, there is very little testing in the literature confirming the theoretical expectations. In one such study[14], the authors performed a series of correlation tests using random noise, scrambling spectral lines or removing one of the seven species from the mix. One of the lessons of such tests is that inaccuracies in line lists are likely to prevent detections with the HRCCS technique, as they would act similarly to 'scrambled' spectral lines.

Even though data such as HITRAN and ExoMol now tend to supply the information about the measurement uncertainties, the common practice is still not to propagate them to the opacity tables and hence to a retrieval, for example to inflate the corresponding posteriors. This limit is related to the need (due to computational reasons) to pre-compute cross-section tables (in wavelength, pressure and temperature) from molecular line lists, and then interpolate through the table every time a specific value is required during the radiative transfer calculations. To implement 'errors' in the transition frequencies is difficult but possible. In fact, ExoMol is currently working on propagating the line list uncertainties into the cross sections.

On the other hand, this Review shows that significant refinement has been made in the line lists of key species such as H_2O and CH_4, and the ultimate aim is to reduce inaccuracies to levels far below the achievable precision, rather than forward modelling the uncertainties. We further note that this opacity table approach is not exclusive to HRCCS, but it is also adopted with lower-resolution space observations, and therefore it is not an issue specific to our Review. At lower resolution, completeness is more important than line position (owing to the reliance on broadband features), but the issue of propagating uncertainties in the line lists still stands.

Perhaps a not directly data-related issue is the non-uniform claims of detections in the literature (see Supplementary Information). For instance, a 5σ detection can reflect substantially different measurables in different papers, and we caution the reader to carefully check the assumptions made to estimate significance.

Finally, the current availability of good-quality HRCCS observations and independent detections of species from space with JWST opens up the possible use of observations to validate the laboratory data, especially for systems for which measurements are difficult or expensive. This could be done by testing for specific transitions, molecular bands or even subsets of strong lines for which good accuracy is achievable. There are many such examples in solar as well as stellar spectroscopic applications, where photospheres of stars provide ideal conditions for laboratory

Glossary

Bayesian retrievals
A statistical approach used to infer atmospheric properties of exoplanets by comparing observed spectra to models, incorporating prior knowledge and uncertainties.

Cross-correlation function
(CCF). A mathematical tool used in HRCCS to measure the similarity between an observed spectrum and a model template, enhancing weak signals buried in noise.

Direct imaging
A method of detecting exoplanets by capturing their light separately from the host star, typically using adaptive optics and coronagraphy to suppress stellar glare.

High-resolution cross-correlation spectroscopy
(HRCCS). A spectroscopic technique that enhances the detection of molecular species in exoplanet atmospheres by cross correlating observed spectra with model templates at high spectral resolution.

Measured active rotational–vibrational energy levels
(MARVEL). A data-driven approach that refines spectroscopic line lists by combining experimental measurements with theoretical calculations to provide highly accurate molecular energy levels.

Resolving power
A measure of a spectrograph's ability to distinguish between closely spaced spectral features, defined as $R=\lambda/\Delta\lambda$, where λ is the wavelength and $\Delta\lambda$ is the smallest detectable difference.

Secondary eclipse
The moment when an exoplanet moves behind its host star, allowing astronomers to isolate and analyse the planet's thermal emission and reflected light.

Transit
The passage of an exoplanet in front of its host star, causing a temporary dip in the star's brightness, which allows for atmospheric characterization through transmission spectroscopy.

spectroscopic data of highest quality, including the spectroscopically rich infrared spectral atlases of the Sun[119] and the very recent detection of the unknown band of methyl cation (CH_3^+) in a protoplanetary disk[120].

Conclusion

In this Review we have highlighted the potential of HRCCS and reviewed its successes in characterizing the atmospheres of exoplanets with ground-based observations. As we are entering the golden era of atmospheric characterization and the first light of the Extremely Large Telescope is just a few years ahead, it is paramount that the technique further matures to a level that can be trusted by the whole community. We have also discussed the most critical needs of both the observational and laboratory communities at the time of writing.

Although the level of detail achieved in determining the chemical inventory of exo-atmospheres is remarkable, we stress the importance of the interplay between the quality and reliability of an astrophysical measurement of a species and the underlying state of a line list and/or laboratory data of such species. Such synergy stems from the dependence of HRCCS on accurate models to detect the atmospheric signatures and extract further information about composition and temperature.

As a closing remark, we would like to reach out to different communities with a few pleas regarding the production and usage of higher-resolution data for and in HRCCS applications, summarized as follows.

To atmospheric modellers

When using line lists in HRCCS applications, there should be an understanding of the source of underlying data, especially the accuracy of line positions, appreciating whether such source is empirical, partly empirical and partly theoretical, or purely theoretical.

As line transitions are now provided with their estimated uncertainties, HRCCS users should only rely on the lines of uncertainty compatible with the resolving power of the instrument, at least for the strongest lines. It is still an open question whether to include (partly or fully) weaker, more uncertain lines, as these contribute the overall opacity of the planet spectrum and therefore must be accounted for when inverting the observed spectrum into abundances and temperature via Bayesian retrievals.

When using molecular data in atmospheric analysis, it is extremely important to cite relevant providers and not only the framework or programme that incorporates these data, at least for the most critical species featured in the discovery or analysis. This applies not only to the database but also the original laboratory or theoretical work. This is not only an ethical point, but highlights the reproducibility and confidence in results reported.

