# ATMOSPHERIC CHEMISTRY FOR ASTROPHYSICISTS: A SELF-CONSISTENT FORMALISM AND ANALYTICAL SOLUTIONS FOR ARBITRARY C/O

Kevin Heng[1], James R. Lyons[2], and Shang-Min Tsai[1]

[1] University of Bern, Center for Space and Habitability, Sidlerstrasse 5, CH-3012, Bern, Switzerland; kevin.heng@csh.unibe.ch
[2] Arizona State University, School of Earth and Space Exploration, Bateman Physical Sciences, Tempe, AZ 85287-1404, USA

## ABSTRACT

We present a self-consistent formalism for computing and understanding the atmospheric chemistry of exoplanets from the viewpoint of an astrophysicist. Starting from the first law of thermodynamics, we demonstrate that the van't Hoff equation (which describes the equilibrium constant), Arrhenius equation (which describes the rate coefficients), and procedures associated with the Gibbs free energy (minimization, rescaling) have a common physical and mathematical origin. We address an ambiguity associated with the equilibrium constant, which is used to relate the forward and reverse rate coefficients, and restate its two definitions. By necessity, one of the equilibrium constants must be dimensionless and equate to an exponential function involving the Gibbs free energy, while the other is a ratio of rate coefficients and must therefore possess physical units. We demonstrate that the Arrhenius equation takes on a functional form that is more general than previously stated without recourse to tagging on ad hoc functional forms. Finally, we derive analytical models of chemical systems, in equilibrium, with carbon, hydrogen, and oxygen. We include acetylene and are able to reproduce several key trends, versus temperature and carbon-to-oxygen ratio, published in the literature. The rich variety of behavior that mixing ratios exhibit as a function of the carbon-to-oxygen ratio is merely the outcome of stoichiometric book-keeping and not the direct consequence of temperature or pressure variations.

Key words: methods: analytical – planets and satellites: atmospheres

## 1. INTRODUCTION

### 1.1. Preamble

Understanding chemistry is indispensable to deciphering the abundances of atomic and molecular species present in an exoplanetary atmosphere. Despite its somewhat late start in the study of exoplanets (e.g., Burrows & Sharp 1999; Zahnle et al. 2009; Moses et al. 2011, 2013a, 2013b; Hu et al. 2012, 2013; Madhusudhan 2012; Line & Yung 2013; Blecic et al. 2015; Venot et al. 2015), atmospheric chemistry has a long and rich history in the Earth and planetary sciences and the study of brown dwarfs (e.g., Prinn & Barshay 1977; Barshay & Lewis 1978; Allen & Yung 1981; Fegley & Lodders 1996; Lodders & Fegley 2002; Ciesla & Charnley 2006, pp. 209–230). Yet, a first-principles, self-consistent formalism that unifies all of the quantities and terminology in a form that is useful for astrophysicists is missing from the literature. For example, there is more than one definition of the "equilibrium constant."

Within the same framework, we demonstrate that the van't Hoff equation (which describes the dimensionless form of the equilibrium constant), the Arrhenius equation (which describes the rate coefficients), and procedures associated with the Gibbs free energy (minimization and scaling) all originate from the first law of thermodynamics. The foundations of atmospheric chemistry are built upon statistical mechanics, since the first law derives from it. To demonstrate the usefulness of our formalism, we use it to compute analytical solutions of chemical systems with pure hydrogen and with carbon, oxygen, and hydrogen (gas phase only); we show that these solutions generalize the work of Burrows & Sharp (1999) and correctly reproduce all of the expected trends.

### 1.2. Survey of Monographs

The novelty of the present study is not in the individual formulae stated, which are mostly previously known, with one exception; we certainly do not claim to be the first to derive these formulae. Rather, it is in the way these results are derived and unified under a common, self-consistent, mathematical formalism that is accessible and palatable to astrophysicists (rather than to chemists). We will now demonstrate this claim of novelty by surveying several textbooks in chemistry.

Specifically, Equations (5), (7), (10), (13), (17), and (22) are commonly stated in textbooks. Our intention is to weave a common mathematical thread between them. We have surveyed the monographs of Slater (1939), Johnston (1966), van Zeggeren & Storey (1970), Moore (1972), Eisenberg & Crothers (1979), Smith & Missen (1982), Steinfeld et al. (1989), Atkins & de Paula (2006), Klotz & Rosenberg (2008), DeVoe (2015), and Glassman et al. (2015) and verified that, while each lists some subset of these formulae, none of them derive and unify all of these formulae in the manner of the present study. None of these monographs derive the generalized form of the Arrhenius equation that we present in Equation (22) or the generalization of the Burrows & Sharp (1999) analytical solutions we present in Section 4.2.

## 2. GENERAL SETUP

In general terms, we consider a chemical reaction involving a pair of reactants ($X_1$ and $X_2$), which produces a pair of products ($Z_1$ and $Z_2$),

$$a_1 X_1 + a_2 X_2 \leftrightarrows b_1 Z_1 + b_2 Z_2, \qquad (1)$$
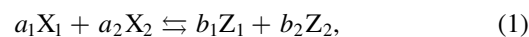
where $a_1$, $a_2$, $b_1$, and $b_2$ are the stoichiometric coefficients. The reactants and products may be atoms or molecules of arbitrary stoichiometry. The forward and reverse reactions are described

by the rate coefficients $k_f$ and $k_r$, respectively. Whenever we discuss something in general terms, it will always be with reference to the preceding chemical reaction.

