Atmospheric regimes and trends on exoplanets and brown dwarfs

Xi Zhang (张曦)

Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, CA 95064, USA; xiz@ucsc.edu

Received 2020 April 14; accepted 2020 June 11

Abstract A planetary atmosphere is the outer gas layer of a planet. Besides its scientific significance among the first and most accessible planetary layers observed from space, it is closely connected with planetary formation and evolution, surface and interior processes, and habitability of planets. Current theories of planetary atmospheres were primarily obtained through the studies of eight large planets, Pluto and three large moons (Io, Titan, and Triton) in the Solar System. Outside the Solar System, more than four thousand extrasolar planets (exoplanets) and two thousand brown dwarfs have been confirmed in our Galaxy, and their population is rapidly growing. The rich information from these exotic bodies offers a database to test, in a statistical sense, the fundamental theories of planetary climates. Here we review the current knowledge on atmospheres of exoplanets and brown dwarfs from recent observations and theories. This review highlights important regimes and statistical trends in an ensemble of atmospheres as an initial step towards fully characterizing diverse substellar atmospheres, that illustrates the underlying principles and critical problems. Insights are obtained through analysis of the dependence of atmospheric characteristics on basic planetary parameters. Dominant processes that influence atmospheric stability, energy transport, temperature, composition and flow pattern are discussed and elaborated with simple scaling laws. We dedicate this review to Dr. Adam P. Showman (1968–2020) in recognition of his fundamental contribution to the understanding of atmospheric dynamics on giant planets, exoplanets and brown dwarfs.

Key words: planets and satellites: atmospheres — planets and satellites: gaseous planets — planets and satellites: terrestrial planets — planets and satellites: physical evolution — stars: brown dwarfs

1 INTRODUCTION

The discovery of planets outside the Solar System has greatly expanded the horizon in planetary science since Mayor & Queloz (1995) discovered the first exoplanet around a Sun-like star (51 Pegasi b). Among more than 4200 exoplanets confirmed to date, the majority of them are larger than our Earth (e.g., Borucki et al. 2011a,b; Howard et al. 2012; Dressing & Charbonneau 2013; Fressin et al. 2013; Petigura et al. 2013). Statistical analysis of the current samples and structural models (e.g., Weiss & Marcy 2014; Lopez & Fortney 2014; Rogers 2015) found that planets with radii larger than about 1.5–1.7 Earth radii are likely to have thick gaseous envelopes made of hydrogen, helium and hydrogen compounds, while the smaller planets are mostly composed of iron and silicates like the Earth. Although the larger planets should hold thick atmospheres, we cannot rule out the possibility of atmospheric existence on smaller planets because Io (~30% of the Earth radius) and Pluto (~20% of the Earth radius) in our Solar System still possess thin atmospheres. To date, most atmospheres that have been characterized are on planets close to the central stars via transit observations or on young planets via the direct imaging technique, meaning that most of these planets are much hotter than the Solar System planets. The current samples of exoplanet atmospheres are considered to be a high-temperature extension of the planetary atmospheres in the Solar System.

In the same year as the discovery of 51 Pegasi b, the first two brown dwarfs, which can be roughly defined as astronomical bodies with a mass between 13 and 80 Jupiter masses, were also confirmed (Gliese 229B in Nakajima et al. 1995 and Teide 1 in Rebolo et al. 1996). The currently known ~2000 brown dwarfs are mostly free-floating, but some are also orbiting star companions. Traditionally, brown dwarfs are considered to be the low-mass branch below M dwarf stars in the substellar main
sequence: L, T and Y sub-categories following the decreasing order of the effective temperature. Nevertheless, to some extent, brown dwarfs behave more like planets. Because their masses are below the hydrogen-burning limit but above the deuterium-burning limit, a brown dwarf’s interior is in the degenerate state, similar to that of gas giant planets, and their radii are around one Jupiter radius. Unlike ordinary stars in which trace elements (other than hydrogen and helium) are mainly in the atomic form, the colder photospheres of brown dwarfs are mostly composed of molecules such as H₂, He, H₂O, CO and CH₄. The emission temperature of the coldest brown dwarf detected so far (WISE 0855) is even below the freezing point of water, and water ice clouds could form there as on Jupiter (Luhman & Esplin 2014; Skemer et al. 2016; Esplin et al. 2016). Thus current samples of brown dwarf atmospheres can be categorized as a high-gravity and high-temperature extension of the hydrogen-dominated giant planetary atmospheres in the Solar System.

To date, observations of the atmospheres of exoplanets and brown dwarfs (hereafter collectively “substellar atmospheres”) mainly focus on transmission, reflection and emission spectra; phase curves as the planets are circling the stars; rotational light curves as the planets spin; and Doppler-shifted atomic or molecular lines by the atmospheric winds and orbital motion. From these data, we can retrieve the distributions of temperature, atmospheric compositions and abundances of gases and clouds, and wind patterns in the atmospheres. Based on the NASA Exoplanet Archive, to date, almost a hundred exoplanets have (some sort of) atmospheric detection (Madhusudhan 2019), among which we have obtained ~50 transmission spectra and ~30 emission spectra. The data quality is, however, not always satisfactory. Arguably, the quality of the observational data of exoplanetary atmospheres is about 40 yr behind its counterpart in the Solar System. For example, spectral coverage, spectral resolution, and noise levels of the dayside emission spectra of a canonical hot Jupiter HD 189733 b from the Spitzer telescope (Grillmair et al. 2008) and Hubble Space Telescope (HST, Swain et al. 2009) are comparable to that of Jupiter spectra in the early era by Gillett et al. (1969). The typical resolving power of exoplanet observation is λ/Δλ ~10 to 100 where λ is the wavelength (Konopacky et al. 2013). Higher-spectral-resolution (λ/Δλ ~10,000 to 100,000) spectra in visible and near-infrared (near-IR) wavelengths from ground-based transmission observations can be achieved to determine the atmospheric composition and even the wind speed using cross-correlation and line-shape analysis techniques (e.g., Snellen et al. 2010; Crossfield et al. 2011; Konopacky et al. 2013). However, it seems more difficult to estimate the uncertainties of ground-based data than those of space-based data (Kreidberg 2018). We expect a great leap in the spectral data quality with future large telescopes such as the James Webb Space Telescope (JWST, e.g., λ/Δλ ~100 to 1000 in space), Atmospheric Remote-sensing Infrared Exoplanet Large-survey (ARIEL), European Extremely Large Telescope (ELT), Giant Magellan Telescope (GMT) and Thirty Meter Telescope (TMT). A spatially resolved image of an exoplanet is expected to remain difficult to achieve in the foreseeable future.

On the other hand, without photon contamination from the host star companions, spectral observations on high-temperature field brown dwarfs typically have much higher data quality. For example, the observational spectrum of a typical T4.5 dwarf 2MASS 0559-14 can achieve a spectral resolving power of λ/Δλ ~2000 and a signal-to-noise ratio (S/N)> 50. The “hot methane” lines in the near-IR can be spectrally resolved and led to the reevaluation of the existing opacity database (Yurchenko et al. 2014). These spectra allow a much better estimate of the atmospheric properties such as temperature and chemical compositions on both L and T dwarfs (e.g., Line et al. 2017). The closest brown dwarf detected so far (Luhman 16 B) can even be spatially resolved (in low resolution) using the Doppler imaging technique (Crossfield et al. 2014).

Despite the current data quality and future challenges in the observations of atmospheres, exoplanet observation has some advantages compared to that in the Solar System. For example, for planets around other stars, we can naturally detect the orbital light curve, i.e., the flux changes at different orbital phases, whereas this can only be done by an orbiting or flyby spacecraft for Solar System planets outside Earth’s orbit. As another example, the giant planets in our Solar System are so cold that water vapor is all trapped below the water cloud layer, and hence water abundances remain unknown. On the other hand, hot exoplanets show clear water vapor signals in the spectra, allowing a better derivation of the carbon-to-oxygen ratio, a crucial parameter for constraining planetary formation and evolution.

Characterizing planetary atmospheres and unveiling the principles underlying their diverse weather and climate—as we have learned from the Solar System studies—is challenging. Atmospheres are fundamentally complex with many interacting processes and a large number of free parameters. A big dataset with sufficient samples is required for comparative planetology to understand the role of each factor. Undoubtedly this dataset could only come from the atmospheres of exoplanets and brown dwarfs. In the dawn of the “third era” in planetary science (Ingersoll 2017), researchers have been astonished by the diversity of atmospheres outside the Solar System. These
substellar atmospheres have provided a wealth of information complementary to their counterparts in the Solar System (Pierrehumbert 2013).

Here are a few examples. Tidally locked planets are synchronized to the central stars with permanent day-side and night-side, a configuration we do not have in our system. Super-Earths and mini-Neptunes are planets with size between the Earth and Neptune. They are a new type of world that is not present in the Solar System but dominates the current, confirmed exoplanetary population. Hot Earth-sized planets are so close to their central stars that their surfaces might be melted or partially melted (e.g., 55 Cancri e or Kepler-10 b). They might be exciting analogs of Jupiter’s moon Io or the early Earth with atmosphere-magma interaction. Planets very close to the central stars are perfect samples to understand how atmospheres are evaporated or blown off. In contrast, worlds that are very far away from their central stars are useful to explore how atmospheres condense on the surfaces. For planets in the “habitable zone” where liquid water could exist on the surface, various climate states are possible, depending on parameters such as planetary rotation rate, central star type, atmospheric composition and orbital configuration. Furthermore, planets at different ages could tell us how planetary atmospheres and climates evolve with time and under different environments.

A number of excellent reviews of atmospheres of exoplanets and brown dwarfs have been published. Some articles generally cover a bit of every aspect (e.g., Seager & Deming 2010; Bailey 2014; Madhusudhan et al. 2014c; Fortney 2018; Madhusudhan 2019) but most of them focus on specific topics such as atmospheric observations (e.g., Tinetti et al. 2013; Burrows 2014; Encrenaz 2014; Pepe et al. 2014; Crossfield 2015; Deming & Seager 2016; Kreidberg 2018; Parmentier & Crossfield 2018; Sing 2018), atmospheric escape (e.g., Lammer et al. 2008; Tian 2015; Owen 2019), atmospheric radiation (e.g., Marley & Robinson 2015; Heng & Marley 2017), atmospheric chemistry (e.g., Lodders 2010; Marley et al. 2013; Moses 2014; Madhusudhan et al. 2016), atmospheric dynamics (e.g., Showman et al. 2010; Showman et al. 2013b; Heng & Showman 2015; Pierrehumbert & Hammond 2019; Showman et al. 2020), space weather (e.g., Airapetian et al. 2020), terrestrial climate (e.g., Forget & Leconte 2014), giant planets (e.g., Marley et al. 2007; Fletcher et al. 2014), brown dwarfs (e.g., Basri 2000; Kirkpatrick 2005; Helling & Fomins 2013; Helling & Casewell 2014; Artigau 2018; Biller 2017) and habitability (e.g., Kasting & Catling 2003; Madhusudhan et al. 2016; Shields et al. 2016a; Kopparapu et al. 2019). However, previous reviews focused less on statistical properties in the emergent ensemble of substellar atmospheres, motivating this article.

In this review, we consider these diverse atmospheres as a systematic test bed for our current understanding of planetary climates. We summarize the statistical “trends” discovered in recent years and discuss various aspects to classify the atmospheres into different climate “regimes.” To be specific, “regimes” and “trends” refer to the dependence of the atmospheric characteristics on the basic planetary parameters. Here “basic planetary parameters” refer to planetary parameters such as the mass, radius, age, gravity, self-rotation rate, escape velocity, semi-major axis, orbital period, eccentricity, obliquity, metallicity (including elemental ratios such as carbon-to-oxygen (C/O) ratio), surface albedo, internal heat flux (internal luminosity) and equilibrium temperature. They could also include host star parameters such as host star type, stellar luminosity, and stellar irradiation spectra. “Atmospheric characteristics” stand for the observed properties of substellar atmospheres such as directly measured broadband photometric fluxes and all kinds of spectral and polarization signatures. It also includes the derived atmospheric properties such as atmospheric existence, atmospheric pressure and mass, bulk luminosity (or effective temperature), albedo, distributions of temperature, gas and particle compositions, wind and waves, and the time variability of those properties from time-domain observations.

We are just beginning to discover and understand those trends and regimes. There are dangers with this approach because of the assumption, as pointed out by Stevenson (2004), that common processes are at work on Solar System planets, including the Earth, exoplanets and brown dwarfs but they yield different and diverse outcomes. If a single fundamental mechanism controls an observable across the sampled planets, we might observe a trend with a typical varying parameter. A typical example is the Hertzsprung-Russell diagram for stars. If a few fundamental mechanisms govern the observables, we might expect a regime shift from one dominant mechanism to another in the parameter space. However, if many factors could lead to similar, almost indistinguishable observable phenomena, the trends or regimes are washed out in a large sample. Given the current data quality for substellar atmospheres, the statistical significance of the trends and regimes in this review will be preliminary and somewhat debatable. However, from a theoretical perspective, this is also a good way to summarize our understanding of substellar atmospheres, highlight fundamental principles underlying essential processes and link back to our knowledge obtained from the Solar System. We will also try to outline some simple analytical scaling laws to help illuminate fundamental processes more intuitively.
This comprehensive review is organized as follows. First, we will make some general remarks on atmospheres. In Section 2, we start with the fundamental equations in planetary atmospheres and elaborate on vital processes and their complex interactions in Section 2.1. We then summarize the difference between the traditional “cold” planetary atmospheres in the Solar System and the currently characterized “hot” atmospheres on exoplanets and brown dwarfs in Section 2.2. Then we feature several important spectral and photometric observations to date for characterizing substellar bodies in Section 2.3. That will help lead into discussion on statistical trends and regimes, summarized in several sub-fields. In each sub-field section, we first introduce the fundamentals and then feature several important regimes and trends. In Section 3, we discuss atmospheric stability with a focus on the atmospheric escape from planets. We highlight the “cosmic shoreline” in Section 3.2 and “planet desert and radius gap” in Section 3.3 in recent observations and underlying mechanisms. In Section 4, we discuss the thermal structure and radiative energy transport, with an emphasis on the radiative-convective boundary (RCB), vertical temperature inversion and mid-infrared (mid-IR) brightness temperature trend on exoplanets in Section 4.2, thermal phase curves on tidally locked exoplanets in Section 4.3, and rotational light curves on brown dwarfs and directly imaged planets in Section 4.4. We talk about atmospheric composition and chemistry in Section 5. In Section 5.2, we discuss gas chemistry, including both thermochemistry and disequilibrium chemistry, followed by a review of hazes and clouds in Section 5.3. In Section 6, we concentrate on the atmospheric dynamics and important regimes classified using non-dimensional numbers. We describe three categories: highly irradiated planets such as the tidally locked planets in Section 6.2, convective atmospheres on directly imaged planets and brown dwarfs in Section 6.3. We only briefly review the terrestrial climates in the habitable zone in Section 6.4 because of its complexity and the lack of data to reveal detailed trends on extrasolar terrestrial atmospheres. We conclude this review with prospects in Section 7.

2 GENERAL REMARKS

2.1 Overview of Important Processes

Atmospheres in and out of the Solar System share similar fundamental physical and chemical processes that should be understood in a self-consistent mathematical framework. The fundamental equation set is composed of a continuity equation, a momentum equation, an energy equation, an equation of state, an equation of radiative transfer, and a series of transport equations for chemical species, including both gas and particles. Equation set (1) lists the governing equations for a three-dimensional (3D), collisional, neutral, inviscid, ideal-gas atmosphere with necessary assumptions.

\[
\begin{align*}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= F^p, \\
\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) + \nabla p + 2\Omega \times \mathbf{u} &+ \nabla \times (\mathbf{\Omega} \times \mathbf{r}) - \rho g = \mathbf{F}_u, \\
\frac{\partial}{\partial t}(\rho \mathbf{u}) &= + \nabla \cdot (\frac{\gamma p}{\gamma - 1} + \frac{1}{2} \rho \mathbf{u} \cdot \mathbf{u} + \rho \Phi) \\
&+ \nabla \cdot \left[\frac{(\gamma p + \frac{1}{2} \rho \mathbf{u} \cdot \mathbf{u} + \rho \Phi)}{\gamma - 1} - \rho \mathbf{g} \right] - K_T \nabla T \\
&+ Q(I_{\nu}) = F_e, \\
p &= \frac{\hbar k T}{m}, \\
\frac{dI_{\nu}}{d\tau_{\nu}} &= I_{\nu} - \omega_{\nu} I^*_{\nu} e^{-\tau_{\nu}} P_{\nu} - (1 - \omega_{\nu})J_{\nu} \\
&- \omega_{\nu} S_{\nu}(\tau_{\nu}, P_{\nu}, I_{\nu}) = 0, \\
\frac{\partial \rho \chi}{\partial t} + \nabla \cdot (\rho \chi \mathbf{u} + \rho D_{\chi} \nabla \left[\frac{\chi}{\chi_e}\right]) &= P_{\chi} - L_{\chi}.
\end{align*}
\]

Here bold represents vector form, \(t\), \(\rho\), \(p\) and \(T\) are time, density, pressure and temperature, respectively. \(\mathbf{u}\), \(\mathbf{r}\) and \(\mathbf{g}\) are the 3D velocity vector, rotational rate vector, radial vector and gravitational acceleration vector, respectively. \(\Phi\) is the gravitational potential energy by mass defined as \(\mathbf{g} = -\nabla \Phi\). \(K_T\) is thermal conductivity. \(k_B\) is the Boltzmann constant. \(m\) is the mean mass of an air molecule. \(\gamma\) is the adiabatic index, i.e., the ratio of the specific heats \(c_p/c_v\). \(Q\) is the radiative heating and cooling terms, \(F^p\), \(F_u\), \(F_e\) are the external forcing terms of density, momentum and energy, respectively. In the radiative transfer Equation (1e), \(\nu\) is the spectral grid (wavelength or frequency), \(\tau_{\nu}\) is slant optical depth, \(I_{\nu}\) is the spectral grid (wavelength or frequency), \(J_{\nu}\) is the self-emission source function, which is the Planck function under Local Thermodynamic Equilibrium (LTE). \(S_{\nu}\) is the scattering source function. \(\omega_{\nu}\) is the single scattering albedo, and \(P_{\nu}\) is the scattering phase function. In the chemical transport Equation (1f), \(\chi\) is the mass mixing ratio of a specific species (either gas or particle). \(D_{\chi}\) is the molecular diffusivity that relaxes the mass mixing ratio towards the equilibrium mass mixing ratio \(\chi_e\). Note that there is no eddy mixing term because the 3D advection term by \(\mathbf{u}\) includes the eddy transport. \(P\) and \(L\) are the chemical/ microphysical production and loss terms, respectively.

The continuity equation, Equation (1a), describes the bulk atmosphere as a compressible fluid and the external forcing term \(F^p\) includes a mass loss to space at the top of the atmosphere, mass exchange with the interior.
(such as volcanism), and surface/ocean and clouds through the condensation and evaporation. The momentum equation, Equation (1b), is a simplified form of the Navier-Stokes equation in fluid mechanics neglecting the molecular and dynamic viscous terms. The external forces include the pressure gradient, Coriolis force, centrifugal force and gravitational force. The latter could spatially vary due to the oblateness of the body. Other external forces in the $F_u$ term include the drag force from surface friction, magnetic interaction, momentum gain or loss due to the mass gain or loss, phase change and gravitational particle settling. The energy equation, Equation (1c), describes the evolution of the atmospheric energy flux, including internal energy, kinetic energy (KE) and gravitational potential energy. The $Q$ term represents diabatic heating and cooling from atmospheric radiation. Thermal conduction via collisions is described in the $KVT$ term. The other energy forcing term $F_e$ includes latent heat and energy exchange during the phase transition, such as cloud formation, Ohmic heating through interaction with the magnetic field, viscous heating due to frictional drag and even the chemical potential energy change during chemical reactions. The equation of state (1d) of the atmosphere approximately follows the ideal gas law, which is valid in most mostospheres. The equation of state needs to be treated carefully in a multi-component atmosphere, especially where clouds form (Li & Chen 2019).

The radiative transfer equation (Eq. (1e)) solves the photon intensity distribution in the atmosphere at each wavelength and angle. The radiative flux divergence is used in the energy equation (the $Q$ term in Eq. (1c)). Also, the actinic flux derived from the intensity is applied to the photochemical calculations. Multiple scattering from the gas (Rayleigh) and particles needs to be considered. Chemical transport Equation (1f) includes advection and molecular diffusion of the chemical species and the chemical production and loss terms. The production and loss come from gas chemistry such as photochemistry, neutral chemistry, ion chemistry and particle microphysics in the haze and cloud formation such as nucleation, coagulation, and condensation aggregation and coalescence processes. The chemical equations are coupled together by the chemical reaction network. Usually, the continuity Equation (1a) would not be altered by the mass-conserved gas chemistry, except that the gas density could change in the condensation and evaporation processes. Note that the chemical reactions do not conserve the total number of molecules. Therefore, the mean molecular mass ($m$ in the equation of state (1d)) could be altered in the chemical and microphysical processes.

In the system described by the equation set (1), the total momentum, mass and energy of the atmosphere do not have to be conserved with time. They depend on the boundary conditions (e.g., whether the atmosphere is escaping to space or condensing at the surface) and internal processes (e.g., cloud formation converting vapor to particles). In most cases, we assume the observed planetary atmospheres have reached a steady state with internal oscillations. In this situation, solving the statistically averaged climate state is a boundary value problem, although setting an appropriate boundary condition is not trivial. In the case of short perturbations, such as comet Shoemaker-Levy 9 impacting Jupiter’s atmosphere in 1994, the giant storm in Saturn’s atmosphere in 2011, dust storm evolution on Mars or climate change in the atmosphere of modern Earth, the above equations could be solved as an initial value problem to understand the evolution of the atmosphere under perturbations.

In this “minimum recipe,” there are several unknown parameters: temperature, pressure, density, wind velocity vector, light intensity (and associated radiative heating and cooling rate and actinic flux), and abundances of chemical tracers including dust, haze and cloud particles. Complexity emerges because of the coupling of parameters and interaction among processes, leading to a high nonlinearity in this system. Realistic atmospheres could only be much more complicated. For example, the equation set (1) does not explicitly include the magnetic field, which becomes important when the atmosphere is so hot that it could be partially ionized. In the high-temperature regime, magnetic field might play a significant role (e.g., Batygin & Stanley 2014; Rogers 2017). Once the magnetic field is coupled with the atmospheric flow, magnetohydrodynamics (MHD) becomes complicated, especially if there is ion chemistry. Maxwell’s equations will need to be solved. We also did not include the collisionless region in the upper atmosphere where the atoms and molecules escape from the planet. In that case, the Boltzmann equation needs to be solved. Interaction between the stellar wind and the atmosphere is also complicated. The near-surface (boundary layer) physics that describes how the lower atmosphere interacts with the surface is not detailed. If one is interested in the deep atmosphere which does not obey the ideal gas law, different equations of state also need to be adopted in the high pressure and high-temperature regime although the available data are sparse.

The climate system contains a wide range of length scales and timescales. Take Earth’s atmosphere as an example. The length scale spans from interactions between electromagnetic waves and atoms/molecules at atomic/molecular scale ($\sim 10^{-10}$ m), to aerosol and cloud microphysics ($10^{-8} - 10^{-3}$ m with particle size), to regional turbulence ($10^{-2} - 10$ m), to convective systems ($10^3$ m for tornados to $10^8$ m for hurricanes), to synoptic weather systems (e.g.,
Tackling all of these length scales and timescales together is impossible, and often investigations need to be simplified and isolated. Also, breaking the system down to many scales with various levels of complexities is the pathway for not only making models or theories viable but to guarantee understanding. Based on the “minimum recipe” equation set (1) and using common simplifications such as hydrostatic balance, large aspect ratio and small density variation, one can formulate simpler equations to describe the behavior of the atmosphere. Some famous forms include the quasi-geostrophic equations, shallow water equations, primitive equations, Boussinesq equations and anelastic equations. See textbooks such as Vallis (2006), Pedlosky (2013) and Holton (2016) for details.

2.2 Cold versus Hot Regimes

To first order, we highlight “cold” versus “hot” atmospheres enlightened by the emerging ensemble of exoplanets and brown dwarfs across a broad range of temperatures. The regime boundary between cold and hot is vaguely defined as the temperature for water vapor-liquid phase transition at 1 bar (∼373 K). Traditional studies on planetary atmospheres in the Solar System mostly focus on the “cold regime” except a few studies such as on the lower atmosphere of Venus and deep atmospheres of giant planets. On the other hand, most characterized exoplanets and brown dwarfs to date would fall in the “hot regime.” Examples include ultra-hot Jupiters with equilibrium temperatures higher than 2200 K, including WASP-121 b, WASP-12 b, WASP-103 b, WASP-33 b, Kepler-13A b, WASP-18 b and HAT-P-7 b (see spectra compiled in Parmentier et al. 2018), and scorching ones KELT-1 b and KELT-9 b. The latter is the hottest known exoplanet to date with a dayside temperature of ∼4600 K (Gaudi et al. 2017). In the past two decades, observations on substellar atmospheres gave birth to a new sub-field in atmospheric science to study “high-temperature atmospheres.” Conventional theories of cold atmospheres in the Solar System might have neglected critical processes in hot substellar atmospheres. In Table 1, we highlight several possible essential differences in the physical and chemical processes between the two regimes.

Compared with low-temperature atmospheres, high-temperature atmospheres become more active so that processes generally operate faster. In atmospheric radiation, the electron states in the atoms and molecules are easier to be excited at a higher temperature. Numerous weak energy transitions in the molecular electronic, vibrational and rotational states—usually negligible in the low-temperature regime—have become significantly stronger to increase the opacity of the atmosphere. The population of quantum states is prone to deviate from the Boltzmann distribution under high temperature, leading to Non-Local Thermodynamic Equilibrium (non-LTE) effect where the gas emission does not obey the simple Planck law anymore. Instead, a complicated vibrational state “chemistry” impacts the atmospheric absorption and emission properties. Third, the radiative timescale is shorter at a higher temperature, implying a faster dissipation rate for the atmospheric heat.

Given sufficient time, chemical reactions proceed in both the forward and backward directions towards thermodynamical equilibrium—the minimum Gibbs free energy state. In reality, because the reaction rates of the forward and backward reactions usually have different temperature dependence, they typically proceed at different speeds. In the cold regime, one direction (namely the “forward reaction”) will proceed much faster than the other direction. Other fast atmospheric processes, such as wind transport, if more rapid than the backward reaction, lead to chemical disequilibrium. In the high-temperature regime, both the forward and backward reactions speed up, and species more easily reach thermochemical equilibrium. Nevertheless, disequilibrium chemistry is still essential because wind transport might also become more potent at a higher temperature. Chemical models seem to support that colder atmospheres show more substantial signs of disequilibrium than hotter atmospheres, but more observations are needed to confirm this hypothesis (e.g., Line & Yung 2013).

Furthermore, in the cold regime, chemistry in the infrared (IR) emission layers is mostly neutral chemistry among molecules. Photoionization and thermal ionization could only be important in the upper thermosphere and the auroral region. In the high-temperature atmo-
sphere, ionization more readily occurs, and atomic neutrals and ions are common (e.g., Lavvas et al. 2014). For example, to date about 15 atomic species have been detected in the atmosphere of a very hot Jupiter KELT-9 b (e.g., Yan & Henning 2018; Cauley et al. 2019b; Jens Hoeijmakers et al. 2018; Hoeijmakers et al. 2019; Pino et al. 2020; Wyttenbach et al. 2020). Photoionization could also be important in the photosphere if the planet is very close to the central star. In these situations, ion-chemistry is also important to understand the observed spectra.

Particles in the atmosphere could also be drastically different between low and high-temperature regimes. The clouds on Solar System planets are mostly molecular solids/liquids maintained by intermolecular forces, such as sulfuric acid on Venus, water on Earth and Mars, CO$_2$ on Mars, ammonia and water on giant planets, methane and hydrogen sulfide on Uranus and Neptune, methane, ethane and hydrogen cyanide on Titan, as well as nitrogen on Pluto and Triton. Observational spectra of hot atmospheres also imply the existence of particles. However, all the above cloud species in the Solar System will remain in the vapor phase in the hot atmospheres. Instead, we expect different compounds with much higher melting temperature, for example, refractive solids maintained by network covalent bonds (e.g., silicate), metallic bonds (e.g., iron) or ionic bonds (e.g., KCl). The formation mechanisms of those mineral and iron clouds in the hot substellar atmospheres are not well understood (see discussion in Sect. 5.3). Atmospheric chemistry will also form organic haze particles such as on Earth, Titan, Pluto, Triton and giant planets. Experiments have shown that organic hazes are able to form in various environment from 300–1500 K (Hörm et al. 2018; He et al. 2018a,b, 2020; Fleury et al. 2019; Moran et al. 2020). Whether the detected particles in the hot substellar atmospheres are organic haze particles or condensational dust clouds is still an open question.

In atmospheric dynamics, a higher-temperature atmosphere usually has a faster speed of a sound wave and other waves, and perhaps higher wind speed too, depending on the rotation and other parameters. While moist convection ubiquitously exists in thick planetary atmospheres in the Solar System due to latent heat release from the cloud condensation, it is less important in the high-temperature regime than dry convection flux (Sect. 6.3). Moreover, because of partial ionization of the atmosphere, the magnetic field will be more easily coupled with the atmosphere. It exerts significant MHD drag on the atmospheric flow or causes significant Ohmic (or Joule) heating. Strong magnetic fields (~20–120 G) on several hot Jupiters have recently been inferred through magnetic star-planet interactions (Cauley et al. 2019b).

Atmospheric loss mechanisms could also be different between the two regimes. In terms of atmospheric escape, most planetary atmospheres in the Solar System are close to the hydrostatic state with a moderate or weak Jeans escape. Hydrodynamic escape (i.e., atmospheric blow-off) could also occur in some cases such as the solar wind and Earth’s polar wind. Pluto’s atmosphere was regarded as a good candidate for ongoing hydrodynamic escape (e.g., Zhu et al. 2014). However, the New Horizons flyby in 2015 discovered a much colder atmosphere on Pluto, and thus the atmospheric loss rate is much smaller (Gladstone et al. 2016; Zhang et al. 2017). On the other hand, for a hot atmosphere close to the central star, the strong stellar flux, and X-ray or ultraviolet (UV) heating in the upper atmosphere could lead to outward hydrodynamic blow-off like a planetary wind. This flow has been detected in recent observations of some exoplanets (e.g., GJ 436 b, Ehrenreich et al. 2015; Lavie et al. 2017b. See Sect. 3.1). The atmosphere can also be lost to the surface or interior. In a very cold atmosphere, the bulk atmospheric component, such as CO$_2$ and N$_2$, could condense onto the surface or even lead to total atmospheric collapse (e.g., Mars, Io, Pluto and Triton). The condensation does not readily occur in hot atmospheres. One exception could be very hot rocky planets tidally locked to central stars. The bulk composition on the dayside might be enriched in silicate vapor such as on 51 Cancri e (e.g., Demory et al. 2016a), Kepler 1520 b (Rappaport et al. 2012; Perez-Becker & Chiang 2013) and K2-22 b (Sanchis-Ojeda et al. 2015). The vapor could con-

<table>
<thead>
<tr>
<th>Process</th>
<th>Low Temperature Regime</th>
<th>High Temperature Regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiation</td>
<td>LTE, less spectral lines</td>
<td>non-LTE, more spectral lines from excited energy levels</td>
</tr>
<tr>
<td>Gas chemistry</td>
<td>one-way reactions dominate</td>
<td>forward and backward reactions, thermal ion chemistry</td>
</tr>
<tr>
<td>Condensed phase</td>
<td>molecular solid/liquid (e.g., H$_2$O, CH$_4$, NH$_3$, N$_2$, CO$_2$)</td>
<td>covalent/ionic/metallic refractive solids (e.g., silicate, Fe, KCl)</td>
</tr>
<tr>
<td>Dynamics</td>
<td>low-speed waves and wind, moist and dry convection, negligible magnetic coupling</td>
<td>high-speed waves and wind, dry convection, MHD effect</td>
</tr>
<tr>
<td>Escape</td>
<td>Jeans escape, non-thermal processes</td>
<td>hydrodynamic escape, non-thermal processes</td>
</tr>
<tr>
<td>Surface interaction</td>
<td>condensation/collapse on surface ocean or ice</td>
<td>gas exchange with magma ocean or melted surface</td>
</tr>
</tbody>
</table>
dense to dust clouds when transported to the nightside and collapse. A hot atmosphere could also melt its rocky surface that leads to interesting interactions (such as ingassing) between the atmosphere and a magma lake/ocean by analogy with Jupiter’s moon Io.

### 2.3 Spectral and Photometric Characterization

Towards a more detailed classification of the substellar atmospheres, photometry and spectroscopy play a central role. Both the atmospheric composition and temperature directly control the broadband magnitudes and colors as well as detailed spectral features in transmission, emission and reflection spectra. Following conventional stellar classification in the Morgan-Keenan (MK) system (Morgan & Keenan 1973), brown dwarfs are classified into several categories according to their spectral colors in the optical and near-IR. The spectral types include L dwarfs (Kirkpatrick et al. 1999), T dwarfs (Burgasser et al. 2002b, 2006, another classification scheme from Geballe et al. 2002 yielded the similar results) and the Y class (Cushing et al. 2011 and Kirkpatrick et al. 2012). See the detailed distinction between the M, L, T and Y spectral classes in the review by Kirkpatrick (2005). In optical wavelengths, the early-L dwarfs are characterized by multiple atomic and molecular lines such as the neutral alkali metals (e.g., Na I, K I, Rb I, Cs I), oxides (TiO and VO) and hydride (e.g., FeH). Both alkali lines and hydrides increase strength in the mid-L, but the oxides TiO and VO disappeared. As the dwarfs become colder such as in late-L and early-T, the spectra show strong water features and alkali lines, whereas hydrides are less important. In the late-T, water dominates the absorption and the line widths of Na I and K I spread widely. Finally, in the cold and faint Y-class, the optical features almost disappeared. The characterization of brown dwarfs in near-IR is also similar. Early-L spectra show features of H₂O, FeH and CO, and atomic metal lines such as Na, Fe and K. CH₄ appears in early-T. CH₄ and H₂O dominate the entire T-type spectra. The Y-dwarfs show up at the cold end of the spectral sequence where the alkali resonance lines disappear and possibly ammonia absorption bands emerge in the near-IR (e.g., Cushing et al. 2011; Kirkpatrick et al. 2012; Line et al. 2017; Zalesky et al. 2019). On the other hand, the mid-IR classification has not been well established yet (Kirkpatrick 2005). The effective temperature of L dwarfs ranges from 1300 K to 2500 K and T dwarfs are typically below 1500 K. The coldest known Y dwarf detected so far is WISE 0855 with an effective temperature of about 235–260 K (Luhman & Esplin 2014) where water could condense as clouds (e.g., Skemer et al. 2016; Esplin et al. 2016; Morley et al. 2018; Miles et al. 2020).

Within the L dwarfs, spectroscopic diversity can be further classified using gravity as in the MK system because both the opacity distribution and vertical temperature profile in the atmosphere are significantly influenced by gravity. For example, the weak FeH absorption and weak Na I and K I doublets indicate low gravity objects (Cruz et al. 2009). For brown dwarfs, gravity is also a good proxy of age. Cruz et al. (2009) proposed a gravity classification scheme for the optical spectra: $\alpha$ for normal gravity, $\beta$ for intermediate gravity and $\gamma$ for very low gravity; those latter two correspond to ages of $\sim$100 and $\sim$10 Myr, respectively. Utilizing equivalent widths for gravity-sensitive features (VO, FeH, K I, Na I and the H band continuum shape) in the near-IR spectra, Allers & Liu (2013) classified the young brown dwarfs into three types. They the low-gravity (VL-G), intermediate gravity (INT-G) and field (FLD-G), corresponding to ages of $\sim$30, $\sim$30–200 Myr and $\sim$200 Myr, respectively. Note that the gravity types are still very uncertain.

Do planets also manifest typical spectral types? Planets are more diverse than brown dwarfs because their temperature and compositions are affected by many factors, such as the distance to the star, metallicity, gravity and internal heat. Despite their complex nature, Fortney et al. (2008) proposed that the dayside atmospheres of hot Jupiters could be classified into two categories: the hotter “pM” and the cooler “pL” classes, by analogy to M and L brown dwarfs respectively. The “pM” planets with an effective temperature greater than 2000 K will exhibit strong thermal inversion (i.e., temperature increases with altitude) in the upper atmosphere caused by the TiO and VO opacity sources and high irradiation from the parent stars, as well as a large day-night temperature difference due to the shorter radiative timescale than the advective timescale. Their dayside spectra are expected to display emission features in photospheres. Note that the existence of TiO or VO in the upper atmosphere might require strong vertical mixing (Spiegel et al. 2009). On the other hand, the cooler class, “pL”, could re-radiate away the incoming stellar radiation more easily and show no thermal inversion in the photosphere. Water absorption features in the near-IR instead dominate their spectra. The search for evidence of the two classes and thermal inversion is still ongoing. Some recently characterized ultra-hot Jupiters with equilibrium temperature greater than 2200 K have been confirmed with temperature inversion and emission features detected, including WASP-121 b (Evans et al. 2017), WASP-18 b (Sheppard et al. 2017; Arcangeli et al. 2018) WASP-33 b (Haynes et al. 2015; Nugroho et al. 2017). Even though the detailed mechanisms of thermal inversion might not exactly be due to the previously proposed TiO/VO opacity, and the transition between planets with and without
inversion might not coincide with 2000K as proposed in Fortney et al. (2008), it seems the exoplanetary atmospheres do show some typical spectral categories that can be further characterized in future spectral observations.

In addition to temperature, gravity might also play a role. Parmentier et al. (2018) classified the hot Jupiter spectra at secondary eclipse using gravity and the dayside temperature. In the higher gravity and/or lower temperature regime, TiO is expected to rain out (e.g., Spiegel et al. 2009; Parmentier et al. 2013; Parmentier et al. 2016; Beatty et al. 2017). In the higher temperature regime (such as on ultra-hot Jupiters with an equilibrium temperature, \(T_{\text{eq}} > 2200\) K), most spectrally relevant molecules, except some with very strong bonds such as \(N_2\), CO and SiO, tend to be thermally dissociated, resulting in spectra with very weak features in general. H\(^+\) opacity becomes an important opacity source in the high-temperature regime as well. More discussions are in Section 4.2.

Photometrically, the substellar bodies can be characterized using color-magnitude diagrams (CMDs) similar to the Hertzsprung-Russell diagram for stars. Dupuy & Liu (2012) compiled a large number of brown dwarfs and illustrated their evolution sequence in both near-IR and mid-IR. One typical diagram is the \(J - H\) color versus \(J\) band magnitude in Figure 1, in which we convert the \(H\) band flux to the HST channels for comparison with data from hot Jupiters and ultra-hot Jupiters (Manjavacas et al. 2019a). In the optical and near-IR sequences, the spectral sequence of brown dwarfs from M, L, T to Y types is evident. As the \(J\) band flux decreases, \(J - H\) color is gradually reddening from M to late L, and suddenly shifts to blue in early-T within an effective temperature range of only about 200 K. It continues bluer to the mid- and late-T but eventually turns back to red in the Y-types. Also, some discrepancies between optical and near-IR types for L dwarfs and the evolution sequences exist (Kirkpatrick 2005). In terms of gravity, the low-gravity objects (VL-G) are systematically redder and brighter than the field brown dwarfs. For more photometric behaviors on the gravity dependence, refer to Liu et al. (2016).

Those spectral trends are statistically robust, but underlying mechanisms are not fully understood. Stephens & Leggett (2003) suggested that the optical sequence primarily came from temperature, but the near-IR diagram is influenced more by clouds or possibly gravity. Kirkpatrick (2005) also argued that the main driver of spectral evolution is temperature, but condensational clouds also play an important role in the spectral change. The inclination angle of those bodies viewed from Earth could also impact the color diversity (e.g., Kirkpatrick et al. 2010; Metchev et al. 2015; Vos et al. 2017). The temperature and cloud formation seem mainly driving the M-L sequence as the objects redden as they cool. The observed temperature and spectral types are only correlated well from early to mid-L. The correlation breaks down in the very sharp transition from mid-L to mid-T as the near-IR color changes blueward in a very narrow effective temperature range (\(\sim 1400 \pm 200\) K) (Kirkpatrick 2005). This transition has been observed for young, old and spectrally peculiar objects in the near-IR (Liu et al. 2016). The underlying mechanisms of the so-called L/T transition problem are not known yet. It was proposed to be relevant to the change of the cloud properties in the atmospheres (e.g., Saumon & Marley 2008; Marley et al. 2010) but alternative mechanisms have also been suggested, such as dynamical regime change driven by gas composition change with temperature (e.g. Tremblin et al. 2015, 2016, 2017b, 2019). We will discuss the details in Section 5.3.1 and Section 6.3.

The mid-IR CMD in the Spitzer/IRAC bands of 3.6 and 4.5 microns is also depicted in Figure 1 (Triaud et al. 2014; Triaud 2014). The mid-IR CMD does not display very distinct spectral types, and the [3.6]–[4.5] color stays roughly the same in M and mid-L. At around 1400 K, as the temperature drops, the [3.6]–[4.5] color exhibits a clear redward shift. This transition is the mid-IR L/T transition. The [4.5]–[5.8] CMD (also see Fig. 1) also shows the sharp L/T transition but towards the blue end as the effective temperature decreases (Triaud 2014). The temperature-driven gas chemistry probably causes both the redward turns in the [3.6]–[4.5] diagram and blueward turn in the [4.5]–[5.8] diagram (Triaud 2014). In the mid-IR, the vibrational-rotational bands of CH\(_4\), CO and H\(_2\)O dominate the absorptions at 3.6, 4.5 and 5.8 microns, respectively. As the temperature drops below the L/T transition temperature, the thermochemical reaction \(\text{CO} + 3\text{H}_2 \rightarrow \text{CH}_4 + \text{H}_2\) favors the production of CH\(_4\) and H\(_2\)O. Consequently, both the absorption at 3.6 and 5.8 microns increase, but CO absorption at the 4.5-micron band decreases, resulting in the color change in the mid-IR L/T transition. Clouds might affect the sharp gradient as well, but it has yet to be investigated in detail.