When using the HRCCS technique in connection with the Bayesian analysis of HRS data to measure temperature and abundances, all spectral information becomes important, including line positions, intensities and shapes. Thus a compromise needs to be found between only using the most accurate lines and incorrectly modelling the line depth, and using all the available lines and mismatching the position of the weaker lines. This is an active challenge in HRCCS and one that will probably require stronger interplay between observers and data providers.

To experimentalists

High-resolution spectra of metal oxides in the infrared, visible and ultraviolet are needed. As the intensity measurements are often not possible in the laboratory, we invite experimentalists to measure lifetimes that provide valuable tests of the strength of transitions for proving theoretical models without the need to know absolute abundances of the species in the experiment.

To computational chemists

Accurate (from high level of theory) first-principles calculations are needed for transition metal diatomics, including TiO, TiH, FeO, FeH, NiO, NiH, CrO, CrH and potentially many more. This includes full bond-length description of the ab initio potential energy, (transition) dipole moment, spin–orbit and other coupling curves with an accurate reporting of relative phases. We realize that studying multiple electronic states and associated dipole transition moments for transition-metal-containing diatomics such as TiO, FeH and FeO is difficult and therefore seems not to have been recognized by the quantum chemists as an important task. Even high-level ab initio calculations are not very useful when only single point calculations or absolute values of non-diagonal properties are provided. While preparing this Review, we were made aware of the recent substantial progress in ab initio calculations of transition metals (FeH) by Ariyarathna et al.[121].

To HRS observers

When claiming a detection, always explain how this is measured and whether the data analysis chosen could bias the outcome. Be informed about the reliability of the underlying model used for cross correlation, and when presented with an unexpected or new

species get in touch with modellers and line list providers to appreciate whether such measurements can be trusted given the current level of accuracy.

We also recognize that different communities use different language and standards. In this case where there is strict interdependency between the observational result and the underlying knowledge of laboratory spectra, it is important that scientists working on these problems understand each other's requirements, needs and limitations. Such calls for interdisciplinary collaboration are timely, as exoplanet science is gradually moving towards the study of smaller and cooler planets, and will eventually tackle objects with temperature and compositions potentially amenable for life. Only by combining the expertise of different disciplines will we be able to achieve not only an exciting scientific result, but also a reliable measurement standing the test of time.