## 3. EQUILIBRIUM CHEMISTRY: GIBBS FREE ENERGY, EQUILIBRIUM CONSTANT(S), VAN'T HOFF'S EQUATION, AND ARRHENIUS'S EQUATION

### 3.1. The Gibbs Free Energy

If we denote the specific internal energy by $U$, the temperature by $T$, and the specific entropy by $S$, then a reasonable guess for the excess energy associated with a chemical reaction is the Helmholtz free energy (Slater 1939; Moore 1972; Eisenberg & Crothers 1979; Atkins & de Paula 2006; Swendsen 2012; DeVoe 2015),

$$F = U - TS. \tag{2}$$

It turns out that this quantity is not general enough because it does not consider the work done on the system. The general quantity is known as the Gibbs free energy (Slater 1939; van Zeggeren & Storey 1970; Moore 1972; Eisenberg & Crothers 1979; Atkins & de Paula 2006; Swendsen 2012; DeVoe 2015),

$$G = F + PV, \tag{3}$$

where $P$ is the pressure, $V = 1/\rho$ is the specific volume, and $\rho$ is the mass density.

The Gibbs free energy plays a role analogous to the Lagrangian of classical mechanics, which is the difference between the kinetic and potential energies of a system. Instead of solving Newton's equation directly, one may minimize the Lagrangian, a technique known as the principle of least action. Gibbs free energy minimization and chemical kinetics are the chemical analogues to these two techniques.

In a chemically active system, the number of particles of each species is generally not a conserved quantity. If we denote the number of particles associated with the $j$th species of the system by $N_j$, then the first law of thermodynamics needs to be modified (van Zeggeren & Storey 1970; Jacobson 2005; Swendsen 2012; Glassman et al. 2015),

$$TdS = dU + P\,dV - \sum_j C_j dN_j, \tag{4}$$

where $C_j$ is the chemical potential associated with each species. The sum is performed over all of the species in the system.

By using the definition of $G$ and the product rule, one may show that (Smith & Missen 1982; Klotz & Rosenberg 2008; DeVoe 2015)

$$dG = VdP - SdT + \sum_j C_j dN_j. \tag{5}$$

We will now proceed to show that a variety of useful quantities originate from this equation, which is essentially still the first law of thermodynamics.

### 3.2. Gibbs Free Energy Minimization

Generally, the entropy of a system increases according to the second law of thermodynamics; at constant temperature and pressure, its Gibbs free energy generally decreases and seeks a minimum. If one is interested in solving for the chemical equilibrium of a network of reactions, then one needs to

minimize the Gibbs free energy of the system (van Zeggeren & Storey 1970). Within the context of our formalism, we will now elucidate the exact expressions involved in this minimization. Generally, we have $C_j = C_j(T, P)$ and Equation (5) cannot be straightforwardly integrated. However, at a constant temperature and pressure—which is the typical circumstance under which one performs Gibbs free energy minimization—Equation (5) reduces to

$$dG = \sum_j C_j dN_j. \tag{6}$$

The integration can be performed trivially to yield (van Zeggeren & Storey 1970; Eisenberg & Crothers 1979; Smith & Missen 1982; Glassman et al. 2015)

$$G = \sum_j C_j N_j. \tag{7}$$

It is not uncommon to see $G$ being *defined* as the product of the chemical potential and the number of particles of a given species. Strictly speaking, it is not a definition—rather, it is the expression for $G$ in the isothermal and isobaric limit.

Equation (7) is the quantity we need to minimize, but we need additional equations to close the system. In the absence of nuclear reactions, this arises naturally from the notion that the elemental building blocks of molecules cannot be created or destroyed. Thus, the number of carbon, hydrogen, oxygen, etc., atoms in a system is invariant between the reactants and the products, whether they exist in their atomic form or are sequestered in molecules. Mathematically, this set of book-keeping equations takes the form (Smith & Missen 1982),

$$\sum_j A_{ij} N_j = N_i'. \tag{8}$$

The matrix $A_{ij}$ states the number of atoms of species $i$ present in the molecular species $j$. The number of atoms of species $i$ is denoted by $N_i'$.

### 3.3. The Equilibrium Constant: More Than One Definition

A persistent source of confusion exists in the literature regarding the definition of the equilibrium constant. Several references list it as being composed of a series of partial pressures associated with the reactants and products and equates it to an exponential term involving the Gibbs free energy (e.g., Burrows & Sharp 1999; Jacobson 2005; Visscher & Moses 2011; Kopparapu et al. 2012; Line & Yung 2013),

$$K_{\mathrm{eq,literature}} = \frac{P_{Z_1}^{b_1} P_{Z_2}^{b_2}}{P_{X_1}^{a_1} P_{X_2}^{a_2}}. \tag{9}$$