Do exoplanets also follow similar color-magnitude sequences? We first consider close-in exoplanets. Because the emission from close-in planets is mainly from the re-radiation of external stellar energy rather than internal energy, this might not be an apples-to-apples comparison. The dayside emissions of close-in planets (mostly hot Jupiters) from HST near-IR (from Manjavacas et al. 2019a) and Spitzer mid-IR channels (from Triaud et al. 2014 and Kammer et al. 2015) are plotted on top of the brown dwarf samples in Figure 1 for comparison. There is no well-characterized hot Jupiter in the T-dwarf temperature range yet. It looks that the planets and brown dwarf-
Crossfield & Kreidberg 2017) and 2M 1207 b from Liu et al. 2014. Cooler and smaller planets such as 51 Eri b from 2015 2011; Fisher & Heng 2018 Liu et al. 2013). Instead, Wakeford et al. Gao et al. 2020; PSO J318 from Liu et al. Wilkins et al. Mallonn & Strassmeier 2016 2013b). For hot Jupiters, HST data (Mandell et al. 2013 2016). The directly imaged planet data are from various sources: Beta-pic b from Bonnefoy et al. (2014), PSO J318 from Liu et al. (2013b) and Liu et al. (2016), 51 Eri b from Macintosh et al. (2015), GJ 504 b from Liu et al. (2016), Ross 458 c from Cushing et al. (2011), the HR 8799 system (b, c, d and e) from Zurlo et al. (2016) and 2M 1207 b from Allers & Liu (2013). In the left panel, both brown dwarfs and directly imaged planets are converted into HST colors based on the scaling relationship in Manjavacas et al. (2019a).

Young, directly imaged planets offer a more direct comparison to brown dwarfs because of their similar self-luminous nature. One would expect those directly imaged planets should be located within the low-mass brown dwarf sequence further towards red. To date, whether the directly imaged exoplanets exhibit a clear L/T transition is inconclusive.

Population studies have shed light on possible statistical properties of an ensemble of exoplanetary spectra and light curves (e.g., Stevenson 2016; Sing et al. 2016; Crossfield & Kreidberg 2017; Fu et al. 2017; Tsiaras et al. 2018; Fisher & Heng 2018; Wakeford et al. 2019; Gao et al. 2020). For example, spectral strengths of metals and water vapor in the transmission spectra of transiting exoplanets can be employed to quantify how cloudy the atmospheres are. The presence of high-altitude condensational clouds and photochemical hazes could significantly weaken the spectral absorption features. Such flattened transmission spectral features have been detected for many hot Jupiters (e.g., Deming et al. 2013; Mandell et al. 2013; Gibson et al. 2013; Sing et al. 2013; Jordán et al. 2013; Wilkins et al. 2014; Mallonn & Strassmeier 2016; Sing et al. 2016; Fu et al. 2017). Cooler and smaller planets such as warm Neptunes and super-Earths are also inferred to possess high altitude aerosols (GJ 1214 b, Kreidberg et al. 2017).
2014a; GJ 436 b, Knutson et al. 2014a; HD 97658 b, Knutson et al. 2014b; GJ 3470 b, Ehrenreich et al. 2014; HAT-P-26 b, Wakeford et al. 2017; HD 106315 c, Kreidberg et al. 2020. HAT-P-11 b is partially cloudy as water vapor can be seen in the HST near IR band (Fraine et al. 2014) and the nearly flat optical transmission spectrum from HST STIS (Chachan et al. 2019). Also, two super-puffs—planets with very low masses but large radii, Kepler 51 b and d—have been observed to show flat transmission spectra in the near IR, indicating abundant atmospheric hazes or dust particles (Libby-Roberts et al. 2020; Wang & Dai 2019; Gao & Zhang 2020). Two other cooler sub-Neptunes, K2-18 b (Benneke et al. 2019b; Tsiaras et al. 2019) and K2-25 b (Thao et al. 2020), might also have hazes or clouds in their atmospheres but the details are not certain yet. A systematic analysis of the transmission spectra in Wakeford et al. (2019) showed that, on most hot Juliers, the amplitudes of the near-IR water spectral features are \((\text{near} \sim 1/3)\) of that expected in clear-sky models, indicating a ubiquitous presence of suspended particles (such as clouds and hazes) on gas giants. An HST campaign (Sing et al. 2016) observed hot Juliers across a broad range of physical parameters and inferred the cloudiness from the strength of water band signals. The spectral strength of water is found to correlate well with the relative absorption strength difference between optical and near-IR and also with that between the near-IR and mid-IR. This provides strong evidence that the clouds and hazes significantly shape both the optical and IR color of transiting giant exoplanets.

Stevenson (2016) proposed a cloudiness metric using the \(J\) band water feature amplitude \((A_H)\) in transmission spectra. The larger the \(A_H\) is, the less cloudy the atmosphere is. In a sample of 14 exoplanets, there is a positive correlation between the cloudiness proxy \(A_H\) and the equilibrium temperature \(T_\text{eq}\) when \(T_\text{eq} < 750\) K. A weak correlation between water signal and gravity was also proposed. Fu et al. (2017) generalized that study to include 34 transiting exoplanets and found a positive correlation between the cloudiness \(A_H\) (similar to that in Stevenson 2016) and \(T_\text{eq}\) between 500–2500 K (also see a recent study in Gao et al. 2020). For Neptune-sized planets, Crossfield & Kreidberg 2017 reported that the \(A_H\) for smaller planets might also correlate with the equilibrium temperature and bulk H/He mass fraction, which they interpreted as a correlation between the metallicity and cloudiness. We will discuss the theories of clouds and hazes on exoplanets and their implications on the above observational trends in Section 5.3.

The transmission spectra on close-in brown dwarf s are difficult to obtain because of their high gravity and small scale height. However, the emission spectra of directly imaged planets and brown dwarfs have high er S/N compared with their close-in counterparts. As we showed in Figure 1, their near IR CMD indicates the existence of clouds in their atmospheres for (e.g., L dwarfs). The spectra of many directly imaged planets might also need clouds or hazes to explain (e.g., HR 8799 planets: Barman et al. 2011; Madhusudhan et al. 2011a; Marley et al. 2012; Currie et al. 2014; Skemer et al. 2014; \(\beta\)-Pic b: Chilcote et al. 2014, Chilcote et al. 2017; 51 Eri b: Macintosh et al. 2015; Zahnle et al. 2016). Moreover, there are also many dusty brown dwarfs typically in the L spectral type, some even displaying a possible silicate feature in the mid-IR (Cushing et al. 2006).

The data quality of emission spectra on close-in exoplanets is generally lower than that on directly imaged planets because of stellar contamination. Nevertheless, their mid-IR broadband emission at 3.6 and 4.5 microns can be observed by Spitzer during their secondary eclipses (e.g., Triaud 2014). Although their mid-IR CMDs manifest large scattering (Fig. 1), recent studies have searched for an internal correlation of the mid-IR brightness temperatures between the 3.6 and 4.5 micron data (Kammer et al. 2015; Wallack et al. 2019; Garhart et al. 2020). The statistical analysis seems to suggest a systematic deviation of the mid-IR spectra from the blackbodies. Moreover, there seems to be a statistically increasing trend of the observed brightness temperatures between 4.5 and 3.6 microns with increasing equilibrium temperature in the range of 800–2500 K (Garhart et al. 2020). This trend is still a puzzle that no current theory can explain. We will discuss it in detail in Section 4.2.

Although it is difficult to resolve surface features on distant substellar bodies, time-domain observations provide clues on their temporal and spatial variations. Horizontal information on substellar atmospheres is primarily obtained from light curve observations. In addition, eclipse mapping (e.g., Rauscher et al. 2007; de Wit et al. 2012) has been suggested to be able to probe the spatial features in future observations. There are three kinds of light curves: reflection, transit and emission. The stellar flux strongly contaminates reflection light curves for close-in planets, and the signals are weak for planets far away from their host stars. For transiting planets, transit light curves in principle could also be used to probe the difference between east limb and west limb (e.g., Line & Parmentier 2016; Kempton et al. 2017; Powell et al. 2019) but the current S/N is still not good enough.

Emission light curves originate from the time evolution of hemisphere-averaged thermal flux emitted from the planets towards the observer. There are two general types of emission light curves. For close-in exoplanets and close-in brown dwarf companions with self-rotations
synchronized with their orbits around central stars due to gravitational tides, emission light curves trace different phases in the orbits and are also called thermal phase curves. Most thermal phase curves are detected through the “warm Spitzer band” at 3.6 and 4.5 microns (see review in Parmentier & Crossfield 2018). For very hot planets, it is also possible to observe emission light curves from the visible band such as Transiting Exoplanet Survey Satellite (TESS, e.g., Shporer et al. 2019). On the other hand, if a cooler planet is bright, the detected light curves in short wavelengths (e.g., HST near-IR band, Kepler band) might include a strong stellar reflection component (e.g., Parmentier et al. 2016). The shape of the phase curves directly probes the photospheric inhomogeneity on these synchronously rotating planets. For example, Knutson et al. (2007) detected a phase offset of the light curve peak, suggesting that the heat redistribution due to atmospheric jets and waves shifts the hot spot away from the substellar point. The temporal variation of the phase curves between different rotations also suggests complicated weather patterns on these planets. One example is Kepler observations targeting HAT-P-7 b, on which the peak brightness offset changes dramatically with time (Armstrong et al. 2016). Another example is Kepler 76 b (Jackson et al. 2019). The phase curve amplitude can vary by a factor of two in tens of days, associated with the peak offset varying accordingly. Population studies have also been performed to understand the statistical properties—such as the albedo and heat redistribution—of an ensemble of exoplanetary phase curves (e.g., Cowan & Agol 2011; Schwartz et al. 2017; Zhang et al. 2018; Keating et al. 2019). Current data samples on transiting planets might have revealed some possible interesting trends of the dayside temperature, nightside temperature, day-night temperature difference, and phase offset on various parameters such as equilibrium temperature and rotation rate. Details will be discussed in Section 4.3.

The emission light curves observed on directly imaged planets and brown dwarfs fall into a different category. On these bodies, IR emission is modulated by planetary self-rotation and in-and-out-of-view of the weather patterns in the photospheres, producing rotational light curves. Photometric variability has been monitored for brown dwarfs since their discovery. Their rotational light curves unveil very active weather associated with temperature and cloud patterns, especially around the L/T transition. The short-term and long-term variations of the rotational light curves can be used to retrieve the surface features (e.g., Apai et al. 2017) and even the wind speed (Allers et al. 2020). Recent progress have been summarized in a series of papers on brown dwarfs from the “Weather on other Worlds” program (Heinze et al. 2013, 2015; Metchev et al. 2011, 2015; Miles-Páez et al. 2017a) and on both low-gravity brown dwarfs and planetary-mass companions from the “Cloud Atlas” program (Lew et al. 2016, 2019, 2020; Manjavacas et al. 2017, 2019a,b; Miles-Páez et al. 2019; Zhou et al. 2019, 2020, 2018), as well as the reviews in Biller (2017) and Artigau (2018). We will discuss the rotational light curves, their variability, and the underlying mechanisms in Section 4.4 and Section 6.3.

3 ATMOSPHERIC STABILITY

3.1 Fundamentals

The stability of a planetary atmosphere primarily depends on the planetary mass, radius and atmospheric temperature. For planets with surfaces, if the surface temperature drops below the main constituents’ saturation temperatures, the atmosphere will collapse. Possible ice-albedo feedback—the condensed ices (e.g., water, CO$_2$, N$_2$, CO, and CH$_4$) could reflect more stellar flux to space and further cools down the surface—accelerates the process. A planet with a collapsed atmosphere enters a snowball climate, with the surface pressure being in thermodynamical equilibrium with the surface ices. Such atmospheric collapse could be common for the terrestrial climate. The current atmospheres of Pluto and Triton are in this state. Earth was in the snowball phase several times. The atmospheres of Mars (Forget et al. 2013) and Titan (Lorenz et al. 1997) might have collapsed in the past. Atmospheric collapse and condensation will also greatly change the compositions of the atmosphere (see Sect. 5.2). The atmosphere could also be absorbed into the magma ocean or the interior in the early age (e.g., Olson & Sharp 2019; Kite et al. 2020).

In this section, we will mainly focus on escape to space (e.g., Jeans 1904; Parker 1958). The atmosphere escapes via both thermal and non-thermal processes. In thermal escape, if the upper atmosphere temperature is so high—either due to strong stellar heating, gravitational energy released during the accretion phase or other heating mechanisms—that the thermal velocities of molecules or atoms exceed the escape velocity of the planet, that volatiles are no longer gravitationally bound. If the atmosphere remains in balance and the velocities of the molecules or atoms still follow the Maxwellian distribution, only a fraction of the molecules in the high-velocity tail of the distribution will be able to escape. This scenario is the Jeans escape. The particles will escape to space from the exobase, which is the altitude above which the atmosphere is no longer collisional. If the temperature of the upper atmosphere is very high, the entire atmosphere can escape hydrodynamically, driven by the pressure gradient. Hydrodynamic escape can still be diabatic. In some situa-
ions such as giant impact by incoming asteroids or comets, if the atmosphere has enough internal and KE per unit mass to escape isentropically, extreme escape could occur as a quick blow-off. The division between the two regimes (Jeans and pressure-driven escape) can be roughly characterized by the Jeans parameter \( \lambda = GM_p m / k_B T R_p \), a dimensionless number that describes the ratio of the gravitational energy \( GM_p m / R_p \) to the thermal energy of the upper atmosphere \( k_B T \), where \( M_p \) and \( R_p \) are the planetary mass and radius (or more precisely, the exobase radius), respectively. \( G \) is the gravitational constant, and \( k_B \) is the Boltzmann constant. \( T \) is the temperature at the exobase. 

\( m \) is the mass of the escaping species. The Jeans parameter is also the ratio of the pressure scale height to the planet radius. Moreover, the square root of \( \lambda \) is roughly equal to the ratio of the escape velocity \( \sqrt{\lambda} v_e = (2GM_p / R_p)^{1/2} \) to the adiabatic sound speed \( \sqrt{k_B T / m} \), where \( k_B \) is the adiabatic index. This expression includes three crucial parameters of the planetary atmosphere: mass, radius and temperature at the exobase and the mass of the escaping particle, which is usually H or He atoms.

Although atmospheric thermal escape has been studied for at least 170 yr (from J. J. Waterson Waterson 1851, also see early works by Jeans 1904; Parker 1958; Hunt et al. 1982; Hunt et al. 1990; Zahnle et al. 1990), the theory and especially numerical simulations are still incomplete. Crudely speaking, the transition between the two end-members—hydrostatic Jeans escape (large \( \lambda \)) to hydrodynamic escape (small \( \lambda \))—is found to occur at around \( \lambda \sim 1 \) (e.g., Volkov et al. 2011a; Volkov et al. 2011b, Tian 2015). The reality is, however, much more complicated. For example, the behavior also depends on the collisional property of the medium characterized by the “Knudsen number” \( Kn \)—the ratio of the mean free path of the escaping gas to the planetary radius. Usually, the transition from hydrodynamic to free molecular flow at the exobase is difficult to resolve without molecular dynamics or Boltzmann numerical simulation. The direct simulation Monte Carlo (DSMC) results (Volkov et al. 2011b) show that, for a single component atmosphere, evaluated using the Jeans parameter and Knudsen number, thermal escape processes fall into different regimes. In the collisional regime (small \( Kn \)), an analytical theory is also consistent with the DSMC results in Gruzinov (2011). The thermal escape at the top of the planetary atmosphere can occur in three regimes: Parker, Fourier and Jeans. In the traditional Parker regime (Parker 1958, 1964a,b) where the Jeans parameter is small (\( \lambda < 2 \)), outflow behaves as a supersonic ideal hydrodynamic wind. Thermal conductivity is negligible, and the temperature structure is controlled by isentropic expansion starting from the sonic surface.

When the Jeans parameter is large (\( \lambda > \sim 4 \sim 6 \)), the atmosphere escapes in a molecule-by-molecule fashion. The escape rates are not significantly different from the traditional Jeans flux if \( \lambda > \sim 6 \). Thermal conduction is important in this regime, and temperature could be nearly isothermal. In between the Parker and Jeans regimes, thermal conduction balances the hydrodynamic expansion. This transition (so-called Fourier regime) occurs in a very narrow range of \( \lambda \sim 2 \sim 4 \). If the atmosphere is not very collisional, such as a low-density medium (large \( Kn \)), thermal conduction is also significant. Otherwise, the traditional Parker wind solution can lead to inaccurate results (Volkov 2016).

An atmosphere could also be lost to space via many non-thermal processes such as photochemically driven escape, ion pick-up by the stellar wind, stellar wind stripping, charge exchange and so on (e.g., Holmstrom et al. 2008; Kislyakova et al. 2013; Kislyakova et al. 2014; Dong et al. 2017). To understand the non-thermal escape processes requires sophisticated photochemical and ion-chemical calculations in the upper atmosphere and a coupled magnetosphere-ionosphere-thermosphere (MIT) simulation for the interaction between the atmosphere and the solar wind. As noted, atmosphere could also be removed by surface weathering and ingassing processes. Typical examples are the silicate-carbonate cycle on Earth (weathering) and helium rain (maybe including neon) in the giant planets. We do not discuss those processes in detail here. Atmospheric escape is not only important for understanding atmospheric mass evolution, but also strongly impacts the atmospheric composition via mass fractionation (e.g., Zahnle & Kasting 1986; Zahnle et al. 1990; Hunt et al. 1987) and altering the planetary redox state over time (e.g., Catling et al. 2001). See the reviews in Lammer et al. (2008), Tian et al. (2013) and Kislyakova et al. (2015) for more details.

Atmospheric escape becomes relevant for exoplanets since multiple species have been detected in their upper atmospheres thanks to the high-resolution facilities in the ultraviolet and visible, such as HST, VLT/ESPRESSO, TNG/HARPS, and GTC/OSIRIS. The first extended hydrogen cloud surrounding a canonical hot Jupiter HD 209458 b was discovered by the Lyman-\( \alpha \) transit technique (Vidal-Madjar et al. 2003). Lyman-\( \alpha \) has also been detected on another hot Jupiter HD 189733 b (e.g., Des Etangs et al. 2010; Des Etangs et al. 2012; Bourrier et al. 2013, two smaller planets including a warm Neptune GJ 436 b (Kulow et al. 2014; Ehrenreich et al. 2015; Lavie et al. 2017b; dos Santos et al. 2019) and GJ 3470 b (Bourrier et al. 2018) and possibly TRAPPIST-1 b and c (Bourrier et al. 2017b), Kepler-444 (Bourrier et al. 2017c) and K2-18 b (dos Santos et al. 2020), suggesting strong hydrogen escape from these bodies. Other hydrogen lines in the Balmer series in
the visible such as Hα (and Hβ in some cases) have also been detected on two hot Jupiters HD 189733 b (e.g., Jensen et al. 2012, Cauley et al. 2016, 2017a,b) and WASP-52 b (Chen et al. 2020), as well as for four ultra-hot Jupiters: MASCARA-2 b (also known as KELT-20 b, Casasayas-Barris et al. 2018), WASP-12 b (Jensen et al. 2018), KELT-9 b (Yan & Henning 2018; Cauley et al. 2019a, Turner et al. 2020, Wytenbach et al. 2020) and WASP-121 b (Cabot et al. 2020). These observations suggested extended hydrogen atmospheres that might originate from the neutral hydrogen escape. Note that some Hα signals from the young, forming planets could instead come from the ongoing accretion, for example, PDS 70 b and PDS 70 c (e.g., Hauff et al. 2019; Aoyama & Ikoma 2019; Hashimoto et al. 2020). On the other hand, extended hydrogen exospheres were not detected on some other planets, such as super-Earths 55 Cnc e (Ehrenreich et al. 2012), HD 97658 b (Bourrier et al. 2017a), GJ 1132 b (Waalkes et al. 2019) and π Men c (García Muñoz et al. 2020).

In extended atmospheres, heavier species including helium and easily ionized metals such as Na, K, Ca, Mg, Si, and Fe have also been detected near or beyond the planetary Roche lobe, for example, on HD 209458 b (Vidal-Madjar et al. 2004; Linsky et al. 2010; Vidal-Madjar et al. 2013; Sing et al. 2019; Cubillos et al. 2020), and other planets including ultra-hot Jupiters (e.g., Fossati et al. 2010; Jens Hoeijmakers et al. 2018; Hoeijmakers et al. 2019; Yan et al. 2019; Cauley et al. 2019b; Turner et al. 2020; Chen et al. 2020). In particular, extended helium atmospheres have also been recently observed on hot Jupiter HD 209458 b (Alonso-Floriano et al. 2019), HD189733 b (Salz et al. 2018), a Jupiter-sized Neptune-mass planet WASP-107 b (Spake et al. 2018; Kirk et al. 2020), Saturn-mass planet WASP-69 b (Nortmann et al. 2018) and a Neptune-sized planet HAT-P-11 b (Allart et al. 2018; Mansfield et al. 2018a). Moreover, the circumstellar gas replenished by mass loss from ablating low-mass planets could absorb stellar chromospheric emission. The Dispersed Matter Planet Project (e.g., Barnes et al. 2020; Haswell et al. 2020; Staab et al. 2020) has recently detected low stellar chromospheric emission around about 40 out of 3000 nearby bright stars, indicating possible existence of highly irradiated, mass-losing exoplanets in these systems. Also, the observed high variability in the transit depths of so-called “super-comets” such as Kepler 1520 b (Rappaport et al. 2012; Perez-Becker & Chiang 2013) and K2-22 b (Sanchis-Ojeda et al. 2015) suggests that they might experience significant evaporation (e.g., Budaj et al. 2020).

The temperature structure of the upper atmosphere is crucial for determining the species escape rates and the resulting transit observations. Atmospheric layers at the very top should be very hot—so-called thermosphere. All thermospheres on Solar System planets are hot, especially on giant planets, but the cause is still debatable (e.g., Yelle & Miller 2004). On exoplanets, it was suggested that intense stellar heating and insufficient cooling—primarily due to thermal dissociation of coolants (e.g., Moses et al. 2011; Koskinen et al. 2013a)—allows the upper atmosphere to reach a temperature of ~ 10000 K. High-resolution observations in the UV and visible provide unambiguous evidence of the hot upper atmospheres for several exoplanets using the Lyman-α (e.g., HD 209458 b, GJ 436 b, GJ 3470 b), hydrogen Balmer series and metal lines (e.g., HD 189733 b, KELT-9 b, KELT-20 b, WASP-12 b, WASP-121 b, WASP-52 b), and Helium line (e.g., HD 209458 b, WASP-107 b, WASP-69 b, HAT-P-11 b). The temperature (and its gradient), density, associated mass loss rate, and even the wind speed at the upper atmosphere can be derived from the powerful high-resolution spectroscopy (e.g., Heng et al. 2015; Wytenbach et al. 2015, 2017, 2020; Fisher & Heng 2019; Welbanks & Madhusudhan 2019; Seidel et al. 2020). If hydrodynamic escape occurs, adiabatic cooling might lead to the temperature decreasing with altitude again. For more discussion on high-resolution spectroscopy, refer to Birkby (2018).

Observations have motivated many theoretical studies that investigated the upper atmospheres and mass loss on hot Jupiters and smaller planets (e.g., Lammer et al. 2003; Yelle 2004; Tian et al. 2005; Erkaev et al. 2007a; García Muñoz 2007; Koskinen et al. 2007a; Schneider et al. 2007; Holmstrom et al. 2008; Penza et al. 2008; Murray-Clay et al. 2009; Stone & Proga 2009; Guo 2011; Guo 2013; Trammell et al. 2011; Lopez et al. 2012; Owen & Jackson 2012; Lopez & Fortney 2013; Owen & Wu 2013; Erkaev et al. 2013; Lammer et al. 2014; Koskinen et al. 2013a,b; Tremblin & Chiang 2013; Bourrier & des Etangs 2013; Bourrier et al. 2013, 2016, 2014; Jin et al. 2014; Kurokawa & Nakamoto 2014; Shaikhislamov et al. 2014; Salz et al. 2015a,b, 2016; Owen & Wu 2017; Dong et al. 2017; Zahnle & Catling 2017; Wang & Dai 2018; Wang & Dai 2019; Jin & Mordasini 2018; Mordasini 2020; Lampón et al. 2020). For close-in exoplanets around Sun-like stars, hydrodynamic escape of atomic hydrogen could occur inside an orbit of about 0.1 AU (e.g., Lammer et al. 2003; Yelle 2004). It was found that the transition between a stable atmosphere and an unstable atmosphere (i.e., escaping) is located around 0.14–0.16 AU, around sub-like stars in 3D simulations (Koskinen et al. 2007a). In the
context of planetary formation and evolution, atmospheric escape could greatly affect the evolution of close-in small planets, especially their planetary size distribution (see reviews in Tian 2015 and Owen 2019). Thus, atmospheric escape has become essential in understanding the current planetary data sample.

In general, there are two important regimes for the thermal escape rate. The escape rate can be “supply-limited” or “energy-limited.” In the supply-limited regime, the “limiting flux principle” (Hunten 1973a, b) states that the thermal escape flux might be limited by several bottlenecks below the exobase such as the cold trap at the tropopause, atmospheric chemistry, cloud formation and vertical diffusion. Take hydrogen escaping Earth as an example. The hydrogen primarily comes from tropospheric water, which condenses as clouds in the upper troposphere before being transported to the stratosphere. This cold trap of water vapor in the tropopause region leads to a very dry stratosphere, with water molar fraction of a few parts per million (ppm). Stratospheric chemistry converts water vapor and some other hydrogen-bearing species such as methane to hydrogen atoms. The conversion rate depends on the chemical pathways and UV photons in the stratosphere. The final bottleneck is diffusion above the homopause, where the species are gravitationally separated according to their molecular weights. Hydrogen atoms rise through the heavier species to the exobase by molecular diffusion and eventually escape into space. The cold trap could effectively limit the hydrogen loss from a terrestrial planet. The efficiency of the cold trap primarily depends on the ratio of latent heat of condensable species (e.g., H₂O) to the sensible heat of the non-condensable species (e.g., N₂, CO₂) at the surface (Wordsworth & Pierrehumbert 2013). If the partial pressure of non-condensable species is small, the cold trap is not efficient, and the upper atmosphere will be moist. In that case, hydrogen escape will eventually lead to severe oxidation of the entire atmosphere on exoplanets. Hydrogen escape on terrestrial planets in the Solar System is generally limited by diffusion (Kasting & Catling 2003). It is also thought that a canonical hot Jupiter HD 209458 b is possibly escaping at the diffusion limit (e.g., Vidal-Madjar et al. 2004, 2003; Yelle 2004; Koskinen et al. 2013b; Zahnle & Catling 2017).

In the second regime, hydrogen escape rate is limited by the energy available for escape. This energy limit could come from the incoming energy itself, but could also from the limiting steps converting the incoming energy into available energy driving the escape such as radiative processes—hydrogen radiative recombination (Murray-Clay et al. 2009) or ionization photons (Owen & Alvarez 2015). In other words, the conversion efficiency of incoming energy to KE is crucial. The energy-limited escape has been widely assumed in hydrogen escape on warm and hot close-in exoplanets (e.g., Lammer et al. 2003, 2009; Vidal-Madjar et al. 2003; des Etangs et al. 2004; Baraffe et al. 2004; des Etangs 2007; Erkaev et al. 2007b; Hubbard et al. 2007b,a; Davis & Wheatley 2009; Leitzinger et al. 2011; Owen & Jackson 2012; Lopez & Fortney 2013; Owen & Wu 2013). For most planets in the cold-temperature regime (see Sect. 2.2), hydrodynamic escape is not violent and the energy supply from stellar heating in the upper atmosphere is sufficient to drive escape under the hydrostatic situation. As temperature increases, the atmospheric escape could rapidly transit from the Jeans regime to the hydrodynamic regime in a rather narrow range of Jeans parameter (Volkov et al. 2011a,b). For hotter hydrogen atmospheres, when hydrodynamic escape occurs, a rapid blow-off of the main constituents requires a large amount of heating from the stellar X-ray and extreme ultraviolet (XUV) or even softer near/far ultraviolet (NUV/FUV) photons (e.g., García Muñoz & Schneider 2019). The partitioning between the two is not well understood at this moment and perhaps varies case by case (Owen & Jackson 2012; Owen & Wu 2013). The energy loss processes in the upper atmosphere (i.e., the thermosphere) are also complicated. Most of the energy could be radiated into space, or thermally conducted to the lower atmosphere. The energy used to drive the intensive planetary wind and atmospheric mass loss is thus limited.

In the energy-limited regime, it is important to quantify both the heating efficiency and wind efficiency. The former measures the X-ray and UV heating and radiative cooling processes in the upper atmosphere, for instance, CO₂ cooling on terrestrial planets (Tian 2013), H⁺ ion cooling for giant planets (Koskinen et al. 2007a) or cooling of hydrogen radiative recombination (Murray-Clay et al. 2009). A careful treatment of the radiative transfer and chemistry is needed. The wind efficiency is a global measure of how much incoming energy is converted to KE for the blow-off. A simple but widely applied energy-limited formula (Watson et al. 1981) for hydrodynamic escape is

\[ \dot{M} \sim \frac{L_{\text{HE}} R_{p}^{3}}{A G M_{p} a^{2}}, \]  

where \( L_{\text{HE}} \) is the high-energy portion of the stellar luminosity, \( a \) is the star-planet distance. Here we neglect the potential energy reduction factor due to the Roche lobe effect (Erkaev et al. 2007a), as well as the difference between the planetary radius (which is vague for giant planets) and the level where the wind is launched. Energy-limited escape is usually valid if the cooling is dominated by adiabatic (or subsonic) expansion with a large escape rate (Johnson et al. 2013). The wind efficiency (\( \eta \)) is...
usually treated as a constant for simplicity, for instance, 10%–20% (e.g., Lopez et al. 2012; Owen & Jackson 2012; Lopez & Fortney 2013; Owen & Wu 2013; Karokoawa & Nakamoto 2014). However, it would be good to keep in mind that the wind efficiency in this simple formula is usually not constant and needs to be utilized with caution. Also, hydrodynamic escape is essentially a self-limiting process because the rapid non-hydrostatic expansion of the atmosphere will adiabatically cool itself down and slow down the wind. The atmospheric structure, including the temperature and density, might change dramatically and thus the heating level and wind efficiency (e.g., Koskinen et al. 2013b). Previous studies (e.g., Erkaev et al. 2013; Lammer et al. 2013) indicate that the simple energy-limited formula might only apply to high-gravity bodies like hot Jupiters where thermospheric expansion is not extreme rather than low-gravity bodies such as cooler Earth-sized planets. Otherwise, escape rates could be significantly overestimated. Moreover, recent hydrodynamic simulations with thermochemistry and radiative transfer (Wang & Dai 2018) demonstrated that for small planets (\(M_p < 10M_E\)), the mass-loss rate scales with radius squared (\(R^2\)) instead of radius cubed in the conventional formula, leading towards a “photo-limited” scenario (e.g., Owen & Alvarez 2015) where EUV photoheating is strong and the gravitational potential is shallow.

Here, instead of focusing on the detailed modeling and theories on the escape of exoplanet atmospheres (see reviews in Lammer et al. 2008; Tian 2015; Owen 2019), we highlight two important regime classifications of currently detected exoplanets from observational statistics known as the “cosmic shoreline” and the “planet desert and radius gap.”

### 3.2 “Cosmic Shoreline”: Irradiation or Impact?

After proto-atmosphere accretion, the long-term existence of an atmosphere is controlled by the planet’s ability to hold its atmosphere. Knowing the fundamental processes such as condensation and escape, one can predict whether a planet has an atmosphere or not. The dominant mechanisms could be statistically tested against existing data. Zahnle (1998) first analyzed the Solar System data and put planets and large moons in a diagram of solar insolation versus escape velocity. An empirical division exists between those bodies with and without apparent atmospheres. Zahnle & Catling (2017) expanded this idea to include asteroids, Kuiper Belt Objects and exoplanets in the same diagram (Fig. 2). Although the escape velocity spans more than two orders of magnitude and the stellar insolation changes about eight orders of magnitude in these \(\sim 600\) samples, an empirical division between atmospheric bodies and airless ones is relatively clear. The regime boundary seems to follow a straight line in log-log space, the so-called “Cosmic Shoreline.” The region below the shoreline is the “atmospheric regime”— planets tend to have atmospheres when insolation is low and gravity is high; planets above the shoreline fall into the “airless regime”— they do not seem to harbor an apparent atmosphere.

The existence of the cosmic shoreline is intuitively understandable, but the detailed mechanisms are not easy to decipher. To first order, escape velocity measures the depth of the gravitational potential on a planet. Stellar insolation represents several external driving forces that lead to atmospheric loss. For example, insolation itself affects the planetary equilibrium temperature and might lead to a thermally unstable state of the entire atmosphere. The high-energy portion of stellar photons in the X-ray, XUV and FUV can directly trigger hydrodynamic escape of the atmosphere. The stellar wind is responsible for many non-thermal processes such as stellar wind stripping, sputtering and ion pick-up. The empirical cosmic shoreline in Figure 2 can be expressed as

\[
I \propto v_e^4, \tag{3}
\]

where \(I\) is stellar insolation at the planetary body, and \(v_e\) is escape velocity.

The underlying principle of this simple scaling law is not obvious. Here we restate the derivation in Zahnle & Catling (2017) using the Jeans parameter \(\lambda\). Since the atmosphere will be lost rapidly through hydrodynamic escape as lambda exceeds unity, we expect the cosmic shoreline corresponds to \(\lambda = 1\). From the Stefan-Boltzmann law, \(I\) scales with \(T^4\). Zahnle & Catling (2017) assumed the molecular weight \(\mu = T^{-1}\) in diverse planetary atmospheres (which is also an empirical observation). Put together, we found \(\lambda \sim I^{-1/2}v_e^2/k_B\), and thus the \(\lambda \sim 1\) corresponds to \(I \propto v_e^4\). An alternative but very similar version to represent the cosmic shoreline is employing the XUV flux as the vertical axis in Figure 2 (see fig. 2 in Zahnle & Catling 2017). One could also obtain a scaling law by assuming that the X-ray and XUV heating primarily drive the hydrodynamic escape. In this scenario, the total fractional mass loss of the atmosphere is the time integral of the energy-limited escape formula (Eq. (2)) proportional to \(I_{\text{xuv}} t_{\text{xuv}} R_p^2/M_p^2\) where \(t_{\text{xuv}}\) is the Kelvin-Helmholtz timescale or cooling timescale in which the planet is under high XUV exposure. Typically the timescale is on the order of a few Myr (e.g., Jackson et al. 2012; Tu et al. 2015). Assume the mass-radius relationship is \(M_p = \rho R_p^3\), which could be problematic because the mass-radius relationship of the planet is not simple, and one can achieve
I_XUV \propto v_e^3 \rho^{1/2}, \text{ which is also roughly consistent with the current sample.}

In the low insolation regime (left-lower corner of Fig. 2), the collapsed atmospheres on Kuiper Belt Objects (including Pluto and Triton) can be divided by another type of cosmic shoreline, which does not follow the simple power-law but manifests as curves in Figure 2. Zahnle & Catling (2017) proposed hydrodynamic thermal escape models for CH₄ and N₂ assuming vapor pressure equilibrium at the surface. Their models could explain the regime division in those low-temperature bodies.

Thermal escape is not the sole explanation. There has been a long-standing hypothesis that a planetary atmosphere can be entirely removed by impact erosion (e.g., Walker 1986; Melosh & Vickery 1989; Zahnle et al. 1992; Zahnle 1993; Zahnle 1998; Griffith & Zahnle 1995; Chen & Ahrens 1997; Brain & Jakosky 1998; Newman et al. 1999; Genda & Abe 2003; Genda & Abe 2005; Catling & Zahnle 2009; Shuvalov 2009; Shuvalov et al. 2014; Korycansky & Zahnle 2011; Schlichting et al. 2015; Zahnle & Catling 2017; Biersteker & Schlichting 2019; Wyatt et al. 2019). The impact erosion scenario has been proposed to understand the early atmosphere of Mars (Melosh & Vickery 1989) and the dichotomy between gas-rich Titan and airless Ganymede/Callisto (Zahnle et al. 1992). Although large uncertainties still remain in evaluating the detailed mechanisms, presumably a thinner atmosphere is easier to be eroded away than a thicker atmosphere, meaning that the impact erosion is a runaway process. Zahnle & Catling (2017) also tested this hypothesis utilizing all planet samples in Figure 2. They simply assumed that impact velocities are proportional to orbital velocities for close-in planets and plotted against the escape velocities of the planets. It was found that, again, there is a regime division between the bodies with and without atmospheres (see their fig. 4). The regime boundary follows \( v_{\text{imp}} = 4 \sim 5 v_e \) where the \( v_{\text{imp}} \) is the impact velocity. Future investigations...
are worth putting forward in this direction and pinning down the uncertainties (Wyatt et al. 2019).

If the cosmic shoreline is real, this empirical law might predict the existence of atmospheres on exoplanets. For example, the recently detected airless body LHS 3844 b (Kreidberg et al. 2019) lies above the cosmic shoreline (Fig. 2). However, there are some exceptions, such as Kepler 51 b and c, very low-density bodies but located above the empirical line, suggesting the cosmic shoreline might also depend on the age of the planet. A more massive, older planet, 55 Cancri e, is also an outlier. Both thermal phase curve observations (Demory et al. 2016b) and HST transmission spectra (Tsiaras et al. 2016) indicated a substantial atmosphere on 55 Cancri e. It would also be interesting to put the future yet-to-be-characterized habitable-zone terrestrial planets such as the Trappist-1 system in the diagram. Current observations can rule out the existence of significant hydrogen atmospheres on TRAPPIST-1 b and TRAPPIST-1 c (de Wit et al. 2016).

Zahnle & Catling (2017) found Proxima Centauri b and Trappist-1f are both near the cosmic shoreline (“on the beach”) and thus we cannot conclude the existence of their atmospheres at this moment. This coincidence is interesting because the known terrestrial planets with atmospheres, including the Earth with life on it, are all located close to the cosmic shoreline.

How did we detect an airless exoplanet? For tidally locked terrestrial exoplanets, an airless body could possess a higher amplitude in the thermal emission light curve due to little heat redistribution between the dayside and the nightside (Kreidberg et al. 2019). On the other hand, the presence of an atmosphere could naturally reduce the dayside thermal emission via cloud formation and heat redistribution (Koll et al. 2019) and also increase the planetary albedo (Mansfield et al. 2019). Until recently, we have found the first indirect evidence of an airless exoplanet LHS 3844 b (Kreidberg et al. 2019) applying the thermal IR light curves from Spitzer. Future observations will further narrow down the cosmic shoreline region’s width and profile the detailed shape of the stability zone among extrasolar terrestrial planets.

3.3 Planet Desert and Radius Gap

If a thick gas envelope is lost via atmospheric escape, the observed planetary size shrinks. If this process occurs commonly on a large number of planets, atmospheric escape might imprint itself in planetary size distributions as a function of insolation or orbital distance. The fractional mass-loss rate on close-in hot Jupiters is small—at around the 1% level (e.g., Hubbard et al. 2007b; Owen & Wu 2013)—and thus the radius change is tiny. On the other hand, smaller planets with lighter gravity could have a significantly large fractional mass loss. In extreme cases, the gas envelope can be completely stripped off, and a bare solid core is left behind. For planets smaller than Neptune, a few percent of hydrogen and helium loss in mass will significantly reduce the planetary size—a radius change that could be observable in the old planet population. Intuitively, one can expect a planet closer to the central star is smaller and denser, and that further away is larger and lighter. Statistically, one might also expect that the occurrence rate of short-period planets drops as the star-planet separation (or the orbital period) decreases.

Early studies have noticed negative correlation between orbital period and planetary mass or gravity for short-period planets and pointed out thermal escape as a possible cause (Mazeh et al. 2005; Southworth et al. 2007). With Kepler data, a deficit in the occurrence rate of close-in small planets is discovered (e.g., Youdin 2011; Szabo & Kiss 2011; Howard et al. 2012; Beauge & Nesvorny 2013; Petigura et al. 2013; Fressin et al. 2013; Burke et al. 2014; Silburt et al. 2015; Mulders et al. 2015; Thompson et al. 2018). It is also found that the inner planets are denser and smaller than the outer ones in multi-planet systems (Wu & Lithwick 2013a; Ciardi et al. 2013). However, for transit observations, uncertainties in star radii can greatly influence the planetary size estimation. Recently, using CKS data together with Gaia astrometry data, Fulton & Petigura (2018) and Fulton et al. (2017) measured the planetary radii down to the 5% precision level. These new data of short-period small planets (within 100d and below 10 Earth radii $R_E$) clearly exhibit two prominent features in the radius-orbit diagram (Fig. 3, from Fulton & Petigura 2018 and Petigura et al. 2018). The first one is the “Planetary Desert” of short-period large planets in the upper corner of the graph (e.g., Wu & Lithwick 2013a; Ciardi et al. 2013). The second one is an intriguing “radius gap” (or “valley”) of that planetary occurrence rate that declines diagonally with increasing semi-major axis. The gap occurs at 1.5–2 $R_E$, separating the planetary population into two groups: super-Earths (1–1.7 $R_E$) and sub-Neptunes (1.7–4 $R_E$).

The radius-orbit diagram with a “desert” and a “gap”—if the planets were not born to be so (for alternatives, refer to Zeng et al. 2019)—shows a strong signature of the atmospheric mass loss after formation. One possibility is the thermal escape via photoevaporation. In fact, the evaporation desert and gap have been theoretically predicted by atmospheric escape models in Owen & Wu (2013) and Lopez & Fortney (2013) and also investigated by subsequent photoevaporation studies (e.g., Jin et al. 2014; Howe & Burrows 2015; Chen & Rogers 2016; Owen & Wu 2017; Jin & Mordasini...
The planetary desert is easy to understand as a result of photoevaporation. For a given initial planetary mass and core mass, planets located closer to the star experience faster erosion, resulting in smaller sizes (e.g., Wu & Lithwick 2013a). Owen & Wu (2013) found that the shape of the upper envelope of the desert could be explained by the photoevaporation of 20 $M_E$ planets with rocky cores of masses 10–15 $M_E$. Simple scaling of the upper envelope can also be crudely estimated here. In the analytical model of Owen & Wu (2017), the atmospheric lifetime scales with multiple factors, including the orbital period, planetary radius and core radius. If the planets below the radius gap are bare cores, the core size is generally less than 2 $R_E$. Planets in the upper envelope of the distribution are also generally larger than twice the core size. In this regime, assuming a constant evaporation timescale, we can rearrange equation (4) in Owen & Wu (2017) and obtain the dependence of the planetary radius on the orbital period:

$$\frac{R_p - R_c}{R_c} \propto P^{-0.83},$$

where $R_c$ is the core radius. An example of this scaling is depicted in Figure 3. Note that many factors could influence the upper envelope. Some of these planets in the “planet desert” are likely to be bare giant cores (e.g., the recently discovered TOI-849 b, Armstrong et al. 2020). However, some might hold a significant gaseous envelope including several very-low-density planets (e.g., “Super-puffs”) and hot Neptune-sized planets around high-metallicity stars (“Hoptunes”, Dong et al. 2018b). Metallicity also plays an important role in the atmospheric cooling process and the escaping mass loss on short-period exoplanets.