Published online: 16 July 2025

References

1. Sing, D. K. et al. A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion. *Nature* **529**, 59–62 (2016).
2. Mansfield, M. et al. A unique hot Jupiter spectral sequence with evidence for compositional diversity. *Nat. Astron.* **5**, 1224–1232 (2021).
3. Kempton, E. M.-R. et al. A framework for prioritizing the TESS planetary candidates most amenable to atmospheric characterization. *Publ. Astron. Soc. Pac.* **130**, 114401 (2018).
4. Lagrange, A.-M. Direct imaging of exoplanets. *Philos. Trans. A Math. Phys. Eng. Sci.* **372**, 20130090 (2014).
5. Currie, T. et al. in *Protostars and Planets VII. Astronomical Society of the Pacific Conference Series* Vol. 534 (eds Inutsuka, S. et al.) (Astronomical Society of the Pacific, 2023).
6. Barstow, J. K., Aigrain, S., Irwin, P. G. J., Kendrew, S. & Fletcher, L. N. Transit spectroscopy with James Webb Space Telescope: systematics, starspots and stitching. *Mon. Not. R. Astron. Soc.* **448**, 2546–2561 (2015).
7. Madhusudhan, N. Exoplanetary atmospheres: key insights, challenges, and prospects. *Annu. Rev. Astron. Astrophys.* **57**, 617–663 (2019).
8. Kempton, E. M.-R. & Knutson, H. A. Transiting exoplanet atmospheres in the era of JWST. *Rev. Mineral. Geochem.* **90**, 411–464 (2024).
9. Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
10. Queloz, D. Echelle spectroscopy by a CCD at low signal-to-noise ratio. In *Proc. 167th Symposium of the International Astronomical Union: New Developments in Array Technology and Applications* (eds Philip, A. G. D. et al.) 221–229 (Kluwer, 1995).
11. Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W. & Albrecht, S. The orbital motion, absolute mass and high-altitude winds of exoplanet HD209458b. *Nature* **465**, 1049–1051 (2010).
12. Brogi, M. et al. The signature of orbital motion from the dayside of the planet τ Boötis b. *Nature* **486**, 502–504 (2012).
13. Snellen, I. A. G. et al. Fast spin of the young extrasolar planet β Pictoris b. *Nature* **509**, 63–65 (2014).
14. Giacobbe, P. et al. Five carbon- and nitrogen-bearing species in a hot giant planet's atmosphere. *Nature* **592**, 205–208 (2021).
15. Brogi, M. & Line, M. R. Retrieving temperatures and abundances of exoplanet atmospheres with high-resolution cross-correlation spectroscopy. *Astrophys. J.* **157**, 114 (2019).
16. Öberg, K. I., Murray-Clay, R. & Bergin, E. A. The effects of snowlines on C/O in planetary atmospheres. *Astrophys. J. Lett.* **743**, 16 (2011).
17. Mordasini, C., van Boekel, R., Mollière, P., Henning, T. & Benneke, B. The imprint of exoplanet formation history on observable present-day spectra of hot Jupiters. *Astrophys. J.* **832**, 41 (2016).
18. Gibson, N. P. et al. Detection of Fe I in the atmosphere of the ultra-hot Jupiter WASP-121b, and a new likelihood-based approach for Doppler-resolved spectroscopy. *Mon. Not. R. Astron. Soc.* **493**, 2215–2228 (2020).
19. Pino, L. et al. Neutral iron emission lines from the dayside of KELT-9b: the GAPS program with HARPS-N at TNG XX. *Astrophys. J. Lett.* **894**, 27 (2020).
20. Brogi, M. et al. The Roasting Marshmallows program with IGRINS on Gemini South I: composition and climate of the ultrahot Jupiter WASP-18 b. *Astron. J.* **165**, 91 (2023).
21. Louden, T. & Wheatley, P. J. Spatially resolved eastward winds and rotation of HD 189733b. *Astrophys. J. Lett.* **814**, 24 (2015).
22. Brogi, M. et al. Rotation and winds of exoplanet HD 189733 b measured with high-dispersion transmission spectroscopy. *Astrophys. J.* **817**, 106 (2016).
23. Wakeford, H. R. et al. HAT-P-26b: a Neptune-mass exoplanet with a well-constrained heavy element abundance. *Science* **356**, 628–631 (2017).
24. Tsiaras, A., Waldmann, I. P., Tinetti, G., Tennyson, J. & Yurchenko, S. N. Water vapour in the atmosphere of the habitable-zone eight-Earth-mass planet K2 18b. *Nat. Astron.* **3**, 1086–1091 (2019).
25. Taylor, J. et al. Awesome SOSS: atmospheric characterization of WASP-96b using the JWST early release observations. *Mon. Not. R. Astron. Soc.* **524**, 817–834 (2023).
26. Ahrer, E.-M. et al. Identification of carbon dioxide in an exoplanet atmosphere. *Nature* **614**, 649–652 (2023).
27. Dyrek, A. et al. SO_2, silicate clouds, but no CH_4 detected in a warm Neptune. *Nature* **625**, 51–54 (2024).
28. Yurchenko, S. N., Tennyson, J., Bailey, J., Hollis, M. D. J. & Tinetti, G. Spectrum of hot methane in astronomical objects using a comprehensive computed line list. *Proc. Natl Acad. Sci. USA* **111**, 9379–9383 (2014).
29. Gaudi, B. S. et al. A giant planet undergoing extreme-ultraviolet irradiation by its hot massive-star host. *Nature* **546**, 514–518 (2017).
30. Sanchis-Ojeda, R. et al. Transits and occultations of an Earth-sized planet in an 8.5 hr orbit. *Astrophys. J.* **774**, 54 (2013).