Taken at face value, the partial pressure has physical units. Since the stoichiometric coefficients of the reactants ($a_1 + a_2$) and the products ($b_1 + b_2$) are generally unequal, $K_{\mathrm{eq,literature}}$ must generally have physical units and cannot be equated to an exponential term (which is by definition dimensionless), unless the partial pressures have somehow been normalized. It is not always explicitly explained that this normalization has been performed. Several monographs have previously mentioned this normalization procedure (Fermi 1936; Moore 1972; Smith & Missen 1982; Atkins & de Paula 2006), but we will now provide a derivation that is consistent with the rest of our formalism.
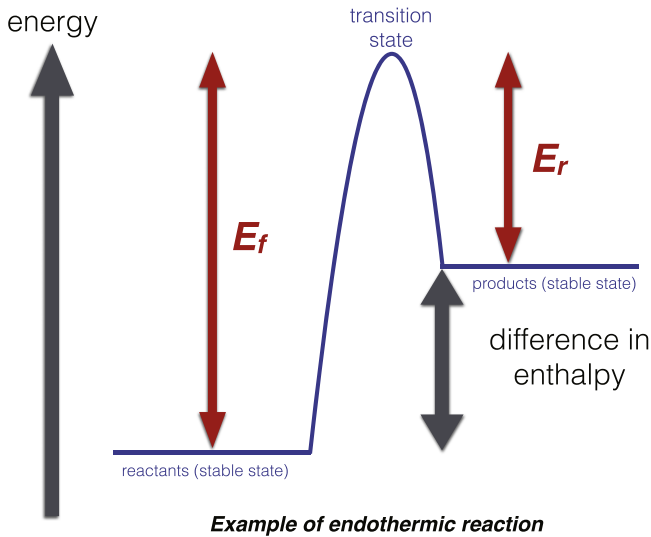
**Figure 1.** Schematic depicting the relationship between the activation energies and the change in enthalpy.

To derive the equilibrium constant, we return to Equation (5) and consider it in the limit of $dT = 0$ and $dN_j = 0$. If we invoke the ideal gas law ($P = \rho \mathcal{R} T$), then we obtain (e.g., Eisenberg & Crothers 1979; DeVoe 2015)

$$G = G_0 + \mathcal{R} T \ln\left(\frac{P}{P_0}\right), \qquad (10)$$

where $P_0$ is a reference pressure, $\mathcal{R}$ is the specific gas constant, and $G_0 \equiv G(P_0, T)$. The preceding equation is useful for scaling the Gibbs free energy to other pressures given its value at a reference pressure—it is exactly the equation one has to use when extracting $G$ for different values of $P$ from thermodynamic databases, which typically tabulate values of $G_0$.

We may use the preceding expression to combine the Gibbs free energy of the reactants and products, weighted by their stoichiometric coefficients (Slater 1939; Eisenberg & Crothers 1979; Smith & Missen 1982; Klotz & Rosenberg 2008; Glassman et al. 2015),

$$\Delta G_0 - \Delta G = -\mathcal{R} T \ln K_{eq}, \qquad (11)$$

where we have defined

$$\Delta G \equiv b_1 G_{Z_1} + b_2 G_{Z_2} - a_1 G_{X_1} - a_2 G_{X_2},$$
$$\Delta G_0 \equiv b_1 G_{Z_1,0} + b_2 G_{Z_2,0} - a_1 G_{X_1,0} - a_2 G_{X_2,0}. \qquad (12)$$

What is interesting is that this first-principles approach naturally yields the definition for the equilibrium constant (e.g., Klotz & Rosenberg 2008),

$$K_{eq} \equiv \frac{(P_{Z_1}/P_0)^{b_1} (P_{Z_2}/P_0)^{b_2}}{(P_{X_1}/P_0)^{a_1} (P_{X_2}/P_0)^{a_2}}. \qquad (13)$$

Notice that this equilibrium constant is naturally dimensionless; its derivation is similar to the ones given in Moore (1972), Smith & Missen (1982), Atkins & de Paula (2006), and DeVoe (2015), who obtained it in terms of chemical potentials and activities. The factors of $P_0$ appear without being inserted in an ad hoc manner.

Physically, the system adjusts itself until it reaches chemical equilibrium, which occurs when $\Delta G = 0$. Let the reference

state, characterized by $P_0$, *not* be in equilibrium, such that $\Delta G_0 \neq 0$. If one is referring to a molecule, then $\Delta G_0$ is the energy needed to construct it from its constituent atoms—it is the Gibbs free energy of formation. If one is referring to mixtures of molecules, then $\Delta G_0$ is the difference in the Gibbs free energies of formation between the reactants and products. Equation (11) naturally yields the relationship between $\Delta G_0$ and $K_{eq}$ (Slater 1939; Johnston 1966; van Zeggeren & Storey 1970; Moore 1972; Eisenberg & Crothers 1979; Smith & Missen 1982; Steinfeld et al. 1989; Atkins & de Paula 2006; Klotz & Rosenberg 2008; DeVoe 2015; Glassman et al. 2015),

$$K_{eq} = \exp\left(-\frac{\Delta G_0}{\mathcal{R} T}\right). \qquad (14)$$

For example, while Equations (40) and (45) of Burrows & Sharp (1999), Equation (7) of Visscher & Moses (2011), Equation (5) of Kopparapu et al. (2012), and Equation (2) of Line & Yung (2013) do not explicitly mention the factors of $P_0$ needed to render $K_{eq}$ dimensionless, it is common practice to omit these reference-pressure terms in standard treatments of chemical equilibria.