The mechanism of a radius gap is less intuitive than the planet desert. The black rectangular box in Figure 2 signifies the region where small exoplanets around the “radius gap” lie in the “cosmic shoreline” plot. The planets with size between 1 and 6 Earth radii and orbital period between

---

**Fig. 3** Size-period distributions of close-in small exoplanets (dots and contours from Fulton & Petigura 2018 and Petigura et al. 2018) for all star types. The points represent the California-Kepler Survey (CKS) planets. Color indicates the relative planetary occurrence rate (normalized by the maximum value) calculated from the simulated 110,733 planets in a sample of 1,000,000 Sun-like stars in Table 9 in Petigura et al. (2018). Note that the number of planets per 100 stars per bin depends on the interval size in the period-radius plane. For reference, the maximum occurrence rate is about four planets per 100 stars having periods within 0.125 dex of 40 d and radii within 0.05 dex of 2.5 $R_E$. The top white line indicates an analytical curve of the planet desert from Eq. (4) with $R_c = 1.5 R_E$, above which planetary occurrence is rare. The lower white line signifies the predicted photoevaporation gap scaling from Owen & Wu (2017): $R_p \propto P^{-0.25}$ (Eq. (5)). Below the line, the planets are assumed to be bare cores in theory, i.e., no H$_2$-He atmosphere, but the theory seems to overestimate the slope of the radius gap from the observations (Van Eylen et al. 2018; Martinez et al. 2019). The ranges of escape velocity and insolation level of the planets around the radius gap are depicted with the rectangular box.
One and 300 Earth days have escape velocity between 10 and 30 km s$^{-1}$ and insolation fluxes between 1 x and 2000 x the Earth’s value. Nevertheless, the “cosmic shoreline” idea merely indicates a boundary line instead of a gap. Owen & Wu (2017) and Owen (2019) elaborated that the radius gap originates from the nonlinear dependence of the mass loss timescale to the envelope mass (see fig. 4 in Owen 2019). The mass loss timescale reaches a maximum (i.e., slowest erosion) when the envelope doubles the core radius (hydrogen mass fraction is about 1%) and creates a stable sub-Neptune group (1.7–4 $R_E$) in the radius-period diagram. Below this critical point, the mass loss timescale drops very quickly below the Kelvin-Helmholtz timescale (in a few Myrs) so that the envelope can be completely stripped, and thus the “evaporation gap” emerges. The bare-cores are thus left behind to form the observed short-period super-Earths group (1–1.7 $R_E$) in Figure 3. Beyond the maximum point, as the envelope mass further increases, the planetary size increases so quickly—and so do the received XUV flux and the mass-loss rate—that it overwhelms the increase in the envelope mass. As a result, the mass loss timescale decreases towards a minimum where the envelope mass matches the core mass. Beyond this minimum point, compression of the atmosphere due to self-gravity maintains a roughly fixed planetary size so that the mass loss timescale increases again. Although the underlying physics of the evaporation valley looks similar to the “cosmic shoreline,” the general cosmic shoreline theory in Zahnle & Catling (2017) does not imply the existence of a clear gap around the regime boundary due to a lack of consideration of the nonlinear behavior of the mass loss timescale. Also, there is a subtle difference. The “cosmic shoreline” separates the planets with atmospheres and airless bodies, whereas the “radius gap” separates the planets with and without primordial hydrogen-helium envelopes. The planets below the gap are not necessarily completely airless but could also possess a significant amount of outgassed secondary atmospheres such as water and carbon dioxide after the early photoevaporation. So-called “bare cores” should be more appropriately understood in the sense that their atmospheres are too thin to have an important impact on their observed radii.

One can also analytically approximate the slope of the radius gap. The bare-core boundary of the evaporation gap might just be another version of the “cosmic shoreline” with a fixed atmospheric composition (like H$_2$-He) in the high XUV scenario in Zahnle & Catling (2017). Following the previous scaling of the “cosmic shoreline” in the XUV case, the energy-limited escape formula (Eq. (2)) gives the fractional mass loss rate that is proportional to $L_{XUV} R_p^3 / M_p^2$. For a solid bare core, the mass-radius relationship (Lopez & Fortney 2014) yields $M_p = \rho R_p^3$. From Kepler’s third law, $L_{XUV} \propto L_{HE} / a^2 \propto P^{-4/3}$ where $P$ is orbital period, one can obtain the size-version of the “cosmic shoreline”: $R_p \propto P^{-4/15}$. If we use the “photon-limited” mass loss rate $\dot{M} \propto L_{XUV} R_p^2 / M_p^2$ from Wang & Dai (2018), the “cosmic shoreline” scaling becomes shallower, $R_p \propto P^{-2/9}$. Both estimates are not very different from the scaling in Owen & Wu (2017) from a detailed treatment on the physics of core and envelope mass evolution

$$R_p \propto P^{-0.25}.$$  

Qualitatively, these analytical scalings indicate a decrease of the transition radius as orbital radius increases, consistent with the gap in the current Kepler-CKS sample (Fig. 3). Quantitatively, the radius gap slope derived from a statistical regression of the observational data (Van Eylen et al. 2018) follows $R \propto P^{-0.09}$. Martínez et al. (2019) reported a similar slope $R \propto P^{-0.11}$. Both are much shallower than the analytical estimates above. Note that the current data showing the radius gap only include short-orbit exoplanets, and hydrodynamic escape might not work well for planets with a period larger than 30 d (Owen & Jackson 2012).

We also emphasize that all the systems are “fossils” that record XUV from an earlier time. Given that most XUV photons were emitted when the star was very young, $L_{XUV}$ is probably not a constant and dependent on the stellar type. Thus the properties of the radius gap are probably different around different types of stars. Future statistics on cooler terrestrial exoplanets are needed to unveil more details.

Besides photoevaporation that takes action after the protoplanetary disk dissipation, alternative hypotheses have been put forward to explain the radius gap. The first one is the core-powered mass loss (Ginzburg et al. 2018; Ginzburg et al. 2016; Gupta & Schlichting 2019, 2020). This mechanism argues that the core luminosity released from the cooling of its primordial energy from planetary formation could drive the atmospheric escape for Gysrs, even without photoevaporation. The core-powered mass loss could also explain the observed radius gap slope (Ginzburg et al. 2018; Gupta & Schlichting 2019). The second hypothesis claims that the radius gap is a natural result of planetary formation pathways—planets above the gap are water-worlds and the ones below are rocky (Zeng et al. 2019). The last one is the impact erosion by planetesimals—planets below the gap were bare cores with their primordial atmospheres stripped away, and the ones above the gap grow enough volatiles to form secondary atmospheres (Wyatt et al. 2019). The impact erosion could not only explain the cosmic shoreline (Zahnle & Catling...
but also reproduce the radius gap, although the details need to be further investigated (Wyatt et al. 2019).

Is the radius gap a result of “nurture” (i.e., photoevaporation, impact erosion or core-powering mass loss) or “nature” (i.e., born to be, late formation in the gas-poor environment)? It is not easy to distinguish these hypotheses. As mentioned above, the observed slope appears shallower than the analytical scalings (Van Eylen et al. 2018; Martínez et al. 2019). In theory, if the evaporation efficiency changes with orbital distance and other factors, the predicted slopes could be different (e.g., Mordasini 2020). The slope is negative instead of positive (Van Eylen et al. 2018) which seems to suggest that the stripped cores do not form in a gas-poor environment after the disk dissipation. In the latter scenario, Lopez & Rice (2018) predicted a positive slope, although no impact delivery or erosion was considered. Investigations on the details of planetary accretion and evolution processes in the disk environment will help improve understanding of the planetary desert and radius gap such as effects of the core mass and compositions (e.g., Owen & Wu 2017; Jin & Mordasini 2018; Mordasini 2020) and stellar and disk metallicity (e.g., Owen & Murray-Clay 2018; Gupta & Schlichting 2020). For example, in the photoevaporation scenario, the radius gap should exhibit a trend with early high-energy emission of stars. However, the current analysis in Loyd et al. (2020) does not show a correlation between the radius gap and stellar activity in near-UV emission. As mentioned before, it is the XUV flux in the early stellar history rather than the current XUV flux that matters for the escape rate. To date, uncertainty in the XUV/X-ray history is large. One can eliminate this uncertainty by analyzing multi-planetary systems. Recent work by Owen & Campos Estrada (2020) found that the current dataset is consistent with their photoevaporation model, with a few exceptions. Moreover, future observations on the atmospheric compositions might also provide clues. For example, it was suggested that planets close to the upper boundary of the radius gap, i.e., the smallest ones in the sub-Neptune population, could have helium-rich atmospheres due to preferential mass loss of hydrogen over helium during photoevaporation (e.g., Hu et al. 2015; Malsky & Rogers 2020).

4 ATMOSPHERIC THERMAL STRUCTURE

4.1 Fundamentals

The equilibrium temperature of a planet depends on the incoming stellar flux, bond albedo and emissivity of the surface and the atmosphere. The atmospheric albedo and emissivity are controlled by the composition in the atmosphere, especially clouds (e.g., Marley et al. 1999). The current dataset of close-in gas giants does not suggest any correlation between the inferred albedo and other planetary parameters (Cowan & Agol 2011; Heng & Demory 2013; Schwartz et al. 2017; Zhang et al. 2018; Keating et al. 2019) although it appears that their bond albedos are systematically low as expected for hot, cloud-free atmospheres (Cowan & Agol 2011). Temperature distribution in the atmosphere is controlled by energy sources, sinks and transport processes. External energy sources on exoplanets include various processes such as stellar irradiation, high-energy particle precipitation and magnetic Ohmic heating. The primary internal energy source on gaseous planets and brown dwarfs is the heat release from gravitational contraction. Geothermal heat from radioactive decay is usually negligible. Deuterium burning is briefly important for young, less massive brown dwarfs (e.g., Burrows et al. 1997; Spiegel et al. 2011). The atmosphere mainly cools down through thermal emission to space. Atmospheric loss processes, such as escape and condensation, can also change the bulk energy of the atmosphere. Energy transport processes fall into three primary types: dynamics, radiation, and conduction. Among all dynamical processes, convection is more important in vertical energy transport, while the horizontal energy transport is controlled by other processes such as large-scale circulation, small-scale eddies and waves, wave breaking and turbulent dissipation. Radiation and conduction are usually important in vertical rather than horizontal energy transport.

Convection represents a large overturning of bulk atmospheric mass and the associated thermal energy and gravitational potential energy. Vigorous convection can be considered as an adiabatic process. As a result, convection tends to vertically smooth out the entropy in the atmosphere, or potential temperature \( \theta \), which is defined as \( \theta = (p/p_0)^{(\gamma - 1)/\gamma} \) where \( p_0 \) is a reference pressure. If the bulk vertical velocity scale is \( w \), the convective timescale is \( \tau_{\text{conv}} \sim H/w \), where \( H = k_B T/m g \) is the scale height of the convective atmosphere. It would be useful to analyze the static stability of the atmosphere, which can be measured by the buoyancy frequency, or Brunt-Väisälä frequency \( N \), given by

\[
N^2 = g \frac{\partial \ln \theta}{\partial z}.
\]

An atmosphere with low static stability (\( N^2 < 0 \)) tends to be convective. The temperature gradient follows the dry adiabatic lapse rate \( dT/dz = g/c_p \) in a dry atmosphere, and follows a shallower moist adiabat including the latent heat release in condensation. If vertical compositional gradient exists (i.e., lighter molecules on top of heavier molecules), tropospheric convection might not behave in a simple Rayleigh-Bénard type. Instead, double-diffusive convection and fingering might occur to result in...
less heat transport efficiency and steeper vertical temperature profile in the deep atmosphere (e.g., Stevenson 1979a; Guillot 1995; Leconte & Chabrier 2012; Tremblin et al. 2015; Leconte et al. 2017; Tremblin et al. 2019).

Radiative energy is transferred via photon exchange among atmospheric layers. Radiation will drive the vertical profile of temperature to radiative equilibrium so that radiative heating and cooling balance each other in each layer. The radiative timescale depends on the temperature and opacities of gas and particle constituents in the atmosphere. In the optically thick limit, usually applicable to the deep atmosphere, thermal radiation can be approximated as a diffusion process. In the gray limit—the atmospheric opacity does not depend on wavelength—the radiative timescale $\tau_{\text{rad,}\infty}$ can be treated as the diffusive timescale of temperature

$$\tau_{\text{rad,}\infty} \sim \frac{p^2 c_p \kappa_R}{g^2 \sigma T^4},$$

(7)

where $\kappa_R$ is the Rosseland mean opacity. $\sigma$ is the Stefan-Boltzmann constant. Here we omit the prefactor close to unity. We can simplify the radiation as a cooling-to-space process in the optically thin limit for the upper atmosphere. The radiative timescale is

$$\tau_{\text{rad},0} \sim \frac{c_p}{\kappa_P \sigma T^4},$$

(8)

where $\kappa_P$ is the Planck-mean opacity. If $\kappa$ and $T$ are vertically constant, the radiative timescale is roughly constant (independent of pressure) in the upper atmosphere ($\tau_{\text{rad},0}$) but increases very rapidly with pressure in the deep atmosphere ($\tau_{\text{rad,}\infty}$). The reason is that, as the pressure increases towards the deep atmosphere, the mean free path of the photon decreases so quickly that the transfer efficiency decreases dramatically. The transition region between the two regimes occurs at the thermal emission level $p_e$ where the mean optical depth $\bar{\kappa}/g \sim 1$. Here $\bar{\kappa} = (\kappa R P)^{1/2}$. At this level, $\tau_{\text{rad},0} \sim \tau_{\text{rad,}\infty} \sim \rho c_p / g \sigma T^2$, which can be considered as the mean radiative timescale of the entire atmosphere (e.g., Showman & Guillot 2002).

Thermal conduction can smooth out the vertical temperature gradient by molecular collisions between adjacent atmospheric layers. It can also be regarded as a thermal diffusion process. The efficiency depends on the mean free path of the bulk gas components. The conduction timescale $\tau_{\text{cond}}$ is given by

$$\tau_{\text{cond}} \sim \frac{p c_P s_m}{g^2} \left( \frac{T}{kg m} \right)^{1/2},$$

(9)

where $s_m$ and $m$ are the mean cross-section and mean mass of the air molecule, respectively. The conduction timescale decreases quickly towards the top of the atmosphere as long as the gas remains collisional.

Crudely speaking, in a globally averaged sense, in the deep atmosphere where the photon mean free path is short and the radiative timescale is long, convection dominates the energy transport process. Temperature follows the adiabat in this region, and the potential temperature is homogenized vertically. Above the convective region, radiation dominates the atmosphere, and the temperature profile follows the radiative equilibrium. In the upper atmosphere where the density is so low that thermal conduction becomes efficient, energy is transported through molecular collision, and the temperature (not potential temperature) gradient tends to be smoothed out. Other dynamical processes could also be important. For instance, wave energy deposition could also heat or cool the upper atmospheric region where waves break, as has been suggested for Jupiter’s thermosphere (Yelle & Miller 2004) and brown dwarf WISE 0855 (Morley et al. 2018). Meridional circulation and waves could also transport the energy from the polar auroral region to the equator on Saturn (e.g., Brown et al. 2020). The vertical temperature profile will be estimated analytically and discussed in Section 4.2.

Vertical structures of temperature and compositions can be obtained from the observed spectra, while the horizontal distribution could be inferred from the light curves. For transmission spectra, transit depth is determined by the line-of-sight optical depth, from which one can invert the vertical optical depth, and density profile of the species through inverse Abel transform (e.g., Phinney & Anderson 1968), and thus derive the temperature from the density profile. However, it is challenging to resolve degeneracies, such as that between the temperature and mean molecular weight (Griffith 2014). For thermal emission spectra, temperature retrieval is a non-trivial inversion problem. The basic principle is that the thermal emission at different wavelengths is sensitive to different vertical layers because the atmospheric optical depth is different. If the spectral resolution is sufficiently high, a vertical profile of temperature can be inverted from the thermal emission spectra. However, in reality, this problem is often ill-defined mathematically due to a finite number of data points, leading to non-unique solutions.

In the last decade, several successful inversion models have been developed to retrieve information on the transmission and emission spectra of exoplanets and brown dwarfs. The grid search method (e.g., Madhusudhan & Seager 2009) is usually computationaly expensive. Bayesian retrieval approaches are widely used, including different techniques such as optimal estimation gradient-descent (e.g., Line et al. 2012, 2013; Lee et al. 2012; Barstow et al. 2017), nested sampling (e.g., Benneke & Seager 2013; Benneke 2015; Waldmann et al. 2015b; Todorov et al.
Regardless of the technical details, the uncertainties of the retrieved temperature and composition primarily come from the following uncertainty sources: measurement uncertainty, relatively low spectral resolution in current data, finite wavelength coverage in the current instruments, laboratory data uncertainties of the optical properties of gas and particles including opacities (e.g., absorption line strength and spectral line shape parameters such as line-broadening width), single scattering albedo and scattering phase functions, abundance uncertainties of the radiatively active gases and particles that affect the observed spectra, and the uncertainties in the forward model of radiative transfer applied in retrieval. Among these factors, (1) and (2) can be improved through more advanced observational techniques and larger telescopes to enhance the S/N; (3) requires instruments with a wider range of IR wavelengths; (4) needs to be significantly improved in future laboratory experiments; (5) depends on the chemistry and microphysics in the atmosphere but the gas and particle abundances can also be jointly retrieved with the temperature profile from observed spectra, and (6) depends on assumptions of the physics and chemistry in the radiative transfer forward model, such as whether the atmosphere can be assumed to be horizontally homogenous (e.g., Line & Parmentier 2016; Feng et al. 2016; Pluriel et al. 2020; Taylor et al. 2020), if the geometry assumes the plane parallel case or a spherical shell (e.g., Caldas et al. 2019), how to treat the gas and particle scattering (Fisher & Heng 2018), and if the non-LTE effect is important. Non-LTE is particularly important for a high-temperature and low-density medium, although the claimed detection of non-LTE emission features on hot Jupiters (e.g., Swain et al. 2010; Waldmann et al. 2011) is still controversial (e.g., Mandell et al. 2011). The inclusion of the non-LTE effect also requires lab information on the collisional deactivation rates of the vibrational energy levels for important radiatively active species.

Longitudinal information of substellar atmospheres is primarily obtained from light curve observations. To date, the most useful data to infer the temperature distribution in the photosphere are emission light curves. The horizontal temperature distribution is primarily controlled by inhomogeneously distributed external energy sources, atmospheric thermal emission and horizontal heat transport. The internal energy release through convection might be higher in the polar region (in the direction of the rotational axis) than the equatorial region on a fast-rotating body (e.g., Showman & Kaspi 2013). However, the resultant temperature difference is much smaller than that from external sources, such as the equator-to-pole temperature contrast due to incoming stellar irradiation. The influence of planetary rotation, eccentricity and obliquity on the external energy source distribution is also essential. Day-night temperature contrast tends to be larger on slowly rotating planets than on fast rotating planets. That can be characterized by the ratio of the radiative timescale to diurnal timescale and that to atmospheric dynamical timescale. In the tidally locked configuration, planets will exhibit permanent dayside and nightside. Seasonally varying temperature patterns are expected on eccentric-tilted planets (more discussion in Sect. 6.2).

Atmospheric dynamics transport heat via including large-scale circulation and small-scale eddies and waves. In the longitudinal direction, zonal jets efficiently redistribute heat between the dayside and nightside, leading to a small longitudinal temperature contrast in the jet region. On the other hand, substellar-to-anti-stellar circulation on slowly rotating planets or tidally locked planets will also redistribute the energy from the dayside to the nightside. The upper atmosphere on the nightside might also be warmed up by compressional heating due to a strong downwelling flow. This dynamical effect has been seen on the nightside of Venus (Bertaux et al. 2007) and might also be crucial on tidally locked exoplanets. Atmospheric waves such as Rossby and Kelvin waves play essential roles in transporting energy in the longitudinal direction. Meridional circulation cells and eddies/waves redistribute heat among latitudes. For instance, on terrestrial planets, Hadley-like circulation transports the excess net heating from the equatorial region to the higher latitudes, but the details could be tricky. Take the Earth as an example. A Hadley cell transports the gravitational potential ener-
gy poleward but sensible and latent heat equatorward (e.g., Shaw & Pauluis 2012). At mid-latitudes (e.g., in the Ferrell cell region), mid-latitude eddies from transient baroclinic waves are responsible for poleward heat transport (e.g., Vallis 2006).

In Section 4.2, we will first discuss important features in the vertical temperature profile, such as the RCB and stratospheric temperature inversion, as well as the influence of controlling factors such as external and internal heat flux and opacity distribution. Then we highlight the classification of close-in exoplanets using temperature inversion and demonstrate spectral statistical trends emerging from the CMD. We will discuss the thermal phase curve for close-in exoplanets and brown dwarfs in Section 4.3 and rotational light curves for self-luminous bodies such as directly imaged planets and free-floating brown dwarfs in Section 4.4.

4.2 Vertical Temperature Profile and Mid-IR Emission

Typical vertical temperature profiles in the atmospheres on planets in the Solar System and several exoplanets and brown dwarfs show several important features (4). First, most thick atmospheres are approximately in radiative-convective equilibrium, characterized by a convective region below and a radiative layer above. The two regions are separated by the RCB. Second, the thin atmospheres on Triton and Pluto are in radiative-conductive equilibrium. Third, some planets develop temperature inversion above their tropospheres, such as Earth, Titan, giant planets and some ultra-hot Jupiters, but some do not, including Venus, Mars, and brown dwarfs and some hot Jupiters. Here we elaborate on the underlying processes governing these behaviors.

To understand the RCB, let us first consider a dry, gray atmosphere with external flux from the top $F_{\text{ext}}$ and internal flux from the bottom $F_{\text{int}}$. The total luminosity of the planet is equal to the sum of the external (incoming stellar irradiation) and internal fluxes (self-luminosity). RCB should depend on the relative strength of the external and internal fluxes. For simplicity, we assume the temperature is roughly isothermal in the radiative zone above the RCB. Below the RCB, the convective atmosphere is characterized by an adiabat following $\nabla_{\text{rad}} = \partial \ln T / \partial \ln p = (\gamma - 1)/\gamma$, where $\gamma$ is the adiabatic index. For self-luminous exoplanets and brown dwarfs where the external flux is negligible, the RCB pressure is where the emission optical depth is unity, $p_{\text{rcb}} = g/\kappa_R$. In the presence of external heat flux, the emission temperature (skin temperature) of the atmosphere increases, and the RCB progresses to the deeper atmosphere. In this case, the RCB pressure level $p_{\text{rcb}}$ can be scaled as

$$p_{\text{rcb}} \sim \frac{g}{\kappa_R}(1 + \frac{F_{\text{ext}}}{F_{\text{int}}} \gamma/4(\gamma - 1)).$$

For self-luminous bodies, $F_{\text{ext}} = 0$, we obtain $p_{\text{rcb}} = g/\kappa_R$. For highly irradiated planets, external flux is much larger than internal flux. For an ideal diatomic gas in a dry atmosphere, $\gamma = 7/5$ and the RCB pressure is roughly proportional to $F_{\text{ext}}/F_{\text{int}}$ (precisely, to the $7/8$ power). Consequently, a factor of two change in the estimated internal heat will lead to a factor of two change in the RCB pressure. Because many hot Jupiters have inflated radii, their internal heat (entropy) might be much higher than a non-inflated Jupiter. Thus, their RCB should be located at a shallower level. Thorngren et al. (2019) pointed out that the internal heat fluxes of inflated hot Jupiters could be much larger than previous estimates, and thus the RCBs are located at lower pressure levels. It is even possible that future high-resolution observations can probe below the RCBs and detect the properties of the convective region of some irradiated gas giants.

In reality, the temperature profile in the radiative zone is usually not isothermal. For a gray, optically-thick atmosphere under radiative equilibrium, the temperature gradient $\nabla_{\text{rad}}$ can be calculated by radiation diffusion

$$\nabla_{\text{rad}} = \frac{\partial \ln T}{\partial \ln p} = \frac{3\kappa_R B}{16g\sigma T^4},$$

where $F_{\text{rad}}$ and $T$ are the radiative flux and temperature at pressure $p$, respectively. If the radiative-equilibrium temperature gradient $\nabla_{\text{rad}}$ is steeper than the adiabat $\nabla_{\text{ad}}$, the atmosphere will be convectively unstable. Therefore, RCB occurs at the pressure level where the two temperature gradients are equal. Including this more realistic consideration would further complicate understanding the RCB.

More importantly, the atmospheric opacity has a significant wavelength dependence. This “non-gray” effect is the key to understanding many features in the vertical temperature structure. First, because the radiative timescale and radiative temperature gradient strongly depend on the emitting flux and opacity (one can get some intuition from Eq. (11) although it is in a gray limit), the atmosphere might exhibit multiple RCBs with alternating radiative and convective zones (e.g., Fortney et al. 2007; Marley & Robinson 2015). Imagine a convective region below the first (top) RCB, and the atmosphere is optically thick. As the temperature increases with pressure, the atmospheric emission peak (e.g., the peak of the Planck function) shifts to shorter wavelengths. It the emission peak happens to overlap with a relatively transparent (low opacity) wavelength region, the atmospheric energy would be carried outward by radiation instead of convection. A second RCB emerges between an upper convective zone
(called the detached convective zone) and a deeper radiative zone. The radiative zone ceases when the atmosphere becomes optically thick again at a deeper level, where a third RCB forms. However, it is difficult to detect the second and third RCBs because they usually lie in very deep atmospheres.

The second important "non-gray" effect is the temperature inversion—increasing with decreasing pressure—in the radiative zone. To elaborate this effect, we adopt the double-gray (or semi-gray) atmosphere assumption, which assumes one gray opacity for the visible and one gray opacity for the IR. The globally averaged radiative equilibrium temperature $T$ can be expressed as a function of pressure $p$ for a classical Milne atmosphere (see derivation in the appendix in Zhang et al. 2013)

$$T^4(p) = \frac{3F_{\text{int}}}{4\sigma\pi} \left( \frac{2}{5} + \tau_{\text{IR}} \right) + \frac{3F_{\text{ext}}}{4\sigma\pi} \left[ 1 + \frac{\alpha}{6\alpha} \right] - \frac{\alpha}{6} E_2(\tau_{\text{vis}}) - \frac{1}{2\alpha} E_4(\tau_{\text{vis}}),$$

(12)

where $E_n(x) = \int_1^\infty e^{-xt}/t^n dt$ is the exponential integral (from the average over angles) and $\alpha = \tau_{\text{vis}}/\tau_{\text{IR}}$ is the ratio of the visible opacity $\tau_{\text{vis}}$ to the IR opacity $\tau_{\text{IR}}$ at pressure level $p$. $F_{\text{int}}$ and $F_{\text{ext}}$ are the internal heat flux and incoming stellar flux, respectively. See similar expressions in other works for pure absorption (e.g., Hubeny et al. 2003; Hansen 2008; Guillot 2010; Robinson & Catling 2012; Parmentier & Guillot 2014; Parmentier et al. 2015) and including scattering (e.g., Heng et al. 2012, 2014).

In this semi-gray framework, if the external flux is negligible, e.g., on brown dwarfs or free-floating planets, the only heat source is from the deep atmosphere. The optically thick lower atmosphere is characterized by low static stability and vigorous vertical mixing due to convection. Above the RCB, the radiative equilibrium temperature profile in the absence of external heat source should decrease with decreasing pressure (Eq. (12)). Also, see the gray atmosphere results (Eq. (11)) and the brown dwarf temperature profiles in Figure 4. Note that this does not mean temperature inversion is impossible in these atmospheres. Other processes than radiative transfer might be critical. For example, breaking of upward propagating waves from the deep atmosphere might deposit the energy and heat the upper atmosphere (e.g., Yelle & Miller 2004; Morley et al. 2018), although the detailed mechanism is complicated because gravity waves might also cool the upper atmosphere (Young et al. 2005).

If there is an external radiative forcing, the temperature profile could develop an inversion profile more easily. Examples are the thick atmospheres on Solar System planets (Fig. 4). In the simple expression in Equation (12), tem-

---

**Fig. 4** Typical temperature-pressure profiles (solid) on Solar System bodies, exoplanets and brown dwarfs. For exoplanets, we display typical profiles for a sub-Neptune (GJ 436 b), cooler hot Jupiter (HD 209458 b) and an ultra-hot Jupiter (WASP-189 b). For brown dwarfs, we show typical radiative-convective temperature profiles for a cooler late-T dwarf and a hotter mid-L dwarf. The dashed lines represent condensation curves of major condensable species. We assume solar metallicity for most condensates. We adopt volume mixing ratios of $10^{-5}, 2 \times 10^{-2}, 1$ and $1$ for $\text{H}_2\text{SO}_4$, $\text{CH}_4$, $\text{N}_2$ and $\text{CO}_2$, respectively.
perature inversion could occur if the visible opacity $\tau_{\text{vis}}$ exceeds the IR opacity $\tau_{\text{IR}}$, i.e., $\alpha > 1$. In this situation, local heating due to the absorption of incoming stellar energy in the visible band is so large that the atmosphere could not emit it away efficiently. As a result, an Earth-like stratosphere forms. If temperature inversion occurs, the static stability of the atmosphere significantly increases with height, and radiative mixing substantially weakens. The level where the temperature inverts is the “tropopause” because tropospheric dynamics such as convective mixing are prohibited below that level. Note that the tropopause and RCB in this context are different because the former and the latter do not always coincide at the same pressure level (e.g., Robinson & Catling 2014).

The nature of the tropopause is influenced by several physical constraints from radiation, dynamics and thermodynamics. From the radiation perspective (e.g., Manabe & Strickler 1964; Held 1982; Thuburn & Craig 1997, 2000; Robinson & Catling 2012, 2014), the tropopause height—to first order—is a solution of the temperature minimum that is consistent with the radiative equilibrium upper atmosphere and the vertically mixed entropy flux from the troposphere below. The lower boundary conditions, such as surface temperature and surface opacity, play an important role. Thermal inversion occurs above the RCB. Most thick atmospheres in the Solar System exhibit temperature inversion at approximately 0.1 bar (Fig. 4), a result related to the atmospheric IR opacity at the surface lying between 1 and 10 in those atmospheres (Robinson & Catling 2014). From the dynamical constraint (e.g., Schneider 2004), large-scale extratropical dynamics, such as horizontal transport of baroclinic eddies, play a dominant role in shaping the temperature profile in the extraterrestrial region and thus the tropopause height. This mechanism could be responsible for the latitudinal distribution of the tropopause height in Earth’s atmosphere. In the moist atmosphere where condensational species could saturate and form clouds, the thermodynamic constraint is as important as other factors for determining the tropopause height (e.g., Thompson et al. 2017). On habitable planets (e.g., Wordsworth & Pierrehumbert 2013), the saturation water vapor pressure in combination with the water vapor radiative cooling greatly affects the temperature profile and thus the tropopause height.

From the radiation constraint and the radiative equilibrium temperature profile (Eq. (12)), UV or visible absorbers are important to create the temperature inversion and stably stratified upper atmosphere. On Solar System planets, the absorbers could be ozone on Earth, methane on giant planets, and also haze particles on Jupiter and Titan (e.g., Robinson & Catling 2014; Zhang et al. 2015). For exoplanets, titanium oxides (TiO) and vanadium oxides (VO)—also major opacity sources that dominate the visible spectra of M-dwarfs—have been proposed to serve as stratospheric absorbers and might cause bifurcation of the temperature profile (e.g., Hubeny et al. 2003; Burrows et al. 2007; Fortney et al. 2008). As mentioned in Section 2.3, Fortney et al. (2008) systematically investigated the atmospheric temperature structures of hot giant planets and suggested that these planets could be classified into two categories: hotter “pM” planets and cooler “pL” planets. The “pM” planets show strong thermal inversion caused by the TiO and VO opacity in the upper atmosphere while the “pL” class does not. Other opacity sources could also lead to thermal inversions, such as sulfur-bearing haze particles (e.g., Zahnle et al. 2009). The other option to create thermal inversion is to have the major coolant vanishing quickly in the upper atmosphere, as in the case of water on ultra-hot Jupiters (e.g., Arcangeli et al. 2018; Parmentier et al. 2018; Lothringer et al. 2018).

However, observational evidence of thermal inversion on exoplanets has remained elusive for years. An isothermal atmosphere will naturally produce blackbody-like spectra. A temperature profile that decreases with height will generally display absorption features. On the other hand, a strong emission feature in the spectra is a possible signal of thermal inversion as it implies that the upper layers are emitting more photons—and thus might be hotter—than the underlying layers. Because of the low-quality thermal emission spectra in a limited range of wavelengths with contaminating star signals as well as strong degeneracy between temperature and atmospheric composition, searching and interpreting the specific emission features in exoplanet spectra have not been very successful (e.g., search for TiO by Sing et al. 2013 and Hoeijmakers et al. 2015). The presumably claimed thermal inversion on HD 209458 b (e.g., Burrows et al. 2007; Knutson et al. 2008) using Spitzer data was later found to be not convincing (e.g., Diamond-Lowe et al. 2014; Schwarz et al. 2015; Line et al. 2016) when more constraints were obtained from the HST Wide Field Camera 3 (WFC3) and ground-based photometry. Other canonical hot Jupiters such as HD 189733 b and WASP-43 b exhibit absorption instead of emission features in their thermal spectra, implying no thermal inversion (Grillmair et al. 2008; Kreidberg et al. 2014b; Stevenson et al. 2014; Line et al. 2016).

Among the ultra-hot Jupiters with equilibrium temperatures higher than 2200 K (see Sect. 2.3), three of them have recently been confirmed with temperature inversion. The observations include TiO and H$_2$O dayside emissions on WASP-121 b (Evans et al. 2017), CO emission on WASP-18 b (Sheppard et al. 2017) and TiO emission on WASP-33 b (Haynes et al. 2015; Nugroho et al. 2017). Another hot Jupiter HAT-P-7 b has also been suggest-
ed with atmospheric thermal inversion (Mansfield et al. 2018b) but no definitive emission feature has been confirmed yet. Some other ultra-hot Jupiters like WASP-12 b and WASP-103 b display absorption spectra that are consistent with blackbodies, indicating possible isothermal atmospheres (e.g., Arcangeli et al. 2018; Parmentier et al. 2018; Kreidberg et al. 2018). The absorbers responsible for thermal inversion on those ultra-hot Jupiters were proposed as TiO/VO (e.g., Arcangeli et al. 2018; Parmentier et al. 2018) or haze and soot particles (e.g., sulfur haze from Zahnle et al. 2009 as suggested for WASP-18 b by Sheppard et al. 2017), or metals such as Na, Fe and Mg. SiO, metal hydrides and continuous opacity like the H− ion (Lothringer et al. 2018, Kitzmann et al. 2018). Thermal inversion is also partly attributed to insufficient cooling of carbon monoxide in the upper atmosphere and H2O depletion due to thermal dissociation (Arcangeli et al. 2018; Parmentier et al. 2018; Lothringer et al. 2018). This suggests that thermal inversion might be common on ultra-hot Jupiters (Arcangeli et al. 2018). The lack of a TiO/VO feature in the spectra of WASP-18 b (Arcangeli et al. 2018) could be due to the thermal dissociation of TiO/VO (Lothringer et al. 2018), strong negative ion opacities such as H− or other metals, or an oxygen-poor atmosphere (Haynes et al. 2015). Thermal dissociation of hydrogen will also shape the day-night temperature contrast on those ultra-hot Jupiters, which will be discussed in Section 4.3.

A systematic investigation of the vertical thermal structure on a statistical sample is also possible. After the Spitzer telescope ran out of cryogen in 2009, the mid-IR bands centered at 3.6 and 4.5 microns have provided the majority of thermal emission observations on warm and hot exoplanets during their secondary eclipses. In cloud-free atmospheres (i.e., no hazes or clouds) on warm and hot H2-dominated planets, the primary opacity sources in the 3.6-micron channel are water and methane gases, while that at 4.5 microns is mainly carbon monoxide with some contribution from water vapor. Therefore, the two channels probe the thermal emission from two different vertical levels in the atmosphere, although the weighting functions (i.e., the contribution of each layer to the outgoing emission) from the two channels have some overlap. We can estimate the temperature of the main emission layer observed at each channel after translating the observed thermal fluxes to the brightness temperatures (TB) using the Planck function. For warm hydrogen planets hotter than 600 K, unless the C/O ratio is so large that the atmosphere is oxygen-poor, atmospheric chemistry naturally favors CO over CH4 (e.g., Moses et al. 2013b, see Sect. 5.2). A more considerable CO opacity than the CH4 implies that the 4.5-micron observation probes at a higher altitude than 3.6 microns. Thus, if TB at 4.5-micron observation is higher than that at 3.6 microns, it might imply a possible thermal inversion in the atmosphere.

Garhart et al. (2020) compiled 78 secondary eclipse depths for a sample of 36 transiting hot Jupiters in the warm Spitzer channels. Most of the planets have smaller brightness temperatures at 4.5 microns than at 3.6 microns. Exceptions include the ultra-hot Jupiters discussed above that show ratio greater than unity, indicating possible thermal inversion. This phenomenon is also consistent with the emission features such as TiO detected in the spectra of those planets. The Spitzer 3.6 and 4.5-micron data have also been applied in the thermal structure retrieval on ultra-hot Jupiters (e.g., WASP-18 b, WASP-103 b and WASP-121 b) and provide important constraints on the determination of thermal inversion in their atmospheres. Several cooler planets also manifest larger-than-unity ratios. HAT-P-26 b is an extreme example that exhibits the brightness temperatures of ~2000 K and ~1400 K (with large uncertainties) at 4.5 and 3.6 microns, respectively. On the other hand, a CH4-rich and CO-poor atmosphere (i.e., large C/O ratio) with the emission level higher at 3.6 than 4.5 microns could also explain the larger-than-unity [4.5]/[3.6] ratio if the atmospheric temperature is not vertically inverted.

Statistical analytics from Garhart et al. (2020) also suggest a weak trend in the ratio of the 4.5 micron TB to the 3.6 micron TB (e.g., [4.5]/[3.6]) as a function of equilibrium temperature (Fig. 5). Previous studies on a small sample of cooler bodies (e.g., Kammer et al. 2015; Wallack et al. 2019) suggested no evident trend of the [4.5]/[3.6] ratio with the equilibrium temperature, planetary-mass or metallicity. When combining with hotter 36 gas giants, the [4.5]/[3.6] ratio seems to increase with equilibrium temperature by 100±24 ppm per Kelvin across the range of 800–2500 K (e.g., Garhart et al. 2020). Figure 5 depicts the overall trend that seems deviated from blackbodies (horizontal dashed line). Despite the uncertainty in the trend slope, it seems the [4.5]/[3.6] ratio is smaller for colder planets and larger for hotter planets. The interpretation of this trend is puzzling as is not expected from current models (e.g., Burrows et al. 2006b; Fortney et al. 2008). This trend might imply that the temperature structure becomes more and more isothermal in the photosphere for hotter planets on which the metallicity plays some unknown roles. The 3D structure of the temperature distribution or haze/cloud particles in the atmosphere could also be possible reasons. More future observations need to confirm this trend and understand the detailed mechanism behind it.
4.3 Thermal Phase Curves on Tidally Locked Exoplanets

Thermal phase curves have been observed on close-in exoplanets, mostly on hot Jupiters. Even without fully resolved orbital phase data, the averaged dayside and nightside flux difference could provide important clues on the heat redistribution on these planets. Parmentier & Crossfield (2018) collected many data and provided a thorough discussion of observational techniques and potential problems. Most thermal phase curve observations come from Spitzer 3.6 and 4.5-micron bands. Some light curves are from the Kepler visible band, but the data have a significant reflection stellar component. Recently observed ultra-hot Jupiters show low albedos and their thermal phase curves can also be directly obtained from the TESS band, such as WASP-18 b (Shporer et al. 2019), WASP-19 b (Wong et al. 2020), WASP-121 b (Daylan et al. 2019; Bourrier et al. 2020), WASP-100 b (Jansen & Kipping 2020), KELT-9 b (Wong et al. 2019) and WASP-33 b (e.g., von Essen et al. 2020). Here we just highlight important trends on available data to date and try to summarize the underlying mechanisms into a simple, self-consistent framework.

To first order, thermal phase curves on close-in exoplanets usually exhibit a sinusoidal shape, characterized by two critical features: phase curve amplitude and phase offset from the secondary eclipse. The thermal phase curve primarily probes the horizontal temperature distribution with contributions from chemical distributions in the photosphere. The temperature pattern is mainly controlled by the day-night irradiation distribution and atmospheric dynamics. The permanent day-night radiative forcing sets the radiative equilibrium temperature distribution, while the atmospheric dynamics such as waves and jets redistribute the heat from the dayside to the night side and cause a deviation from equilibrium. Compared with the radiative equilibrium baseline, the regulated day-night temperature contrast decreases, and the longitude of the temperature maximum is shifted from the substellar point. Thus the light curve amplitude and phase offset can be utilized to diagnose the interplay between radiation and dynamics in the atmosphere. Non-uniformly distributed chemical species due to local chemistry, and dynamical transport will further complicate the analysis. The detailed atmospheric dynamics on close-in exoplanets will be discussed in Section 6.2. Here we present some simple scaling theories to elaborate the underlying mechanisms governing the horizontal temperature distribution to understand the thermal phase curves. We will discuss the phase curve amplitude and day-
night temperature contrast in Section 4.3.1 and then address the phase offset in Section 4.3.2.