31. Tennyson, J. Accurate variational calculations for line lists to model the vibration rotation spectra of hot astrophysical atmospheres. *WIREs Comput. Mol. Sci.* **2**, 698–715 (2012).
32. Tennyson, J. & Yurchenko, S. N. The ExoMol project: software for computing molecular line lists. *Int. J. Quantum Chem.* **117**, 92–103 (2017).
33. Hoeijmakers, H. J. et al. A search for TiO in the optical high-resolution transmission spectrum of HD 209458b: hindrance due to inaccuracies in the line database. *Astron. Astrophys.* **575**, 20 (2015).
34. Regt, S., Kesseli, A. Y., Snellen, I. A. G., Merritt, S. R. & Chubb, K. L. A quantitative assessment of the VO line list: inaccuracies hamper high-resolution VO detections in exoplanet atmospheres. *Astron. Astrophys.* **661**, 109 (2022).
35. Tennyson, J. et al. The 2024 release of the ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J. Quant. Spectrosc. Radiat. Transf.* **326**, 109083 (2024).
36. Furtenbacher, T., Császár, A. G. & Tennyson, J. MARVEL: measured active rotational–vibrational energy levels. *J. Mol. Spectrosc.* **245**, 115–125 (2007).
37. Al-Derzi, A. R. et al. An improved rovibrational linelist of formaldehyde, $H_2{}^{12}C^{16}O$. *J. Quant. Spectrosc. Radiat. Transf.* **266**, 107563 (2021).
38. Bowesman, C. A., Qu, Q., McKemmish, L. K., Yurchenko, S. N. & Tennyson, J. ExoMol line lists — LV: hyperfine-resolved molecular line list for vanadium monoxide ($^{51}V^{16}O$). *Mon. Not. R. Astron. Soc.* **529**, 1321–1332 (2024).
39. Tennyson, J. et al. The 2020 release of the ExoMol database: molecular line lists for exoplanet and other hot atmospheres. *J. Quant. Spectrosc. Radiat. Transf.* **255**, 107228 (2020).
40. Zhang, J., Hill, C., Tennyson, J. & Yurchenko, S. N. ExoMolHR: a relational database of empirical high-resolution molecular spectra. *Astrophys. J. Suppl. Ser.* **276**, 67 (2025).
41. Parker, L. T. et al. Into the red: an M-band study of the chemistry and rotation of β Pictoris b at high spectral resolution. *Mon. Not. R. Astron. Soc.* **531**, 2356–2378 (2024).
42. Madhusudhan, N. et al. Carbon-bearing molecules in a possible Hycean atmosphere. *Astrophys. J. Lett.* **956**, 13 (2023).
43. Wogan, N. F. et al. JWST observations of K2-18b can be explained by a gas-rich mini-Neptune with no habitable surface. *Astrophys. J. Lett.* **963**, 7 (2024).
44. Holmberg, M. & Madhusudhan, N. Possible Hycean conditions in the sub-Neptune TOI-270 d. *Astron. Astrophys.* **683**, 2 (2024).
45. Carleo, I. et al. The GAPS programme at TNG XXXIX. Multiple molecular species in the atmosphere of the warm giant planet WASP-80 b unveiled at high resolution with GIANO-B*. *Astron. J.* **164**, 101 (2022).
46. Biassoni, F., Borsa, F., Haardt, F. & Rainer, M. High-resolution transmission spectroscopy of the hot-Saturn HD 149026b. *Astron. Astrophys.* **691**, 283 (2024).
47. Brogi, M., Line, M., Bean, J., Désert, J.-M. & Schwarz, H. A framework to combine low- and high-resolution spectroscopy for the atmospheres of transiting exoplanets. *Astrophys. J. Lett.* **839**, 2 (2017).
48. Miles, B. E. et al. The JWST Early Release Science Program for Direct Observations of Exoplanetary Systems II: a 1 to 20 μm spectrum of the planetary-mass companion VHS 1256-1257 b. *Astrophys. J. Lett.* **946**, 6 (2023).
49. Madhusudhan, N. C/O ratio as a dimension for characterizing exoplanetary atmospheres. *Astrophys. J.* **758**, 36 (2012).
50. Zhang, Y. et al. The ^{13}CO-rich atmosphere of a young accreting super-Jupiter. *Nature* **595**, 370–372 (2021).
51. Li, G. et al. Rovibrational line lists for nine isotopologues of the CO molecule in the $X^1\Sigma^+$ ground electronic state. *Astrophys. J. Suppl. Ser.* **216**, 15 (2015).
52. Birkby, J. L. et al. Detection of water absorption in the day side atmosphere of HD 189733 b using ground-based high-resolution spectroscopy at 3.2 μm. *Mon. Not. R. Astron. Soc.* **436**, 35–39 (2013).
53. Wakeford, H. R. et al. HST hot Jupiter transmission spectral survey: detection of water in HAT-P-1b from WFC3 near-IR spatial scan observations. *Mon. Not. R. Astron. Soc.* **435**, 3481–3493 (2013).
54. Rothman, L. S. et al. The *HITRAN* 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **110**, 533–572 (2009).
55. Rothman, L. S. et al. HITEMP, the high-temperature molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **111**, 2139–2150 (2010).
56. Polyansky, O. L. et al. ExoMol molecular line lists XXX: a complete high-accuracy line list for water. *Mon. Not. R. Astron. Soc.* **480**, 2597–2608 (2018).
57. Barber, R. J., Tennyson, J., Harris, G. J. & Tolchenov, R. N. A high accuracy computed water line list. *Mon. Not. R. Astron. Soc.* **368**, 1087–1094 (2006).