We now seek another possible definition of the equilibrium constant. Let the number density be generally represented by $n$; self-explanatory subscripts relate it to the appropriate reactant or product. In chemical equilibrium, we expect the forward and reverse rate coefficients to be related as follows (Moore 1972),

$$k_f n_{X_1}^{a_1} n_{X_2}^{a_2} = k_r n_{Z_1}^{b_1} n_{Z_2}^{b_2}. \qquad (15)$$

A plausible, alternative definition for the equilibrium constant is (Johnston 1966; Moore 1972; Steinfeld et al. 1989)

$$K'_{eq} \equiv \frac{k_f}{k_r}. \qquad (16)$$

Note that since $k_f$ and $k_r$ generally do not possess the same physical units, $K'_{eq}$ is expected to be dimensional.

We may relate our two definitions of the equilibrium constant (Atkins & de Paula 2006),

$$K'_{eq} = K_{eq} (k_B T)^{a_1+a_2-b_1-b_2} P_0^{b_1+b_2-a_1-a_2}, \qquad (17)$$

with $k_B$ being the Boltzmann constant. As has been pointed out by Visscher & Moses (2011), the "pressure correction term" (which is really a temperature correction term) is sometimes missed by other workers. It is less well-known that this correction term has already been elucidated by Fermi (1936). It vanishes when $a_1 + a_2 = b_1 + b_2$ and we have $K_{eq} = K'_{eq}$ (which is commonly, but not always, true).

In other words, $K'_{eq}$ is used to "reverse" the forward rate coefficients, but it is $K_{eq}$ that relates it to the Gibbs free energy. Equation (17) relates them properly. In the literature, what we have defined as $K_{eq}$ and $K'_{eq}$ are often denoted, respectively, by $K_P$ and $K_{eq}$ instead (e.g., Visscher & Moses 2011), although such an approach is not universally adopted (e.g., Line & Yung 2013).

### 3.4. The van't Hoff Equation

If we differentiate Equation (11) with respect to the temperature, we obtain the van't Hoff equation (Smith &

Missen 1982; Jacobson 2005),

$$\frac{\partial(\ln K_{eq})}{\partial T} = \frac{\Delta G_0}{\mathcal{R}T^2}. \tag{18}$$

Note that we are allowed to go from Equation (11) to (18) only because we have *constructed* $\Delta G_0$ to be isothermal.

In most incarnations of the van't Hoff equation, it is the change in enthalpy, rather than the Gibbs free energy, that is stated (Moore 1972; Smith & Missen 1982; Jacobson 2005; DeVoe 2015). If the system is isothermic and adiabatic, then these two statements are equivalent (Slater 1939).

### 3.5. The Arrhenius Equation: Rate Coefficients and Activation Energies

We next derive the expressions for the rate coefficients. At this point, we need to invoke the notion of the activation energy, which is the energy barrier associated with a forward or reverse reaction. One may think of the reactants and products as being two different stable states residing at different energy levels. To transition from one state to the other requires that one surmounts an energy barrier, which is the activation energy (Figure 1). The barrier of the activation energy originates from the need to overcome bond strengths and the requirement that the reactants have specific orientations during a collision. For a single reaction, the difference between the activation energies of the forward and reverse reactions is the change in the enthalpy,

$$\Delta G_0 = E_f - E_r - T\Delta S_0, \tag{19}$$

where $E_f$ and $E_r$ are the activation energies associated with the forward and reverse reactions, respectively, and $\Delta S_0$ is the change in entropy at the reference pressure. The preceding expression allows us to cast the adjectives "exothermic" and "endothermic" in more precise, mathematical terms. If the activation energy of the forward reaction exceeds that of the reverse one, then one needs to inject energy into the system for it to proceed, i.e., $E_f - E_r > 0$. One refers to this as an endothermic reaction. Reactions with $E_f - E_r < 0$ are exothermic.

By combining the expressions for $K_{eq}$ and $K'_{eq}$, we obtain

$$\ln k_f - \ln k_r = -\frac{E_f - E_r}{\mathcal{R}T} + \frac{\Delta S_0}{\mathcal{R}}$$
$$+ (a_1 + a_2 - b_1 - b_2) \ln\left(\frac{k_B T}{P_0}\right). \tag{20}$$

The symmetries inherent in the preceding equation suggest that it may have been constructed from two independent governing equations for the rate coefficients (Upadhyay 2006). Mathematically, "splitting" this equation is a degenerate endeavor and is not rigorous. To persist in this endeavor, we have to appeal to physics. First, we expect that the governing equations for $k_f$ and $k_r$ must enjoy a large degree of symmetry between them. Second, we expect $k_f$ and $k_r$ to be associated with $E_f$ and $E_r$, respectively. Thus, a plausible guess is that the preceding equation originated from the difference between these two equations,

$$\ln k_f = -\frac{E_f}{\mathcal{R}T} + c_f \ln T + \frac{c'_f \Delta S_0}{\mathcal{R}} + c''_f,$$
$$\ln k_r = -\frac{E_r}{\mathcal{R}T} + c_r \ln T + \frac{c'_r \Delta S_0}{\mathcal{R}} + c''_r. \tag{21}$$

The coefficients $c_f$ and $c_r$ cannot be stated uniquely. For example, we can have $c_f = a_1 + a_2$ and $c_r = b_1 + b_2$; we may also have $c_f = -b_1 - b_2$ and $c_r = -a_1 - a_2$. This mathematical freedom implies that $c_f$ and $c_r$ may take on a range of values and may be positive or negative.