4.3.1 Phase Curve Amplitude and Day-night Temperature Contrast

In principle, we can achieve scaling for the day-night temperature contrast by combining the horizontal momentum equation, thermodynamic equation and continuity equation (see appendix A in Zhang & Showman 2017). In steady state, the scaling equations can be represented as

\[
\frac{R \Delta T \ln(p_s/p)}{2L} \sim \frac{U^2}{L} + \frac{U}{\tau_{\text{drag}}} \quad (13a)
\]

\[
\frac{\Delta T_{\text{eq}} - \Delta T}{2\tau_{\text{rad}}} \sim \frac{wN^2H}{R} + \frac{qU}{c_pL} \quad (13b)
\]

\[
\frac{U}{L} \sim \frac{w}{H} \quad (13c)
\]

Here \( R \) is the gas constant in units of \( J \, kg^{-1} \, K^{-1} \). \( \Delta T_{\text{eq}} \) is the day-night temperature contrast under radiative equilibrium, and \( \Delta T \) is the actual contrast. \( \ln(p_s/p) \) is different in log-pressure from the deep pressure \( p_s \) where the temperature is horizontally homogeneous. This term approximates the layer thickness in the log-pressure coordinate. \( U \) and \( w \) are the typical horizontal and vertical wind scales, respectively. \( L \) and \( H \) are the typical horizontal length scale and pressure scale height, respectively. \( \alpha \) is the buoyancy frequency (Eq. (6)).

Equation (13a) is a scaling of the horizontal momentum equation in which we assumed a simple linear frictional term characterized by \( \tau_{\text{drag}} \). On a close-in exoplanet with a broad superrotating wind (i.e., eastward wind) pattern, the pressure gradient force balances the nonlinear inertial term, the Coriolis force and the drag force. This form essentially tries to combine different momentum balance regimes discussed in Komacek & Showman (2016). Equation (13c) is the scaling of the continuity equation in which the vertical divergence balances the horizontal divergence of the mass.

Equation (13b) is the thermodynamic equation where the horizontal advection of the temperature is assumed to be smaller than the vertical entropy advection, so-called “weak-temperature gradient (WTG) approximation” (Sobel et al. 2001). This assumption has been demonstrated to be valid for typical close-in hot Jupiters (Komacek & Showman 2016) and almost all cool terrestrial exoplanets (e.g., Pierrehumbert & Hammond 2019). In a traditional WTG framework, the radiative heating rate balances the vertical entropy advection. But on the right-hand side of Equation (13b), we include an additional heating source \( qU/c_pL \). This term collects several different possibilities other than the traditional radiative heating, such as the thermal dissociation of hydrogen on the dayside and recombination on the nightside. In this case, \( q = L_q \chi_H \) where \( L_q \) is the bond energy of hydrogen molecules, and \( \chi_H \) is the mass mixing ratio of the hydrogen atoms that recombine on the nightside and release energy. This mechanism was suggested by Showman & Guillot (2002) and has recently been considered in the context of ultra-hot Jupiters (e.g., Bell & Cowan 2018; Komacek & Tan 2018; Tan & Komacek 2019). Other possibilities have not been well investigated in the context of exoplanetary atmospheres, for another example, the photodissociation of species (mainly hydrogen) and recombination. This case is similar to the previous one except that \( \chi_H \) depends on UV intensity from the central star. Another possibility is downwellling compressional heating on slowly rotating planets in analogy with the upper atmosphere of Venus (e.g., Bertaux et al. 2007). In the pressure coordinate, this term should be regarded as the non-hydrostatic effect of adiabatic cooling and compressional heating. Also, dissociation of hydrogen changes the atmosphere’s molecular weight and leads to expansion cooling on the dayside. On the nightside, the recombination results in molecular weight increase and subsequent compression heating. This mechanism has been included in recent simulations of Tan & Komacek (2019) but was not quantified separately. Another heating mechanism could be shock heating with the dissipating KE associated with mean flows (e.g., Heng 2012). The final possibility is heat transport via interactions with the magnetic field, but which might be too complicated to be represented by a simple term like \( qU/c_pL \).

For a scaling theory, one can assume \( \Delta T_{\text{eq}} \sim T_{\text{eq}} \) for tidally locked planets and horizontal length scale \( L \) is the planetary radius \( R_p \). One can also use isothermal sound speed \( (RT_{\text{eq}})^{1/2} \), which differs from the adiabatic sound speed by a factor of \( \gamma \), to approximate the fastest gravity wave speed \( NH \) in the isothermal limit and the cyclostrophic wind speed induced by the day-night temperature difference in radiative equilibrium \( U_{\text{eq}} = (RT_{\text{eq}} \ln(p_s/p)/2)^{1/2} \). If we simply assume the depth of the temperature variation \( \ln(p_s/p) \sim 2 \), the dynamical timescales of both wave propagation and wind advection across the planet are comparable and can be approximated by \( \tau_{\text{dyn}} = R_p(RT_{\text{eq}})^{-1/2} \). The solution of the scaling equation set (13) is

\[
\frac{\Delta T}{\Delta T_{\text{eq}}} \sim 1 - 2\alpha_1^{-1}(\sqrt{1 + \alpha_2^2} - \alpha_2), \quad (14a)
\]

\[
\frac{U}{U_{\text{eq}}} \sim \sqrt{1 + \alpha_2^2} - \alpha_2. \quad (14b)
\]
The non-dimensional parameters $\alpha_1$ and $\alpha_2$ are defined as

$$\alpha_1 = \frac{\tau_{\text{dyn}}}{\tau_{\text{rad}}} (1 + \frac{q}{c_p T_{\text{eq}}})^{-1}, \quad (15a)$$

$$\alpha_2 = \alpha_1^{-1} + (\Omega + \tau_{\text{drag}}^{-1}) \tau_{\text{dyn}}/2. \quad (15b)$$

If $q$ is zero (i.e., no extra heating mechanism), the solution is consistent with that in Zhang & Showman (2017). Although the detailed dynamical and thermodynamical mechanisms are not elucidated in this simple scaling theory, the solution implies that the bulk atmospheric behavior such as the temperature and wind is governed by dimensionless numbers: $\Omega \tau_{\text{dyn}}, \tau_{\text{dyn}}/\tau_{\text{drag}}, \tau_{\text{dyn}}/\tau_{\text{rad}}$ and $q/c_p T_{\text{eq}}$. An additional number $\tau_{\text{vis}}/\tau_{\text{IR}}$ is important for the vertical temperature profile and radiative transfer (the term $\alpha$ in Eq. (12)). These five parameters highlight important processes of planetary rotation, wave dynamics, frictional drag, radiative transfer and (hydrogen) latent heat. The first three numbers $\tau_{\text{dyn}}, \tau_{\text{drag}}$, and $\tau_{\text{rad}}$ come from comparing the dynamical timescale (flow advection timescale) $\tau_{\text{dyn}}$ with that in other processes such as the rotational timescale $\Omega^{-1}$, drag timescale $\tau_{\text{drag}}$ and radiative timescale $\tau_{\text{rad}}$, respectively. Based on these numbers we can decouple the atmospheric dynamics on tidally locked exoplanets into several regimes in Section 6.2.

The first number $\Omega \tau_{\text{dyn}}$ is also equivalent to the inverse of the “WTG parameter” $\Lambda = c_q/\Omega R_p$ introduced in Pierrehumbert & Hammond (2019) for terrestrial planets. Here $c_q$ is the isothermal sound speed based on the potential temperature of the adiabatic region above the surface. $\Lambda$ can be considered as the ratio of the Rossby deformation radius to the planetary radius. For synchronously rotating terrestrial planets, scaling laws are different because the surface flux needs to be taken into account to estimate the heat redistribution between the dayside and nightside. Scalings in the rocky planet regime have been derived in previous studies (e.g., Wordsworth 2015; Koll & Abbot 2016; Koll & Abbot 2018).

The third dimensionless number $\tau_{\text{dyn}}/\tau_{\text{rad}}$ is particularly important in understanding the thermal phase curve on tidally locked exoplanets. A strong radiative relaxation tends to maintain the day-night thermal contrast towards the radiative equilibrium state, while atmospheric winds and waves redistribute the heat and reduce the thermal contrast. With other factors unchanged, both the radiative timescale $\tau_{\text{rad}} \propto T_{\text{eq}}^{-3}$ and the dynamical timescale $\tau_{\text{dyn}} \propto T_{\text{eq}}^{1/2}$ decrease with increasing temperature, but with a different dependence—the former decreases faster than the latter. Thus a hotter exoplanet tends to be more radiatively controlled, leading to a larger day-night temperature contrast. This trend has been confirmed by two-dimensional (2D) and 3D numerical simulations of hot Jupiters (e.g., Perez-Becker & Showman 2013; Komacek & Showman 2016; Komacek et al. 2017). The theory can also be applied to tidally locked planets in the habitable zone. Because of their relatively low temperatures, they should have small day-night temperature contrast in the free atmospheres, and thus are located in the WTG regime. 3D terrestrial climate simulations on those planets have also confirmed this behavior (e.g., Wordsworth 2015; Koll & Abbot 2016; Haqq-Misra et al. 2018; Pierrehumbert & Hammond 2019).

If we take $q$ as the latent heat, the fourth dimensionless number $q/c_p T_{\text{eq}}$ can also be considered as the inverse Bowen ratio (Bowen 1926). The Bowen ratio is the ratio of sensible heat flux (heat transfer flux between the surface and the atmosphere) to the latent heat flux and depends on details such as the temperature gradient and condensation processes. Here we generalize this concept to compare the latent heat with the atmosphere’s thermal energy utilizing $q/c_p T_{\text{eq}}$. This number is useful for diagnosing the importance of latent heat release from condensable species such as water and silicate (see Sect. 6.3).

Figure 6(A) summarizes the observed day-night temperature contrasts on tidally locked gas giants that were inferred from the amplitude of thermal phase curves (see summary in Komacek et al. 2017 and more data from Parmentier & Crossfield 2018). It looks like there might be an increasing trend with equilibrium temperature in the current sample (Fig. 6A), especially the 4.5-micron data. However, data uncertainties are too large to confirm the trend, which has also been pointed out in the day-night flux contrast in Parmentier & Crossfield (2018). The curves from the simple scaling theory are displayed mainly for illustrative purposes. In reality, these planets likely have different detailed properties such as temperature profile, opacity (especially clouds) and frictional drag. The realistic mechanisms should be more complicated than the discussion here (see Parmentier et al. 2016), and more observations with better data quality are needed for further analysis.

In order for the non-traditional heat source $q$ to take effect, $q$ needs to be comparable to or larger than the thermal energy $c_p T_{\text{eq}}$ (Eq. (15a)), which is about 0.6 eV ($c_p \sim 3.5 R$) for a 2000 K hot hydrogen atmosphere. The latent heat of most condensable species released from the intramolecular bonds is at the 0.1–10 eV level and is generally unimportant in hot, solar-metallicity atmospheres because the species is not abundant (see Sect. 6.3). On the other hand, hydrogen bond energy is $\sim 4.5$ eV. If a large fraction of hydrogen molecules are dissociated on the dayside and recombine on the night side, the heat release could exceed the thermal energy $c_p T$ by a factor of 10 or more. Bell & Cowan (2018) quantified this effect by calculating the hydrogen atom fraction due to thermal dissociation and
found that this mechanism is important for planets hotter than \( \sim 2500 \) K. Some recently characterized ultra-hot Jupiter such as WASP-33 b (\( T_{\text{eq}} \sim 2723 \) K, Zhang et al. 2018) and KELT-9 b (dayside \( T_{\text{eq}} \sim 4600 \) K, Gaudi et al. 2017; Wong et al. 2019; Mansfield et al. 2020) might fall into this regime as they show smaller day-night temperature contrast than expected. Komacek & Tan (2018) included this term in a scaling theory. They found that the ultra-hot Jupiters could have a lower day-night temperature contrast than the cooler ones. Figure 6(A) illustrates an example (dashed curve) if one includes the thermal dissociation and recombination in the scaling theory and the day-night temperature contrasts on planets in the ultra-hot Jupiter regime (\( T_{\text{eq}} \sim 2200 \) K) decrease with equilibrium temperature and seem to explain the day-night contrast of WASP-33 b and KELT-9 b qualitatively.

The simple scaling theory also implies a nonlinear dependence of the day-night temperature contrast on the equilibrium temperature in the ultra-hot Jupiter regime. After about 2200 K, the day-night temperature contrast first decreases with \( T_{\text{eq}} \) but increases again after \( T_{\text{eq}} > 3000 \) K. This is because the thermal dissociation has reached the limit beyond about 3000 K and atomic hydrogen dominates the entire dayside atmosphere in our sim-
ple scaling. In other words, as the atomic hydrogen fraction \( \chi_H \) increases with \( T_{eq} \) (Saha equation, Bell & Cowan 2018), the hydrogen latent heat term \( q = L_q \chi_H \) first increases but is saturated at about 3000 K when \( \chi_H \sim 1 \). Beyond this temperature, a hotter planet is more radiative dominated (see the previous discussion on \( \tau_{dyn}/\tau_{rad} \)) and the day-night temperature contrast increases with \( T_{eq} \) again. This nonlinear behavior is signified as the dashed curve in Figure 6(A), which needs to be confirmed by more realistic dynamical simulations and observational data for planets with \( T_{eq} > 3000 \text{ K} \). 3D numerical models have been applied to investigate the effects of hydrogen thermal dissociation and recombination (Tan & Komacek 2019).

The models indicate that hydrogen atoms produced by the thermal dissociation on the dayside mostly recombine at the terminators before being transported to the nightside. Although the nightside atmosphere also increases due to this mechanism, the terminators are heated significantly in a 3D model. Consequently, the decrease of day-night temperature contrast is mainly due to the dayside cooling rather than the nightside warming. The photodissociation of hydrogen (not included in current models) due to high-energy UV stellar flux might also be important but is probably limited to pressure level less than \( 10^{-5} \text{ bar} \) (see an example atomic hydrogen profiles in Moses et al. 2011). It looks unlikely to impact the photospheric temperature distribution. However, a strong magnetic effect might also occur on the zonal flow in this high-temperature regime, as the atmospheres should be partially ionized, the influence of which on the day-night temperature contrast for ultra-hot Jupiters has yet to be explored.

The dayside and nightside brightness temperatures can also be separately derived from well-characterized phase curves. Figure 7 displays 16 hot Jupiters and an irradiated (close-in) brown dwarf across a large range of \( T_{eq} \) at 3.6 and 4.5 microns from Spitzer (Keating et al. 2019; Keating et al. 2020; Jansen & Kipping 2020). Their dayside temperatures roughly scale linearly with the equilibrium temperature (assuming zero albedo) from 1000–3000 K, but the nightside temperatures remain the same (~1100 K) when \( T_{eq} \) changes from 1100 to 2500 K, and then increase after \( T_{eq} > 2500 \text{ K} \) (Keating et al. 2019). The hottest planet KELT-9 b (TESS channel data from Wong et al. 2017 and the 4.5 micron data from Mansfield et al. 2020) also seem to follow this trend. The different temperature trends in the two hemispheres can be roughly understood as a result of increasing heat redistribution efficiency between the day and nightsides as the planet gets hotter (Keating et al. 2019). Based on the scaling theory in Equation (13), we can further separate the dayside and nightside temperatures assuming \( 2T_{eq} = T_{day} + T_{night} \) and

\[
T_{day} - T_{night} \sim \Delta T/\Delta T_{eq} \\
T_{day} \sim (1 + \epsilon^4)^{1/4}T_{eq}, \quad (16a) \\
T_{night} \sim \epsilon T_{day}, \quad (16b)
\]

where \( \epsilon \) is the ratio of the nightside to dayside temperatures and defined as \( \epsilon = 2\alpha_1^{-1}(\sqrt{1+\alpha_2^2} - \alpha_2) \), and \( \alpha_1 \) and \( \alpha_2 \) are from Equation (14a). The scaling results demonstrate a decent explanation of the increasing trend of the dayside temperature with \( T_{eq} \) (Fig. 7). In fact, given that the dayside temperature is generally much larger than the nightside (\( \epsilon < 1 \)), the dayside temperature in Equation (16a) can be well approximated by \( T_{day} \sim 2^{1/4}T_{eq} \) in the limit of \( \epsilon^4 \sim 0 \). The predicted nightside temperature also stays roughly constant below \( T_{eq} = 2500 \text{ K} \), implying that the deposited stellar energy on the dayside is not efficiently transported to the nightside as the stellar flux increases. This is because the increasing dominance of radiation over wave dynamics as \( T_{eq} \) increases \( (\tau_{rad}/\tau_{dyn} \propto T_{eq}^{5/2}) \). The observed increasing nightside temperature after \( T_{eq} > 2500 \text{ K} \) implies that the redistribution efficiency increases again.

Including the hydrogen dissociation and recombination mechanism in the ultra-hot Jupiter regime can explain the nightside temperature trend better. Nevertheless, after about 3200 K, the nightside temperature decreases again in our simple theory because the heating efficiency has reached its maximum, which needs to be confirmed in future 3D models and observations. Note the explanation of WASP 121 b and KELT 9 b is not very good by our simple theory. Using the 3D GCM from Tan & Komacek (2019), Mansfield et al. (2020) could explain the day-night temperature difference of KELT 9 b although the phase offset was not explained. Again, we emphasize the analytical theories here are mainly for illustrative purposes because the theory is oversimplified without including details of the radiative transfer, tracer transport and opacity sources, in particular, clouds. For example, as stated before, 3D simulations from Tan & Komacek (2019) show that the hydrogen dissociation and recombination could not significantly heat the nightside atmosphere of ultra-hot Jupiters. Alternately, some studies (Keating et al. 2019, Beatty et al. 2019) also proposed that nightside thick clouds could mitigate the emission temperature variation across the equilibrium temperature range and help maintain the uniform brightness temperature on the nightside of cooler hot Jupiters (see Sect. 5.3 for discussion). Realistic 3D dynamical simulations with cloud formation and future observations at different wavelengths could elucidate the underlying mechanism through the analysis of temperature and cloud distributions as well as the wavelength dependence of the thermal emission.
It is also interesting to show a possible trend of day-night temperature contrast with the planetary orbital period (e.g., Stevenson et al. 2017). Two factors are influencing this trend. For a fixed star type, the orbital period correlates with equilibrium temperature. As these close-in exoplanets are expected to be tidally locked and synchronously rotating, the orbital period correlates with the rotation period. For a shorter-period planet, the equilibrium temperature is higher, and the rotation is also faster. From the scaling theory, one should also expect that a faster rotation and a stronger drag maintain a larger spatial temperature gradient. Also, a hotter planet tends to have a smaller heat redistribution efficiency. That implies that the day-night temperature contrast for a fixed stellar type is higher for a planet with a shorter orbital period. For a Sun-like star, one can approximate the equilibrium temperature as $T_{eq} = 1380(P/3)^{-1/3}$ K. Put in the analytical theory and Figure 6(B) illustrates a decreasing trend of the day-night temperature contrast with increasing orbital period. This conclusion is also consistent with recent simulations on tidally locked planets with very rapid rotation rates (Tan & Showman 2020b) that show larger day-night contrast (see more discussion in Sect. 6.2). However, one should also be cautious because the planets in Figure 6 are orbiting around different types of stars, and there is no clear correlation between the equilibrium temperature and orbital period in this sample. So if there is any trend, it might be more related to the rotation rate ($\Omega$) dependence instead of the temperature ($T_{eq}$) dependence. Future observations need to separate the two factors (i.e., $\Omega$ and $T_{eq}$) by employing more statistically significant data for planets around each stellar type.

4.3.2 Phase Offset

Thermal phase curves on tidally locked exoplanets usually exhibit phase offset. In the absence of dynamics, the hottest spot at the same pressure level is located at the substellar point, and the peak of the thermal phase curve occurs right at the secondary eclipse. Heat redistribution by atmospheric jets and waves shifts the hot spot away from the substellar point and leads to a phase offset of the light curve peak before the secondary eclipse. This behavior was first
predicted in a 3D atmospheric model (Showman & Guillot 2002) and later was detected in the observation of a hot Jupiter (Knutson et al. 2007). In a kinematic picture (e.g., Cowan & Agol 2011; Zhang & Showman 2017), the phase offset is controlled by the horizontal heat transport and radiative relaxation. Strong radiative damping maintains the horizontal temperature distribution in the equilibrium substellar-to-anti-stellar pattern, leading to a small phase offset in the thermal phase curve. A strong longitudinal heat transport would likely advect the hot spot away from the substellar point, thus increasing the phase offset. A more realistic analysis using a dynamical model in Hammond & Pierrehumbert (2018) found that the hot spot phase shift is a result of zonal flow Doppler shifting the stationary wave response. Strong damping reduces the forced wave response and brings the response in phase with the forcing while in a weak damping case, the Doppler shift by the zonal jet leads to a large phase offset. Quantitatively, the phase offset $\delta$ can be estimated based on the relative magnitude of the radiative timescale and dynamical timescale

$$\delta \sim \tan^{-1}(\tau_{rad}/\tau_{dyn}). \tag{17}$$

Zhang & Showman (2017) proposed a more complicated formula, but the idea is similar. The theory can explain the idealized 3D GCM results in Zhang & Showman (2017). Because the ratio $\tau_{rad}/\tau_{dyn} \propto T_{eq}^{5/2}$ decreases with the equilibrium temperature, the phase offset becomes smaller as the planets get hotter and thus more radiatively controlled. This scaling predicts a trend that generally agrees with the IR observational data (Fig. 6(C)). Exceptions will be discussed later.

For a fixed stellar type, the equilibrium temperature decreases with an increasing orbital period, and the phase offset increases with the orbital period. See an analytical curve assuming a Sun-like star in Figure 6(D) for illustration. However, there is no clear dependence of the phase offset on the orbital period for currently characterized planets orbiting around different types of stars (also see fig. 3 in Parmentier & Crossfield 2018). A larger size of the planet sample is needed to analyze the statistical properties for planets around each stellar type. In general, the analytical scaling in both Figures 6(C) and 6(D) predicts a larger phase offset than the observations because we have estimated the dynamical timescale based on the isothermal sound wave speed $\tau_{dyn} = R_p (RT_{eq})^{-1/2}$, and the heat transport might be overestimated, leading to a larger phase offset than in the real atmospheres.

Despite a large scattering in the current data, this simple, first-principle scaling seems to do a decent job to explain the first-order, systematical behavior of the thermal phase curves on tidally locked planets. However, the caveats of this scaling are also evident. First, it does not include the feedbacks between radiation and dynamics. It only considers the horizontal heat transport and neglects the vertical entropy advection that seems essential for many gas giants. 3D GCM simulations with a realistic radiation scheme overestimate the phase offset (Parmentier & Crossfield 2018), implying more complicated dynamics therein. For rapidly rotating tidally locked planets, the hot spot could also be shifted westward by the off-equatorial Rossby waves in addition to the eastward offset by the eastward propagating Kelvin waves and mean flow at the equator (e.g., Lee et al. 2020; Tan & Showman 2020b). These effects were not considered in the simple scaling Equation (17).

Second, this theory predicts the amplitude of the light curve should correlate with the phase offset—a flatter phase curve is associated with a larger phase offset. However, we do not observe a clear correlation between the day-night contrast and phase offset in the current dataset (not shown here). More precise observations with smaller errors are needed to unveil any potential correlation here. When including the reanalysis of Spitzer data of WASP-43 b from Mendonça et al. (2018), Beatty et al. (2019) further suggest that there is no clear trend in the phase offset as a function of $T_{eq}$. Instead, the atmospheric clouds might have significantly altered the thermal phase curves.

Third, there are some outliers. Some hot Jupiters show westward phase offsets (i.e., the peak of the light curves occurs after the secondary eclipse) at the thermal wavelengths. For instance, CoRoT-2 b exhibits a clear westward hot spot phase shift at 4.5 microns (Fig. 6(C), Dang et al. 2018). The possible westward phase offsets on HD 149026 b at Spitzer 3.6 microns is a bit suspicious (Zhang et al. 2018). Qatar-1 b might also show display westward offset at Spitzer 4.5 microns but the data are also consistent with zero (Keating et al. 2020). Moreover, HAT-P-7 b exhibits strong variability between the eastward and westward offset in the phase curve (Armstrong et al. 2016), as do Kepler 76 b (Jackson et al. 2019), WASP-12 b (e.g., Bell et al. 2019) and WASP-33 b (Zhang et al. 2018; von Essen et al. 2020). These first-order data-model discrepancies suggest several missing physical processes in the current understanding of the phase offset on those planets, such as non-synchronous rotation dynamics (Rauscher & Kempton 2014; Showman et al. 2015a), magnetic effects (Batygin & Stanley 2014; Rogers 2017; Hindle et al. 2019), partial cloud coverage in the east hemisphere of CoRoT-2 b or planetary obliquity (e.g., Rauscher 2017; Ohno & Zhang 2019a; Adams et al. 2019b).

Other than the considerable variation between eastward and westward offset, some ultra-hot Jupiters manifest a much larger eastward phase offset than expect-
ed. For example, WASP-33 b shows large phase shifts in both warm Spitzer channels, implying an arguably increasing trend of the phase offset with equilibrium temperature beyond some critical value (Zhang et al. 2018; von Essen et al. 2020). Again WASP-33 b is strange because the recent TESS observations from von Essen et al. 2020 also indicated a large westward offset (28.7±7.1 degrees, see Fig. 6(C)), implying a sign of time variability. Recently, Jansen & Kipping (2020) analyzed the TESS phase curves of WASP-100 b and reported an eastward hotspot offset of 71±2 degrees. It suggests that some new physics might affect the flow pattern in the ultra-hot regime, albeit a slightly cooler, irradiated brown dwarf KELT-1 b (Beatty et al. 2019) is also showing a puzzlingly large phase offset (Fig. 6(C)). The recently characterized KELT-9 b is also puzzling. While the large phase curve amplitude is expected in the high-temperature regime, the large phase shift (18.7 degrees) at 4.5 microns is not expected (Mansfield et al. 2020). The recent 3D model, including hydrogen dissociation and recombination (Tan & Komacek 2019), could explain the day-night contrast of this planet, but the model could not explain the large phase shift. The magnetic effect may play an essential role in the heat redistribution on this hot and ionized planet. More ultra-hot Jupiter observations are needed to fully reveal any possible statistical correlation of the phase offset with equilibrium temperature in the high-temperature regime.

Amplitude and phase offset on the observed thermal phase curves also depend on wavelengths, which probe different vertical layers in the atmospheres. Theoretically, both the radiative timescale and dynamical timescale change vertically, leading to different $\tau_{\text{rad}}/\tau_{\text{dyn}}$ at different layers and the resulting thermal phase curves at different wavelengths. Phase-resolved emission spectroscopy is promising to probe the phase shift of the phase curves among different wavelengths (Stevenson et al. 2014). Furthermore, in short wavelengths such as HST near-IR band 1.5 micron (Fig. 6(C)), the phase curves are influenced by the reflection of the stellar light (e.g., Shporer & Hu 2015; Parmentier et al. 2016). For the light curves observed in the Kepler band at 0.6 micron, significant westward phase offsets are detected on cooler planets (<1800 K), indicating brightest spots on the western hemispheres due to cloud reflection of the stellar light at short wavelengths (see purple dots in Fig. 6(C), Parmentier et al. 2016). The nightside and western hemispheres are the coldest regions on the tidally locked exoplanets, facilitating mineral and metal cloud formation there. For hotter planets (e.g., >2000 K), the thermal emission components dominate the thermal phase curves, and thus the peak phases are shifted before the secondary eclipse, even in the optical wavelengths. This has also been seen in recently observed phase curves of ultra-hot Jupiters using TESS (pink dots in Fig. 6(C), Shporer et al. 2019; Shporer et al. 2020; Daylan et al. 2019; Bourrier et al. 2020; Wong et al. 2019; Jansen & Kipping 2020). For example, the TESS phase curve of KELT-9 b (Wong et al. 2019) shows a smaller eastward phase shift (4.4 degrees) than the Spitzer band data (18.7 degrees, Mansfield et al. 2020), implying that the two wavelengths probe different vertical levels on that planet.

Non-synchronized rotation, orbital eccentricity and planetary obliquity could further complicate the thermal structure evolution and thermal phase curve behaviors. If the planet is orbiting far away from the star where the gravitational tidal effect is weak and the tidal circularization timescale is long, the planet is not expected to be tidally locked. Eccentricity is not easy to be damped by the gravitational tides as the planet migrates inward. The planetary rotation axis might also be misaligned with the orbital normal, resulting in non-zero planetary obliquities, as evidenced by many planets in the Solar System. In a multi-planet system, even the close-in planets might have non-zero obliquities in highly compact systems (Millholland & Laughlin 2019). Non-synchronized rotation induces a movement of the substellar point along the longitude and alters the diurnal cycle of the stellar forcing (e.g., Showman et al. 2015b; Penn & Vallis 2017, 2018). A fast rotation could also homogenize the longitudinal temperature distribution. The details of the dynamics will be discussed in Section 6.2.

Both the eccentricity and obliquity have large impacts on the observed thermal phase curve. Orbital eccentricity causes the “eccentricity season” in which the star-planet distance, and thus the atmospheric temperature, changes with the orbit phase. Planetary obliquity could also lead to a strong seasonal cycle due to the tilt of the rotation axis. Planets orbiting in a highly eccentric orbit sweep very fast near the periastron due to Kepler’s second law, resulting in a highly skewed thermal phase curve and some possible oscillation pattern due to the pseudo-synchronous rotation (i.e., Langton & Laughlin 2007; Lewis et al. 2010; Kataria et al. 2014). An extreme example of this case could be HD 80606 b with an eccentricity of ~0.93 (Kataria et al. 2013). On an oblique planet, the substellar point migrates back and forth between the northern and southern hemispheres in one orbit. As a result, the peak of the thermal emission in the phase curve varies from case to case, depending on the obliquity and viewing geometry (e.g., Rauscher 2017; Ohno & Zhang 2019a,b; Adams & Laughlin 2018). In some cases, the phase offset could occur after the secondary eclipse, i.e., westward phase shift. Ohno & Zhang (2019b) provided an intuitive
understanding that the summed fluxes control the light curve shape from the shifted hot spot projected onto the orbital plane and the pole heated at the summer solstice. 

Adams & Laughlin (2018) developed a thermal radiative model to explore the full phase light curves and suggested high-obliquity signatures might be linked to the recently detected abnormal phase offset signals on some exoplanets such as WASP-12 b, CoRoT-2 b and (possibly) HD 149026 b. Future observations on those abnormal signals might be useful to constrain the planetary obliquity and eccentricity.

4.4 Rotational Light Curves on Brown Dwarfs and Directly Imaged Planets

For directly imaged exoplanets and brown dwarfs, disk-integrated photometric modulation has been observed and studied for two decades, starting from Tinney & Tolley (1999) shortly after the first detected brown dwarfs (see Biller 2017 and Artigau 2018 for a more detailed review). Rotational light curves are not only useful for constraining the rotational rates, but also the weather on those worlds. Regarding the light curve behaviors, there are two aspects. The first one is photometric light curves in emission caused by self-rotation and spatial heterogeneity. The amplitude of the light curves ranges from sub-percent to tens of percent on brown dwarfs (e.g., Artigau et al. 2009; Metchev et al. 2011; Radigan et al. 2012; Apai et al. 2013; Heinze et al. 2013; Yang et al. 2014a; Radigan 2014; Metchev et al. 2015; Heinze et al. 2015; Buenzli et al. 2015; Lew et al. 2016; Yang et al. 2016a; Miles-Páez et al. 2017a; Apai et al. 2017; Manjavacas et al. 2017; Zhou et al. 2018; Lew et al. 2020) and on directly imaged exoplanets (e.g., Biller et al. 2015; Zhou et al. 2016; Lew et al. 2020; Manjavacas et al. 2019a,b; Miles-Páez et al. 2019; Zhou et al. 2019, 2020). The shapes of the light curves are sometimes irregular rather than a simple sinusoidal shape, with more than one peak in the curves within one rotation. The second one is the temporal variability of the rotational light curves in both amplitude and shape (e.g., Apai et al. 2013; Yang et al. 2016a; Apai et al. 2017). There are both short-term and long-term variabilities, associated with the weather change in the photospheres. Apai et al. (2017) summarized three important puzzling behaviors in the temporal variability: (1) single-peaked light curves splitting into double-peaked (e.g., Radigan et al. 2012; Yang et al. 2016a), (2) rapid transitions from very low-amplitude (<0.5%) to high-amplitude (~5%) (Yang et al. 2016a), and (3) recurring features embedded in the irregularly evolving light curves (e.g., Karalidi et al. 2015).

Observed light curves on brown dwarfs depend on the spectral type and the observed wavelength. It looks like almost all L and T dwarfs are variable with amplitudes larger than 0.2%. Early surveys with limited data sample could not conclude whether the fraction of objects showing rotational variability is uniformly distributed across the spectral type or not (e.g., Wilson et al. 2014; Radigan et al. 2014). Later statistical studies seem to support that brown dwarfs in the L/T transition region tend to exhibit stronger variability and higher amplitude than the objects outside the transition (e.g., Radigan 2014; Eriksson et al. 2019). L dwarfs with IR variability larger than 2% are generally limited within the red, low-gravity objects (Metchev et al. 2015). The wavelength dependence implies a pressure-dependent behavior in the photosphere because different wavelengths probe at different layers. The most noticeable phenomenon is the so-called “phase offset” between 0 and 180 degrees (e.g., Buenzli et al. 2012; Radigan et al. 2012; Biller et al. 2013; Yang et al. 2016a). Here the “phase” is loosely defined because the shapes of the light curves are not always sinusoidal, especially in high-resolution observations. The “phase offset” phenomenon just states that the peaks at different wavelengths occur at a different observational times in the same rotational period. Moreover, on some objects the phase lags correlate with the probed pressure level at different wavelength (Buenzli et al. 2012; Yang et al. 2016a). The dependence of the rotational variability on the gravity or rotational period is still an open question.

The mechanisms behind the observed behaviors of the light curves are not completely understood. The rotational light curve itself is a result of the inhomogeneous distributions of temperature and opacity sources, i.e., the weather patterns in the photosphere rotating in and out of the view to the observer (e.g., Apai et al. 2013; Zhang & Showman 2014; Crossfield 2015). The spatial distribution of the clouds is the leading hypothesis (e.g., Radigan et al. 2012; Apai et al. 2013, 2017) such as the “patchy clouds scenario” with completely depleted cloud holes (e.g., Ackerman & Marley 2001; Burgasser et al. 2002b; Marley et al. 2010; Morley et al. 2014) or the “thin-and-thick cloud scenario” (e.g., Apai et al. 2013). Dust-cloud break-up was also proposed to cause the L/T transition (see Sect. 5.3.1 for discussion about clouds on brown dwarfs). Alternatively, temperature could also vary with longitude due to moist convection (e.g., Zhang & Showman 2014; Tan & Showman 2017), cloud radiative feedback (Tan & Showman 2019; Tan & Showman 2020a), thermal wave propagation (Robinson & Marley 2014), trapped waves in the bands (Apai et al. 2017) and dynamical modulations due to upward propagating waves (Showman et al. 2019). It is also important to keep in mind that clouds and temperature are tightly coupled together due to condensation processes, radiative feedback and atmospheric cir-
Morley et al. 2014) conducted a comprehensive study. Most of these studies focused on reflectational modulation and light curve variability. Polarimetric measurements on brown dwarfs should be able to provide more clues on the underlying mechanisms of the rotational modulation and its time variability (e.g., Apai et al. 2013; Morley et al. 2014). On the other hand, cloud-free mechanisms have also been suggested to cause variability at the L/T transition. Tremblin et al. (2016) claimed that the brown dwarf variability could be a result of surface heterogeneity of carbon monoxide or temperature due to atmospheric waves although the details were not elucidated. If this were true, gas (e.g., CH$_4$) absorption bands should exhibit abnormal amplitudes in L/T transition objects. This seems not consistent with recent observations (e.g., Buenzli et al. 2015; Biller 2017).

How to test possible mechanisms underlying the rotational light curves? It is hard to spatially resolve the weather patterns on these distant objects except the very close ones, such as Luhman 16B, which has been mapped using the Doppler imaging technique (Crossfield et al. 2014), although the data are still much noisier than the bright stellar counterparts. To date, continuous monitoring of objects over multiple rotations is a successful method to break degeneracies in surface brightness distribution and time-evolution (e.g., Apai et al. 2019). Apai et al. (2017) compiled light curves of 32 rotations for six brown dwarfs from Spitzer, along with simultaneous HST time-evolving spectra for some of the rotations. The analysis showed that beating patterns of the planetary-scale waves—rather than large bright spots—modulate the cloud thickness in the zonal bands on L/T brown dwarfs and produce the rotational modulation and light curve variability. Polarimetric observations (e.g., Goldman et al. 2009; Miles-Páez et al. 2017b) often provide new insights on the oblateness of the body, cloud grain properties and atmospheric banding, but it was difficult to achieve sufficient sensitivity for brown dwarfs. Recently, Millar-Blanchaer et al. (2020) successfully detected polarized signals from Luhman 16AB. The degree of linear polarization is about 300 ppm for Luhman 16A and about 100 ppm for 16B. The data imply cloud patchiness and banded structures on 16A, but the interpretation for 16B is still ambiguous.

On the other hand, giant planets in the Solar System might provide clues because their rotational light curves can be understood together with corresponding global maps. For example, Jupiter and Neptune exhibit strong rotational modulations (e.g., Gelino & Marley 2000; Karalidi et al. 2015; Simon et al. 2016; Stauffer et al. 2016; Ge et al. 2019). The photometric variability also depends on the wavelength, with the amplitude ranging from less than one percent to tens of percent. With sufficient data, the atmospheric patterns and jet speed can be derived from rotational light curves (e.g., Karalidi et al. 2015; Simon et al. 2016). Most of these studies focused on reflective lights. Based on high-resolution spatial maps of Jupiter from UV to mid-IR (e.g., Simon et al. 2015; Fletcher et al. 2016), Ge et al. (2019) conducted a comprehensive study on Jupiter including both reflective light curves (UV and visible) and emission light curves (mid-IR). The peak-to-peak amplitudes of Jupiter’s light curves range from less than 1% up to 4% at most wavelengths, but the amplitude exceeds 20% at 5 microns. The rotational modulations originate mainly in the cloudless belts instead of the cloudy zones. Important discrete patterns responsible for the rotational modulation include the Great Red Spot (GRS), expansions of the North Equatorial Belt (NEB), patchy clouds in the North Temperate Belt (NTB) and a train of hot spots in the NEB. The temporal variation of the light curves is caused by periodic events in the belts and longitudinal drift of the GRS and patchy clouds in the NTB.

The thermal emission light curves on Jupiter shed light on brown dwarfs and directly imaged planets. There are two mechanisms found for modulating Jupiter’s light curves. For small rotational variability (i.e., 1% level in the mid-IR), the surface inhomogeneity is induced by the spatial distribution of temperature and opacities of gas and aerosols. On the other hand, the vertical distribution of clouds is important for the 20% variation at 5 microns. At this wavelength, the large photometric modulation is induced by holes in the upper clouds at wavelengths of atmospheric windows where the gas has little opacity. Note that all giant planets in the Solar System are zonally banded. Whether this is true for brown dwarfs and directly imaged exoplanets is still uncertain although some recent observations have shed light on it (e.g., Apai et al. 2017; Millar-Blanchaer et al. 2020). See Section 6.3 for discussion.

Most information obtained from the rotational light curves is the surface inhomogeneity across longitude. However, the latitudinal properties of brown dwarfs and directly imaged planets can also be inferred, in the statistical sense, if we know their inclination angles to the observ-
er. The inclination angles of the brown dwarfs with known rotation periods can be derived from the line-of-sight rotational velocities (e.g., Radigan et al. 2014; Vos et al. 2017). Both the light curve amplitude and $J-K$ color anomaly (i.e., after subtraction of the mean color in the same spectral type) correlate with the inclination. The viewing geometry might mostly explain the former because a higher inclined (pole-on) object should exhibit smaller rotational modulation since fewer features can rotate in and out of view. Also, the analysis of Jupiter implies that Jupiter has larger rotational modulation at lower latitudes (both at NEB and SEB) than the higher latitudes. If the brown dwarfs are also banded like Jupiter, this will contribute to the amplitude-inclination trend observed by Vos et al. (2017).

The latter behavior ($J-K$ anomaly vs. inclination)—an equator-on object tends to be redder—is interesting. Why does an object tend to be redder at lower latitudes? If the temperature is not systematically lower in the equatorial region, it is perhaps due to more clouds or larger cloud particle sizes forming at lower latitudes (e.g., Kirkpatrick et al. 2010; Vos et al. 2017). However, the physical mechanism for this cloud distribution has yet to be explored in detail. One possibility is surface gravity at the equator is smaller than the polar region on a fast-rotating, oblate object with a larger equatorial radius than the poles. Thus the cloud might extend to lower pressure levels (also depends on the mixing). Alternatively, if we again use Jupiter as an analog, Jupiter shows more ammonia clouds at lower latitudes (e.g., the equatorial zone and the south tropical zone). The reason appears to be related to the global circulation pattern in the deep atmosphere of Jupiter recently revealed by the Juno mission (e.g., Bolton et al. 2017). The microwave observations (Li et al. 2017), as well as the VLT radio observations (de Pater et al. 2019, also see Showman & de Pater 2005), found that ammonia gas is enriched in the equatorial region where thick ammonia clouds form, while it is largely depleted in the off-equatorial region, the mechanism of which is still unknown. If this mechanism (for a different kind of cloud) also occurs on those early-L and early-T brown dwarfs in Vos et al. (2017), it could also explain the observed reddening at lower latitudes. We will discuss more global dynamics on brown dwarfs and directly imaged exoplanets in Section 6.3, where we will see simulations on rapidly rotating brown dwarfs that indeed show lower emission at the equator (Fig. 15).