58. Melin, S. T., Sanders, S. T. & Nasir, E. F. Comparison of ExoMol simulated spectra for H_2O to high-temperature low-pressure gas cell measurements at 1723K in the 7321–7598 cm^{-1} range. *J. Quant. Spectrosc. Radiat. Transf.* **253**, 107079 (2020).
59. Gandhi, S. et al. Molecular cross-sections for high-resolution spectroscopy of super-Earths, warm Neptunes, and hot Jupiters. *Mon. Not. R. Astron. Soc.* **495**, 224–237 (2020).
60. Partridge, H. & Schwenke, D. W. The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data. *J. Chem. Phys.* **106**, 4618–4639 (1997).
61. Flowers, E., Brogi, M., Rauscher, E., Kempton, E. M.-R. & Chiavassa, A. The high-resolution transmission spectrum of HD 189733b interpreted with atmospheric Doppler shifts from three-dimensional general circulation models. *Astron. J.* **157**, 209 (2019).
62. Fortney, J. J., Lodders, K., Marley, M. S. & Freedman, R. S. A unified theory for the atmospheres of the hot and very hot Jupiters: two classes of irradiated atmospheres. *Astrophys. J.* **678**, 1419–1435 (2008).
63. Parmentier, Vivien et al. From thermal dissociation to condensation in the atmospheres of ultra hot Jupiters: WASP-121b in context. *Astron. Astrophys.* **617**, 110 (2018).
64. Burrows, A. & Sharp, C. M. Chemical equilibrium abundances in brown dwarf and extrasolar giant planet atmospheres. *Astrophys. J.* **512**, 843–863 (1999).
65. Sedaghati, E. et al. Detection of titanium oxide in the atmosphere of a hot Jupiter. *Nature* **549**, 238–241 (2017).
66. Nugroho, S. K. et al. High-resolution spectroscopic detection of TiO and a stratosphere in the day-side of WASP-33b. *Astrophys. J.* **154**, 221 (2017).
67. Schwenke, D. W. Opacity of TiO from a coupled electronic state calculation parametrized by ab initio and experimental data. *Faraday Discuss.* **109**, 321–334 (1998).
68. McKemmish, L. K. et al. ExoMol molecular line lists — XXXIII. The spectrum of titanium oxide. *Mon. Not. R. Astron. Soc.* **488**, 2836–2854 (2019).
69. Serindag, D. B. et al. Is TiO emission present in the ultra-hot Jupiter WASP-33b? A reassessment using the improved ExoMol TOTO line list. *Astron. Astrophys.* **645**, 90 (2021).
70. Prinoth, B. et al. Titanium oxide and chemical inhomogeneity in the atmosphere of the exoplanet WASP-189 b. *Nat. Astron.* **6**, 449–457 (2022).
71. Witsch, D. et al. The rotationally resolved infrared spectrum of TiO and its isotopologues. *J. Mol. Spectrosc.* **377**, 111439 (2021).
72. Cameron, W. D. & Bernath, P. Visible opacity of M dwarfs and hot Jupiters: the TiO $B^3\Pi$–$X^3\Delta$ band system. *Astrophys. J.* **926**, 39 (2022).
73. Bernath, P. & Cameron, D. Line lists for TiO minor isotopologues for the $A^3\Phi$–$X^3\Delta$ electronic transition. *J. Quant. Spectrosc. Radiat. Transf.* **310**, 108745 (2023).
74. McKemmish, L. K. et al. A hybrid approach to generating diatomic line lists for high resolution studies of exoplanets and other hot astronomical objects: updates to ExoMol MgO, VO and TiO line lists. *RAS Tech. Instrum.* **3**, 565–583 (2024).
75. Pelletier, S. et al. Vanadium oxide and a sharp onset of cold-trapping on a giant exoplanet. *Nature* **619**, 491–494 (2023).
76. McKemmish, L. K., Yurchenko, S. N. & Tennyson, J. ExoMol line lists — XVIII. The high-temperature spectrum of VO. *Mon. Not. R. Astron. Soc.* **463**, 771–793 (2016).
77. Bowesman, C. A., Akbari, H., Hopkins, S., Yurchenko, S. N. & Tennyson, J. Fine and hyperfine resolved empirical energy levels for VO. *J. Quant. Spectrosc. Radiat. Transf.* **289**, 108295 (2022).
78. Qu, Q., Yurchenko, S. N. & Tennyson, J. A method for the variational calculation of hyperfine-resolved rovibronic spectra of diatomic molecules. *J. Chem. Theory Comput.* **18**, 1808–1820 (2022).
79. Qu, Q., Yurchenko, S. N. & Tennyson, J. A variational model for the hyperfine resolved spectrum of VO in its ground electronic state. *J. Chem. Phys.* **157**, 124305 (2022).
80. Qu, Q., Yurchenko, S. N. & Tennyson, J. An empirical spectroscopic model for eleven electronic states of VO. *J. Mol. Spectrosc.* **391**, 111733 (2023).
81. Bowesman, C. A., Yurchenko, S. N. & Tennyson, J. A hyperfine-resolved spectroscopic model for vanadium monoxide ($^{51}V^{16}O$). *Mol. Phys.* **122**, 2255299 (2024).
82. Simonnin, A. et al. Time resolved absorption of six chemical species with MAROON-X points to strong drag in the ultra hot Jupiter TOI-1518 b. Preprint at https://arxiv.org/abs/2412.01472 (2024).
83. Maguire, C. et al. High resolution atmospheric retrievals of WASP-76b transmission spectroscopy with ESPRESSO: monitoring limb asymmetries across multiple transits. *Astron. Astrophys.* **687**, 49 (2024).
84. Nugroho, S. K. et al. First detection of hydroxyl radical emission from an exoplanet atmosphere: high-dispersion characterization of WASP-33b using Subaru/IRD. *Astrophys. J. Lett.* **910**, 9 (2021).
85. Brooke, J. S. A. et al. Line strengths of rovibrational and rotational transitions in the $X^2\Pi$ ground state of OH. *J. Quant. Spectrosc. Radiat. Transf.* **138**, 142–157 (2016).
86. Wright, S. O. M. et al. A spectroscopic thermometer: individual vibrational band spectroscopy with the example of OH in the atmosphere of WASP-33b. *Astron. J.* **166**, 41 (2023).
87. Mitev, G. B., Yurchenko, S. N. & Tennyson, J. Predissociation dynamics of the hydroxyl radical (OH) based on a five-state spectroscopic model. *J. Chem. Phys.* **160**, 144110 (2024).
88. Cont, D. et al. Exploring the ultra-hot Jupiter WASP-178b — constraints on atmospheric chemistry and dynamics from a joint retrieval of VLT/CRIRES+ and space photometric data. *Astron. Astrophys.* **688**, 206 (2024).
89. Mitev, G. B. et al. ExoMol photodissociation cross-sections – II. Continuum absorption and pre-dissociation spectra for the hydroxyl radical. *Mon. Not. R. Astron. Soc.* **539**, 3732–3740 (2025).