Finally, we end up with the Arrhenius equations,

$$k_f = A_f \, T^{c_f} \exp\left(-\frac{E_f}{\mathcal{R}T}\right),$$
$$k_r = A_r \, T^{c_r} \exp\left(-\frac{E_r}{\mathcal{R}T}\right), \tag{22}$$

where we necessarily have

$$c_f - c_r = a_1 + a_2 - b_1 - b_2,$$
$$c'_f - c'_r = 1,$$
$$c''_f - c''_r = (a_1 + a_2 - b_1 - b_2) \ln\left(\frac{k_B}{P_0}\right). \tag{23}$$

The pre-exponential factor $A_f$ absorbs terms associated with $c'_f$, $c''_f$, and $\Delta S_0$; its counterpart, $A_r$, does the same for $c'_r$, $c''_r$, and $\Delta S_0$. Absorbing the entropy into the pre-exponential factors was previously noted by Yung & DeMore (1999), but our derivation is more general as it involves $c'_f$ and $c'_r$. We note that one may also use the van't Hoff equation as a starting point for the derivation.

Traditionally, derivations or statements of the Arrhenius equation include only the exponential term involving the activation energy (Johnston 1966; Moore 1972; Steinfeld et al. 1989; Yung & DeMore 1999; Jacobson 2005; Atkins & de Paula 2006). They omit the power-law terms and tag them on, after the fact (e.g., Jacobson 2005), partially as a means of using them as fitting functions for experimental data. Our derivation demonstrates that there is a sound basis to including these terms. Thus, the Arrhenius equations attain a status that is elevated above that of mere ad hoc fitting functions. Typically, $c_f = 0$ and $c_r = 0$ suffice for low temperatures; "non-Arrhenius" behavior, where $c_f \neq 0$ and $c_r \neq 0$, is important at high temperatures (Glassman et al. 2015).

Kinetic theory states that the rate coefficient is given by $\langle \sigma_{coll} v_{rel} \rangle$, where $\sigma_{coll}$ is the cross section for collisions between the reactants and $v_{rel}$ is the relative velocity between them. If $\sigma_{coll}$ is independent of the relative velocity, then $\langle \sigma_{coll} v_{rel} \rangle \propto T^{1/2}$ (Glassman et al. 2015). If $\sigma_{coll}$ depends on the relative velocity, then more general power-law dependences on $T$ are possible.

The Arrhenius equations do not account for three-body reactions. When the number density of the third body is low, the reaction rate is linearly proportional to it. As it increases, a point is reached where the reaction rate saturates to a limiting value. Fitting functions for implementing this saturation effect have previously been given by, for example, Visscher & Moses (2011).
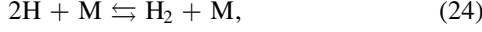
## 4. ANALYTICAL MODELS OF ATMOSPHERIC CHEMISTRY

The formalism and concepts we have established may be highlighted via a set of analytical models.

### 4.1. Pure Hydrogen

For completeness and as the simplest example, we consider a system consisting purely of hydrogen in its atomic and molecular forms,

$$2H + M \leftrightarrows H_2 + M, \qquad (24)$$

where M is a third body of arbitrary stoichiometry.

Using our formalism for the evolution equations (see Appendix A), we may write down

$$\frac{1}{2}\frac{\partial n_H}{\partial t} = -n_H^2 n_M k_f + n_{H_2} n_M k_r,$$

$$\frac{\partial n_{H_2}}{\partial t} = n_H^2 n_M k_f - n_{H_2} n_M k_r. \qquad (25)$$

If we add these equations and perform the integration, we end up with

$$n_H + 2n_{H_2} = n_{total}. \qquad (26)$$

This is already a demonstration that the correction factor (the reciprocal of the stoichiometric coefficient) is essential (Johnston 1966; Steinfeld et al. 1989), if one desires to get the book-keeping between the hydrogen atoms and molecules correct. Here, $n_{total}$ is the total number of particles in the system.

In chemical equilibrium, the (dimensional) equilibrium constant of the reaction is

$$K'_{eq} = \frac{n_{H_2}}{n_H^2}. \qquad (27)$$

If we plug this expression back into Equation (26) and define $K' \equiv K'_{eq} n_{total}$, we may solve for the (normalized) number density of atomic hydrogen (Gail & Sedlmayr 2014),

$$\tilde{n}_H \equiv \frac{n_H}{n_{total}} = \frac{-1 + (1 + 8K')^{1/2}}{4K'}. \qquad (28)$$

The preceding expression is similar, but not identical, to that presented in Barshay & Lewis (1978).