5 ATMOSPHERIC COMPOSITION

5.1 Fundamentals

A planetary atmosphere is mostly composed of gas molecules and atoms. Suspended solids and liquid particles, so-called aerosols, are also ubiquitous. In the hot upper atmosphere where ionization processes play a role, the plasma phase is a significant fraction. In principle, atmospheric compositions are determined by the accreting volatile materials during planetary formation and by subsequently atmospheric evolution. The primordial atmosphere (proto-atmosphere) is formed through the accretion of gas and dust from the forming disk environment and should be mostly composed of hydrogen and helium with minor constituents depending on the location and composition of the formation feeding zone. Significant amounts of ices in various forms can be accreted onto planets residing outside the snow lines. The subsequent evolution track is very different between small and large planets—how to define “large” and “small” is not very clear, perhaps related to the radius gap discussed in Section 3.3. In general, large planets (such as gas giants, ice giants and possibly sub-Neptunes) retain their primordial hydrogen and helium envelopes, whereas planets with the size of Earth or Mars lost their proto-atmospheres. The secondary atmospheres on small planets might be formed and evolve through either various ingassing and degassing processes after the loss of primordial atmospheres (Catling & Kasting 2017). Examples of degassing processes are outgassing from the magma ocean, volcanic eruptions on terrestrial planets and core erosion on giant planets. Examples of ingassing processes include surface weathering, subduction during plate tectonics on terrestrial planets, and helium rains on giant planets (e.g., Stevenson & Salpeter 1977). Exchange of chemical compositions between atmospheres and planetary interiors implies that the secondary atmospheric compositions are closely related to the redox state of planetary interiors. There are also tertiary processes to exchange the atmospheric composition with the space environment, such as atmospheric escape, stellar wind injection, asteroid and comet impacts, and late disk accretion in long-lived disks (e.g., Kral et al. 2020).

Atmospheric compositions provide clues on planetary formation and evolution processes. Three categories seem within observational reach. The first one is metallicity—the relative ratios of heavy elements to hydrogen. The second one is the ratio of carbon and oxygen. The third one is the ratio between the refractory materials (rocks and real metals) and the volatiles (e.g., ices). This ratio determines the planet type. The ratio of carbon to oxygen is particularly important for atmospheric chemistry as it affects the redox of the planets. In general, the elemental ratios change with distance in the protoplanetary disk, suggesting the bulk composition of the planets should also change if they were born at different locations in the disk. The C/O ratio might be inferred from atmospheric measurements, but that on gas giants in the Solar System is highly uncertain.
because the main oxygen-bearing species, H$_2$O, condenses as clouds, and CH$_4$ condenses on Uranus and Neptune (Atreya et al. 2020).

On the other hand, for high-temperature exoplanets, carbon and oxygen abundances in the photosphere can be retrieved directly from the spectra. However, to get the bulk composition in the entire planet, one still needs to assume that the elements are well mixed in the planetary interiors and atmospheres. This assumption could be violated due to interactions among transport, chemistry and phase change processes. Noble gases are typically important among all trace species because of their chemically inert and non-condensable nature. However, directly detecting the abundance of a noble gas except helium is very difficult without an entry probe. Helium, on the other hand, has been detected from the Helium 10830 Å triplet lines in possibly evaporating atmospheres on several exoplanets (see Sect. 3.1). There is also an issue with possible depletion of helium due to helium rain that could also dissolve neon (Niemann et al. 1996).

From transmission and emission spectroscopy on both transit and directly imaged planets, many species have been detected in substellar atmospheres, including several atomic species in the UV such as H, He, C, O, K, Na, Si, Mg, Ti, Fe, Ca, Li, and metallic oxides TiO, VO and AlO, as well as H$_2$O, CO, CH$_4$, NH$_3$ and HCN detected in the IR. Check Table 1 in Madhusudhan (2019) for the latest summary and references therein. On high S/N spectra of brown dwarfs, more metallic species have been detected, such as Rb, Cs, as well as the hydrides MgH, CaH, CrH and FeH (Kirkpatrick 2005). The currently inferred metallicities of substellar atmospheres from the observed abundances of photospheric H$_2$O, CO and CH$_4$ seem to show a decreasing trend with increasing planetary mass (Fig. 8). Also see Welbanks et al. (2019) for the individual mass-metallicity relations derived from each species such as H$_2$O, CH$_4$, Na and K. Despite large uncertainties in the data, the trend in Figure 8 seems consistent with four giant planets in the Solar System. The retrieved metallicities of several brown dwarfs are roughly consistent with the solar value (Line et al. 2017). This trend probably implies that smaller planets in general accreted less hydrogen and helium fraction during their formation across different disk environments.

Isotopic compositions are particularly useful in understanding the evolution of the atmosphere. Enhancement of the deuterium to hydrogen (D/H) ratio on Uranus and Neptune relative to Jupiter and Saturn by a factor of 2-3 indicates the icy giants accreted more deuterium-rich icy blocks in the protoplanetary disk (Hartogh et al. 2011; Atreya et al. 2020). The exceedingly large D/H ratio in Venus’ atmosphere is evidence of past atmospheric escape (Donahue et al. 1982; McElroy et al. 1982). Exchange processes between the interior/surface and the atmosphere can
be inferred from the isotopic signatures of helium, argon, carbon, oxygen, sulfur and so on. Detecting isotopes on exoplanet atmospheres is still difficult using current facilities, but it will be possible to infer the D/H ratio from CH$_3$D or HD in the mid-IR thermal emission spectra (Morley et al. 2019). To compare isotopic abundances between the planets and their formation environment, one must also understand the atmospheric isotopic compositions on stars (e.g., Crossfield et al. 2019).

Given the elemental abundances, atmospheric abundances are controlled by temperature, chemistry and transport processes. Equilibrium chemistry drives the atmosphere towards thermodynamical equilibrium, given a sufficiently long time. Disequilibrium processes—including photochemistry, ion chemistry, biochemistry (for life-bearing planets) and phase change—force the atmosphere out of thermodynamic equilibrium. Atmospheric tracer transport by winds, waves and turbulences also results in chemical disequilibrium. In one-dimensional (1D) chemical models, vertical transport is conventionally approximately by a diffusion process (Andrews et al. 1987; Yung & DeMore 1998), the strength of which is characterized by vertical eddy diffusivity ($K_{zz}$). The chemical transport timescale is $\tau_{\text{trans}} = H^2/K_{zz}$, where $H$ is usually taken as the pressure scale height. Figure 9 illustrates several $K_{zz}$ profiles that were empirically determined for Solar System planets, as well as the theoretical predictions from models of some typical exoplanets. There seems no obvious trend for the $K_{zz}$ within Solar System planets, but hotter exoplanets may have larger diffusivities than colder Solar System planets.

The diffusive approximation generally works well for 1D models but has some caveats because the physical underpinning of $K_{zz}$ is elusive. There are approximately three regimes from the bottom of the atmosphere to the top. In a bottom convective atmosphere, vertical transport probably is well-approximated by eddy diffusion according to the traditional Prandtl mixing length theory (e.g., Prandtl 1925; Gierasch & Goody 1968; Gierasch & Conrath 1985; Smith 1998; Ackerman & Marley 2001; Bordwell et al. 2018). $K_{zz}$ is approximately a product of a convective velocity and a typical vertical length scale. This scaling applies to convective atmospheres on directly imaged planets and brown dwarfs, as well as the deep convective part on close-in irradiated exoplanets. The convective velocity scaling will be discussed in Section 6.3. Basically, in the slow rotation regime (e.g., Clayton 1968; Stevenson 1979b; Showman & Kaspi 2013), flows tend to be radially isotropic and mixing length theory predicts:

$$K_{zz,\text{slow}} \sim \left( \frac{\alpha g F^4}{\rho c_p} \right)^{1/4},$$

where $\alpha$ is the thermal expansivity and equals $1/T$ for isobaric expansion, $g$ is gravity, $F$ is the convective heat flux (internal heat flux), $l$ is the mixing length, $\rho$ is the air density and $c_p$ is the specific heat at constant pressure. The vertical length scale $l$ is usually assumed to be the pressure scale height $H$ but it also changes with the chemical timescale—the shorter-lived species have smaller length scales (e.g., Smith 1998; Bordwell et al. 2018). The $K_{zz,\text{slow}}$ can also be applied to the equatorial region on rapidly rotating planets because the rotational effect (Coriolis effect) is not important at the equator (Wang et al. 2015).

In the rapid rotation regime (e.g., Golitsyn 1981, 1980; Boubnov & Golitsyn 1986, 1990; Fernando et al. 1991; Showman et al. 2011; Showman & Kaspi 2013), large-scale flows tend to align along columns parallel to the rotation axis (Hough 1897; Proudman 1916; Taylor 1917). Both the velocity scaling and vertical length scale are different from those in the slowly rotating regime. If we take a characteristic length scale as $l = w/\Omega$ where $w$ is the vertical velocity, the $K_{zz}$ scaling is (e.g., Wang et al. 2015)

$$K_{zz,\text{rapid}} \sim \frac{\alpha g F^4}{\rho c_p \Omega^2},$$

where $\Omega$ is the rotational rate.

In the low-density upper atmosphere (e.g., Earth’s mesosphere), waves such as gravity waves generated from the lower atmosphere propagate vertically and break, leading to strong vertical mixing of the chemical tracers. Lindzen (1981) first parameterized $K_{zz}$ from the turbulence and stress from those breaking gravity and tidal waves (also see Schoeberl & Strobel 1984; Strobel 1981; Strobel et al. 1987). For energy-conserved waves, wave amplitude increases as density drops, suggesting that $K_{zz}$ decreases with pressure in a fashion of $\rho^{-1/2}$. For an isothermal atmosphere, approximately $K_{zz} \propto \rho^{-1/2}$.

The behavior in the middle part—the stably stratified atmosphere—is also complicated. For example, in Earth’s stratosphere, tracer transport is controlled by both large-scale overturning circulation and vertical wave mixing (Hunten 1975; Holton 1986). The eddy diffusivity should be considered to be an effective parameter for global-mean tracer transport. The magnitude of $K_{zz}$ depends on many other factors and may differ from planet to planet. Just like the eddy diffusivity in the convective medium depends on the chemical tracer itself (e.g., Smith 1998; Bordwell et al. 2018), that in a stratified atmosphere has similar behavior. Several studies (e.g., Holton 1986; Parmentier et al. 2013; Zhang & Showman 2018a, b; Komacek et al. 2019b) found that the parameterized $K_{zz}$ depends not only on atmospheric dynamics but also the tracer itself such as the tracer chemistry and trace distributions. Also, in strati-
fied atmospheres on hot planets, turbulent vertical transport driven by vertical shear instabilities could contribute to the vertical mixing (e.g., Fromang et al. 2016; Menou 2019). Menou (2019) derived the eddy mixing coefficients in double-diffusive shear instabilities using the secular Richardson number and turbulent viscosity. The derived $K_{zz}$ is inversely proportional to pressure squared (i.e., $K_{zz} \propto p^{-2}$).

Although the 1D model is still useful for a first-order global-mean situation (e.g., Yung & DeMore 1998), several efforts have been put forward at simulating the tracer distributions in 3D dynamical models with simplified chemical schemes (e.g., Drummond et al. 2018, 2016, 2020; Steinrueck et al. 2019). Based on the tracer distributions in 3D models, one could also derive 1D equivalent eddy diffusion coefficients (e.g., Parmentier et al. 2013; Charnay et al. 2015b). Zhang & Showman (2018a,b) specifically investigated the regimes of global-mean vertical tracer mixing in stratified planetary atmospheres. They found that $K_{zz}$ strongly depends on the large-scale circulation strength, horizontal mixing due to eddies and waves, and local tracer sources and sinks due to chemistry and microphysics. The first regime is for a short-lived tracer with chemical equilibrium abundance uniformly distributed across the globe, and global-mean vertical tracer mixing behaves diffusively. Unlike the traditional assumption, different chemical species in a single atmosphere should, in principle, have different eddy diffusion profiles. The second regime is for a short-lived tracer with a non-uniform distribution of the chemical equilibrium abundance. A significant non-diffusive component in this regime might lead to a negative $K_{zz}$ under the diffusive assumption. In the third regime where the tracer is long-lived with the tracer material surface significantly controlled by dynamics, global-mean vertical tracer transport is also largely influenced by non-diffusive effects.

Zhang & Showman (2018a,b) derived an analytical solution of $K_{zz}$ and validate that against 2D and 3D global-mean vertical mixing properties over a wide parameter space. For stably stratified atmospheres on tidally locked exoplanets, if chemical equilibrium abundance is uniformly distributed, the analytical solution of $K_{zz}$ can be approximated using the continuity Equation (13c) and Equation (14b)

$$K_{zz,\text{strat}} \sim (RT_{eq})^{5/2} (\sqrt{1 + \alpha_2^2} - \alpha_2) (1 + \zeta)^{-1},$$

where $\alpha_2$ is given in Equation (15b) and $\zeta$ is the ratio of the vertical transport timescale $H/w$ to the chemical timescale.
The formation environment, would further complicate the chemical process. Lastly, for tidally locked planets, the large day-side and nightside temperature difference would imply very different chemistry and cloud compositions (e.g., Parmentier et al. 2016; Venot et al. 2020b; Powell et al. 2018, 2019), but the horizontal transport would try to homogenize, or even quench, the tracer distributions in the horizontal direction. Chemical-transport models in 2D (e.g., Agúndez et al. 2014a; Venot et al. 2020b) and 3D (e.g., Cooper & Showman 2006; Parmentier et al. 2013; Lines et al. 2018b; Drummond et al. 2018, 2016, 2020; Steinrueck et al. 2019) have shed light on those behaviors but remain to be confirmed by observations in the future. Non-uniformly distributed chemical tracers, if they are radiatively active, would impact the transmission and emission spectra, transit light curves, and thermal phase curves on close-in exoplanets (e.g., Venot et al. 2020b) but we did not discuss it in detail in this review. In the following, we will first talk about the gas chemical species and the atmospheric compositional diversity in Section 5.2 and focus on clouds and hazes in Section 5.3.

5.2 Gaseous Compositional Diversity

In this section, we first discuss the bulk compositions and then talk about the minor species in the atmospheres, as well as the important controlling factors. Even though recent studies show that the overall ratios of C/O and magnesium-to-silicon (Mg/Si) in solar-metallicity stars are not very compositionally diverse (Bedell et al. 2018), the ratios in the protoplanetary disks significantly change with the radial distance due to the ice lines of condensable species such as water, carbon monoxide and carbon dioxide (Madhusudhan et al. 2014a). The formation environment of the planets and their subsequent migration, as well as the associated atmospheric formation and evolution processes such as accretion, outgassing, impact, condensation and escape, could lead to a wide range of elemental ratios and metallicities in the atmospheres (e.g., Elkins-Tanton & Seager 2008; Schaefer & Fegley 2010; Schaefer et al. 2012; Lupu et al. 2014). To first order, we can simply categorize planetary atmospheres into several regimes in terms of their bulk compositions across the entire parameter space of planetary mass, temperature, metallicity and elemental ratios. The currently confirmed exoplanets with estimated masses, radii and equilibrium temperatures, as well as large Solar System bodies, are displayed in Figure 10. Planets within different size ranges are color-coded. Here we crudely summarize them in terms of escape velocity, equilibrium temperature, and ratios of hydrogen/carbon/oxygen (H/C/O) and highlighted several important aspects related to the bulk compositional diversity.

We first consider the condensation and evaporation processes in which temperature plays a key role. The bulk compositions are normally simple chemical compounds made of hydrogen, carbon, oxygen, nitrogen and sulfur. The condensational temperatures for those compounds are usually low due to their weak intermolecular bonds or hydrogen bonds. In Figure 10, we roughly mark
their condensational temperatures (triple point temperatures) with vertical dashed lines. $N_2$ and CH$_4$ condense below 100 K. NH$_3$, CO$_2$, H$_2$O and H$_2$SO$_4$ condense at around 200–300 K. If planets are colder than their condensational temperatures, corresponding compounds will be primarily locked in the ice or liquid phase and thus their abundances in the atmosphere will be limited by the saturation vapor pressure. Take Earth as an example. Earth’s atmosphere will be a steam atmosphere dominated by several hundred bars of water vapor if the ocean completely evaporates. On the other hand, in the high-temperature regime beyond 1000 K, elements are not tied up in condensates. It is possible to evaporate the rocks and metals at the surface and form a silicate atmosphere, as proposed for 55 Cancri e (e.g., Demory et al. 2016a) and the proposed “super comets” such as Kepler 1520 b (Rappaport et al. 2012; Perez-Becker & Chiang 2013) and K2-22 b (Sanchis-Ojeda et al. 2015). In gas-melt equilibrium with the magma ocean, the atmospheric composition could be dominated by Na, K, Fe, Si, SiO, O and O$_2$ as the major atmospheric species (Ito et al. 2015).

A thermal escape of species is likely to have a large impact on atmospheric composition. The escape rate of an individual species depends on its molecular weight. The ad-hoc orange lines in Figure 10 are by no means quantitative boundaries. The H$_2$&He line also seems to divide the larger planets (radius larger than four Earth radii, red dots) and the smaller ones into two groups. Larger planets are reasonably represented by hydrogen or hydrogen-helium atmospheres. For planets smaller than one Earth radius, the compositional candidates are mostly restricted to higher molecular weight in the atmosphere, if there is any. From the perspective of atmospheric thermal escape, atmospheres could be dominated by many possible molecules such as water, N$_2$, O$_2$, CO, CO$_2$ and SO$_2$ (and even argon?). Other escape mechanisms such as solar-wind stripping might further constrain the atmospheric composition in this small terrestrial planet regime.

The mid-size planets between one and four Earth radii, namely sub-Neptunes, mostly reside between the H$_2$&He line and the H$_2$O&CH$_4$ line. Note that the H$_2$&He line also goes through the sub-Neptunes, as does the “Cosmic Shoreline.” Therefore H$_2$&He atmospheres are still possible on these bodies. Although we do not distinguish the mini-Neptunes (if we define them as hydrogen-dominated) and super-Earths (non-hydrogen-dominated), it looks to be more challenging for smaller and hotter sub-Neptunes to retain a low-molecular-weight atmosphere than the bigger and colder ones. We expect the atmospheric composition in the sub-Neptune regime might be highly diverse since al-
most all kinds of compositions are possible on those planets.

To further classify these atmospheres, we introduce thermoequilibrium chemistry, which assumes the atmosphere composition is solely dependent on temperature and elemental abundances. This has been investigated by a number of works (e.g., Lodders & Fegley 2002; Visscher et al. 2006, 2010; Kempton et al. 2012; Moses et al. 2013b, 2011; Line et al. 2011; Venot et al. 2012; Hu & Seager 2014; Mbarek & Kempton 2016; Tsai et al. 2017). The most important three elements are hydrogen, carbon and oxygen. The dependence of the composition on the H/C/O ratios is summarized in a ternary plot in Figure 11 based on simulation results in Hu & Seager (2014) for a typical sub-Neptune temperature range (500–1200 K). The low-metallicity atmospheres are probably still hydrogen-dominated. As the metallicity increases from top to bottom, the composition starts to diversify. The atmosphere would be more oxygen-rich to the left and more carbon-rich to the right. If oxygen dominates over the carbon but not hydrogen, a water world (steam atmosphere) is a possibility; the other end member is a hydrocarbon-dominated atmosphere if carbon dominates over oxygen. Higher-order (more than two carbons in a molecule) hydrocarbon atmospheres (CnHm) are thermochemically favorable and there is no need to invoke photochemistry to break the chemical bonds in methane (Hu & Seager 2014).

In the high metallicity regime, hydrogen compounds are no longer important. If the C/O ratio is low, the atmosphere might be dominated by molecular oxygen without biochemistry. Further photolysis could produce ozone. As the C/O ratio increases, the bulk composition shifts to CO and CO2, similar to the atmospheres on Venus and Mars. If the C/O ratio is high, the extra carbon atoms will not be able to combine with other elements so that the bulk composition could be dominated by carbon (graphite?). In this regime, graphite is actually stable for a large range of temperature and metallicity conditions (e.g., Moses et al. 2013b). If graphite is abundant, a large fraction of carbon would be required in the condensed graphite form, reducing the C/O ratio to near unity and resulting in a CO- or CO2-rich atmosphere. Condensed graphite might also be a source of the haze particles (Sect. 5.3). Nevertheless, details on the graphite chemistry have yet to be explored.

The above scenarios are just end-members. If the bulk metallicity has a relatively balanced H/C/O ratio, there will be a range of possible atmospheric compositions. Temperature plays a crucial role in determining what the atmosphere is made of (the “misc.” regime in Fig. 11). Including less abundant elements such as nitrogen (e.g., Moses et al. 2013b; Moses et al. 2013a), sulfur (e.g., Zahnle et al. 2009) and silicon would further complicate the classification. For example, C3H4-dominated atmospheres might not exist because species like HCN can dominate instead. A N2-rich atmosphere could be common. If silicon is present, SiC or SiO might also dominate the high metallicity atmospheres under certain conditions (Fig. 11). Moreover, disequilibrium chemistry such as photochemistry, ion-chemistry and vertical mixing will change the abundances of trace species in the atmosphere, but whether these processes could alter the bulk composition is an open question.

In addition to the bulk composition discussed above, minor constituents are also important and have a notable impact on the spectra, light curves and radiative energy balance of the substellar atmospheres. Key gaseous species providing very important opacities include H2O, CH4 and other hydrocarbons, CO, CO2, NH3, HCN, O3 and some sulfur-bearing species such as H2S and SO2. Again, the minor species also depends on many factors. Other than temperature and elemental ratios, the abundances and distributions of minor chemicals also crucially depend on disequilibrium processes induced by photochemistry and transport.

In general, for low-metallicity hydrogen-dominated atmospheres, the minor species are usually hydrogen compounds like H2O, CH4, and NH3 with photochemically generated hydrocarbons and nitriles. For high-metallicity atmospheres composed of H2O, CO, CO2, or N2, molecules with more than one heavy atoms per molecule and oxidized photochemical products such as O2, O3 and NO are abundant. Here we mainly focus on the hydrogen-helium atmosphere. The C/O ratio in the hydrogen atmosphere is an important factor. Madhusudhan (2012) proposed a classification scheme based on irradiation (essentially the temperature) and the C/O ratio in hydrogen atmospheres. The boundary between a C-rich atmosphere and an O-rich atmosphere is C/O ~ 1. It was claimed that C-rich atmospheres are not likely to have the thermal inversion because TiO and VO are not abundant. The O-rich, haze-free atmospheres could only develop thermal inversion in the high-temperature regime, but the low-temperature regime does not, similar to the pM and pL classes in Fortney et al. (2008). Note that the calculations for C-rich atmospheres in Madhusudhan (2012) have neglected the possible contribution of carbon-based aerosols, which could easily produce atmospheric inversion.

The detection of the C/O ratio in an exoplanetary atmosphere is important. From the formation point of view, it is expected that the C/O ratios for most stars should be less than unity because oxygen is more cosmically abundant than carbon (e.g., Fortney 2012; Brewer et al. 2016), but the formation and evolution of the planetary atmospheres
The thermochemical carbon cycle can be summarized as $\text{CH}_4 + \text{H}_2\text{O} \leftrightarrow \text{CO} + 3\text{H}_2$. The rates of $\text{CH}_4 \rightarrow \text{CO}$ and $\text{CO} \rightarrow \text{CH}_4$ conversion depend on the efficiency to form and break the strong C-O bond, respectively. It was proposed that the rate limiting step of $\text{CH}_4 \rightarrow \text{CO}$ is the reaction of the OH and CH$_3$ radicals, e.g., $\text{CH}_3 + \text{OH} \rightarrow \text{CH}_2\text{OH} + \text{H}$ or $\text{CH}_3 + \text{OH} + \text{M} \rightarrow \text{CH}_2\text{OH} + \text{H} + \text{M}$ (Moses et al. 2011; Tsai et al. 2018), where M is the ambient bulk gas molecule. Also, that of CO $\rightarrow$ CH$_4$ is perhaps $\text{CH}_2\text{OH} + \text{M} \rightarrow \text{CH}_3 + \text{OH} + \text{M}$ (Moses et al. 2011) or $\text{CH}_2\text{OH} + \text{H} \rightarrow \text{CH}_3 + \text{H}_2\text{O}$ (e.g., Venot et al. 2014, 2015, 2020a; Zahnle & Marley 2014). The carbon interconvers-
sion cycle $\text{CH}_4 + \text{H}_2\text{O} \leftrightarrow \text{CO} + 3\text{H}_2$ is also considered as the main pathway controlling water abundances.

For $\text{N}_2 \leftrightarrow \text{NH}_3$ interconversion, the net cycle can be written as $\text{N}_2 + 3\text{H}_2 \leftrightarrow 2\text{NH}_3$, but the rate limiting steps are highly uncertain (see discussion in Moses 2014). For $\text{N}_2 \rightarrow \text{NH}_3$, the rate-limiting step is speculated as $\text{NH} + \text{NH}_2 \rightarrow \text{N}_2\text{H}_2 + \text{H}_2$. The last possibility is debatable because the hot thermosphere might absorb most of the incoming high-energy flux. The hot interior hypothesis is particularly interesting because it implies some unknown heat source that might be related to the tidal heating due to the non-zero eccentric orbit (Agúndez et al. 2014b) or obliquity-induced tides (Millholland 2019). The interconversion between CO and CH$_4$ has also been suggested to play a very important role in substellar atmosphere evolution on brown dwarfs (e.g., Tremblin et al. 2017b, 2019), which will be discussed in Section 6.3.

The photolysis can also be important for directly imaged planets (Moses et al. 2016). Photolysis of CO, CH$_4$ and NH$_3$ could produce a significant amount of CO$_2$, HCN and hydrocarbons like C$_2$H$_2$ and even photochemical hazes (e.g., Lavvas & Koskinen 2017; Gao & Zhang 2020), which will be discussed in the following Section 5.3. For a detailed review of the photochemistry, refer to Moses (2014) and reference therein. Photochemistry has been shown to significantly alter the chemical compositions of terrestrial exoplanetary atmospheres and the interpretation of their spectra (e.g., Selsis et al. 2002; Segura et al. 2005; Segura et al. 2007; Hu et al. 2012, 2013; Rugheimer et al. 2015; Arney et al. 2016, 2017; Meadows et al. 2018; Lincowski et al. 2018; Chen et al. 2019). One interesting fact is that photo-
chemistry could produce false positives of gaseous biosignatures. For example, there is a variety of ways to produce abiotic molecular oxygen from the photolysis of water and CO$_2$ by strong UV flux (Gao et al. 2013; Tian et al. 2014; Wordsworth & Pierrehumbert 2014; Luger & Barnes 2015; Harman et al. 2015, 2018) on terrestrial exoplanets, especially on planets around M-dwarfs with high-energy fluxes. On the other hand, ion-chemistry is particularly important to understand the composition and energy balance in the upper thermosphere and the detailed mechanisms of atmospheric escape (Yelle 2004; García Muñoz 2007; Scheucher et al. 2020), but many of the chemical reaction coefficients have large uncertainties at this moment.

5.3 Clouds and Hazes

A growing body of evidence suggests that spectra of exoplanets and brown dwarfs are significantly affected by the presence of aerosols—condensational clouds and chemical hazes—that are also ubiquitous in all substantial atmospheres of planets in the Solar System. The most prominent evidence is from the muted spectral features. For example, if an exoplanetary atmosphere is cloud-free, its transmission spectra at optical wavelengths would exhibit a Rayleigh scattering slope with sharp spectral features from alkali metals like sodium and potassium if hot enough. In the presence of high-altitude aerosols, however, the spectral slope and metal absorption peaks are significantly reduced and may even disappear. Similarly, in the IR, the predominant gas (e.g., water, methane) rotational-vibrational features seen in a clear atmosphere could also be blocked by the presence of aerosols.

As noted in Section 2.3 and reference therein, such flattened transmission spectra have been seen for many hot Jupiters and cooler and smaller planets. The mean particle size and cloud top pressures have been retrieved from some of their spectra (e.g., Kreidberg et al. 2014a; Knutson et al. 2014a; Morley et al. 2015; Benneke 2015; Benneke et al. 2019a). For example, the cloud tops on GJ 1214 b (Kreidberg et al. 2014a), GJ 436 b (Knutson et al. 2014a) and HD 97658 b (Knutson et al. 2014b) are as high as the 0.1 mbar pressure level. For GJ 3470 b, the cloud top is at a lower altitude (Benneke et al. 2019a). These high-altitude aerosols cause trouble in retrieving atmospheric compositions on sub-Neptunes. For example, GJ 1214 b could be made of water, hydrogen or other heavier elements (e.g., Miller-Ricci & Fortney 2010; Rogers & Seager 2010), but the flattened spectra are not useful for distinguishing among these candidates due to the lack of detected molecular features (Kreidberg et al. 2014a). Atmospheric windows with lower cloud opacity at longer wavelengths are needed to solve this problem.

In addition to the spectral evidence, spatial information on the substellar atmospheres also indicates the existence of aerosols. The rotational light curves of brown dwarfs are a good example of the influence of clouds on thermal emission (see Sect. 4.4). The reflection light curves in the Kepler bands also demonstrate the importance of aerosols. For example, significant westward phase offsets in the visible wavelengths in Figure 6(C) probably come from cloud reflection (Parmentier et al. 2016). The depletion of condensable vapors could also be a result of cloud condensation. A recent observation on an ultra-hot Jupiter WASP-76 b ($T_{\text{eff}} \sim 2190$ K) using high-dispersion transit spectroscopy found asymmetry in the atomic iron signature in the atmosphere (Ehrenreich et al. 2020). The iron absorption is absent on the nightside close to the morning terminator in contrast to the other limb, indicating that the iron is possibly condensing on the nightside.

Aerosols on exoplanets and brown dwarfs have both direct and indirect sources. The direct sources include dust emission from the surface or dust infall from space. The surface sources are common on terrestrial planets, such as volcanic ash and sea salt on Earth, and dust storms on Mars. For example, it was proposed that radiatively active mineral dust emitted from the surface could postpone planetary water loss and impact the habitability of Earth-like exoplanets (Boutle et al. 2020). The atmospheres of “Super-puffs” Kepler 51 b and 51 d might also be dusty because of the outflow of tiny grains from the surface (Wang & Dai 2019). The infalling dust could come from meteoric dust sources (see Gao et al. 2014 for the Venus case) or directly from the protoplanetary disk (e.g., PDS 70 b and c, Wang et al. 2020).

The indirect sources refer to atmospheric condensation and chemical processes. For example, hazes and clouds on exoplanets and brown dwarfs have been predicted to form from either condensation of salt, silicate and metal vapors (e.g., Ackerman & Marley 2001; Morley et al. 2012; Ohno & Okuzumi 2018; Gao & Benneke 2018; Gao et al. 2018; Ormel & Min 2019; Ohno et al. 2020), or coagulation of particles generated by atmospheric chemistry (e.g., Lavvas & Koskinen 2017; Hörst et al. 2018; Fleury et al. 2019; He et al. 2020; Moran et al. 2020). In the condensate cloud scenario, clouds form when the condensable species become supersaturated, ranging from KCl and ZnS in the cooler regime to Mg$_2$SiO$_4$, TiO$_2$, MnS, Cr, Fe, corundum ($\text{Al}_2\text{O}_3$), calcium-aluminates and calcium-titanates (e.g., perovskite CaTiO$_3$) in hotter atmospheres (e.g. Visscher et al. 2006, 2010; Lodders 2010). Some L-dwarf spectra manifest possible broad absorption features
at around 9 µm that could result from the SiO vibrational band (Cushing et al. 2006).

Photochemical hazes on exoplanets have been hypothesized to form via atmospheric photochemistry and ion chemistry of methane, nitrogen and sulfur (e.g., Kempton et al. 2012; Morley et al. 2013; Zahlhe et al. 2016; Lavvas & Koskinen 2017; Gao et al. 2017a; Hörst et al. 2018; He et al. 2018a,b, 2020; Kawashima & Ikoma 2018, 2019; Kawashima et al. 2019; Lavvas et al. 2019; Fleury et al. 2019; Moran et al. 2020; Adams et al. 2019c; Gao & Zhang 2020; Ohno & Kawashima 2020), as analogues of hazes in Solar System atmospheres. These chemical haze particles may be highly porous like those on Titan and Pluto, where chemically produced “macromolecules” or “monomers” coagulate into large fluffy aggregates (e.g., Lavvas et al. 2013, 2011; Gao et al. 2017b). This production of photochemical hazes in warm atmospheres has been confirmed by laboratory experiments (e.g., Hörst et al. 2018; He et al. 2018a,b, 2020; Fleury et al. 2019; Moran et al. 2020), in the relevant temperature range from 300–1500 K. Yet, particle formation mechanisms and their impacts on substellar atmospheres and observations remain poorly understood.

Detailed simulations have also been conducted to understand aerosol formation in the warm and hot regime. Pioneering work from Ackerman & Marley (2001) and subsequent works (e.g., Saumon & Marley 2008; Marley et al. 2010; Morley et al. 2012, 2014, 2015) simulated 1D cloud profiles in substellar atmospheres based on an idealized, homogeneous chemical equilibrium framework. They assumed that the species immediately condense to particles once supersaturated but did not simulate the particle growth such as coagulation in Ormel & Min (2019) and Ohno et al. (2020). A sophisticated, kinetic, brown dwarf and exoplanet grain chemistry model is described in a series of papers by Helling and collaborators (e.g., Helling et al. 2008a; Witte et al. 2009, 2011; Woitke & Helling 2003, 2004; Woitke et al. 2020; Helling et al. 2006, 2008b, 2001; Helling & Woitke 2006; Helling et al. 2016, 2019; Samra et al. 2020; Helling et al. 2020, see reviews in Helling & Caswell 2014 and Helling 2019). The model considered more complicated mixtures of dust grains, but it does not include certain important grain growth processes, such as the vital effect of grain surface energy on the condensation process (“Kelvin effect”, e.g., Rossw 1978; Pruppacher & Klett 1980; Seinfeld & Pandis 2016). More recently, microphysical models originating from Earth science have been successfully applied to hot Jupiters (Powell et al. 2018, 2019; Gao et al. 2020) and smaller planets (Gao & Benneke 2018) to simulate multiple cloud layers via processes such as nucleation, coagulation, condensation, sedimentation, evaporation and transport. These models can predict the particle size distributions that have shown to be important for spectral simulations, but they do not consider the mixture of condensable species.

Photochemical haze models including coagulations microphysics (e.g., Lavvas & Koskinen 2017; Gao et al. 2017a; Kawashima & Ikoma 2018; Kawashima et al. 2019; Kawashima & Ikoma 2019; Lavvas et al. 2019; Adams et al. 2019c; Gao & Zhang 2020) have been applied to both Jupiter-sized planets and smaller planets. Spectral slopes in optical transmission spectra on some exoplanets are observed to be steeper than the Rayleigh slopes (so-called “super-Rayleigh slopes”, Sedaghati et al. 2017; Pinhas et al. 2019 Welbanks et al. 2019; May et al. 2019). It was suggested that photochemical haze produced in the upper atmosphere could result in an increasing trend of atmospheric opacity with altitude, which might explain the super-Rayleigh slopes of transit depth toward blue in optical wavelengths (e.g., Lavvas et al. 2019; Ohno & Kawashima 2020). The sub-Neptune GJ 1214 b has been regarded as the test bed for those haze models owing to the surprisingly flat spectra observed in the near-IR (e.g., Kreidberg et al. 2014a). Models reached a consensus that a very high metallicity is required to explain the spectral flatness of this planet. The detected radio emission from a close-in planet HAT-11 b (Des Etangs et al. 2013) indicates the existence of lightning inside the clouds, suggesting that the particle charge could be important (e.g., Helling et al. 2013; Helling & Rimmer 2019; Hodosán et al. 2017). However, the effect of charging on particle microphysical growth has not been investigated in detail for exoplanets and brown dwarfs.

Several computationally expensive 3D simulations have also been performed with particles for exoplanets (e.g., Parmentier et al. 2013; Charnay et al. 2015a,b; Oreshenek et al. 2016; Lee et al. 2015, 2016, 2017; Lines et al. 2018a,b; Roman & Rauscher 2019) and for brown dwarfs (Tan & Showman 2017; Tan & Showman 2020a), but the microphysics is usually simplified to increase the computational efficiency. A few models with fully coupled cloud microphysics, radiative and transfer and 3D dynamics can only perform very short-term integrations (∼40-60 d, Lee et al. 2016; Lines et al. 2018a). In 3D models where the large-scale dynamics can be resolved, particles are advected by atmospheric circulation and their sizes are found distributed inhomogeneously across the globe (e.g., Lines et al. 2018a,b). However, the 3D model results typically produce flat spectra in the UV and visible, in contrast with observations which display slopes across the wavelengths. Detailed diagnosis is still
needed to identify the underlying mechanisms. See a brief discussion in Section 6.2 and for detailed discussions refer to a recent review by Helling (2019).

Although individual objects deserve investigation in detail, statistical trends from observations could put strong constraints on the complicated aerosol formation processes on brown dwarfs and exoplanets in a single framework. Clouds have been proposed to significantly influence not only the spectral sequence of the emission light from brown dwarfs and directly imaged planets, but also the near-IR water signals in transmission spectra of close-in giant planets and sub-Neptunes ( Sect. 2.3). We first discuss the brown dwarfs and directly imaged planets in Section 5.3.1 and close-in exoplanets in Section 5.3.2.

5.3.1 Spectral Trends on Brown Dwarfs and Directly Imaged Planets

As mentioned in Section 4.4, clouds are important for understanding the rotational light curves and their variability of an individual body. The impact of clouds also shows up in well-characterized near-IR CMD (Fig. 12). For example, TiO conversion to TiO$_2$ and TiO and VO condensation (such as perovskite CaTiO$_3$) from thermochemical models (e.g., Lodders 1999; Burrows & Sharp 1999; Allard et al. 2001; Lodders & Fegley 2002) have been suggested to cause the M/L transition. On the other hand, the cloudless model could not explain the reddening of the M-L trend (Fig. 12), due to the onset of H$_2$ collision-induced absorption and CH$_4$ bands as the atmospheres cool down (e.g., Saumon & Marley 2008). As mentioned before, observations on L dwarfs show possible evidence for silicate grain absorption in Spitzer IRS data (e.g., Cushing et al. 2006), indicating the importance of clouds in controlling the spectral sequence. On the other hand, unlike warm and cloudy L dwarfs, T dwarfs are generally thought to be cold and clear in their photospheres. In the late-T to Y, the color reversal (blue to red) is probably due to both the disappearance of $J$ band alkali metal opacity because the metals are bound into molecules (Liu et al. 2010; Kirkpatrick et al. 2012) and the emergence of NH$_3$ absorption in the H band (e.g., Lodders & Fegley 2002; Burrows et al. 2003b). Chloride and sulfide clouds (e.g., Morley et al. 2012; Beichman et al. 2014) might also contribute to the T and Y sequences.

1D numerical models (e.g., Tsuji 2002; Tsuji et al. 2004; Tsuji & Nakajima 2003; Tsuji 2001; Allard et al. 2001; Ackerman & Marley 2001; Saumon & Marley 2008; Marley et al. 2010) with silicate clouds could explain the redward sequence of L dwarfs. Cloud-free models agree with the blueward sequence of T dwarfs (also plotted in Fig. 12). The very red colors of very low-gravity objects (VL-G, see Fig. 1) are still difficult to explain (e.g., Charnay et al. 2018). In a low-gravity environment, how could the cloud particles be lofted in the very high photosphere resulting in red color? One possible explanation is that those cloud particles are tiny so that sedimentation is not as efficient as vertical mixing, but they are still big enough to affect the near-IR emission. However, a detailed microphysical model with a more realistic treatment of the particle size distribution and cloud radiative feedback is needed to investigate this possibility in the VL-G regime.

A big unsolved puzzle is the L/T transition where brown dwarfs exhibit almost the same effective temperature ($\sim 1400 \pm 200$ K), but their color changes abruptly from red to blue (Kirkpatrick 2005). While for directly imaged planets, it is not obvious if a similar sharp transition exists or not (Fig. 12, also see Fig. 1). Traditional 1D models predict much more gradual color change than the observed sharp L/T transition on brown dwarfs. The hypotheses with clouds to explain the sharp L/T transition can be grouped into two main categories. The first one can be called the “rain out” or “downpour” scenario (e.g., Knapp et al. 2004; Tsuji & Nakajima 2003; Tsuji et al. 2004; Burrows et al. 2006a; Cushing et al. 2008; Saumon & Marley 2008; Stephens et al. 2009). During the L/T transition, the cloud particle size changes and might cause a sudden cloud drop and clear the atmosphere. Saumon & Marley (2008) combined a cloud model from Ackerman & Marley (2001) with planetary evolutionary models to simulate the near-IR CMD. In the Ackerman & Marley (2001) model, a parameter called $f_{\text{sed}}$ is prescribed as the efficiency of sedimentation compared with vertical mixing. If $f_{\text{sed}}$ is kept constant for all objects, the color change is too gradual to explain the sharp L/T transition. Saumon & Marley (2008) changed the $f_{\text{sed}}$ rapidly during the L/T transition to simulate the change of particle size. They could successfully reproduce the near-IR CMD. However, the detailed microphysics of this cloud particle size evolution has not been elucidated.

The second hypothesis can be called “patchy cloud” scenario (e.g., Ackerman & Marley 2001; Burgasser et al. 2002a; Marley et al. 2010), which might be analogous to the belt-zone structure and 5-micron hot spots in the NEB on Jupiter. If the cloud fraction in the atmosphere changes rapidly during the L/T transition, such as dust breakup forming cloud holes, hot air emits from the deep atmosphere and the disk-averaged color could suddenly transition blueward. This patchy cloud scenario could also produce large rotational light curves of brown dwarfs, consistent with observations. Moreover, light curve variability is often observed in the L/T transition objects, suggesting a pretty dynamic weather pattern regime. Recent observations (e.g., Apai et al. 2013, 2017) suggested oth-
er mechanisms than cloud holes to cause the light curve variability, such as thin-thick cloud distribution, spots and trapped waves (see discussion in Sect. 4.4). Whether these mechanisms could lead to a rapid color change during the L/T transition has yet to be investigated. A thorough understanding of the patchy cloud scenario requires a 3D convective model with cloud radiative feedback (e.g., Tan & Showman 2020a).

Alternatively, cloud-free models have been proposed to explain brown dwarf spectra. Atmospheric retrieval work found that a nearly isothermal photosphere could explain the muted near-IR features in L dwarfs (e.g., Burningham et al. 2017). The problem is that an isothermal atmosphere might strongly violate convective-radiative equilibrium in brown dwarf atmospheres such as in Ackerman & Marley (2001). With strongly pressure-dependent opacities like H$_2$-H$_2$ collision-induced absorption (CIA) and broadening of various molecular and atomic (like K and Na) lines, the atmospheric lapse rate is usually large. It was recently proposed that fingering convection (Tremblin et al. 2015, 2017b) or thermo-chemical instability (Tremblin et al. 2016) might cause the shallower temperature gradient. Tremblin et al. (2016) also claimed that cloud-free models could explain the spectral sequence as the result of thermochemical instabilities in the CO/CH$_4$ transition in the case of the L/T boundary and the N$_2$/NH$_3$ transition in the case of the T/Y boundary. The details of this mechanism have not been completely worked out. As noted in Section 4.4, it looks like this mechanism could not be responsible for the light curve variability seen in the L/T transition because the observed light curves and their variability do not show very different behaviors between the gas absorption cores and the outside continuum. To date, photospheric clouds and cloud variability remain the most probable mechanism for the observed rapid L/T color change and the light curve variability. This proposed cloud-
free mechanism is still under debate. See more discussion on atmospheric dynamics and convection in Section 6.3.