90. Bell, T. J. et al. Methane throughout the atmosphere of the warm exoplanet WASP-80b. *Nature* **623**, 709–712 (2023).
91. Benneke, B. et al. JWST reveals CH_4, CO_2, and H_2O in a metal-rich miscible atmosphere on a two-Earth-radius exoplanet. Preprint at https://doi.org/10.48550/arXiv.2403.03325 (2024).
92. Wiedemann, G., Deming, D. & Bjoraker, G. A sensitive search for methane in the infrared spectrum of τ Bootis. *Astrophys. J.* **546**, 1068–1074 (2001).
93. Gordon, I. E. et al. The *HITRAN* 2016 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **203**, 3–69 (2017).
94. Hargreaves, R. J. et al. An accurate, extensive, and practical line list of methane for the HITEMP database. *Astrophys. J. Suppl. Ser.* **247**, 55 (2020).
95. Guilluy, G. et al. The GAPS Programme at TNG. XXXVIII. Five molecules in the atmosphere of the warm giant planet WASP-69b detected at high spectral resolution. *Astron. Astrophys.* **665**, 104 (2022).
96. Rey, M., Nikitin, A. V. & Tyuterev, V. G. Theoretical hot methane line lists up to T=2000K for astrophysical applications. *Astrophys. J.* **789**, 2 (2014).
97. Rey, M., Nikitin, A. V. & Tyuterev, V. G. Accurate theoretical methane line lists in the infrared up to 3000K and quasi-continuum absorption/emission modeling for astrophysical applications. *Astrophys. J.* **847**, 105 (2017).
98. Yurchenko, S. N. & Tennyson, J. ExoMol line lists IV: the rotation–vibration spectrum of methane up to 1500K. *Mon. Not. R. Astron. Soc.* **440**, 1649–1661 (2014).
99. Yurchenko, S. N., Owens, A., Kefala, K. & Tennyson, J. ExoMol line lists — LVII. High accuracy ro-vibrational line list for methane (CH_4). *Mon. Not. R. Astron. Soc.* **528**, 3719–3729 (2024).
100. Kefala, K., Boudon, V., Yurchenko, S. N. & Tennyson, J. Empirical rovibrational energy levels for methane. *J. Quant. Spectrosc. Radiat. Transf.* **316**, 108897 (2024).
101. Madhusudhan, N. & Seager, S. A temperature and abundance retrieval method for exoplanet atmospheres. *Astrophys. J.* **707**, 24–39 (2009).
102. Pelletier, S. et al. Where is the water? Jupiter-like C/H ratio but strong H_2O depletion found on τ Boötis b using SPIRou. *Astron. J.* **162**, 73 (2021).
103. Line, M. R. et al. A solar C/O and sub-solar metallicity in a hot Jupiter atmosphere. *Nature* **598**, 580 (2021).
104. Gandhi, S. et al. Retrieval survey of metals in six ultrahot Jupiters: trends in chemistry, rain-out, ionization, and atmospheric dynamics. *Astron. J.* **165**, 242 (2023).
105. Kirk, J. et al. BOWIE-ALIGN: JWST reveals hints of planetesimal accretion and complex sulphur chemistry in the atmosphere of the misaligned hot Jupiter WASP-15b. *Mon. Not. R. Astron. Soc.* **537**, 3027–3052 (2025).
106. Owens, A., Yurchenko, S. N. & Tennyson, J. ExoMol line lists — LVIII. High-temperature molecular line list of carbonyl sulphide (OCS). *Mon. Not. R. Astron. Soc.* **530**, 4004–4015 (2024).
107. Underwood, D. S. et al. ExoMol line lists XIV: a line list for hot SO_2. *Mon. Not. R. Astron. Soc.* **459**, 3890–3899 (2016).
108. Azzam, A. A. A., Yurchenko, S. N., Tennyson, J. & Naumenko, O. V. ExoMol line lists — XVI. The rotation–vibration spectrum of hot H_2S. *Mon. Not. R. Astron. Soc.* **460**, 4063–4074 (2016).
109. Gorman, M. N., Yurchenko, S. N. & Tennyson, J. ExoMol molecular line lists — XXXVI. $X^2\Pi$–$X^2\Pi$ and $A^2\Sigma^+$–$X^2\Pi$ transitions of SH. *Mon. Not. R. Astron. Soc.* **490**, 1652–1665 (2019).
110. Brady, R. P., Yurchenko, S. N., Tennyson, J. & Kim, G.-S. ExoMol line lists — LVI. The SO line list, MARVEL analysis of experimental transition data and refinement of the spectroscopic model. *Mon. Not. R. Astron. Soc.* **527**, 6675–6690 (2024).
111. Zilinskas, M. et al. Observability of evaporating lava worlds. *Astron. Astrophys.* **661**, 126 (2022).
112. Sakellaris, C. N., Miliordos, E. & Mavridis, A. First principles study of the ground and excited states of FeO, FeO^+, and FeO^-. *J. Chem. Phys.* **134**, 234308 (2011).
113. Cheung, A. S.-C., Lyyra, A. M., Merer, A. J. & Taylor, A. W. Laser spectroscopy of FeO: rotational analysis of some subbands of the orange system. *J. Mol. Spectrosc.* **102**, 224–257 (1983).
114. Allen, M. D., Ziurys, L. M. & Brown, J. M. The millimeter-wave spectrum of FeO in its $X^5\Delta_i$ state (v=0): a study of all five spin components. *Chem. Phys. Lett.* **257**, 130–136 (1996).
115. Owens, A., Conway, E. K., Tennyson, J. & Yurchenko, S. N. ExoMol linelists — XXXVIII. High-temperature molecular line list of silicon dioxide (SiO_2). *Mon. Not. R. Astron. Soc.* **495**, 1927–1933 (2020).
116. Cheverall, C. J., Madhusudhan, N. & Holmberg, M. Robustness measures for molecular detections using high-resolution transmission spectroscopy of exoplanets. *Mon. Not. R. Astron. Soc.* **522**, 661–677 (2023).
117. Grimm, S. L. et al. HELIOS-K 2.0 and an open-source opacity database for exoplanetary atmospheres. *Astrophys. J. Suppl. Ser.* **253**, 30 (2021).
118. Chubb, K. L. et al. The ExoMolOP database: cross-sections and K-tables for molecules of interest in high-temperature exoplanet atmospheres. *Astron. Astrophys.* **646**, 21 (2021).
119. Wallace, L., Livingston, W., Hinkle, K. & Bernath, P. Infrared spectral atlases of the Sun from NOAO. *Astrophys. J. Suppl. Ser.* **106**, 165–169 (1996).
120. Berné, O. et al. Formation of the methyl cation by photochemistry in a protoplanetary disk. *Nature* **621**, 56–59 (2023).
121. Ariyarathna, I. R., Leiding, J. A., Neukirch, A. J. & Zammit, M. C. Ground and excited electronic structure analysis of FeH with correlated wave function theory and density functional approximations. *J. Phys. Chem. A* **128**, 9412–9425 (2024).
122. Schwarz, H. et al. The slow spin of the young substellar companion GQ Lupi b and its orbital configuration. *Astron. Astrophys.* **593**, 74 (2016).