Since $K' \propto K'_{eq} P/T$, one may argue that increasing $K'$ values correspond to decreasing temperatures. Figure 2 shows the curves of $\tilde{n}_H$ and $\tilde{n}_{H_2} \equiv n_{H_2}/n_{total}$. As expected, molecular hydrogen prevails at low temperatures. So far, our toy model does not allow us to define what "low" is, as we have not related $K'$ to $T$ and $P$. Appendix B lists the Gibbs free energies used to make this conversion. In Figure 2, we include a separate set of calculations where $\tilde{n}_H$ and $\tilde{n}_{H_2}$ are shown as functions of temperature and pressure. At $T \lesssim 3000$ K, hydrogen exists predominantly in its molecular form.

### 4.2. Carbon, Hydrogen, and Oxygen: Methane, Water, Carbon Monoxide, and Acetylene

Inspired by the work of Burrows & Sharp (1999), we seek to generalize our toy model of a system with pure hydrogen to one that contains carbon (C), hydrogen (H), and oxygen (O), albeit only in gaseous form, and any carbon-to-oxygen ratio (C/O).
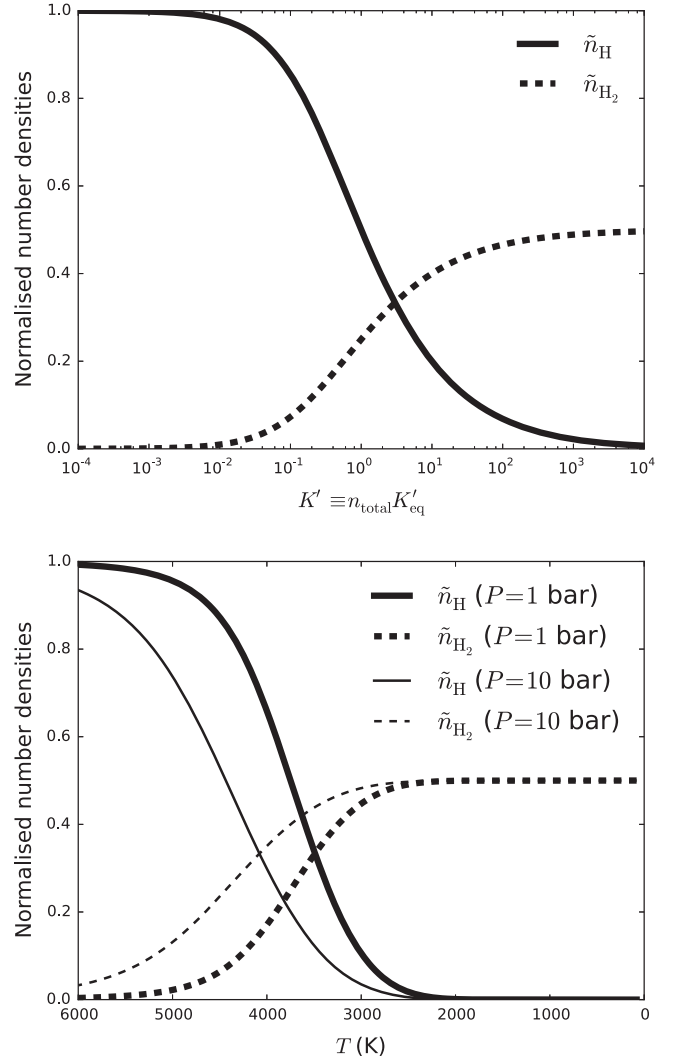


Figure 2. Abundances of atomic and molecular hydrogen, normalized by the total number density, as a function of the normalized equilibrium constant (top panel) and temperature (bottom panel). In the top panel, $K'$ is a proxy for the temperature; larger $K'$ values correspond to lower temperatures.

We wish to compute the relative abundances of the resulting molecules: methane ($CH_4$), water ($H_2O$), carbon monoxide (CO), and acetylene ($C_2H_2$).

We consider the reaction of methane with water to form carbon monoxide and molecular hydrogen (Burrows & Sharp 1999; Lodders & Fegley 2002; Moses et al. 2011),