5.3.2 Spectral Trends on Close-in Exoplanets

Statistical clear-to-cloudy trends for transiting exoplanets began to emerge thanks to the increasing number of observed transmission spectra. Sing et al. (2016) found that the near-IR water feature amplitude is correlated with two spectral indices. The first one is the relative strength between optical scattering and near-IR absorption. The second one is the relative absorption strength between the near-IR and mid-IR. These observations suggest that clouds play an important, systematic role in shaping the transmission spectra. They also pointed out that hot Jupiters do not exhibit a strong relationship between temperature and cloud signatures, whereas brown dwarfs have a very obvious spectral sequence. Figure 1 shows the hot Jupiters on top of the brown dwarf near-IR CMD. The color of hot Jupiters has a much larger scatter, indicating their cloud formation is more complicated and diverse than brown dwarfs. Sing et al. (2016) attributed the reason to the difference in vertical temperature structures between hot-Jupiter atmospheres and those on brown dwarfs. Because of intense stellar irradiation, hot Jupiters possess much steeper pressure-temperature profiles than do field brown dwarfs. However, because cloud condensation curves are also steep (i.e., stronger dependence of pressure than temperature, see Fig. 4), a small temperature change will significantly change the cloud base pressure to a much larger extent on hot Jupiters than on brown dwarfs. Also, due to the diversity of planetary metallicity, gravity and radiative properties, the cloud materials could be cold trapped at depth (at ∼1–100 bar) on some hot-Jupiters but not the others (e.g., Parmentier et al. 2013; Powell et al. 2018), which would also increase the cloud variability in exoplanetary photospheres.

Planetary clouds are diverse and complicated, but current hot Jupiter data do suggest some possible cloudiness trend as a function of temperature (e.g., Stevenson 2016; Fu et al. 2017; Gao et al. 2020). Using a larger size of samples, Fu et al. 2017 found that the near-IR water spectral strength \( A_H \)—defined as the transit depth difference between the near-IR water band and the underlying continuum in units of the scale height—increases with the equilibrium temperature \( T_{eq} \) from 500–2500 K. A further analysis in Gao et al. (2020) indicates a seemingly non-monotonic trend among hot Jupiters. \( A_H \) increases with \( T_{eq} \) when \( T_{eq} < 1300 \) K and \( T_{eq} > 1600 \) K while the opposite trend seems to exist for planets located within \( 1300 < T_{eq} < 1600 \) K (Fig. 12).

Using a 1D aerosol microphysics model, Gao et al. (2020) proposed a mechanism for the non-monotonic \( A_H - T_{eq} \) trend on hot Jupiters (Fig. 12). They showed that aerosol opacity in the HST WFC3 channel is dominated by silicates when \( T_{eq} > 950 \) K, while iron and sulfur clouds do not form efficiently due to their higher nucleation energy barriers. The kinetic model results are different from the prediction from thermochemical models. The atmospheres are relatively clear when \( T_{eq} > 2200 \) K, which is too hot for global-scale silicate clouds to form, although clouds might still be present on the nightside (see the last paragraph of this section). As \( T_{eq} \) decreases, the formation of high-altitude silicate increases the cloudiness. Meanwhile, as the planets get cold, the cloud layers also progressively move to the deeper atmosphere, resulting in relatively clear atmospheres. Below 950 K, due to rising methane abundances and photodissociation rates, high-altitude photochemical hazes form and damp the near-IR water features. The future search of possible spectral features of the aerosols at longer wavelengths, such as the silicate feature at 10 microns, is the key to test this hypothesis (e.g., Ormel & Min 2019; Powell et al. 2019; Gao et al. 2020).

The dominant role of silicate clouds in the high-temperature range echoes the earlier work on brown dwarfs (e.g., Saumon & Marley 2008; Marley et al. 2010) that utilizes the silicate clouds to explain the spectral sequence evolution and L/T transition. The above hot Jupiter cloudiness trend (the right panel in Fig. 12) seems to share similarities with the near-IR CMD of brown dwarfs (the left panel on Fig. 12), despite that the former diagnoses the transmission properties of the atmospheres and the latter probes the emission. First, both data sets are not consistent with the clear-sky models in the high-temperature regime. Second, the 39 hot Jupiter samples seem to also exhibit a turn in the cloudiness index \( A_H \) at around \( T_{eq} \sim 1400 \) K, which is reminiscent of the brown dwarf L/T transition although the turn of the exoplanet curve looks weaker. Note that 1400 K is also roughly the effective temperature when the brown dwarfs change their self-emission color (also related to the cloud opacity). Such similarity might not be a coincidence. Instead, it looks like a smoking gun that both the brown dwarf L/T transition and hot Jupiter cloudiness trend share some common behaviors of high-temperature clouds, although the underlying mechanism has yet to be investigated in detail. On the other hand, it would also be interesting to analyze if there is any statistical trend in the reflective spectra on those hot Jupiters because silicate clouds could significantly increase planetary albedo (e.g., Marley et al. 1999).

Cooler and smaller planets might also show some statistical clear-to-cloudy trends with equilibrium temper-
From hot to cold, these planets include HAT-P-26 b, HD 106315 c, HAT-P-11 b, HD 97658 b, GJ 436 b, GJ 3470 b, GJ 1214 b, Kepler 51 b, Kepler 51 d, K2-25 b, and K2-18 b. Among those, HAT-P-26 b shows a relatively clear atmosphere (Wakeford et al. 2017). HD 106315 c displays some weak water feature in the near IR band (Kreidberg et al. 2020). HAT-P-11 b manifests water features (Fraine et al. 2014) but might be partially cloudy, as indicated by a nearly flat optical transmission spectrum from HST STIS recently (Chachan et al. 2019). HD 97658 b (Knutson et al. 2014b), GJ 436 b (Knutson et al. 2014a), GJ 3470 b (Ehrenreich et al. 2014), and GJ 1214 b (Kreidberg et al. 2014a) are cloudy. Super-puffs Kepler 51 b and d show very flat transmission spectra in the near IR (Libby-Roberts et al. 2020), perhaps due to aloft tiny dust particles (Wang & Dai 2019) or high-altitude photochemical hazes (Gao & Zhang 2020). The current data on K2-25 b are consistent with a flat spectrum (Thao et al. 2020), implying a cloudy atmosphere or a high-molecular-weight atmosphere. The coolest one, K2-18 b ($T_{eq} \sim 255 \text{ K}$), exhibits water features in the near-IR but could also be partially cloudy (Benneke et al. 2019b; Tsiaras et al. 2019). It looks that the atmosphere might become clear again when the temperature drops below about 400 K. Is there another non-monotonic trend from 1000 K to 300 K? Using the six planets in this list, Crossfield & Kreidberg (2017) first hypothesized that the water band amplitude changes either with the hydrogen and helium mass fraction or the equilibrium temperature. Fu et al. (2017) analyzed both Jupiter- and Neptune-sized planets together. The entire sample exhibits the $A_H$ dependence on the equilibrium temperature. With the new super-puff data, Libby-Roberts et al. (2020) revisited this statistical trend and concluded that the clear-to-cloudy trend is more consistent with the equilibrium temperature dependence (Fig. 12) instead of the metallicity dependence (see their fig. 16).

The underlying mechanisms behind this seemingly clear-to-cloudy trend of equilibrium temperature on cooler and smaller exoplanets have not been investigated. The mechanism is not likely the same as the high-temperature “condensation clouds” scenario proposed by Gao et al. (2020) for giant planets as the clouds (e.g., ZnS or KCl) tend to condense at a deeper atmosphere as temperature decreases (Crossfield & Kreidberg 2017). It might be more consistent with the photochemical hazes as methane becomes more important in the low-temperature regime (e.g., Morley et al. 2015; Gao et al. 2020). Morley et al. (2015) pointed out that atmospheres on sub-Neptunes change from haze-free to hazy atmospheres at around 800–1100 K due to the onset of CH$_4$ (see their fig. 9). In principle, the photochemical haze formation depends on UV intensities and plasma environment. Both factors depend on the star-planet distance and are likely to be positively correlated with the planetary equilibrium temperature. Therefore the total haze precursors (and presumably the total haze abundances) decrease as the irradiation level decreases, as does the temperature if other factors are fixed. However, the observed data show that colder atmospheres between 400–800 K appear to be hazier (Fig. 12), except for the coldest one, K2-18 b, where the planetary atmosphere is located in the habitable zone and appears relatively clear. This planet thus might be explained by its low UV irradiation level. Crossfield & Kreidberg (2017) conjectured that variations in haze formation altitude (cloud top) could play a role. Systematic, microphysical modeling of photochemical haze formation from 300–1000 K that takes into account variations in the rates and locations of haze production, haze transport and the impact of condensate clouds is needed to understand any possible trends better and explore the role of parameters such as metallicity, temperature and stellar UV fluxes.

Inhomogeneous aerosol distributions on exoplanets also impact observations on tidally locked exoplanets. The detailed microphysical simulations in Powell et al. (2018, 2019) imply that hot Jupiters might have very distinct transmission spectra between the eastern and western limbs: the eastern one has sloped spectra, and the western has flatter spectra. It remains to be confirmed because the current techniques can only observe limb-averaged spectra. Inhomogeneous aerosol coverage would cause a distorted transit light curve due to the different absorption radius on the eastern and western limbs (e.g., Line & Parmentier 2016; Kempton et al. 2017; Powell et al. 2019).

As noted in Section 4.4, orbital phase curves and rotational light curves are also heavily modulated by clouds on brown dwarfs and exoplanets. Clouds emit IR light as well as scatter and reflect incoming starlight, such that light curves, including both thermal emission and optical reflection, can be greatly affected by the spatial distribution of clouds. Light curve observations (e.g., Parmentier et al. 2016; Parmentier & Crossfield 2018) suggest that clouds on tidally locked exoplanets are inhomogeneously distributed and cause phase offset with respect to the secondary eclipse. Therefore, the simple explanation in Section 4.4 should be much more complicated because of the existence of clouds. Besides, thermal phase curves of some hot Jupiters are difficult to interpret without invoking clouds on the nightsides (e.g., WASP-43 b, Kataria et al. 2015; Stevenson et al. 2017). On the other hand, asymmetries in the optical light curve are useful for constraining particle distributions and properties. The optical light curves of Kepler-7 b constrain the spatial distribution of aerosols and their composition and particle size (e.g., García Muñoz & Isaak 2015; also see Demory et al. 2013;
Heng & Demory 2013; Webber et al. 2015; Esteves et al. 2015; Esteves et al. 2015; Parmentier et al. 2016). Several other exoplanets also exhibit a westward offset bright spot, indicating cloud reflection (e.g., Shporer & Hu 2015; Kepler-7 b, 8 b, 12 b and 41 b). A recently observed westward offset bright spot on CoRoT-2 b (Dang et al. 2018) at 4.5 microns might also be related to cloud distributions. Understanding the inhomogeneous aerosol distribution on exoplanets requires simulating microphysics coupled with the dynamical transport of aerosols under various conditions. Recent 3D modeling efforts have made some progresses but could not explain the data (e.g., Lines et al. 2018a,b). However, the lack of laboratory data on the model input parameters further hinders quantitative conclusions about the mechanisms. Also, although clouds could be bright and increase the planetary albedo, to date there is no apparent correlation between the geometric albedo and the incident stellar flux (e.g., Cowan & Agol 2011; Heng & Demory 2013; Schwartz et al. 2017; Zhang et al. 2018; Keating et al. 2019).

Keating et al. (2019) derived the dayside and nightside temperature of hot Jupiters from the Spitzer phase curves (Fig. 7). The nightside brightness temperatures across a broad range of $T_{eq}$ ($<2500$ K) are roughly the same ($\sim 1100$ K). Although atmospheric theory predicts that the nightside temperature could behave more uniformly than the dayside (see discussion in Sect. 4.3), this trend might also be explained by the ubiquitous existence of clouds on the nightside (Keating et al. 2019). If the clouds—a strong opacity source that affects the emission temperature—form at roughly the same temperature across the parameter space, the outgoing thermal flux might just be controlled by, but not necessarily equal to, the cloud base temperature. This theory also seems consistent with the above hypothesis from Gao et al. (2020) that thick clouds on close-in gas giants are dominated by a single component, such as silicates. The silicate condensation temperature is about 1400 K near the cloud base, depending on the actual pressure-temperature profile. That also implies the Spitzer channels can only probe the emission near the thick cloud top, or wherever the cloud opacity reaches unity, rather than the cloud base emission. Again, future observations at other thermal wavelengths are needed to disentangle the contributions of the uniformity of brightness temperature on hot Jupiters from the dynamical heat transport and that from the nightside clouds.

6 ATMOSPHERIC DYNAMICS

6.1 Fundamentals

The atmospheric flow pattern is primarily controlled by differential heating, drag and planetary rotation. The external energy source comes from the top (i.e., the stellar irradiation) or the bottom (i.e., convection or surface fluxes outside the domain). Depending on the energy flux distribution and atmospheric energy transport processes (e.g., radiation, conduction and convection), spatially inhomogeneous heating causes temperature anomalies and pressure gradient and drives the atmospheric movement. Thus the chemistry of the opacity sources from radiatively active gas and particles greatly influences the atmospheric dynamics. Drag exerts the momentum (and energy) exchange with the atmospheric flow via surface friction (on terrestrial planets), magnetic effect (for deep and hot ionized flow) or small-scale dissipative viscous processes.

In a rotating frame, the Coriolis effect plays an important role in shaping the fluid motion. Consider a deep, convective atmosphere on a fast-rotating giant planet. Rotation and convection tend to homogenize the entropy, leading to a barotropic fluid regime—small density variation on an isobar (constant pressure surface). The Taylor-Proudman theorem (Hough 1897; Proudman 1916; Taylor 1917) predicts that the wind flow behaves constant as vertical columns paralleled with the rotation axis and net exchange across the columns is not permitted. Atmospheric flows move freely with the Taylor columns in the east-west directions around the rotation axis. On the other hand, in a shallow, stratified, rotating atmosphere, the horizontal motion is usually much larger than the vertical case because the vertical velocity is suppressed due to the large aspect ratio, vertical stratification and rapid rotation (e.g., Showman et al. 2010). The atmosphere is approximately in hydrostatic equilibrium on a large scale. With appropriate approximations—hydrostatic, shallow-fluid and traditional approximations (see Holton 2016; Vallis 2006), the equation set (1) introduced in Section 2.1 can be reduced into the so-called “primitive equations.” In this simplified system, hydrostatic equilibrium implies that the fluid parcel is incompressible in the pressure coordinate, and gravity disappears in the equations (e.g., Vallis 2006). The effect of gravity is thus only limited in the radiation via determining the column density and opacity between two pressure levels but not on the fluid dynamics directly (see numerical examples in Kataria et al. 2016 with different gravities).

From the force balance point of view, one can characterize planetary atmospheric dynamics in several regimes using a dimensionless number: the Rossby number $Ro = U/\Omega L$, where $U$ is the typical wind speed, $L$ is the characteristic length scale of the atmospheric flow and $\Omega$ is the rotational rate. In a slowly rotating atmosphere such as on Venus, $Ro \gg 1$, the horizontal motion is controlled by the balance between the inertial force (centrifugal force for Venus) and the pressure gradient, residing in the cyclostrophic regime. As the rotation rate increases, the
Coriolis force becomes as important as the inertial terms. In the intermediate regime where \( \mathrm{Ro} \sim 1 \), multi-way force balance applies among pressure gradient, Coriolis force, nonlinear advection and atmospheric drag. For example, large hurricanes are balanced by the pressure gradient, Coriolis force and centrifugal force (“gradient wind balance”). In a fast-rotating atmosphere, \( \mathrm{Ro} \ll 1 \), the pressure gradient tends to be balanced with the Coriolis force, leading to the geostrophic regime. In this regime, the latitudinal temperature gradient from equator to pole is associated with a positive zonal wind shear leading to a faster east (or slower west) zonal wind at higher altitude. This is called the thermal wind balance.

The characteristic length scale \( L \) is important. The typical length scale is the Rossby deformation radius \( L_R \sim N H / \Omega \) where \( N \) is the buoyancy frequency (Eq. (6)), and \( H \) is the pressure scale height. Flow with length scale larger than the Rossby deformation radius is influenced by planetary rotation, whereas small-scale flow is typically affected more by local processes (such as buoyancy). If one takes the wind scale \( U \) as \( NH \), \( L_R \) corresponds to a length scale where the \( \mathrm{Ro} \) is equal to one. Rossby deformation radius is a natural length scale of many atmospheric phenomena such as geostrophic adjustment, baroclinic instabilities and the interaction of convection with the environment (Vallis 2006).

In a shallow atmosphere, planetary sphericity also plays a role because the vertical component of the Coriolis force is changing with latitude \( \phi \), characterized by the Coriolis parameter \( f = 2\Omega \sin \phi \). The local Rossby number can be written as \( U / f L \). The local Rossby number is larger at lower latitudes and smaller at higher latitudes. The dynamical regime consequently could be different from latitude to latitude. Not only does the Rossby deformation radius change with latitude, but also the horizontal fluctuations of pressure, density or potential temperature—they are proportional to each other—might be controlled by different mechanisms at different latitudes. At low latitudes, the local \( \mathrm{Ro} \) is large; the pressure gradient is balanced by the inertial term (Charney 1963). The horizontal potential temperature fluctuation \( \theta_h \) is estimated as

\[
\frac{\Delta \theta_h}{\theta} \sim \frac{U^2}{gD} \sim \frac{\mathrm{Fr}}{\mathrm{Ro}} \quad \mathrm{Ro} > 1. \tag{22}
\]

Here \( \mathrm{Fr} \) is the Froude number \( \mathrm{Fr} = U^2/gD \) for a flow depth of \( D \). \( \mathrm{Fr} \) can be described as the square of the ratio of wind speed to the gravity wave speed. It characterizes the relative strength of the inertia of a fluid particle to the gravity. In the extratropics or middle-latitude where the local \( \mathrm{Ro} \) is small, geostrophy leads to potential temperature fluctuation (Charney 1963)

\[
\frac{\Delta \theta_h}{\theta} \sim \frac{fUL}{gD} \sim \frac{\mathrm{Fr}}{\mathrm{Ro}} \quad \mathrm{Ro} < 1. \tag{23}
\]

Because the local Rossby number is smaller than one in this regime, the density perturbation in the middle latitudes is expected to be larger than in the tropics (Charney 1971). For the tropical regime, one can invoke the WTG approximation as described in Section 4.3.1 (e.g., Sobel et al. 2001).

From the view of vorticity—the curl of the velocity field, planetary rotation is a fundamental vorticity in the system. The vertical component of the planetary vorticity changes with latitude as the Coriolis parameter \( f \) changes. This is called the \( \beta \)-effect where \( \beta = \partial f / \partial \phi \) is the meridional gradient of \( f \). Because the fluid parcel tries to conserve its total vorticity (more precisely, potential vorticity, see Holton 2016), the vorticity gradient provides a restoring force for the meridional disturbance, producing Rossby waves. A Rossby wave plays a significant role in the formation of zonal jets via interaction with the mean flow. This “eddy-driven” jet formation mechanism causes multiple jet streams in the middle latitudes on giant planets and terrestrial planets. The characteristic length scale of the jet width is naturally related to \( \beta \). Rhines (1975) pointed out the jet width should be scaled as \( L_{\text{jet}} \sim \pi (2U_e/\beta)^{1/2} \), where \( U_e \) is the eddy velocity scale. This Rhines scale, although it is primarily from a 2D turbulent flow argument, could be related to the jet width on the multiple jets on 3D giant planets. There is another jet width scale that is more associated with the zonal jet velocity \( (U_{\text{jet}}) \) and potential vorticity gradients \( L_{\text{jet}} \sim \pi (2U_{\text{jet}}/\beta)^{1/2} \) (e.g., Williams 1978; Lian & Showman 2008; Scott & Dritschel 2012). The low latitudes have a larger \( \beta \), leading to larger anisotropy and large waves, whereas at high latitudes, the inertial advection dominates over the \( \beta \) effect resulting in a more turbulent atmosphere. This has been demonstrated in 2D shallow-water simulations (e.g., Showman 2007; Scott & Polvani 2008).

From an energetics point of view, the local energy imbalance from the external or internal sources leads to fluctuations of temperature and density on isobars that create the available potential energy (APE)—only a small fraction of the total potential energy that is then converted into KE. Qualitatively, in the framework of the classical Lorenz energy cycle (Lorenz 1955, 1967), both the APE and KE are partitioned into zonal mean and eddy (deviation from the zonal mean) components. The energy cycle starts from the production of the mean APE and eddy APE, the conversion among the four energy components, and the eventual loss of KE through frictional dissipation. The energy cycle could be complicated and requires a detailed analysis of the entire system, in particular, radiative energy flow in the atmosphere. One can see the discussion in Peixoto & Oort (1992) for Earth and Schubert & Mitchell (2013) for other Solar System planets. For exoplanets, this cycle has not
been analyzed in detail yet. The mean APE can be converted into the mean KE through the formation of thermally direct, overturning meridional circulation as well (e.g., Li et al. 2007). For zonal jet formation, another source of mean KE is from the mean APE to the eddy APE, then to the eddy KE and eventually the mean KE. Conversion from the mean to the eddy APE is done by generation and growth of eddies such as non-axisymmetric waves and other disturbances through many processes, such as convection, shear instabilities and baroclinic instability. The conversion from eddy KE to mean KE is through the eddy momentum convergence into the mean flow. Those eddy energies can be cascaded into a larger scale through the “inverse cascade” process in a quasi-2D regime for a large-aspect-ratio fluid like a shallow atmosphere, in contrast to the 3D turbulence where KE is cascaded into the smaller scale and eventually lost via viscous dissipation. For detailed discussion, refer to textbooks such as Vallis (2006).

The atmosphere can be considered to be a heat engine or a refrigerator. A classic heat engine extracts energy from a hot region and transfers it to a cold region. In this process, it converts part of the energy into work. The heat engine efficiency is the ratio of work it has done to the input heat. In a convective atmosphere on terrestrial planets (or even a local weather system such as a hurricane), the flux is carried upward from the hot boundary layer near the surface and emitted in the top, cold atmosphere. In this process, the atmosphere is doing work to produce KE, which is eventually lost in frictional or viscous dissipation (e.g., Emanuel 1986; Peixoto & Oort 1992; Rennó & Ingersoll 1996; Emanuel & Bister 1996; Schubert & Mitchell 2013). The adiabatic processes in those systems can be analogous to the classic Carnot engine. The Carnot efficiencies for the terrestrial atmospheres in the Solar System are estimated to be less than 27.5%, 13.2%, 4.4% and 4.1% for Venus, Earth, Mars and Titan, respectively (Schubert & Mitchell 2013). On tidally locked terrestrial planets, the day-night temperature difference in the boundary layer is large, but the temperature in the free atmosphere is roughly homogenous because of small wave-to-radiative timescales (see Sect. 4.3.1). As a result, the heat engine mainly works by the day-night overturning circulation. Koll & Abbot (2016) analyzed this system and found out the heat engine efficiency could be estimated as \( (T_d - T_{eq}) / T_d \) where \( T_d \) is the dayside temperature and \( T_{eq} \) is the equilibrium temperature.

Hot Jupiter atmospheres can also be regarded as a heat engine (e.g., Goodman 2009; Koll & Komacek 2018). However, the atmospheres are highly irradiated from the top and thus highly stratified. A Carnot cycle, which assumes adiabatic expansion and compression and isothermal heat addition and removal, is not a good analogy. Instead, Koll & Komacek (2018) proposed that one could approximate hot Jupiter atmospheres using the Ericsson cycle, which assumes isothermal expansion and compression and isobaric heat addition and removal. The heat engine efficiency of the Ericsson cycle is always smaller than that of the Carnot cycle. However, note that the heat engine concept is just a crude analogy. Circulation in some parts of the atmosphere could behave as a refrigerator, a reverse model of a heat engine. It might occur in those thermally indirect circulations (forced motions), for example, the wave-forced circulation in the lower stratosphere (e.g., Newell 1964). The anti-Hadley-like behavior in the equatorial region seen in some dynamical models on tidally locked planets (e.g., Charnay et al. 2015a) might also act more like a refrigerator, rather than a classic heat engine.

Because both forcing and rotation play key roles in atmospheric motion, in Figure 13 we classify the atmospheres in terms of the two parameters. First, utilizing the ratio of external stellar flux to internal flux or surface flux from below, we can classify the atmospheres into three regimes (Showman 2016): externally forced, internally forced and forced by both external and internal sources. Most close-in planets such as tidally locked hot Jupiters and sub-Neptunes are mainly forced by the external source from the central star. Most brown dwarfs and directly imaged planets are mainly forced from their internal fluxes. For planets and brown dwarfs located at an intermediate distance from the star, both the stellar forcing and internal flux are important. All Solar System atmospheres seem to fall in this regime (Fig. 13) but with different reasons. We consider the surface flux as the internal flux for terrestrial atmospheres, comparable to the external solar flux. For giant planets and brown dwarfs, low-mass, self-luminous, substellar evolution models show that their internal luminosity highly depends on their mass and age (e.g., Burrows et al. 2001; Phillips et al. 2020). As these objects become older, their radii decrease very slowly after about 1 Gyr due to Coulomb and electron degeneracy, but their internal luminosity continues decreasing via radiative cooling to space. The giant planets in the Solar System happen to have a roughly similar magnitude of external and internal fluxes (maybe except Uranus) at their current ages. Young hot Jupiters and highly irradiated brown dwarfs could also lie in this regime, for example, the recently discovered close-in brown dwarf rotating around a white dwarf (J1433, Santisteban et al. 2016).

To first order, if an atmosphere is mostly irradiated by external flux from the top, the photosphere is stably stratified (\( N^2 > 0 \)). This is generally the case for close-in exoplanets such as hot Jupiters. However, as noted in Section 4.2, inflated hot Jupiters could have much higher inter-
Fig. 13 Classification of substellar atmospheric dynamics using global Rossby number $\mathcal{R}_\Omega = U/\Omega R_p$ and the ratio of the external to internal fluxes ($F_{\text{ext}}/F_{\text{int}}$, see Eq. (10)). For terrestrial atmospheres in the Solar System, the internal fluxes (surface fluxes) are calculated utilizing blackbody emission based on the surface temperature. The Uranus value is the upper limit from Voyager (Pearl et al. 1990). For exoplanets, only planets with size larger than two Earth radii are considered. All sub-Neptunes are assumed to have hydrogen atmospheres. For self-luminous, directly imaged exoplanets, we calculated their internal fluxes by subtracting incoming stellar fluxes from their observed emission fluxes. For close-in exoplanets, the internal fluxes are difficult to estimate and might correlate with their ages. Here we made an assumption (probably an oversimplification) by applying the current internal flux of Jupiter (7.485 W m$^{-2}$) from Li et al. (2018). For planets less than 0.2 AU from their host stars, we assume they are tidally locked; for those located more than 0.2 AU, we estimated the rotational period using the mass scaling $v_e \sim v_0 (M/M_J)^{1/2}$, where $v_0 = 10 \text{ km s}^{-1}$ and $M_J$ is the mass of Jupiter. Field brown dwarfs and rogue planets without host stars have no external fluxes. Global Rossby numbers of Solar System bodies are calculated based on realistic winds, but that of the exoplanets are estimated from the isothermal sound speed using equilibrium temperature for tidally locked planets and effective temperature (based on their observed luminosity) for non-tidally locked planets (see text).

To characterize the rotational effect of the entire planet, we define a “global Rossby number” $\mathcal{R}_\Omega = U/\Omega R_p$ using the planetary radius $R_p$ as the length scale $L$. Adopting typical wind speeds on Solar System bodies, in Figure 13 we show that slowly rotating planets such as Venus (may also include Titan) are in the “tropical regime” ($\mathcal{R}_\Omega \gg 1$). Earth, Mars and giant planets are in the “geostrophic regime” ($\mathcal{R}_\Omega \ll 1$). Triton and Pluto fall in the intermediate regime ($\mathcal{R}_\Omega \sim 1$). To estimate $\mathcal{R}_\Omega$ of exoplanets and brown dwarfs, here we approximate $U$ using a typical isothermal sound speed $(R T_{\text{eq}})^{1/2}$. Thus the global Rossby number is a ratio of the isothermal sound speed to the equatorial velocity $\Omega R_p$ of the planet. $\mathcal{R}_\Omega$ is also the inverse of the dimensionless number $\Omega \tau_{\text{dyn}}$, we introduced in Section 4.3.1. When we utilize the isothermal sound speed
to approximate the wind velocity, $\mathcal{R}_D$ is also similar to the “WTG parameter" ($\Lambda = c_0/\Omega R_p$, see Sect. 4.3.1) introduced in Pierrehumbert & Hammond (2019) for terrestrial planets. If $\Lambda \gg 1$, we expect a global WTG behavior, i.e., weak horizontal temperature gradients on the entire planet. On the other hand, temperature gradients are strong in the regime of $\Lambda \ll 1$. If $\Lambda$ is order unity, one expects WTG behavior near the equator, but strong temperature gradients in the extratropics. Note that some studies also apply the thermal wind expression $U \sim R(\Delta \theta/\Omega R_p)$ to estimate the wind speed so that the Rossby number is redefined as a thermal Rossby number (e.g., Mitchell & Vallis 2010; Wang et al. 2018).

To estimate the rotation rate $\Omega$ of the planets, we first consider planets that are not greatly slowed down by the tidal effect. The equatorial velocity $v_e$ of giant planets and brown dwarfs seems to follow an empirical scaling law with the planetary mass (e.g., Snellen et al. 2014; Allers et al. 2016; Bryan et al. 2018)

$$v_e \sim v_0(M/M_J)^{1/2},$$

where $v_0 = 10 \text{ km s}^{-1}$ and $M_J$ is the mass of Jupiter. Thus the global Rossby number can be empirically expressed as

$$\mathcal{R}_D \sim 0.08(H/H_J)^{1/2}(R_g/R_p),$$

where $H$ is the pressure scale height and $H_J = 25 \text{ km}$ is roughly the pressure scale height of Jupiter’s upper troposphere. $R_J$ is the radius of Jupiter. As visible in Figure 13, self-luminous brown dwarfs almost certainly lie in the geostrophic regime. For young, hot giant planets, even though the scale height could be ten times larger than Jupiter’s, $\mathcal{R}_D$ could be smaller than unity for a large range of temperatures. For terrestrial planets in the habitable zone, $\mathcal{R}_D$ is generally smaller than unity, even for a low-mass small planet like Mars with a nitrogen atmosphere.

For a synchronously rotating planet, the rotation period is the same as the orbital period and related to the equilibrium temperature via Kepler’s third law. In this regime, the global Rossby number is

$$\mathcal{R}_D \sim \frac{a}{0.03 \text{ AU}}^{5/4}\left(\frac{R_J}{R_p}\right)^{5/4}\left(\frac{m_a}{m_H}\right)^{1/2},$$

where $a$ is the semi-major axis and $m_a/m_H$ is the ratio of mean molecular mass of the atmosphere to hydrogen. The global Rossby number does not depend on the stellar mass if we use the stellar mass-luminosity relationship $L \propto M^4$. The reason is that, given the same semi-major axis, as the stellar mass increases, the planetary orbital period decreases and the stellar luminosity increases. Both the planetary rotation rate and equilibrium temperature increase; their effects almost cancel out in the global Rossby number.

As Figure 13 demonstrates, $\mathcal{R}_D$ is around unity for a hot Jupiter in a typical 3-day orbit. For smaller planets such as hot sub-Neptunes, $\mathcal{R}_D$ could be larger, and many of their atmospheres are in the tropical regime. However, those with fast rotation (e.g., very close-in planets) lie in the geostrophic regime (Fig. 13). Also, for Earth-like planets with an atmosphere of heavier molecules, the global Rossby number (and thus the WTG parameter $\Lambda$) could exceed unity if the small planet is relatively far from the star and is still tidally-locked. For example, GJ 1132 b and LHS 1140 and the Trappist I planets in the habitable zone might have global Rossby numbers larger than unity and thus are in the tropical regime (the WTG regime in Pierrehumbert & Hammond 2019). This regime is different from the mid-latitude climate on Earth in which geostrophy is important but could resemble the tropics on Earth where the WTG approximation is applicable (e.g., Sobel et al. 2001).

In the following discussion, we will summarize our understanding of two specific populations of exoplanets and brown dwarfs in terms of their forcing pattern. The first category is the close-in, highly irradiated gaseous planets such as hot Jupiters and sub-Neptunes, warm Jupiters and warm Neptunes. The observable atmospheres of this type are mostly stably stratified. The second population is weakly irradiated planets that are located far away from their host stars, such as directly imaged planets and brown dwarfs, on which the internal flux plays an important role. We will focus more on the convective nature of this category. For highly irradiated brown dwarfs or young hot Jupiters with comparable external and internal fluxes, we only briefly discuss here due to the lack of sufficient constraints from observations yet. Again, note that most Solar System planets are in this regime (Fig. 13). Lastly, we will briefly discuss the terrestrial planets in the habitable zone and highlight the uniqueness of this climate regime in the presence of liquid water.

### 6.2 Highly Irradiated Planets

The most famous examples in the highly irradiated exoplanet population are the synchronously rotating planets locked by stellar tides. The observational characterization of this type has been discussed in Section 4.3. For the flow pattern, one can infer the jet speed (dynamical timescale) by analyzing the temperature distribution, as revealed by the hot spot phase shift in the thermal phase curve (e.g., Showman & Guillot 2002; Knutson et al. 2007), but the presence of clouds greatly complicates the thermal e-
mission. Eclipse mapping techniques (e.g., Rauscher et al. 2007; de Wit et al. 2012) could also be useful to map the spatial inhomogeneity on those distant objects. On the other hand, directly probing the wind speed on those planets is possible using Doppler techniques (e.g., Snellen et al. 2010; Showman et al. 2013a). An ultra-high resolution cross-correlation method has been applied for a specific atom or a molecule (such as CO, Mg and Fe) to measure the planet radial velocity and even wind-induced redshift/blueshift for both transiting (e.g., Snellen et al. 2010; de Kok et al. 2013; Birkby et al. 2013; Brogi et al. 2013; Wyttchenbach et al. 2015; Louden & Wheatley 2015; Brogi et al. 2016; Birkby et al. 2017; Salz et al. 2018; Flowers et al. 2019; Ehrenreich et al. 2020) and non-transiting planets (e.g., Brogi et al. 2012; Rodler et al. 2012; Brogi et al. 2014). For example, a blueshift of 2 ± 1 km s\(^{-1}\) was reported on HD 209458 b utilizing the CO lines by Snellen et al. (2010) and considering both CO and H\(_2\)O by Brogi et al. (2016). A blueshift of several km s\(^{-1}\) was detected on HD 189733 b utilizing the atomic sodium doublet (Wyttchenbach et al. 2015), the H\(_2\)O and CO infrared lines (Brogi et al. 2016, Flowers et al. 2019), and the Helium I triplet (e.g., Salz et al. 2018).

Employing time-resolved ultra-high resolution spectra, one can even derive the wind speed at separate limbs on tidally locked planets. For example, on HD 189733 b, Louden & Wheatley (2015) resolved a redshift of 2.3\(^{+1.5}_{-1.3}\) km s\(^{-1}\) on the leading limb and a blueshift of 5.3\(^{+4.0}_{-1.8}\) km s\(^{-1}\) on the trailing limb, suggesting an equatorial super rotating jet. The entire 3D wind structure on this planet could be complicated. A recent reanalysis showed that the sodium doublet data on HD 189733 b are consistent with a super-rotating wind, a day-to-night flow, or a very strong vertical wind (close to the escape velocity) in the upper atmosphere (Seidel et al. 2020). The vertical flow structure could be better constrained by future observations in a broader range of wavelengths. However, cloud condensation of those metals could limit the application of this technique using metals as the tracers. A recent effort on an ultra-hot Jupiter WASP-76 b only detected the wind speed on the trailing limb from iron lines. Iron vapor was not detected due to a significant depletion in the nightside and around the morning terminator, probably a result of cloud condensation in the lower atmosphere (Ehrenreich et al. 2020).

The day-night temperature difference and wind speed can be estimated using the scaling equation set (13) in Section 4.3.1. The thermal phase offset for a clear atmosphere was also estimated utilizing kinematic wind transport. Nevertheless, these scaling theories do not provide insights on the detailed mechanisms of the dynamics, such as the origin of the equatorial superrotating jet, development of the day-night flow pattern, wave-adjustment dynamics, eddy-eddy interaction, eddy-mean flow interaction and turbulent energy transfer. The weather of tidally locked planets is further complicated owing to the interaction between dynamics and radiation, chemistry, cloud microphysics and electromagnetic field. For a detailed review of the dynamics on tidally locked giant planets, refer to Showman et al. (2010), Heng & Showman (2015) and Showman et al. (2020) and that for terrestrial planets refer to Showman et al. (2013b) and Pierrehumbert & Hammond (2019).

Here we just briefly summarize our current understanding of the mechanisms under different regimes. In particular, in light of the scaling in Section 4.3.1, the bulk atmospheric flow of a tidally locked exoplanet is governed by dimensionless numbers: \(\Omega r_{\text{dyn}}/r_{\text{rad}}, \tau_{\text{dyn}}/\tau_{\text{rad}}\) and \(q/c_p T_{\text{eq}}\). We can roughly characterize the atmospheric dynamics on tidally locked gas giants into four regimes with an emphasis of each dimensionless number: canonical tropical regime (\(r_{\text{dyn}}/r_{\text{rad}}\), “nominal”), fast-rotating geostrophic regime (\(\Omega r_{\text{dyn}}, \tau_{\text{ultrafast}}\)), strong drag regime (\(r_{\text{dyn}}/r_{\text{drag}}\), “drag”) and ultra-hot regime (\(q/c_p T_{\text{eq}}\), “ultrahot”). In addition, to highlight the importance of opacity sources, we have two more regimes: high metallicity regime (“high metallicity”) and cloud regime (“cloud”). For cool and small planets, compositional diversity could also greatly impact the dynamics. Figure 14 summarizes all the six regimes with a representative dynamical pattern (temperature, flow or cloud tracer) for each regime from 3D GCM simulations.

Canonical drag-free hot Jupiter simulations from the 3D general circulation models exhibit a strong, broad eastward (superrotating) jet at the equator and westward wave patterns off the equator (Fig. 14, “nominal” regime). The temperature pattern is shifted to the east compared with the stationary day-night radiative forcing pattern centered at the substellar point. These temperature and flow patterns showed up in the first hot Jupiter GCM results from Showman & Guillot (2002) and the temperature offset was later confirmed by observations in Knutson et al. (2007). The subsequent 3D hot Jupiter models qualitatively agree with the Showman and Guillot results (e.g., Cooper & Showman 2006; Dobbs-Dixon & Cowan 2017; Showman et al. 2009; Rauscher & Menou 2010; Heng et al. 2011; Perna et al. 2012; Mayne et al. 2014; Mendonca et al. 2016; Carone et al. 2019; Mayne et al. 2019; Deitrick et al. 2020; Ge et al. 2020). The underlying mechanism of the equatorial superrotating wind is, however, not easy to understand. According to Hide’s theorem (Hide 1969; Schneider 1977), in a steady axisymmetric atmosphere with diffusion, a local maximum in absolute angular momentum cannot be maintained away from bound-
aries by the mean flow. Thus, a local maximum in angular momentum such as equatorial superrotation must imply upgradient eddy momentum fluxes that balance the diffusion of angular momentum.

Where do the upgradient eddy momentum fluxes come from in hot Jupiter atmospheres? The key to understanding the mechanism dates back to the “Matsuno-Gill model” in the Earth’s tropics. Matsuno (1966) considered freely propagating, linear wave modes on a β plane. Gill (1980) analyzed the atmospheric modes in response to the stationary, longitudinal forcing in the tropics. The excited large-scale wave modes, including Kelvin waves, Rossby waves, mixed Rossby-gravity waves and gravity waves, can be trapped in an equatorial width characterized by the equatorial Rossby deformation radius \((NH/\beta)^{1/2}\). The canonical hot Jupiters with the global Rossby number around unity exhibit those modes under stationary day-night stellar forcing. In particular, the standing, eastward Kelvin modes and westward Rossby modes form a chevron-shaped feature, with northwest-southeast tilts in the northern hemisphere and southwest-northeast tilts in the southern hemisphere. These horizontal eddy patterns have been shown to feed angular momentum to the equatorial superrotating jet and maintain it on tidally locked planets using a simpler one and a half layer shallow water model (Showman & Pollmani 2010; Showman & Pollmani 2011).

Later on, Tsai et al. (2014) performed a 3D analysis of the resonance of the Rossby waves and vertical wavefront tilt that provides vertical eddy-momentum flux to influence the jet acceleration and deceleration. Debras et al. (2020) further extended the analysis to arbitrary drag and radiative timescales and highlighted nonlinear feedbacks in the system on the onset of the prograde jet. Mayne et al. (2017) argued the angular momentum transferred by mean meridional circulation, aside from the eddies, is also important for the jet maintenance. Motivated by the roles of thermal tides in generation of the superrotating jets on slowly rotating planets in the Solar System such as Venus and Titan, recently Mendonça (2020) performed a detailed wave analysis and pointed out that the semi-diurnal tides excited by the stellar forcing play an important role in jet generation. Semi-diurnal eddy features have been seen in previous 3D models. For example, Showman & Pollmani (2011) emphasized that an important difference between their 2D shallow-water cases and full 3D cases is that the 3D models develop pronounced mid-latitude Rossby-wave anticyclonic gyres on the dayside and cyclonic gyres on the nightside in both hemispheres, resulting from feedback from the mean flow on the eddies (also see discussion in Pierrehumbert & Hammond 2019).