123. Sánchez-López, A. et al. Water vapor detection in the transmission spectra of HD 209458 b with the CARMENES NIR channel. *Astron. Astrophys.* **630**, 53 (2019).
124. Guilluy, G. et al. Exoplanet atmospheres with GIANO — II. Detection of molecular absorption in the dayside spectrum of HD 102195b. *Astron. Astrophys.* **625**, 107 (2019).
125. Hawker, G. A., Madhusudhan, N., Cabot, S. H. C. & Gandhi, S. Evidence for multiple molecular species in the hot Jupiter HD 209458b. *Astrophys. J. Lett.* **863**, 11 (2018).
126. Barber, R. J. et al. ExoMol line lists — III. An improved hot rotation–vibration line list for HCN and HNC. *Mon. Not. R. Astron. Soc.* **437**, 1828–1835 (2014).
127. Cabot, S. H. C., Madhusudhan, N., Hawker, G. A. & Gandhi, S. On the robustness of analysis techniques for molecular detections using high-resolution exoplanet spectroscopy. *Mon. Not. R. Astron. Soc.* **482**, 4422–4436 (2019).
128. Coles, P. A., Yurchenko, S. N. & Tennyson, J. ExoMol molecular line lists XXXV. A rotation–vibration line list for hot ammonia. *Mon. Not. R. Astron. Soc.* **490**, 4638–4647 (2019).
129. Chubb, K. L., Tennyson, J. & Yurchenko, S. N. ExoMol molecular linelists — XXXVII. Spectra of acetylene. *Mon. Not. R. Astron. Soc.* **493**, 1531–1545 (2020).
130. Flagg, L. et al. ExoGemS detection of a metal hydride in an exoplanet atmosphere at high spectral resolution. *Astrophys. J. Lett.* **953**, 19 (2023).
131. Burrows, A., Ram, R. S., Bernath, P., Sharp, C. M. & Milsom, J. A. New CrH opacities for the study of L and brown dwarf atmospheres. *Astrophys. J.* **577**, 986–992 (2002).
132. Bernath, P. F. MoLLIST: molecular line lists, intensities and spectra. *J. Quant. Spectrosc. Radiat. Transf.* **240**, 106687 (2020).
133. Landman, R. et al. Detection of OH in the ultra-hot Jupiter WASP-76b. *Astron. Astrophys.* **656**, 119 (2021).
134. Smith, P. C. B. et al. A combined ground-based and JWST atmospheric retrieval analysis: both IGRINS and NIRSpec agree that the atmosphere of WASP-77A b is metal-poor. *Astron. J.* **167**, 110 (2024).
135. Gordon, I. E. et al. The HITRAN2020 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **277**, 107949 (2022).
136. Hargreaves, R. J. et al. Spectroscopic line parameters of NO, NO_2, and N_2O for the HITEMP database. *J. Quant. Spectrosc. Radiat. Transf.* **232**, 35–53 (2019).
137. Jacquinet-Husson, N. et al. The 2015 edition of the GEISA spectroscopic database. *J. Mol. Spectrosc.* **327**, 31–72 (2016).
138. Kurucz, R. L. Including all the lines. *Can. J. Phys.* **89**, 417–428 (2011).
139. Ryabchikova, T. et al. A major upgrade of the VALD database. *Phys. Scr.* **90**, 054005 (2015).
140. Rey, M., Nikitin, A. V., Babikov, Y. L. & Tyuterev, V. G. TheoReTS — an information system for theoretical spectra based on variational predictions from molecular potential energy and dipole moment surfaces. *J. Mol. Spectrosc.* **327**, 138–158 (2016).
141. Huang, X., Freedman, R. S., Tashkun, S., Schwenke, D. W. & Lee, T. J. AI-3000K infrared line list for hot CO_2. *J. Mol. Spectrosc.* **392**, 111748 (2023).
142. Richard, C., Boudon, V. & Rotger, M. Calculated spectroscopic databases for the VAMDC portal: new molecules and improvements. *J. Quant. Spectrosc. Radiat. Transf.* **251**, 107096 (2020).
143. Endres, C. P., Schlemmer, S., Schilke, P., Stutzki, J. & Müller, H. S. P. The Cologne Database for Molecular Spectroscopy, CDMS, in the Virtual Atomic and Molecular Data Centre, VAMDC. *J. Mol. Spectrosc.* **327**, 95–104 (2016).
144. Pickett, H. M. et al. Submillimeter, millimeter, and microwave spectral line catalog. *J. Quant. Spectrosc. Radiat. Transf.* **60**, 883–890 (1998).
145. Venot, O. et al. A chemical model for the atmosphere of hot Jupiters. *Astron. Astrophys.* **546**, 43 (2012).
146. Moses, J. I. Chemical kinetics on extrasolar planets. *Philos. Trans. A Math. Phys. Eng. Sci.* **372**, 20130073 (2014).