$$CH_4 + H_2O \leftrightarrows CO + 3H_2. \qquad (29)$$
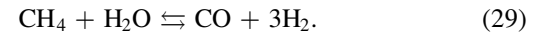
The formulae presented in the appendix of Burrows & Sharp (1999) consider only this reaction and thus are unable to represent carbon-rich atmospheres, where a variety of hydrocarbons are present at high temperatures (Lodders & Fegley 2002; Madhusudhan 2012; Venot et al. 2015). If these hydrocarbons are excluded, then one gets the spurious result that methane is always the dominant carbon carrier at high temperatures and in carbon-rich situations. Our desire for an analytical model does not allow us to include all of the hydrocarbons that are expected to form. Instead, we assume that acetylene is the dominant hydrocarbon and include it via the following reaction (Lodders & Fegley 2002; Moses et al.
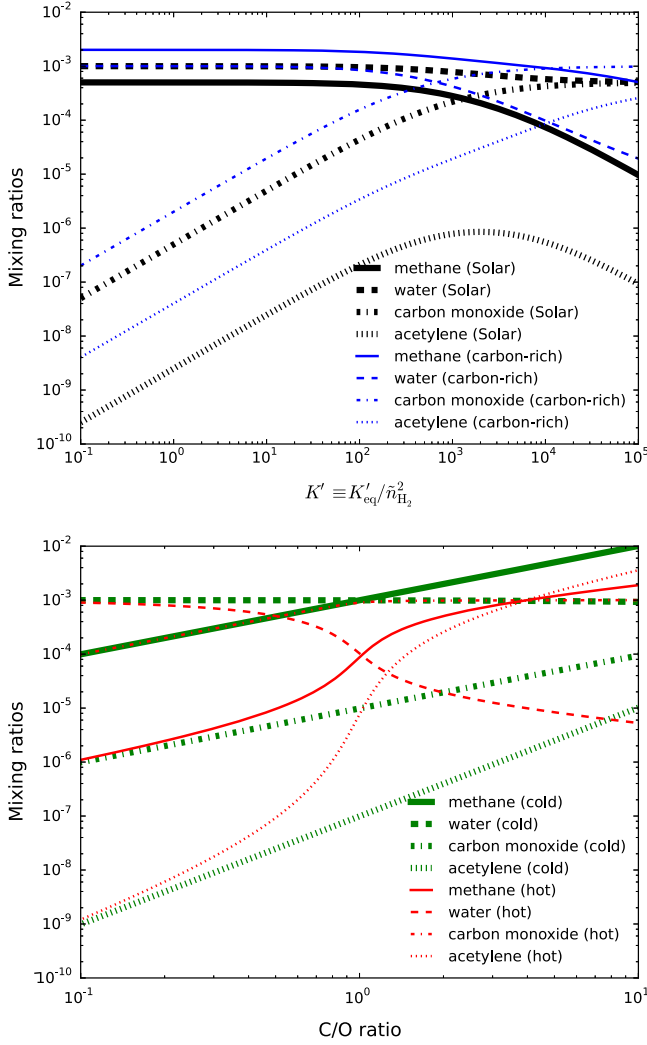
**Figure 3.** Mixing ratios of methane, water, carbon monoxide, and acetylene. Again, $K'$ is a proxy for the temperature, but larger $K'$ values correspond to higher temperatures. We have set $\tilde{n}_O = 5 \times 10^{-4}$ because this is the approximate value of the Sun's photospheric oxygen abundance (Lodders 2003). Top panel: mixing ratios as a function of $K'$. The "Solar" and "carbon-rich" cases correspond to $\tilde{n}_C/\tilde{n}_O = 0.5$ and $\tilde{n}_C/\tilde{n}_O = 2$, respectively. Bottom panel: mixing ratios as a function of the carbon-to-oxygen ratio ($\tilde{n}_C/\tilde{n}_O$). The "cold" and "hot" cases correspond to $K' = 10$ and $K' = 10^5$, respectively. For illustration, we have set $K_2'/K' = 10^{-2}$.

2011),

$$2CH_4 \leftrightarrows C_2H_2 + 3H_2. \tag{30}$$
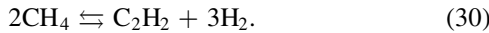
If nitrogen is present, we expect hydrogen cyanide (HCN) to form as well (Madhusudhan 2012), but in the interest of algebraic tractability we will not include it. Furthermore, Venot et al. (2015) have shown using calculations of chemical kinetics that acetylene and hydrogen cyanide are the dominant hydrocarbons in carbon-rich atmospheres.

In reality, both reactions are net reactions that consist of large networks of individual reactions, some of which produce transient species en route to the products. We assume that hydrogen exists mostly in its molecular form, such that the partial pressure of $H_2$ is, to a good approximation, the total pressure ($P$) of the system. Atomic hydrogen is expected to introduce only a small correction to $P = n_{H_2}k_BT$. This

simplification essentially removes the need for an additional equilibrium constant to account for the atomic to molecular transition (and vice versa) of hydrogen, as was described in Section 4.1.

The dimensional equilibrium constant of the reaction described in Equation (29) is

$$K_{eq}' = \frac{n_{CO}n_{H_2}^3}{n_{CH_4}n_{H_2O}} = \frac{\tilde{n}_{CO}n_{H_2}^2}{\tilde{n}_{CH_4}\tilde{n}_{H_2O}}, \tag{31}$$

while that of the reaction in Equation (30) is

$$K_{eq,2}' = \frac{n_{C_2H_2}n_{H_2}^3}{n_{CH_4}^2} = \frac{\tilde{n}_{C_2H_2}n_{H_2}^2}{\tilde{n}_{CH_4}^2}. \tag{32}$$

Analogous to the case study of pure hydrogen, we have defined

$$K' \equiv \frac{K_{eq}'}{n_{H_2}^2}, \quad K_2' \equiv \frac{K_{eq,2}'}{n_{H_2}^2}, \tag{33}$$

but we note that since $K' \propto K_{eq}'T^2/P^2$, we expect $K'$ to increase with temperature, opposite from the trend associated with the pure-hydrogen system. We will again use $K'$ as a proxy for the temperature. An important limitation of our model is the difficulty with relating $K'$ and $K_2'$, because this requires us to explicitly state the functional forms of the change in Gibbs free energies of the two reactions. We first make the simplest assumption: that $K_2'/K'$ is a constant; we will discuss the implications of this assumption later. While we could certainly specify the temperature dependence of $K'$ and $K_2'$— which is what Burrows & Sharp (1999) did—we initially choose not to so as to understand what such a simple model would teach us. We will see shortly that the simplicity yields an important insight, which is that we recover most of the qualitative trends simply by using the equilibrium constants as proxies for the temperature.