The day-night temperature difference and hot spot phase offset on tidally locked planets are caused by eastward group propagation of Kelvin waves and the equatorial superrotating flow (e.g., Showman & Pollmani 2011; Perez-Becker & Showman 2013; Komacek & Showman 2016). More precisely, the hot spot offset is caused by the zonal flow Doppler shifting the stationary wave response because the mean zonal wind on hot Jupiters has a horizontal velocity close to those waves. Hammond & Pierrehumbert (2018) demonstrated this mechanism in a 2D system with a horizontally shearing flow. The kinematic scaling of day-night temperature difference and hot spot phase shift using the dimensionless number \(\tau_{\text{dyn}}/\tau_{\text{rad}}\) in Section 4.3.2 still holds true (e.g., Cowan & Agol 2011; Zhang & Showman 2017), but there is some difference in the effect of radiative damping. In the kinematic theories, the radiative damping directly controls the temperature distribution, relaxing it to the day-night equilibrium temperature pattern. In the theory of Hammond & Pierrehumbert (2018), the damping weakens the forced wave response, relaxing the phase of the response to the forcing phase. Thus, the dynamical models from Hammond & Pierrehumbert (2018) highlight the importance of the wave response to the mean flow. Note that the mean flow generation is also closely related to the waves. To date, the complete 3D picture of nonlinear wave-mean-flow interaction and the influence on heat redistribution has not been thoroughly analyzed. Note that the magnetic field, clouds and other complicated factors play roles in real hot Jupiter atmospheres, and may be more important roles than the simple dynamics argument above.

Some fast-rotating tidally locked planets (Fig. 14, “ultrafast”) lie in the geostrophic regime with small Rossby numbers. This emerging regime includes both planets with an orbital period around one Earth day or smaller (e.g., WASP-12 b, WASP-103 b, WASP-18 b, WASP-19 b, NGTS-7A b and TOI 263.01) and several super-fast rotating brown dwarfs with rotation period within two hours (e.g., NLTT 5306, WD 0137–349, EPIC 21223532 and WD 1202–024). Not surprisingly, some of these planets are very close to their host stars, so they are also ultra-hot Jupiters, but the temperature on rapid rotators could also be mild if the host stars are cool. Lee et al. (2020) presented the first 3D simulation in this regime on the atmosphere of WD 0137–349B around a white dwarf. They found a large day-night temperature contrast and multiple, alternating east-west jet patterns. Tan & Showman (2020b) performed 3D simulations to explore the atmospheric dynamics in this geostrophic regime systematically. As expected, geostrophic adjustment is important. Because the equatorial Rossby deformation radius is small, the meridional extent of the temperature pattern is confined within a very narrow region around the equator. A big difference from the nominal case is in the zonal mean zonal wind pat-
The formation mechanism of which is as- 

In par- 

Interestingly, westward equatorial winds (in the 

The basic planetary parameters, such as size and 

Heaton et al. 2017; Sheppard et al. 2015; Nogroho et al. 2017). In par- 

The atmospheric metallicity is increased to 10× solar in the “high metallicity” case. The equilibrium temperature is 3000 K in the “ultrahot” case. The “cloud” case assumes a magnesium silicate cloud (Mg$_2$SiO$_4$) with cloud radiative feedback. The rotational period of the “ultrafast” case is 2.5 h.

Ultra-hot Jupiters are also a recent emerging popu- 

Ultrahot Jupiters are also a recent emerging popu- 

Instead of a broad equatorial jet in the tropical regime, multiple off-equatorial jets emerge on a fast-rotating hot Jupiter (Fig. 14), the formation mechanism of which is as- 

As summarized in Section 4.2, high-temperature chem- 

Fig. 14 Typical patterns of tidally locked giant planets for six dynamical regimes: “nominal,” “drag,” “high metallicity,” “ultrahot,” “cloud” and “ultrafast.” We plotted the horizontal maps at 100 Pa of the temperature for the first four cases, the cloud mass mixing ratio for the “cloud” case, and the wind pattern for the “ultrafast” case. Simulations were performed using the 3D global MITgcm with a gray radiative transfer scheme (from Xianyu Tan) except the “ultrafast” case that implemented a Newtonian cooling scheme (Tan & Showman 2020b). The basic planetary parameters, such as size and gravity, are similar to those of HD 209458 b. The “normal” case is assumed solar metallicity and drag-free. Based on the “nominal” case, a linear frictional drag is applied in the “drag” case. The atmospheric metallicity is increased to 10× solar in the “high metallicity” case. The equilibrium temperature is 3000 K in the “ultrahot” case. The “cloud” case assumes a magnesium silicate cloud (Mg$_2$SiO$_4$) with cloud radiative feedback. The rotational period of the “ultrafast” case is 2.5 h.

Fig. 14 Typical patterns of tidally locked giant planets for six dynamical regimes: “nominal,” “drag,” “high metallicity,” “ultrahot,” “cloud” and “ultrafast.” We plotted the horizontal maps at 100 Pa of the temperature for the first four cases, the cloud mass mixing ratio for the “cloud” case, and the wind pattern for the “ultrafast” case. Simulations were performed using the 3D global MITgcm with a gray radiative transfer scheme (from Xianyu Tan) except the “ultrafast” case that implemented a Newtonian cooling scheme (Tan & Showman 2020b). The basic planetary parameters, such as size and gravity, are similar to those of HD 209458 b. The “normal” case is assumed solar metallicity and drag-free. Based on the “nominal” case, a linear frictional drag is applied in the “drag” case. The atmospheric metallicity is increased to 10× solar in the “high metallicity” case. The equilibrium temperature is 3000 K in the “ultrahot” case. The “cloud” case assumes a magnesium silicate cloud (Mg$_2$SiO$_4$) with cloud radiative feedback. The rotational period of the “ultrafast” case is 2.5 h.
tion and recombination are included. The westward winds become more pronounced when the temperature exceeds 2400K. In the simulations, westward winds are accelerated by vertical eddies that overcome the eastward forcing by horizontal eddies, but the detailed mechanism has yet to be explored. Also, for the same stellar type, an ultra-hot Jupiter is usually rotating faster than a cooler Jupiter, and thus the rotational effect also needs to be taken into account. When the rotational effect is included, some of the previously seen westward jets at the low pressure disappear (Tan & Komacek 2019).

If frictional drag is strong, the flow pattern of a tidally locked planet can be significantly altered (“drag” regime in Fig. 14). The drag force could come from multiple sources. For terrestrial planets, frictional drag from the surface sets the lower boundary condition of the flow. For hot Jupiter atmospheres that could be partially ionized, Lorentz force due to magnetic field should play a role (e.g., Perna et al. 2010a). Small-scale vertical turbulent mixing (e.g., Li & Goodman 2010; Ryu et al. 2018) and breaking gravity waves (e.g., Lindzen 1981) could also be considered as drag forces exerted on the large-scale flow. If we simplify the drag effect as linear friction, one can understand the flow pattern in terms of force balance (e.g., Showman et al. 2013a). A three-way balance of the frictional drag, Coriolis force and the pressure gradient causes the horizontal eddy wind to rotate clockwise in the northern hemisphere and counterclockwise in the southern hemisphere, leading to equatorward-eastward and poleward-westward velocity tilts and thus driving equatorial superrotation, in addition to its direct damping effect from the wind itself. If the frictional drag is stronger than both the Coriolis force (i.e., large $\Omega \tau_{\text{drag}}$) and non-linear inertial force (i.e., large $\tau_{\text{dyn}}/\tau_{\text{drag}}$), the horizontal wind is directly controlled by the balance of the drag force and the pressure gradient and exhibits a day-night divergent flow pattern instead of an east-west jet pattern. This strong drag effect has been investigated in all current dynamical models (e.g., see a large grid of idealized 3D simulations in Komacek & Showman 2016). Therefore, by observing the horizontal wind pattern on tidally locked giant planets (e.g., through the cross-correlation technique), one might infer the strength of the atmospheric drag. For very rapidly rotating tidally locked planets, if drag is strong, the thermal phase curves could actually show a near alignment of peak flux to secondary eclipse (Tan & Showman 2020b), as observed in close-in brown dwarfs orbiting white dwarfs, for example, NLT 5306 (Steele et al. 2013), WD 0137–349 (Casewell et al. 2015; Longstaff et al. 2017), EPIC 21223532 (Casewell et al. 2018) and WD 1202–024 (Rappaport et al. 2017).

The magnetic drag effect could be particularly important for hot planets. Strong magnetic fields ($\sim$20–120G) have been detected on four hot Jupiters from the energy released in the Ca II K line during star-planet interactions (HD 179949 b, HD 189733 b, τ Boo b and ν And b, by Cauley et al. 2019b). Crudely, one can estimate the importance of magnetic effect in an ionized medium using a non-dimensional number called “plasma-$\beta$”—the ratio of the plasma pressure ($p$) to the magnetic pressure $(B^2/8\pi)$ where $B$ is the background magnetic field. The magnetic effect dominates when $\beta \ll 1$, such as in the solar corona. The plasma pressure dominates when $\beta \gg 1$, such as in the solar interior. We can further use the Alfvén Mach number $M_A$ for subsonic flow. $M_A^2 = U^2/U_A^2 = 4\pi\rho M_A^2/B^2$ is the ratio of the flow speed $U$ to the Alfvén speed $U_A = (B^2/4\pi\rho)^{1/2}$ where $\rho$ is the plasma density. $M_A = U/(p/\rho)^{1/2}$ is the Mach number to the isothermal sound speed. If $M_A$ is small, the magnetic field controls the flow. For example, for a strong magnetic field strength of 100 G, if the wind velocity is subsonic with $M_a \sim 0.1$, $M_A$ reaches unity at 0.1 bar where the magnetic force (Lorentz force) could be as important as the pressure force in a fully ionized atmosphere.

The realistic MHD effect could only be more complicated than a simple drag effect because of partial ionization and feedbacks of the flow pattern to the magnetic field. In particular, the non-ideal MHD effects from Ohmic resistivity, Hall effect and ambipolar diffusion could play a role in a partially ionized medium like hot Jupiter atmospheres, leading to incomplete coupling between the atmosphere and magnetic field. Most published studies to date have neglected feedbacks (Perna et al. 2010a,b, 2012; Menou 2012; Rauscher & Menou 2013; Hindle et al. 2019) but there are some recent efforts considering more realistic MHD situation (e.g., Batygin & Stanley 2014; Rogers & Komacek 2014; Rogers & Showman 2014; Rogers 2017).

Exoplanets have a large range of metallicities (Fig. 8) that could influence the atmospheric dynamics. For example, the simulated temperature pattern on a hot Jupiter using ten times solar metallicity looks different from the nominal case (Fig. 14, “high metallicity” case). The metallicity effect could be more influential for small planets as their bulk composition might not be hydrogen. As depicted in Figure 8, for sub-Neptunes and smaller planets, compositional diversity of the bulk atmosphere greatly increases, ranging from low molecular mass atmospheres of H$_2$ to higher molecular atmospheres of water, CO$_2$, N$_2$ or other species (see Sect. 5.2). Compared with the hydrogen case, three important effects need to be taken into account in these higher-metallicity atmospheres: molecular weight, heat capacity and radiative opacity (Zhang & Showman...
Take GJ 1214 b as an example. The simulated atmospheric flow pattern greatly changes with different assumptions on the bulk composition (H$_2$, H$_2$O or CO$_2$) or metallicity in hydrogen atmospheres, or the presence of cloud or haze particles in the atmosphere (e.g., Kataria et al. 2014; Charnay et al. 2015a,b). A detailed characterization of the dynamics on those planets requires further observations in longer wavelengths to probe deep below the high-altitude particle layers to determine the atmospheric composition. From an atmospheric dynamics point of view, it would be good to keep in mind that the global Rossby numbers for smaller and hot planets (e.g., warm Neptunes) are likely to be higher than for their Jovian-sized counterpart (e.g., warm Jupiters) and thus their climate states lie in the large-Rossby-number regime (Fig. 13), which means their day-night contrast is generally smaller than their gas giant counterpart.

Clouds on tidally locked exoplanets might need a separate discussion, given their importance in observations. As mentioned in Section 4.3.2, cloud particles in the atmospheres could significantly distort the phase curves and might cause a westward offset of the bright spot in the Kepler band (Fig. 6) and regulate the emission temperature on the nightside (Fig. 7). Moreover, as a large opacity source, clouds could impact the radiative flux exchange in the atmospheric layers and influence the dynamics. Roman & Rauscher (2019) investigated the cloud radiative feedback to the atmospheric temperature patterns in a 3D GCM using a parameterized cloud scheme but without tracer transport. To fully understand their effects, 3D distribution of the cloud tracers needs to be resolved in fully coupled radiative hydrodynamical simulations with cloud physics and tracer transport. To date, only two planets have been simulated in the fully coupled fashion: HD 189733 b (Lee et al. 2016) and HD 209458 b (Lines et al. 2018b), but the simulations were so computationally expensive that only short-term integrations were performed. We demonstrate a simple, fully coupled case in Figure 14, the “cloud” case, implementing a gray radiative transfer scheme and assuming constant particle size, but with tracer transport and cloud radiative feedback. It can be seen that cloud mass distribution is highly non-uniform across the globe and seems to follow the temperature distribution well in this case. In more realistic simulations in Lee et al. (2016) and Lines et al. (2018b), particle size seems to anticorrelate with temperature, with smaller particles in the equatorial region on the dayside, and larger particles in high latitudes and the nightside. The cloud distributions are also largely shaped by the circulation pattern. For example, the cloud simulations for HD 209458 b show three distinct zonal bands with one at the equator and two off-equatorial bands (e.g., Lines et al. 2018b), roughly correlating with the zonal-mean zonal wind pattern. Clouds are also highly variable, which might provide temporal evolutions of detected spectral features. See Helling (2019) for more discussions.

Unlike the classic tidally locked planets in circular orbits, the climates on other irradiated planets located further from their host stars are significantly influenced by orbital eccentricity, self-rotation and planetary obliquity. Due to stellar tidal effect, the timescale of the orbital circularization scales as $a^{13/2}$ where $a$ is the semi-major axis (Goldreich & Soter 1966) and that of the spin synchronization depends on $a^6$ (Bodenheimer et al. 2001). As a result, the orbit of a close-in planet is not necessarily circular. In fact, observations indicate that some short-period exoplanets have high eccentricities, for example, hot Jupiter HAT-P-2 b ($e \sim 0.51$), HD 80606 b ($e \sim 0.93$) and sub Neptune GJ 436 b ($e \sim 0.15$). The large difference of this regime compared with the planets in circular orbits originates from the large temporal variation of the stellar flux. If the radiative timescale is short, the atmosphere will experience a significant “eccentricity season.” Several studies (e.g., Langton & Laughlin 2008; Lewis et al. 2010; Kataria et al. 2013; Lewis et al. 2014; Ohno & Zhang 2019a,b) have investigated the atmospheric dynamics of eccentric exoplanets. They found that spatial patterns of the atmospheric temperature and circulation are qualitatively similar to that of planets in circular orbits, although the magnitudes of the temperature fluctuation and wind velocity could change with time. The eccentric orbit significantly influences the shape of the thermal light curve because of intense stellar heating during perihelion and non-uniform orbital velocity of the planet passage.

Outside the synchronization zone, planets have much faster self-rotation rates and are likely to lie in the geostrophic regime (Fig. 13). Warm Jupiters are good examples. Note that these planets are still highly irradiated by the central star. The stellar flux is still stronger than the expected interior flux by several orders of magnitude. Showman et al. (2015b) investigated the influences of planetary rotation on 3D atmospheric dynamics on non-synchronized giant planets. A non-synchronized planet with a slow rotation rate and a high incoming stellar flux is dominated by an equatorial superrotating jet like a canonical hot Jupiter, whereas a planet with a fast rotation rate and a low stellar flux develops mid-latitude jets, like on our Jupiter and Saturn. Nevertheless, these fast rotators manifest a westward flow at the equator instead of equatorial superrotation like on our Jupiter, perhaps due to a lack of moist processes in the high-temperature regime such as water condensation or insignificance of the internal heat from below, which were proposed to be important to drive the equatorial eastward
flow on Jupiter and Saturn (e.g., Schneider & Liu 2009; Lian & Showman 2010). Penn & Vallis (2017, 2018) investigated non-synchronized terrestrial exoplanets. They demonstrated that the substellar point moves westward due to rapid rotation, and the hot spot is shifted eastward from the substellar point. But if the gravity waves are faster than the substellar point movement, the hot spot could be shifted westward, resulting in different thermal phase curves.

Planetary obliquity is much more easily damped by stellar tides than eccentricity (Peale 1999). The non-synchronized planets could have non-zero eccentricities and obliquities. Rauscher (2017) investigated the 3D dynamics on planets in a circular orbit and demonstrated that the atmospheric flow pattern significantly varies with obliquity. Using a 2D shallow-water model, Ohno & Zhang (2019a,b) unified previous studies of non-synchronized planets with different orbital eccentricities, rotation rates and planetary obliquities. They classified the atmospheric dynamics into five regimes considering the radiative timescale and obliquity (see Fig. 1 in Ohno & Zhang 2019a). If the radiative timescale is shorter than the rotation period, the atmosphere displays a time-varying day-night contrast and a day-to-night flow pattern (regime I). When the radiative timescale is longer than the rotation period but shorter than the orbital period, the temperature pattern is controlled by diurnal mean insolation. For obliquity smaller than \( \sim 18^\circ \) (regime II), the temperature distribution is longitudinally homogeneous with an equator-to-pole gradient. An eastward flow is dominant in this regime. For obliquity larger than \( 18^\circ \) (regime III), the atmosphere is heated in the polar region, resulting in a westward wind on the heated hemisphere but an eastward flow on the other hemisphere. If the radiative timescale is longer than the orbital period, the temperature field is dominated by the annual mean insolation. For obliquity smaller than \( 54^\circ \) (regime IV), the atmosphere exhibits an equator-to-pole temperature gradient and an eastward flow on the entire planet. For obliquity larger than \( 54^\circ \) (regime V), the temperature gradient is from the pole to the equator, and a westward flow dominates. Compared with the complicated dynamical behavior, the behaviors of thermal phase curves in the entire parameter space can only be more complex because of the wide range of view geometry (e.g., Rauscher 2017; Ohno & Zhang 2019b; Adams et al. 2019b), as discussed in Section 4.3.

Observations indicate that atmospheres of highly irradiated planets appear to be dynamically variable. Dramatic short-term variability of the peak brightness offset in the Kepler light curve has been observed on hot Jupiter HAT-P-7 \( b \) (Armstrong et al. 2016). Recently, another hot Jupiter Kepler 76 \( b \) has been observed to exhibit large variability in reflection and emission on a timescale of tens of days (Jackson et al. 2019). Hot Jupiter atmospheric flows should be generally rotationally stable (e.g., Li & Goodman 2010; Menou 2019) but the equatorial jets could also be potentially unstable due to barotropic Kelvin-Helmholtz instability and vertical shear instabilities (e.g., Fromang et al. 2016). In general, transients in the atmosphere could also come from several mechanisms such as barotropic and baroclinic instabilities (Pierrehumbert & Hammonds 2019), large-scale atmospheric waves (Komacek & Showman 2019), large-scale oscillations due to wave-mean-flow interactions (Showman et al. 2019), as well as mean-flow interaction with the magnetic field (Rogers 2017). Hot Jupiter simulations found that the globally averaged temperature can be time-variable at the 0.1%-1% level and the variation of globally averaged wind speeds is at the 1%-10% level (Komacek & Showman 2019). The abundances of atmospheric chemical tracers, either gas or clouds, could also vary significantly with time (e.g., Parmentier et al. 2013). Relatively long-term variability could result from the eccentricity and obliquity seasons, as seen in the thermal phase curves on eccentric planets. The long-term variation of the climate is related to orbital dynamics such as the Milankovitch cycles, including periodic changes of obliquity, axial precession, apsidal precession and orbital inclination (e.g., Spiegel et al. 2010; Deitrick et al. 2018a; Deitrick et al. 2018b), but these timescales might be too long for observations. Some close-in planets might be experiencing rapid orbital decay (e.g., WASP 12 \( b \), Maciejewski et al. 2016; Patra et al. 2017), which could also induce interesting time variability in a decadal timescale.

### 6.3 Weakly irradiated planets and brown dwarfs

For a distant planet located far from its host star, the internal heat flux (i.e., self-luminosity) plays a dominant role in the atmospheric dynamics. Extreme cases in this regime are free-floating planets and field brown dwarfs. To date, direct imaging is the best observational technique to characterize these atmospheres, inherited from traditional stellar astronomy. Compared with the close-in planets, observational data on directly imaged planets have a much better quality because of much less stellar contamination. High-resolution spectra provide clues about the vertical distributions of temperature and opacities from chemical tracers, while time-domain photometry, such as rotational light curve and Doppler imaging, can be utilized to unveil their horizontal distributions. The steady patterns in the rotational light curves suggest the mean state of the surface inhomogeneity; the temporal variability of those curves indicates the short-term and long-term
weather patterns—both are closely related to atmospheric dynamics. High-resolution spectroscopy has been applied to measure the rotational line broadening and infer the self-rotation rate of planets and brown dwarfs (e.g., Snellen et al. 2014; Allers et al. 2016; Bryan et al. 2018) but with this technique alone we are not at the stage to separate the surface wind from the internal solid-body rotation. Recently, combining with radio wave observations to infer the rotational rate of the internal magnetic field, Allers et al. (2020) successfully detected differential rotation between the photosphere (from the IR rotational light curves) and the deep interior (from the periodic radio burst, e.g., Williams & Berger 2015) of 2MASS J10475385+2124234, a T6.5 brown dwarf 10.6 pc away. The atmosphere is rotating faster than the interior, suggesting a strong superrotating (eastward) wind in the photosphere with a speed of 650 ± 310 m s\(^{-1}\). This behavior is similar to Jupiter and Saturn, where the global-mean zonal wind is also superrotating, mostly from the broad eastward jet at the equator. The same analysis implies that the Jupiter wind speed at its equator is about 106 km s\(^{-1}\) (Allers et al. 2020), close to the cloud tracking results from Cassini (e.g., Porco et al. 2003). The global-mean zonal wind on Saturn is probably also superrotating based on its strong eastward equatorial wind, but the value is not well constrained due to the uncertainty of Saturn’s solid-body rotation rate.

The atmospheres of planets and brown dwarfs in this regime are fast-rotating and strongly convective. A naïve picture of these atmospheres is an adiabatic temperature profile in the deep atmosphere, rotationally symmetric weather pattern, and homogeneously distributed chemical tracers due to dynamical quenching. The realistic picture is much more complicated and significantly deviates from the above description. The existence of rotational light curves on these bodies implies strong spatial inhomogeneity in their photospheres. Ammonia in the deep troposphere of Jupiter retrieved from the Juno spacecraft data also shows substantial variation across latitude (e.g., Bolton et al. 2017; Li et al. 2017) and also hints that the traditional quenching framework might not be sufficient to understand the tracer transport behavior in the convective atmosphere.

When studying the dynamics of distant directly imaged planets, the first problem is the bottom boundary. These planets do not have surfaces at the bottom, raising the question of whether the flow on these atmospheres is “shallow” or “deep.” This question is twofold. First, is the dominant atmospheric motion horizontal, vertical or intrinsically 3D? Second, is the weather pattern in the upper atmosphere connected to the deep atmosphere?

We first discuss the first question related to the preferential direction of heat transport in the atmosphere. One can analyze the potential temperature gradient in the horizontal direction versus the vertical one. For the large-scale dynamics in the photosphere, the horizontal gradients of potential temperature in the regimes of low and high Rossby numbers were discussed using Equations (22) and (23) (Charney 1963) respectively. By scaling the entire momentum equation, we can combine the two regimes applying one unified scaling of the horizontal (latitudinal) potential temperature gradient \(\Delta \theta_\text{h}/\theta \sim Fr(1 + Ro^{-1})\). The vertical potential temperature gradient is related to the static stability \(N^2 \sim g\Delta \theta_\text{v}/\theta D\), where \(D\) is the flow depth. Thus the ratio of the horizontal to vertical potential temperature contrasts, or sometimes the “baroclinic criticality” \(\xi\), can be scaled as

\[
\xi \sim \frac{1}{1 + Ro^{-1}}, \quad (27)
\]

where the Richardson number \(\text{Ri} \sim \frac{N^2 D^2}{U^2}\) characterizes the atmospheric stratification versus the vertical wind shear. The atmosphere is subject to free convection if \(\text{Ri}\) is smaller than 0.25. For reference, the baroclinic criticality \(\xi\) for Earth’s atmosphere is about unity. This scaling is consistent with an alternative derivation in Allison et al. (1995). It implies that in the tropical regime, the horizontal to vertical potential temperature slope ratio is not dependent on the rotation rate, \(\xi \sim \text{Ri}^{-1}\). For rapidly rotating planets (geostrophic regime), \(\xi \sim \text{Ri}^{-1} \text{Ro}^{-1} \sim \frac{1}{\Omega R_p U/g H}\). Komacek et al. (2019a) used the turbulent cascade scaling (e.g., Held & Larichev 1996) and achieved a more detailed scaling of baroclinic criticality to the planetary parameters for rapidly rotating planets, (see their eq. (6)). It shows a strong dependence of \(\xi\) on the rotation rate, scale height and planetary size, qualitatively consistent with our simple scaling described above.

Allison et al. (1995) classified the dynamics of planetary atmospheres in the Solar System in a \(\text{Ri} – \text{Ro}\) diagram. They found three regimes. Slow-rotating planetary atmospheres like Venus and Titan have both large Richardson (\(\sim 10\)) and Rossby (\(\sim 100\)) numbers and small \(\xi\) (note that Allison et al. 1995 used the ratio \(1/\xi\)). In this regime, a large vertical potential temperature gradient with a small horizontal temperature contrast is developed in a relatively stratified atmosphere due to large-scale Hadley-like circulation. The atmospheres of Earth and Mars lie in the geostrophic regime with a smaller \(\text{Ro}\) (\(\sim 0.1–1\)) but a similar \(\text{Ri}\) with Venus and Titan. In this regime, eddies transport heat effectively in both upward and poleward directions. As a result, the baroclinic criticality \(\xi\) is around unity. The third regime is the giant planet regime with both small Richardson (\(\sim 1\)) and Rossby (\(\sim 0.01\)) numbers.
Strong vertical convection transports heat efficiently, leading to almost vertical isentropes and a large baroclinic criticality $\xi$. The fluid motion in those atmospheres behaves more like the 2D case.

Without knowing the exact temperature and wind structures, it is difficult to estimate the exact Richardson number for an atmosphere outside the Solar System. The tidally locked giant planets likely lie in a different regime from the directly imaged planets and brown dwarfs. Tidally locked giant planets generally have higher Rossby numbers (Fig. 13), and their photospheres are more stably stratified. These planets should mostly occupy a similar corner to Venus and Titan in the $\text{Ri} - \text{Ro}$ diagram. On the other hand, for directly imaged planets and brown dwarfs, the global Rossby numbers appear to be comparable to that of Jupiter (Fig. 13). Their Richardson number could be as large as the gas giants in the Solar System as well. Thus directly imaged planets and brown dwarfs have both lower $\text{Ri}$ and $\text{Ro}$ because of their convective nature and fast rotation. They are likely to overlap with Jupiter and Saturn in the diagram. Their horizontal and vertical potential temperature contrast $\xi$ should be small even with a strong horizontal motion because strong vertical convection could efficiently homogenize the entropy in the vertical direction, or along the direction of the rotational axis if the flow is deep.

How deep is the atmospheric flow? This is not an easy question to answer. Insights come from recent observations from the Juno spacecraft and Cassini “Grand Finale” mission. A deep zonal flow could perturb the gravitational fields of giant planets. The perturbation signals could be measured by precise spacecraft tracking during the orbit (e.g., Kaspi et al. 2010; Kaspi 2013; Liu et al. 2013a; Cao & Stevenson 2017a; Kong et al. 2018a,b,c). The latest data suggest that the surface zonal jets could extend to the interiors, a depth of about 3000 km on Jupiter (e.g., Kaspi et al. 2018; Guillot et al. 2018) and about 9000 km on Saturn (e.g., Less et al. 2019; Galanti et al. 2019). The winds are coupled with the magnetic field and might be damped in the region where Lorentz drag becomes important (e.g., Cao & Stevenson 2017b; Kaspi et al. 2018). The same deep winds have also been inferred on ice giants Uranus and Neptune from Voyager data, to a depth of about 1000 km (Kaspi et al. 2013). If this is a generic behavior on giant planets, the atmospheric flows on tidally locked planets, directly imaged planets and brown dwarfs are potentially deep. This deep circulation might be the key for explanation for inflated radii of hot Jupiters (e.g., Showman & Guillot 2002; Youdin & Mitchell 2010; Tremblin et al. 2017a; Sainsbury-Martinez et al. 2019). On the other hand, because those atmospheres are much hotter than the cold gas giants in the Solar System, the thermal ionization rate should be much higher at the same pressure level, implying that the influence of the magnetic field could be much more important, as we briefly discussed in Section 6.2. Therefore, the deep flows on hot planets might cease at a shallower level than that on the cold gas giants. For example, if an electric conductivity of $1 \text{ S m}^{-1}$ is sufficient to influence the zonal jets in the interior of Jupiter at about $5 \times 10^4$ bar (Guillot et al. 2018), the magnetic breaking might effectively impact the zonal jets at a pretty shallow level on a hot gas giant (could be as shallow as $\sim 100$ bar, see fig. 4 in Wu & Lithwick 2013b). A 3D MHD simulation coupling the realistic radiative photosphere and the convective interior is required to investigate the details but is challenging with current computational facilities.

Neglecting the effect of MHD, purely hydrodynamic simulations of the deep atmospheres have been performed on directly imaged planets and brown dwarfs by Showman & Kaspi (2013). In these models with convective heat flux from below, rotation plays an important role in organizing the large-scale flow pattern. In the slowly rotating regime ($\text{Ro} > 1$), rotation is not important. The rising convective plumes originate from the bottom and rise upward quasi-radially. Convective also appears globally isotropic. In this regime, the traditional mixing length theory (e.g., Clayton 1968; Stevenson 1979b, also see Section 5.1) predicts that the vertical velocity scales as $w \sim (\alpha g F^2/\rho C_p)^{1/3}$ (see the notations under eq. (18)). The temperature fluctuation in this regime is $\Delta T \sim (F^2/\rho^2 c_p^2 g)_{\alpha gl}^{1/3}$. On the other hand, if the body is rapidly rotating ($\text{Ro} < 1$), planetary rotation organizes the large-scale flow to align along with columns parallel to the rotation axis, i.e., the Taylor columns. Assuming the buoyancy force balances vertical Coriolis forces in an isotropic flow, one can deduce the velocity scale $w \sim (\alpha g F^2/\rho c_p)_{\Omega}^{1/2}$ (e.g., Golitsyn 1980, 1981; Boubnov & Golitsyn 1986, 1990; Fernando et al. 1991). The horizontal temperature perturbation is $\Delta T \sim (F^{\alpha} g/\rho c_p)_{\Omega}^{1/2}$. For instance, for a rapidly rotating brown dwarf with a heat flux of $10^4 \text{ W m}^{-2}$, the temperature fluctuation at 1 bar is about 2 K (Showman & Kaspi 2013). The simulations predict that the polar temperature is larger by about one Kelvin than the equatorial temperature because convection occurs more efficiently at high latitude.

In addition to the Rossby number $\text{Ro}$, several dimensionless numbers are useful to characterize the behaviors of convection and deep flow structure on weakly irradiated giant planets and brown dwarfs. The Rayleigh number $Ra$, Ekman number $E$ and Prandtl number $Pr$ are defined.
as follows

\[ \text{Ra} = \frac{\rho g F D^4}{\kappa c_{\text{p}}^{1/2} \nu} \]

\[ E = \frac{\nu}{\Omega D^2} \]

\[ \Pr = \frac{\nu}{\kappa} \]

where \( D \) is the thickness of the convective layer, \( \kappa \) is the thermal conductivity and \( \nu \) is the kinematic viscosity. The Rayleigh number measures the strength of the thermal convection. The Ekman number measures the significance of rotation. The Prandtl number evaluates the relative importance between thermal conduction and momentum transport. Thermal convection occurs if the Rayleigh number is larger than the critical Rayleigh number \( \text{Ra}_{cT} \sim O(E^{4/3}) \) (e.g., Roberts 1968, but note that the Roberts’ paper used the Taylor number \( \text{Ta} = E^{-2} \)). The Prandtl number for gas is on the order of unity. Viscosity of the gas giants is very low, leading to a high Rayleigh number and low Ekman number. The viscosity is also very uncertain. For example, the viscosities from molecular diffusion and that from turbulent diffusion could differ by several orders of magnitude. The extremely low viscosity also imposes a great computational challenge in realistic 3D models on gas giants. In fact, the dynamical regime in current simulations is far from realistic situations (e.g., Showman et al. 2011).

Nevertheless, if the underlying physics governing the thermal convection in a rapidly rotating atmosphere is universal, dimensionless numbers provide useful insights into the dynamical regimes of exoplanets and brown dwarfs. For small Ekman number and Prandtl number of order unity, Schubert & Zhang (2000) classified four important regimes in terms of the ratio of \( \text{Ra} \) to \( \text{Ra}_{cT} \): (1) if \( \text{Ra} / \text{Ra}_{cT} < 1 \), convection is inhibited and geostrophic flows along the azimuthal direction are possible. (2) For \( 1 < \text{Ra} / \text{Ra}_{cT} < O(1) \), convection occurs in the form of azimuthally propagating waves in a columnized configuration parallel to the rotation axis. (3) For \( O(1) < \text{Ra} / \text{Ra}_{cT} < \text{Ra}^{*} / \text{Ra}_{cT} \) where \( \text{Ra}^{*} \) is another critical number. In this regime, small-scale convection disturbs the columns chaotically. (4) \( \text{Ra} / \text{Ra}_{cT} > \text{Ra}^{*} / \text{Ra}_{cT} \), the strong nonlinear advection significantly dominates over the Coriolis effect. Showman et al. (2011) derived the scaling of mean jet speed with heat flux and viscosity in two regimes. In the regime where convection is weakly nonlinear, the jet speed scales approximately with \( F / \nu \). On the other hand, if the convection is strongly nonlinear, the jet speed has a weaker dependence on the heat flux in the form of \((F / \nu)^{1/2}\).

We can scale these numbers from Jupiter’s values to the exoplanet and brown dwarf regime to roughly estimate their behaviors compared to the highly nonlinear, vigorous and chaotic Jovian atmosphere. Here we assume the viscosity on these bodies is the same as that of Jupiter, and the depth of the convection zone is the planetary size that is roughly the Jupiter radius (e.g., \( D \sim R_J \)). Similar to the previous work by Schubert & Zhang (2000), we obtain the dependence of \( \text{Ra} \) and \( E \) on the mass and temperature: \( \text{Ra} \propto g F \propto M T^4_{cT} \) and \( E \propto \Omega^{-1} \propto M^{-0.5} \). For the rotation rate, we have applied the velocity scaling (Eq. (24)). Thus the \( \text{Ra} \) values on hot, massive exoplanets and brown dwarfs are orders of magnitude higher than Jupiter’s value, while the \( E \) does not change too much with the mass. In other words, the effect of rotation on these bodies is not much stronger than that on Jupiter, while the thermal convection could be very different. Schubert & Zhang (2000) claimed that for very massive, rapidly rotating bodies, the \( \text{Ra} \) is large, and the convection should be fully 3D and chaotic. On these bodies, bands of alternating zonal winds like on Jupiter may not be expected.

The horizontal wind speed in the convective region may not be large, but Showman & Kaspi (2013) argued that the upward propagating waves could drive the mean flow in the overlying stratified layers, leading to large-scale circulation and fast horizontal flows. This wave-derived flow impacts the temperature, wind and tracer distributions, and thus the observational signatures in the photosphere. The shallow weather layer in the photosphere, in the simplest picture, can be understood as a forced-dissipative system. The stratified layer is forced mechanically by the convection below the RCB, and the momentum and energy are dissipated by viscous friction and radiation. Unlike the unforced freely-evolving turbulent fluid (e.g., Cho & Polvani 1996), the behavior of the forced-dissipative system strongly depends on the relative strength of the forcing and dissipation. 2D model simulations in Zhang & Showman (2014) show that banded zonal flow patterns spontaneously emerge from the interaction between turbulence and planetary rotation if the bottom heat flux is strong or radiative dissipation is weak. On the other hand, if the internal forcing is weak or radiative dissipation is strong, atmospheric turbulence damps quickly before self-organizing into large-scale jets. Transient eddies and isotropic turbulences dominate the weather pattern. Jupiter appears to lie in the first regime (strong-forcing and weak-damping), but some hot brown dwarfs or directly imaged exoplanets might lie in the latter (weak-forcing and strong-damping). In a more detailed picture, a variety of waves generated in a 3D convective, rotating atmosphere such as gravity waves and Rossby waves propagate upward and dump the momentum in the stratified layer. Those waves could force a large-scale circulation pattern (Showman & Kaspi 2013) as well as multiple zonal jets in the photosphere (Showman et al. 2019).
In particular, the equatorial region could exhibit vertically stacked eastward and westward jets that emerge at the top of the atmosphere and migrate downward over time in a periodic fashion. This behavior resembles the oscillations that were observed in the equatorial regions on Earth (quasi-biennial oscillation, QBO, with a period of ~2yr, Baldwin et al. 2001) and Saturn (quasi-quadrennial oscillation, QQO, with a period of ~4yr, Leovy et al. 1991) and Jupiter (quasi-quadrennial oscillation, QOQ, with a period of ~15yr, Orton et al. 2008; Fouchet et al. 2008). These peculiar wave-mean-flow interaction behaviors on brown dwarfs and directly imaged exoplanets are potentially detectable in future observations.

Although the formation of zonal jets and banded structures on directly imaged exoplanets and brown dwarfs is theoretically possible, searching for banded structures is challenging. Nevertheless, several lines of observational evidence suggest that brown dwarfs might exhibit zonally banded patterns in the photospheres. The first one is the recent detection of differential rotation using IR photometry and radio observations (Allers et al. 2020). The second one is the light curve variability observed by Apai et al. (2017) that suggests the beating of trapped waves in the atmospheric bands on brown dwarfs. The third one is the recently detected polarimetric signals from Luhman 16A (Millar-Blanchaer et al. 2020). The linear polarization signal of 300 ppm on Luhman 16A prefers banded structures rather than oblateness. It would be interesting to apply the above techniques to other brown dwarfs as well as the directly imaged exoplanets and reveal possible banded weather patterns.

If we neglect the photochemistry for weakly irradiated objects (which is not necessarily true, see Moses et al. 2016), dynamical mixing is the primary mechanism that drives the chemical tracers out of thermochemical equilibrium. In general, convection, large-scale circulation, and associated waves and eddies all contribute to the tracer mixing. Empirically, chemical models can be applied to constrain the vertical mixing under diffusive approximation (e.g., Moses et al. 2016; Miles et al. 2020). Theoretically, as discussed in Section 5.1, Miles et al. 2020. The conventional prescription of eddy mixing in a convective atmosphere has been parameterized using the Prandtl mixing length theory (e.g., Prandtl 1925; Smith 1998). However, recent observations show that turbulent transport should homogenize the potential temperature (entropy) instead of the temperature and increase—instead of decrease—the temperature gradient. A recent paper by Tremblin et al. 2019) argued that the radiative source (or other energy sources) was not negligible in the photosphere, which is therefore diabatic instead of adiabatic. In this radiative-convective regime, compositional convection could indeed lead to a reduction of the temperature gradient, but their simulations are limited to a small domain. A large-scale simulation including necessary physics is further needed to explore this idea of “diabatic convection,” and quantify whether or not a cloudless model can indeed explain the redness of very low-G bodies and the L-T transition. On the other hand, it is elusive how this cloudless theory could explain the rotational light curves of brown dwarfs and directly imaged planets. To date, clouds remain a better candidate responsible for the observed rotational modulations.

2D and 3D hydrodynamic simulations with tracer transport in a non-rotating convective atmosphere. They modified the traditional mixing length theory by introducing a new length scale—the scale height of the reacting species under chemical equilibrium—and achieved a better scaling of the averaged 1D eddy mixing strength. This conclusion is in line with that from 2D and 3D global simulations for stably stratified atmospheres in Zhang & Showman (2018a,b) and Komacek et al. (2019b) that found that tracer chemistry needs to be taken into account in the estimate of eddy diffusivity.

Chemical species could also have important feedbacks on dynamics. Radiatively active species modulate the IR opacity distribution and change the radiative energy distribution in the atmosphere. Besides the radiative effect, the atmospheric dynamics and temperature structure could be affected by the change of mean molecular weight during the chemical processes on brown dwarfs and young giant planets (Tremblin et al. 2015, 2016, 2017b, 2019). Chemical transitions of CO ↔ CH4 and N2 ↔ NH3 were proposed to change the vertical gradient of the mean molecular weight in the atmosphere and trigger thermochemical instability, similar to the fingering convection (compositional convection) in Earth oceans. The resulting turbulent transport could lead to a reduction of the thermal gradient near the photosphere and might explain the NIR-band reddening of very low-gravity bodies and the onset of L-T transition in a cloud-free atmosphere (Tremblin et al. 2015, 2016, 2017b). However, Leconte (2018) subsequently pointed out that turbulent transport should homogenize the potential temperature (entropy) instead of the temperature and increase—instead of decrease—the temperature gradient. A recent paper by Tremblin et al. (2019) argued that the radiative source (or other energy sources) was not negligible in the photosphere, which is therefore diabatic instead of adiabatic. In this radiative-convective regime, compositional convection could indeed lead to a reduction of the temperature gradient, but their simulations are limited to a small domain. A large-scale simulation including necessary physics is further needed to explore this idea of “diabatic convection,” and quantify whether or not a cloudless model can indeed explain the redness of very low-G bodies and the L-T transition. On the other hand, it is elusive how this cloudless theory could explain the rotational light curves of brown dwarfs and directly imaged planets. To date, clouds remain a better candidate responsible for the observed rotational modulations.

It looks like clouds are inevitable in interpreting many kinds of observations of exoplanets and brown dwarfs from previous sections. Also, of all chemical tracers, cloud species might play the most complicated role in atmospheric dynamics. Condensable cloud species have three main
effects on the atmospheric dynamics: virtual effect, latent heat effect and radiative effect. Some theoretical models have tried to explore these effects in the context of the directly imaged planets and brown dwarfs (Fig. 15).