147. Shabram, M., Fortney, J. J., Greene, T. P. & Freedman, R. S. Transmission spectra of transiting planet atmospheres: model validation and simulations of the hot Neptune GJ 436b for the James Webb Space Telescope. *Astrophys. J.* **727**, 65 (2011).
148. Yurchenko, S. N. et al. ExoMol linelists — XXIV: a new hot line list for silicon monohydride, SiH. *Mon. Not. R. Astron. Soc.* **473**, 5324–5333 (2018).
149. Irwin, P. G. J. et al. Analysis of gaseous ammonia (NH_3) absorption in the visible spectrum of Jupiter — update. *Icarus* **321**, 572–582 (2019).
150. Barton, E. J., Yurchenko, S. N., Tennyson, J., Béguier, S. & Campargue, A. A near infrared line list for NH_3: analysis of a Kitt Peak spectrum after 35 years. *J. Mol. Spectrosc.* **325**, 7–12 (2016).
151. Basilicata, M. et al. The GAPS programme at TNG. *Astron. Astrophys.* **686**, A127 (2024).
152. Xue, Q. et al. JWST transmission spectroscopy of HD 209458b: a supersolar metallicity, a very low C/O, and no evidence of CH_4, HCN, or C_2H_2. *Astrophys. J. Lett.* **963**, 5 (2024).
153. Yurchenko, S. N., Mellor, T. M., Freedman, R. S. & Tennyson, J. ExoMol line lists XXXIX. Ro-vibrational molecular line list for CO_2. *Mon. Not. R. Astron. Soc.* **496**, 5282–5291 (2020).
154. Schaefer, L. & Fegley, B. Chemistry of silicate atmospheres of evaporating super-Earths. *Astrophys. J.* **703**, 113–117 (2009).
155. Kite, E. S., Fegley, B., Schaefer, L. & Gaidos, E. Atmosphere–interior exchange on hot, rocky exoplanets. *Astrophys. J.* **828**, 80 (2016).
156. Kesseli, A. Y., Snellen, I. A. G., Alonso-Floriano, F. J., Mollière, P. & Serindag, D. B. A search for FeH in hot-Jupiter atmospheres with high-dispersion spectroscopy. *Astron. J.* **160**, 228 (2020).
157. Pavlenko, Y. V. Theoretical modelling of optical and IR spectra of brown dwarfs and ultracool dwarfs. *Astron. Nachr.* **326**, 934–939 (2005).
158. Reiners, A., Homeier, D., Hauschildt, P. H. & Allard, F. A high resolution spectral atlas of brown dwarfs. *Astron. Astrophys.* **473**, 245–255 (2007).
159. Crozet, P. et al. Correlations between laboratory line lists for FeH, CrH, and NiH and M-star spectra collected with ESPaDOnS and SPIRou. *Astron. Astrophys.* **679**, 116 (2023).
160. Lyulin, O., Vasilchenko, S., Mondelain, D. & Campargue, A. The CRDS spectrum of acetylene near 1.73 µm. *J. Quant. Spectrosc. Radiat. Transf.* **234**, 147–158 (2019).
161. Jacquemart, D., Soulard, P. & Lyulin, O. M. Recommended acetylene $^{12}C_2H_2$ line list in 13.6 µm spectral region: new measurements and global modeling. *J. Quant. Spectrosc. Radiat. Transf.* **256**, 107200 (2020).
162. Lyulin, O. M., Vasilchenko, S. S. & Perevalov, V. I. High sensitivity absorption spectroscopy of acetylene near 770 nm. *J. Quant. Spectrosc. Radiat. Transf.* **294**, 108402 (2023).
163. Currie, T. et al. Direct imaging and spectroscopy of extrasolar planets. Preprint at https://arxiv.org/abs/2205.05696 (2023).

Acknowledgements

The idea for this Technical Review arose from the Royal Society Discussion Meeting on Exoplanet Spectroscopy at High Resolution held near Northampton, UK, in 2023. We thank the other attendees at this meeting for conversations and the Royal Society for funding under the Theo Murphy meeting programme. The work of J.T. and S.Y. and the ExoMol project received support from the European Research Council under the European Union's Horizon 2020 research and innovation programme through Advance Grant numbers 267219 and 883830. S.Y. also acknowledges Science and Technology Facilities Council (STFC) Project No. ST/Y001508/1 and the use of the DiRAC HPC services at Cambridge and Leicester funded by BEIS, UKRI and STFC capital funding and STFC operations grants.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42254-025-00839-z.

Peer review information *Nature Reviews Physics* thanks Stefanie Milam and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.