In the following, we discuss these effects sequentially. The virtual effect, or the mass-loading effect, originated from the fact that condensable cloud species often have different molecular weights from the background air. The mean molecular weight of the mixed air is different due to the mixing ratio change of condensable vapor during the condensation and evaporation processes. The change in local density would also change the static stability of the atmosphere and affect the dynamics. The virtual effect is particularly significant in hydrogen atmospheres because the background hydrogen is much lighter than all condensable species ranging from water to silicate. For Solar System planets, water condensation and evaporation near the cloud layers could change the vertical density gradient—which is characterized by a quantity called “virtual temperature”—and stabilize the atmosphere against convection (e.g., Guillot 1995; Li & Ingersoll 2015).

The virtual effect critically depends on metallicity in the atmosphere. If condensable vapor is not abundant, the stabilization effect is weak. Imagine an air parcel at the top of the atmosphere that is cooled down by radiation. If there are not many condensable species, the air density will just increase due to cooling, and the parcel will sink. If the condensable species is abundant—quantitatively, its mixing ratio is larger than the critical value given by equation (17) in Leconte et al. (2017)—as the parcel cools, it will first unload the heavier condensable via precipitation, and the air parcel could actually become lighter than the surroundings and stay aloft. Stable stratification is developed to suppress the convection. The convective APE is accumulated below the stratification until the top-layer temperature further drops so that the air parcel is denser than the environment. Then convection starts, and a large amount of the stored energy is suddenly released, resulting in a sizeable erupting storm. This mechanism has been proposed to explain the quasi-periodic giant storms in Saturn’s troposphere occurring about every 30 yr (Li & Ingersoll 2015). Whereas, water in the deep atmosphere of Jupiter is probably not abundant enough to trigger this periodic behavior, explaining the lack of observed giant water storm eruptions on Jupiter. For hot giant planets and brown dwarfs where the silicate could be the major condensates, it is also possible to have similar periodic storms—with a period related to the radiative cooling timescale—if the metallicity is sufficiently high. Detailed numerical simulations have yet to be performed to explore this possibility.

The virtual effect on the convection suppression via cloud formation can also significantly impact the temperature structure. Normally, if a dry convection is suppressed with molecular weight gradient but heat is still allowed to transport, the atmosphere could go into the double-diffusive convection regime (Stern 1960; Stevenson 1979a; Rosenblum et al. 2011; Leconte & Chabrier 2012; Garaud 2018) in which the temperature gradient can be greatly reduced. However, in the presence of cloud formation, Leconte et al. (2017) show that if the condensation occurs much faster than the vapor diffusion, local condensable vapor abundance is almost instantaneously controlled by the temperature change as if the heat and vapor diffuse at the same efficiency. Condensation thus suppresses the double-diffusive instability. As a result, the heat is transported through the slower radiation process near the cloud formation level in the stable layer. The temperature in the deep atmosphere below the clouds could be much hotter than the conventional estimate.

The second effect of the clouds on atmospheric dynamics is the latent heat effect. Latent heat release during cloud formation facilitates moist convection in the atmosphere. As mentioned in Section 4.3.1, the significance of the latent heat effect of different condensable species can be evaluated using the inverse Bowen ratio of $q/c_p T$ (Bowen 1926). Here we adopt the temperature as the condensational temperature $T_c$ and let $q = \chi L$ where $\chi$ is the mass mixing ratio of the condensable species that is primarily determined by metallicity (or surface condition). $L$ and $c_p T_c$ are the latent and thermal heat energy at $T_c$. The inverse Bowen ratio $\chi L/c_p T_c$ for water is about 0.02 if we take $\chi \sim 1\%$. But for silicates (e.g., MgSiO$_3$) on hotter hydrogen atmospheres, the ratio is about $10^{-6}$ if we take $\chi \sim 0.05\%$. Therefore, water has a much larger energetic effect on the moist convection in a cold atmosphere than silicate in a hot atmosphere. Tan & Showman (2017) investigated the importance of silicate latent heating on the atmospheric dynamics on brown dwarfs. They found the latent heat from silicate condensation is small, but the produced eddies in the moist convection could still form zonal jets and storms. The storms are patchy with a temporal evolution on a timescale of hours to days. But the temperature perturbation due to the silicate condensation is localized and only on the order of 1 Kelvin. When averaged over the observed disk, the moist convective storms seem difficult to reproduce the observed large amplitude of the rotational light curves from one to tens of percent on variable brown dwarfs.

Instead, the cloud radiative effect might be the key to understanding the atmospheric dynamics and cloud variability on hotter giant planets and brown dwarfs. Unlike the close-in planets where clouds both reflect the stellar light and interact with the atmospheric IR emission, clouds affect the directly imaged planets and brown dwarfs, most-
Fig. 15 Simulations of atmospheric dynamics and MgSiO$_3$ cloud formation in a convective, self-luminous atmosphere on a Jupiter sized body. Top: 2D local non-hydrostatic simulations of a low-gravity (left, $g = 10^2 \text{m/s}^{-2}$) and a high-gravity (right, $g = 10^3 \text{m/s}^{-2}$) object. The plots are from Zhang et al. (2019b) employing the SNAP model (Li & Chen 2019; Ge et al. 2020). The white-grey patches are clouds and color represents the mass fraction of the silicate vapor. Given other parameters being fixed, the vertical extent of the clouds is more compact as gravity increases. Bottom (from Xianyu Tan, also see his local 3D simulations in Tan & Showman 2020a): 3D global GCM simulations of a rapidly rotating (left, period 2.5 hours) and a slowly rotating (right, period 20 hr) atmosphere. The color represents the outgoing thermal radiation flux at the top of the atmosphere. The patchy structure size increases as the rotation period increases.

ly via IR opacity. Spatially inhomogeneously distributed clouds such as cloud patchiness and vertical extent could strongly impact the radiative budget and modify the horizontal and vertical temperature distributions, which will substantially influence the atmospheric dynamics.

Two types of cloud radiative feedbacks could exist. The first one is 1D, local, spontaneous variability. Consider a local, optically thick cloud column that is perturbed to a higher altitude, resulting in a lower emission to space at the cloud top and a larger heating rate trapped inside the clouds. Vertical motion is enhanced to balance excess heating. Consequently, cloud condensate is mixed upward to increase the cloud top height further. It is positive feedback. 1D simulations in Tan & Showman (2019) coupled the radiative transfer with cloud formation and mixing demonstrated that cloud radiative instability could produce temperature variability up to hundreds of Kelvins on a timescale of one to tens of hours in brown dwarf atmospheres. This 1D, spontaneous variability looks to be a promising mechanism to explain the light curves and their variability. This type of radiative feedback might occur only on a scale smaller than the Rossby deformation radius and in the convective portion of the atmosphere. On a larger scale, the geostrophic adjustment takes the role due to rapid rotation. The second type of cloud radiative feedback occurs on a larger scale and for a relatively stratified atmosphere, for example, on some L dwarfs where clouds only condense in the upper stratified part of the atmosphere (e.g., Tsuji 2002; Morley et al. 2012). In the absence of convection, 1D spontaneous variability would not occur. In this scenario, large-scale cloud radiative instability occurs in the form of 2D or 3D flows with a range of unstable modes (Gierasch et al. 1973). Imbalanced radiative heating and cooling lead to strong temperature contrast, which could drive an overturning circulation. This circulation would transport the clouds horizontally and vertically as feedback. It might be a mechanism to maintain the patchy clouds on directly imaged planets and brown dwarfs (Tan & Showman 2020a).

Several theoretical steps are required to fully understand the cloud radiative effect in 3D convective atmospheres on directly imaged planets and brown dwarfs. First, local, non-hydrostatic models with a simple cloud formation scheme and radiative feedback can shed light on the system's details. For example, how is the horizontally inhomogeneous heating in the IR opaque clouds pro-
duced that drives the turbulence and circulation? How do the circulation feedback to both the tracer transport and the cloud formation? Is there cloud self-aggregation occurring in this high-temperature regime? Does the radiative instability play a role in episodic storms in 3D? Preliminary 2D results from Zhang et al. (2019b) show richness of the physics in these local tests (Fig. 15), including highly variable cloud fraction change, severely depleted vapor in the downwelling region and strong dependence of the gravity with many compact clouds in the high-gravity regime.

Second, 3D large-scale simulations with cloud formation in rotating atmospheres are also important in understanding how rotation impacts the clouds and storms. Local simulations by Tan & Showman (2020a) on an $f$-plane (constant Coriolis parameter) demonstrate that vigorous circulation can be driven and self-sustained by cloud radiative feedback. The local wind speeds can reach $10^3$ m s$^{-1}$, and horizontal temperature contrast could be up to a few hundred Kelvin. Strong rotation suppresses the vertical extent of the clouds. The 3D global simulations (Fig. 15) found that storm size is generally larger at low latitudes and smaller at high latitudes. Cloud thickness also reaches its maximum at the equator and decreases toward high latitudes. At mid and high latitudes, the storm size scales inversely proportionally to the Coriolis parameter $f$, i.e., storm size is smaller in higher latitudes and on more rapid rotators (Fig. 15). Equatorial waves greatly modulate thick clouds and clouds holes in the low latitudes. As a result, brightness variability originates from the inhomogeneously distributed thick and thin clouds and cloud holes, as well as the propagation of equatorial storms. This finding is consistent with the mechanisms proposed for the observed light curve change on Jupiter (Ge et al. 2019). The outgoing thermal radiation could vary locally by a factor of two due to variations in cloud opacity and temperature structure. In an equator-on geometry, the disk-integrated variation could be large enough to explain the observed light curve amplitudes on brown dwarfs (Fig. 15). If these distant bodies are observed from the pole-on geometry, the rotation modulation is mainly contributed by the evolution of turbulent eddies and storm themselves. Detailed discussions refer to a recent review by Showman et al. (2020).

Lastly, because the cloud radiative properties critically depend on the shape and size distribution of the cloud particles (see Sect. 5.3), microphysics in the cloud formation is essential. As mentioned before, integrating the microphysical calculation in 3D dynamical simulations is very computationally expensive and, to date, can only be integrated at a short timescale (e.g., Lee et al. 2016 for simulations on a tidally locked planet). Future work is needed to achieve an efficient parameterization of the cloud microphysics for 3D dynamical models.

6.4 Terrestrial Climates in the Habitable Zone

Last but certainly not the least, we briefly talk about the climate regimes on habitable terrestrial planets. We only keep it brief due to a lack of sufficient observational data for exoplanets at this moment. We will also focus more on the dynamics and climate patterns rather than detailed radiation or chemistry. The habitable zone is defined as the region where the planetary temperature or planetary climate allows liquid water on the surface. Sunlit liquid water is the key to habitability as we understand it today. For example, water is an excellent solvent to allow many chemical and biological reactions due to its polar arrangement of oxygen and hydrogen atoms in the molecular structure. From the atmospheric perspective, sunlit water is also the key to understanding the climates of habitable planets.

Water makes the climate on habitable planets fundamentally different from the other terrestrial planets outside the habitable zone, such as that on the hot terrestrial planet 55 Cancri e, and planets like Venus or the present-day Mars. Currently, the big challenges in the field are not only to observe and characterize these mild terrestrial atmospheres but also to understand the complex behaviors of a moist climate in theory.

The complexity primarily comes from the fact that water has three important phases: vapor, liquid and ice. Each phase has significant but different roles in the terrestrial climate system. In the vapor phase, water vapor is a strong greenhouse gas with strong IR opacity across most wavelengths. In the liquid phase, the ocean regulates the climate on a timescale of decades and longer. Also, liquid water is the catalyst in the carbon cycle, including $\text{CO}_2$ dissolution, surface erosion and probably plate tectonics, and thus plays a part in regulating the climate on a geological timescale. In the ice phase, water ice floats on top of liquid water, contributing to the planetary surface albedo. Besides, both liquid and ice clouds significantly impact the energy budget of the system through cloud albedo and also through cloud opacity. The phase transition between water vapor and liquid/ice is associated with substantial latent heat exchange. As mentioned in Section 6.3, water ranks among the top of all species in terms of the inverse Bowen ratio.

Consequently, the climate on a habitable planet with a large amount of liquid water is naturally in the moist regime, in which the water latent heat changes and hydrological cycle dominates many climate behaviors. One must understand water before we understand the moist climate on habitable planets. The fact is, even though Earth is a well-observed planet, we have not fully understood the moist climate dynamics yet. For example, the fundamental mechanism of the Madden-Julian oscillation in the Earth’s tropics has not been fully explained.
The classical habitable zone boundaries for inner and outer edges are estimated from 1D radiative-convective models (e.g., Kasting et al. 1993; Selsis et al. 2007; Kitzmann et al. 2010; Wordsworth et al. 2010; Kaltenegger et al. 2011; Zsom et al. 2012; Kopparapu et al. 2013; Rugheimer et al. 2013; Grenfell et al. 2014; Kopparapu et al. 2014; Rugheimer et al. 2015; Turbet et al. 2016; Yang et al. 2016b; Meadows et al. 2018). For example, the estimates in Figure 16 (from Kopparapu et al. 2013) show the dependence of insolation required for the habitable zone on stellar temperature. The inner and outer boundaries are not merely controlled by the insolation itself. The second-order dependence comes from different stars emitting different spectral energy distributions to which the atmosphere responds differently. In the atmosphere, water vapor absorption is strong in the near-IR, while ice/snow/cloud albedo and atmospheric scattering are strong in the short wavelengths. As a consequence, given the same insolation, a planet orbiting a hotter star will absorb less stellar light in the atmosphere than a planet orbiting a colder star, pushing the habitable zone closer to the hotter star (Fig. 16).

3D theoretical dynamical models have been put forward to understand the moist climate in a larger parameter space for Earth-like planets (non-tidally-locked planets) in the habitable zone (e.g., Abe et al. 2011; Madden & Julian 1971), and cloud feedbacks remain among the largest sources of uncertainties contributing to the current climate model diversity (Webb et al. 2013). Also, one should keep in mind that desert planets with little water on their surfaces could also be abundant in the habitable zone around M dwarfs (e.g., Tian & Ida 2015). Whether the role of water on the climates and habitability on these arid planets is important or not needs further analysis.
Boschi et al. 2013; Shields et al. 2013; Yang et al. 2013; Leconte et al. 2013; Yang et al. 2013; Shields et al. 2013; Wolf & Toon 2014; Shields et al. 2014; Kaspi & Showman 2015; Yang et al. 2014b; Shields et al. 2014; Wolf & Toon 2015; Pierrehumbert & Ding 2016; Shields et al. 2016b; Popp et al. 2016; Godolt et al. 2016; Wang et al. 2016; Ding & Pierrehumbert 2016; Wolf et al. 2017; Kilic et al. 2017; Kopparapu et al. 2017; Way et al. 2017; Way et al. 2018; Adams et al. 2019a; Yang et al. 2019c, also see a recent review in Shields et al. 2016a and Kopparapu et al. (2019) and the white paper by Wolf et al. 2019b). When complicated 3D climate dynamics are considered, the boundaries could be drastically different from the 1D model predictions (Fig. 16). The 1D and 3D models show that the outer edge is dominated by many factors including albedo, greenhouse gas inventory, CO₂ collapse, clouds, carbonate-weather feedback, surface pressure and so on (e.g., Kasting et al. 1993; Kopparapu et al. 2013; Turbet et al. 2017). The inner edge could also be strongly affected by the planetary rotation rate that largely modulates the cloud distribution and planetary albedo (e.g., Yang et al. 2014b; Way et al. 2016).

Climate on fast-rotating terrestrial planets can be classified into several regimes (e.g., Goldblatt 2015; Wolf et al. 2017). Using the global mean surface temperature (\(T_s\)) on 3D simulations as a proxy, Wolf et al. (2017) defined four potentially stable climate states that are separated by abrupt climatic transitions (Fig. 16): snowball (\(T_s < 235 \text{ K}\)), waterbelt (235 K < \(T_s < 250 \text{ K}\)), temperate (275 K < \(T_s < 315 \text{ K}\)) and moist greenhouse (\(T_s < 330 \text{ K}\)). Those states are in stable equilibrium where the incoming stellar flux balances the outgoing thermal radiation, and the states are resilient against any small perturbation. In the snowball state that might have occurred in Earth’s Neoproterozoic glaciations (0.75 to 0.54 billion years ago), the ocean surface is globally covered by ice (e.g., Kirschvink 1992; Hoffman et al. 1998; Liu et al. 2020). In the waterbelt state, the equatorial ocean can be ice-free, although the middle and high latitudes are fully glaciated (e.g., Hyde et al. 2000; Abbot et al. 2011). The current Earth is in the temperate state. When the surface temperature further increases, the planet could enter the moist greenhouse state in which the atmosphere is warm enough so that water is no longer trapped by the cold tropopause, resulting in a moist stratosphere (e.g., Kasting et al. 1993). In this scenario, hydrogen loss to space is efficient due to the photolysis of water vapor. The specific rate of water photolysis likely depends on the stellar activity (e.g., Chen et al. 2019). Also, there is a strong hysteresis (or bistability) between the snowball and temperate climate states, meaning that there exist two stable climate solutions given the same stellar flux (e.g., Budyko 1969), although the stellar flux range allowing the hysteresis is sensitive to other parameters such as rotation rate and surface pressure. For example, it decreases or even disappears if we increase the rotation rate of the planet (Abbot et al. 2018) or on tidally locked exoplanets (Checlair et al. 2017).

According to the definition of habitability, a habitable climate includes waterbelt, temperate and cooler moist greenhouse states. If the stellar flux is small, the climate enters the snowball state. Snowball state can be habitable in some conditions. A good example is the Earth’s climate in the Neoproterozoic era. For exoplanets, in local regions such as high geothermal heat flux or thin ice, photosynthesis life can still exist. Another possibility is that although the ocean is covered entirely by ice and snow, some continents are ice-free and are warm enough to maintain liquid water (e.g., Paradise et al. 2019).

If the stellar flux is large, the entire atmosphere will enter the thermal runaway process, and there is no stable climate solution until the ocean water is all evaporated into the atmosphere. The underlying mechanism of the so-called runaway greenhouse climate is owing to the limit of thermal emission in the moist atmosphere. There are two limits (Nakajima et al. 1992). The moist stratosphere limit is the “Komabayashi-Ingersoll limit” (Komabayasi 1967; Ingersoll 1969) where the stratospheric emissivity is set by 100% relative humidity at the tropopause. Before the stratospheric limit is reached, the thermal emission of the troposphere itself can be limited because the moist adiabat in the troposphere has to follow the saturation vapor pressure of water when the entire atmosphere becomes water-dominated. This lower limit—Goldblatt & Watson (2012) termed it “Simpson-Nakajima limit” (Simpson 1927; Nakajima et al. 1992)—is more relevant to runaway greenhouse effect. Venus might have experienced the runaway greenhouse process (Ingersoll 1969, or at least moist greenhouse, Kasting 1988), resulting in very little water and a high D/H ratio in the atmosphere.

The climate state classification might depend on specific models because of the complexity of the 3D climate models. For example, the simulations in Wolf et al. 2017 found that there is no stable climate solution in the temperature range of 250 K < \(T_s < 275 \text{ K}\) and 315 K < \(T_s < 330 \text{ K}\). However, other models could produce stable climate solutions between 250 and 275 K with oceanic dynamics (e.g., Yang et al. 2011). Also, using a different cloud scheme, other models found stable climate solutions between 315 and 330 K (e.g., Leconte et al. 2013). Moreover, a moist greenhouse state does not even exist in some other Earth-like models. For example, Leconte et al. 2013 found an abrupt transition from dry stratosphere to runaway greenhouse state. Future model intercomparison
is needed to understand the differences among 3D climate models.

Planets in orbits close to low-mass stars (such as M dwarfs) could also be habitable. They are also important targets in future observational surveys due to their proximity to the host stars so that transits are more common and the S/N is larger. These planets are likely to be synchronously rotating or in spin-orbit resonances (such as Mercury) due to stellar tides. The tidal locking semi-major axis is empirically found to scale with the stellar mass to the 1/3 power (e.g., Peale 1999; Kasting et al. 1993; Dobrovolskis 2009; Edson et al. 2011; Haqq-Misra & Kopparapu 2015).

Assuming the stellar mass-luminosity relationship $L \propto M^\alpha$ and the mass-radius relationship $R \propto M^{\beta}$, the insolation $I$ at the tidal locking distance is dependent on the stellar effective temperature in the following relation

$$I \propto T^7/\alpha.$$ \hspace{1cm} (29)

Taking $\alpha \sim 3.7$ and $\beta \sim 0.724$ (Demircan & Kirkham 1991), we get the “Tidal Locking Radius” curve: $I \propto T^{7/\alpha}$ (Fig. 16). This implies that the incoming stellar flux at the tidal locking distance depends strongly on stellar temperature. Basically, any star hotter than the M (or late K) type is not likely to have Earth-like planets (in terms of instellation) in synchronous rotation (Haqq-Misra & Kopparapu 2015).

For planets in this tidal locking regime, a number of 3D climate models have also been adapted from the Earth to study circulation patterns (e.g., Joshi et al. 1997; Joshi 2003; Merlis & Schneider 2010; Showman et al. 2010; Wordsworth et al. 2011; Pierrehumbert 2011; Edson et al. 2011; Showman et al. 2013b; Leconte et al. 2013; Yang et al. 2013; Hu & Yang 2014; Yang et al. 2014c; Wang et al. 2014; Carone et al. 2014, 2015, 2016; Wordsworth 2015; Koll & Abbot 2015; Koll & Abbot 2016; Haqq-Misra & Kopparapu 2015; Turbet et al. 2016; Pierrehumbert & Ding 2016; Kopparapu et al. 2016; Kopparapu et al. 2017; Boutle et al. 2017; Way et al. 2017; Wolf et al. 2017; Wolf 2017; Fujii et al. 2017; Noda et al. 2017; Fauchez et al. 2019; Wolf et al. 2019a; Del Genio et al. 2018; Kopparapu et al. 2017; Haqq-Misra et al. 2018; Lewis et al. 2018; Chen et al. 2019; Komacek & Abbot 2019; Yang et al. 2019b; Yang et al. 2019c; Ding & Wordsworth 2020; Yang et al. 2020; Suissa et al. 2020, also see a recent review in Shields et al. 2016a, Pierrehumbert & Hammond 2019 and Kopparapu et al. 2019 and the white paper by Wolf et al. 2019b). Some of those models also investigated the planets in other possible spin-orbit resonance states (e.g., Wordsworth et al. 2011; Yang et al. 2013; Wang et al. 2014; Way et al. 2017; Boutle et al. 2017; Yang et al. 2019a; Suissa et al. 2020). Pierrehumbert & Hammond (2019) provided a detailed review of the atmospheric dynamics on tidally locked terrestrial planets. Because of the tidally locked configuration, the climate behaves differently from that on non-tidally locked planets. In particular, interesting day-night differences have been suggested in this regime compared with the zonally homogenous fast rotators like Earth. One of the noticeable behaviors is that strong convective clouds form on the dayside and significantly increase the planet’s albedo, thus allowing the planet to be habitable under higher insolation levels (e.g., Yang et al. 2013; Kopparapu et al. 2016; Way et al. 2016). Also, a relatively thin atmosphere could lead to atmospheric collapse on the nightside on tidally locked planets (e.g., Kite et al. 2011; Wordsworth 2015). In the nearly snowball state, the climate of tidally locked planets could show a peculiar “eyeball” state with an open ocean near the substellar point (Pierrehumbert 2011) but a recent study indicates that the open ocean might also be closed by sea-ice drift under certain circumstances (Yang et al. 2020).

In terms of classification, several regimes have been demarcated using 3D simulations with Earth’s atmospheric compositions. A series of papers by Carone et al. (2014, 2015, 2016, 2018) has investigated the terrestrial troposphere circulation regimes as functions of planetary radius and orbital period (while still being tidally locked). The wind field is influenced by tropical Rossby waves that lead to equatorial superrotation and by extratropical Rossby waves for two high-latitude wind jets. They demarcated four circulation regimes, including the troposphere and stratosphere, in terms of the rotational period $P$ (see fig. 1 in Carone et al. 2018). (1) For $P < 3\,\text{d}$, there is a strong mixture of equatorial superrotation and high-latitude wind jets in the troposphere. The stratosphere circulation from the equator to the pole is inefficient due to an anti-Brewer-Dobson-circulation induced by tropical Rossby waves. (2) For $3\,\text{d} < P < 6\,\text{d}$, the tropospheric wind pattern is either equatorial or high-latitude wind jets. The stratospheric equator-to-pole transport is a bit more efficient because extratropical Rossby waves counterbalance the effect of tropical Rossby waves. (3) For $6\,\text{d} < P < 25\,\text{d}$, there is weak superrotation in the troposphere, and stratospheric transport could be efficient if there is stratospheric wind breaking. (4) If $P > 25\,\text{d}$, the tropospheric wind exhibits radial flow structures. The stratospheric transport is efficient because of thermally driven circulation.

Similarly, in an idealized GCM study Noda et al. (2017) found that the tropospheric circulation patterns change as the planetary rotation rate increases. They identified four regimes. Type I is a day-night thermally direct circulation, Type II shows a zonal wavenumber one resonant Rossby wave on an equatorial westerly jet. Type III
exhibits long timescale north-south asymmetric variation and Type IV manifests a pair of mid-latitude westerly jets.

Haqq-Misra et al. (2018) further characterized the circulation patterns in more realistic 3D simulations into three dynamical regimes using the non-dimensional equatorial Rossby deformation radius and the non-dimensional Rhines length (to the planetary radius). The “slow rotation regime” occurs when both the Rossby deformation radius and the Rhines length exceed the planetary radius, and the circulation has a mean zonal circulation from the dayside to the nightside. In the “rapid rotation regime” with Rossby deformation radius less than the planetary radius, the circulation displays a mean zonal circulation that partially spans a hemisphere but with banded clouds over the sub-stellar point. In the third regime, “Rhines rotation regime,” which occurs when the Rhines length is greater than the radius but the Rossby deformation radius is less than the radius, a thermally direct circulation emerges from the dayside to the nightside, but midlatitude jets also exist. These dynamical regimes can be characterized by thermal emission phase curves from future observations.

As mentioned in Section 6.1, the “WTG parameter” \( \Lambda \) in Pierrehumbert & Hammond (2019) can be used to characterize the dynamical regimes for tidally locked terrestrial planets. Their 3D simulations revealed that if \( \Lambda > 5 \), the horizontal temperature distributions are homogeneous (e.g., WTG behavior) due to energy-transporting circulations. However, nonlinearity could occur that breaks the WTG behavior if the circulations become very strong. As for time variability, the circulation usually exhibits small temporal fluctuation in the large \( \Lambda \) regime (e.g., \( \Lambda > 5 \)). The planetary-scale transients such as strong eddies and wave disturbances appear to be important if \( \Lambda \lesssim 2 \). A likely source of these transients is the baroclinic instability.

The parameter space of climate on habitable planets is vast. There is still much to explore to characterize all the climate regimes. Other than the important parameters such as stellar flux and planetary rotation rate that we have discussed above, other parameters such as total atmospheric pressure, atmospheric composition (especially the radiatively active species such as CO\(_2\), CH\(_4\) and clouds, or condensible atmospheres, Ding & Pierrehumbert 2016), and the ocean and ice dynamics (Hu & Yang 2014; Yang et al. 2020) have also been shown to be important. Furthermore, other less explored parameters might be crucial as well, for example, planetary eccentricity (e.g., Wang et al. 2014; Adams et al. 2019a), obliquity (e.g., Dobrovolskis 2009; Wang et al. 2016; Kilic et al. 2017; Kang et al. 2019; Kang 2019a,b), planetary gravity and radius (e.g., Yang et al. 2019a; Yang & Yang 2019), tidal heating (e.g., Barnes et al. 2013), magnetic field (e.g., Dong et al. 2017, Dong et al. 2018a), stellar activity (e.g., Badhan et al. 2019; Chen et al. 2019; Airapetian et al. 2020), initial water inventory (e.g., Ding & Wordsworth 2020), interior and surface processes (e.g., Walker et al. 1981; Charnay et al. 2017) and so on. Check the recent review by Shields et al. (2016a) and Kopparapu et al. (2019) for a detailed summary of these parameters and discussions.

7 Future Prospects

Looking back at the dawn of exoplanet and brown dwarf science, we are impressed by the amount of information we have been able to retrieve from these distant, unresolved faint pixels. Although it is difficult to characterize their atmospheres without any bias due to limited data quality and our a priori knowledge from the Solar System bodies, preliminary analysis of the relationship between the observed atmospheric characteristics and fundamental planetary parameters has identified interesting yet somewhat arguable regimes and trends in the current substellar atmosphere sample. The stellar and planetary parameters considered here include mass, radius, gravity, rotation rate, metallicity, surface albedo, internal luminosity, stellar luminosity, stellar spectra and orbital parameters. From these parameters, one can derive several fundamental scales in the atmospheres. For length scales, there are the pressure scale height, planetary radius, Rossby deformation scale and Rhines scale. For timescales, there are radiative timescale, conductive timescale, wind transport timescale, eddy transport timescale, chemical timescale and mass loss timescale. For velocity scales, there are light speed, sound speed, rotational velocity, jet velocity, eddy velocity, escape velocity and thermal velocity. For energy scales, there are thermal energy, photon energy, latent heat, gravitational potential energy, KE and convective potential energy. In this review, we have shown that atmospheric trends and regimes can be linked back to these scales in the atmospheres and thus to the stellar and planetary parameters. Based on these scales, we can derive several dimensionless numbers such as the Jeans parameter, Rossby number, Froude number, Mach number, Alfvén Mach number, Richardson number, Rayleigh number, Ekman number, Taylor number, Prandtl number, WTG parameter, \( \Omega \tau_{\text{dyn}}, \tau_{\text{dyn}}/\tau_{\text{drag}}, \tau_{\text{dyn}}/\tau_{\text{rad}}, \) inverse Bowen ratio \((q/c_p T_{\text{eq}}), \tau_{\text{vis}}/\tau_{\text{RT}}, \tau_{\text{dyn}}/\tau_{\text{chem}} \) and \( F_{\text{ext}}/F_{\text{int}} \). We demonstrated that these numbers are important to understand the behaviors of various planetary climate. Simple scaling laws shed light on the underlying mechanisms. There are two key aspects of linking the atmospheric characteristics to bulk planetary parameters. First, in this review, we have mainly focused on the influence of the planetary parameters on the atmospheric behaviors and resulting signals. Second, atmospheric processes could also significantly impact the planetary parame-
ters, which were not discussed in this review. Here we briefly mentioned some important aspects. As the outer boundary of a planetary body, planetary atmospheres directly exchange mass and energy with space. A number of studies have signified that radiative processes in the outer gaseous envelopes (i.e., atmospheres) could greatly influence planetary-mass accretion rates (e.g., Lee et al. 2018; Ginzburg & Chiang 2019) and subsequent radius evolution (e.g., Chabrier & Baraffe 2000; Burrows et al. 2003a; Fortney et al. 2010). Photoevaporative and core-powered mass loss processes have been proposed to explain the observed “radius gap” in low-mass planets (e.g., Lopez & Fortney 2013; Owen & Wu 2013; Lopez & Fortney 2013; Ginzburg et al. 2018) as discussed in Section 3.3. A significant mass loss would shrink the planetary size and change the rotational and orbital angular momentum and, thus, the planetary rotation rate and orbital parameters. For example, if the planetary mass loss is larger than 10%, Matsumoto & Ogihara (2020) showed that planetary orbits in resonant chains could be destabilized. Several mechanisms related to atmospheric dynamics have been proposed to explain the unexpected inflated sizes of some hot Jupiters (e.g., Showman & Guillot 2002; Youdin & Mitchell 2010; Tremblin et al. 2017a; Sainsbury-Martinez et al. 2019). The ionization rate in the atmosphere is vital for the proposed Ohmic heating inflation mechanisms on hot Jupiters (e.g., Batygin & Stevenson 2010; Huang & Cumming 2012; Wu & Lithwick 2013b) and the terminal rotation rates of giant planets (e.g., Batygin 2018). For smaller planets, distribution of atmospheric compositions such as dusty outflow or high-altitude photochemical haze might greatly enlarge the apparent planetary radii in short wavelengths (e.g., Sekiya et al. 1980; Cubillos et al. 2017b; Lammer et al. 2016; Wang & Dai 2019; Gao & Zhang 2020). Atmospheric escaping flux from ablating planets might pollute the stellar emission (e.g., Haswell et al. 2020). This pollution might not only change the received stellar flux in the atmosphere but also create a possible new way to detect the atmospheric compositions even without a single photon from the planet itself. Taking into account the feedback of atmospheric processes on the basic stellar and planetary parameters during planetary formation and evolution would further complicate the characterization of exoplanets and brown dwarfs. New trends and regimes might also emerge in these interesting climate systems.

The future is challenging, but promising. For a long time we are likely to lack spatially-resolved images or in situ information from entry probes or flyby missions. That will put a cap on our understanding of the detailed weather and climate on specific bodies, as we have learned lessons from Solar System science in the spacecraft age. However, the explosion of empirical data from dedicated telescopic observations and large surveys would unveil, at least to first order, the nature of diversity of planetary climate and the main regimes in large parameter space. The population of exoplanets continues rapidly growing from past and future space-based transit surveys such as Kepler, K2, TESS and CHEOPS as well as ground-based surveys like WASP, KELT, MASCARA, HAT and TrEs for hot and warm gas giants, and NGTS, MEarth, Trappist, SPECULOOS, and ExTrA for smaller and cooler planets. Most data about exoplanet atmosphere observations will come from current and future space telescopes like HST, CHEOPS, TESS, JWST, PLATO, ARIEL, WFIRST, OST, HabEx, LUVOIR and high-precision ground-based facilities such as VLT/FORS2, VLT/ESPRESSO, VLT/CARMENES, TNG/HARPS, GTC/OSIRIS, VLT/Sphere, Gemini/GPI, Subaru/SCExAO, Magallan/MagAO(-X) and VLTI-GRAVITY. Several Chinese space missions on exoplanets such as CSST, Miyin and CHES are also in preparation. In the upcoming decade, JWST and ARIEL might be the two most important IR telescopes for characterizing substellar atmospheres (see discussions in Greene et al. 2016; Tinetti et al. 2016). The detailed, high-contrast observations of atmospheres on habitable terrestrial planets and possible biosignatures might need to await the next generation of large ground-based facilities such as the ELT, GMT and TMT, and space-based observatories such as HabEx and LUVOIR.

To aid the interpretation of telescopic observations, laboratory experiments are critically important to improve the accuracy of input parameters in atmospheric modeling (Fortney et al. 2019). In terms of atmospheric radiative properties, spectroscopic data include spectral line intensities, line broadening parameters, line mixing parameters, collisional deactivation parameters for different quantum states for Non-LTE calculations, collision-induced absorption for various gas mixtures, aerosol optical properties such as absorption and scattering coefficients, single scattering albedo and the scattering phase function. These spectroscopic data need to be improved to cover more wavelengths, higher resolutions, and various temperature and pressure environments to reveal diverse substellar atmospheric conditions. In particular, high-resolution spectroscopy might require a precise measurement of the line core location and line shape that might deviate from the conventionally assumed Voigt profile. For the chemical properties, chemical kinetic data such as reaction rates and their dependence on temperature, pressure and quantum states are largely uncertain. To model the detailed formation process of clouds and chemical hazes, one needs to measure the mechanical, thermal and electronic properties of the particulate matters in the atmosphere, such
as surface tension, coagulation properties, surface reaction rates of dusty grains, heat capacity, latent heat and electronic charge. Laboratory chamber simulations of the gas chemistry, photochemical haze formation and cloud condensation are crucial to understanding these very complicated processes. The electrical properties of the gases are also important in understanding ion chemistry, lightning, and possible energy processes such as Ohmic heating. The equation of state of gas mixtures under high temperature and pressure are also important to model the deep atmospheres of giant planets. In the situation where lab experiments are not available, ab initio calculations could provide alternatively useful information for the model input, such as the molecular line information from molecular physics simulations (e.g., Exomol, Tennyson & Yurchenko 2012) and the equation of state of hydrogen-helium mixtures in the high-temperature and high-pressure regimes from density functional molecular dynamics simulations (e.g., Militzer & Hubbard 2013).

Theoretical advancement is also required to synthesize new knowledge to improve the current framework. It might proceed on three fronts. First, refining the data retrieval techniques for the most robust information to be derived from observations. This task includes both designing the most efficient observational mode using available facilities and improving numerical techniques for inverse modeling. Second, exploring the parameter space and unveiling new mechanisms. Some of the new physics might come from testing the current theory in extreme conditions, such as super-fast rotators, atmospheric collapses and strong star-planet interactions. Some might come from the analysis of interactions among different processes such as dynamics, radiation and chemistry (including grain chemistry and cloud microphysics), such as the thermal dissociation and recombination of hydrogen on ultra-hot Jupiters. Could other exothermic and endothermic chemical reactions also significantly affect the temperature and dynamical structures? Is radiation pressure important for dynamics on very hot exoplanets? Third, improving detailed numerical simulations. There are still many challenging technical problems. For example, how to appropriately represent the radiative transfer, chemistry and cloud processes in 3D dynamical simulations while not significantly slow down the computational efficiency? What are the proper upper and lower boundary conditions in atmospheric simulations? Also, as the model becomes more and more complicated, a hierarchical approach to understanding the detailed mechanisms using models with different levels of complexity would be appreciated.

The community should sustain intensive interaction to broaden and deepen our understanding of the substellar atmospheres. We encourage the observers, experimentalists and theorists work together to collaborate and participate in long-term workshops about exoplanets and brown dwarfs such as Exoplanets, ExoClimes, Cool Stars, Cloud Academy, UCSC OWL Exoplanet Summer Workshop and many more (e.g., check Future Meetings on Extrasolar Planets). We also encourage data sharing activities such as open-source data and software (e.g., BART; ExoTransmit; ExoClimes Simulation Platform; VPLanet). Because atmospheric models are highly technical, model intercomparison projects should be promoted to understand the theoretical consistency and differences in the field (e.g., Yang et al. 2019c; Barstow et al. 2020; Fauchez et al. 2020).

In the end, we should not forget to connect the new knowledge learned from those exotic substellar atmospheres back to the Solar System bodies where in-situ data are available. With rapidly evolving data and theory—the two prongs in Hume’s fork—a unified, first-principle climate theory for diverse exoplanets and brown dwarfs can be established. It will place the Earth and Solar System in the large charts of atmospheric regimes and trends across the entire multi-dimensional parameter space, which are yet to be explored, in the universe.

Acknowledgements The stellar and planetary parameters were obtained from the NASA Exoplanet Archive, the Extrasolar Planets Encyclopaedia and ExoKyoto. I thank Yunwen Duan for help with the data compilation and analysis. I thank editor Jianghui Ji for inviting and motivating me to complete this review. Thanks to Elena Manjavacas, Kevin Zahnle, Erik Petigura, Mike Line, Hannah Wakeford, Mark Marley, Didier Saumon, Peter Gao, Xianyu Tan, and Eric Wolf for the figure data. Special thanks to Kevin Zahnle, Xianyu Tan, Thaddeus Komacek, Julianne Moses, Mark Marley, Pete & Dinah Showman, Eric Wolf, and Jun Yang for their detailed comments. I also thank Justin Erwin, Jonathan Fortney, Vivien Parmentier, Michael Liu, Katelyn Allers and Linfeng Wan for helpful discussions. Lastly, I specifically dedicate this review article to my postdoc advisor Dr. Adam Showman (1968–2020) for his help in these years. Brown dwarfs never die, they simply fade away.

References

Artigau, É. 2018, in Handbook of Exoplanets, eds. H. J. Deeg & J. A. Belmonte (Cham: Springer International Publishing), 555
Barnes, R., Mullins, K., Goldblatt, C., et al. 2013, Astrobiology, 13, 225
Biller, B. 2017, Astronomical Review, 13, 1
Birkby, J. L. 2018, in Handbook of Exoplanets, eds. H. J. Deeg & J. A. Belmonte (Cham: Springer International Publishing), 1485
X. Zhang: Atmospheres on Exoplanets and Brown Dwarfs

Popp, M., Schmidt, H., & Marotzke, J. 2016, Nature Communications, 7, 1
Proudman, J. 1916, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 92, 408
Pruppacher, H. R., & Klett, J. D. 1980, Nature, 284, 88
Roberts, P. H. 1968, Royal Society of London Philosophical Transactions Series A, 263, 93
Rogers, T. M. 2017, Nature Astronomy, 1, 0131
Rossow, W. B. 1978, Icarus, 36, 1
Samra, D., Helling, C., & Min, M. 2020, A&A
Schaefer, L., & Fegley, B. 2010, Icarus, 208, 438
Schlichting, H. E., Sari, R., & Yalinezhin, A. 2015, Icarus, 247, 81
Schneider, E. K. 1977, Journal of the Atmospheric Sciences, 34, 280
Schneider, T. 2004, Journal of Atmospheric Sciences, 61, 1317
Schneider, T., & Liu, J. 2009, J. Atmos. Sci., 66, 579
Schobeiri, M. R., & Strobel, D. F. 1984, Dynamics of the Middle Atmosphere, 45
Schubert, G., & Mitchell, J. L. 2013, Comparative Climatology of Terrestrial Planets, 181
Taylor, G. I. 1917, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 93, 99
Thompson, D. W. J., Bony, S., & Li, Y. 2017, Proceedings of the National Academy of Sciences, 114, 8181
Tian, F., Chassefière, E., Leblanc, F., & Brain, D. 2013, cctp, 567
Tian, F., & Ida, S. 2015, Nature Geoscience, 8, 177
Tinetti, G., Encrenaz, T., & Coustenis, A. 2013, A&AS, 21, 63
Tsuji, T. 2001, in Ultracool Dwarfs, eds. H. R. A. Jones, & I. A. Steele (Berlin, Heidelberg: Springer), 9
Vallis, G. K. 2006, Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation (Cambridge, UK: Cambridge Univ. Press)
Wang, Y., Read, P. L., Tabataba-Vakili, F., & Young, R. M. B. 2018, Quarterly Journal of the Royal Meteorological Society, 144, 2537
Watson, J. J. 1851, in Abstracts of the Papers Communicated to the Royal Society of London No. 5 (The Royal Society London), 604
Webb, M. J., Lambert, F. H., & Gregory, J. M. 2013, Climate Dynamics, 40, 677
Williams, G. P. 1978, Journal of Atmospheric Sciences, 35, 1399