The conservation of particles, as described in Equation (8), states that

$$n_{CH_4} + n_{CO} + 2n_{C_2H_2} = n_C,$$
$$n_{H_2O} + n_{CO} = n_O,$$
$$4n_{CH_4} + 2n_{H_2O} + 2n_{C_2H_2} + 2n_{H_2} = n_H. \tag{34}$$

These equations may be manipulated to obtain

$$\tilde{n}_{CH_4} + \tilde{n}_{CO} + 2\tilde{n}_{C_2H_2}$$
$$= \tilde{n}_C(4\tilde{n}_{CH_4} + 2\tilde{n}_{H_2O} + 2\tilde{n}_{C_2H_2} + 2),$$
$$\tilde{n}_{H_2O} + \tilde{n}_{CO} = \tilde{n}_O(4\tilde{n}_{CH_4} + 2\tilde{n}_{H_2O} + 2\tilde{n}_{C_2H_2} + 2). \tag{35}$$

Note that the number densities of the molecules marked by tildes have been normalized by $n_{H_2}$, while those of the atoms have been normalized by $n_H$. The former are the mixing ratios, while the latter are the normalized elemental abundances.

With two particle conservation equations and the expressions for $K'$ and $K_2'$, we have four equations and four unknowns. They can be manipulated to yield a cubic equation for the mixing ratio of methane,

$$\mathcal{C}_0\tilde{n}_{CH_4}^3 + \mathcal{C}_1\tilde{n}_{CH_4}^2 + \mathcal{C}_2\tilde{n}_{CH_4} + \mathcal{C}_3 = 0, \tag{36}$$
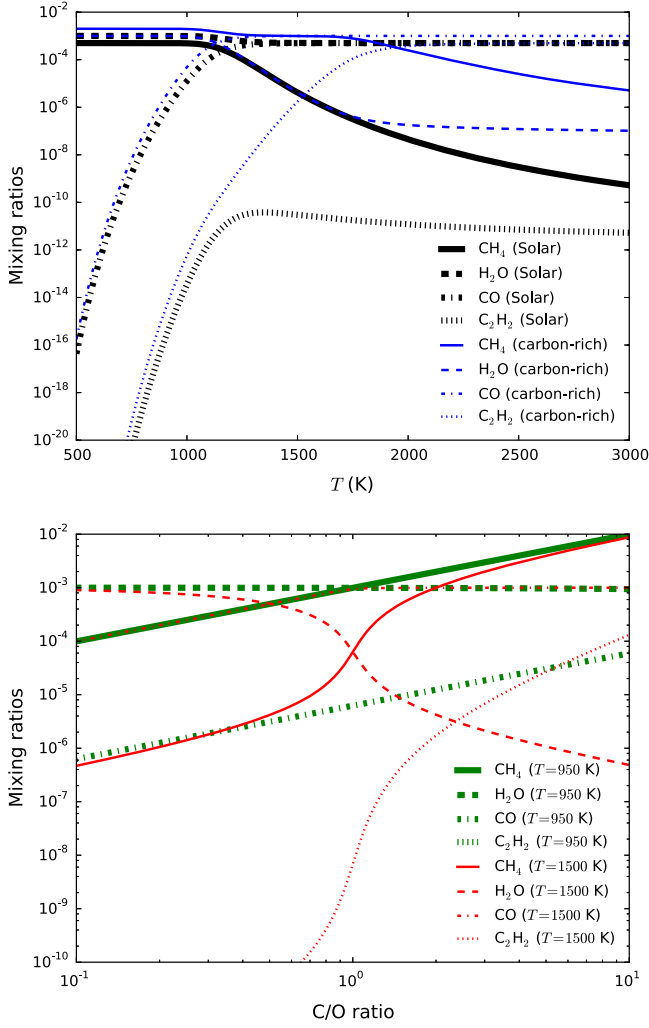
**Figure 4.** Same as Figure 3, but with the equilibrium constants being related to temperature and pressure via the Gibbs free energies taken from the JANAF database. For illustration, we set $P = 1$ bar. Here, the "cold" and "hot" cases correspond to $T = 950$ K and $T = 1500$ K, respectively. For the bottom panel, note that acetylene has a mixing ratio below $10^{-10}$ (the lower limit of the vertical axis) for the cold case.

which has the coefficients

$$
\begin{aligned}
\mathcal{C}_0 &= 2K'K_2'(\tilde{n}_O - \tilde{n}_C + 1), \\
\mathcal{C}_1 &= K'(4\tilde{n}_O - 4\tilde{n}_C + 1) \\
&\quad - K_2'[4\tilde{n}_O\tilde{n}_C + 2(1 - 2\tilde{n}_O)(\tilde{n}_C - 1)], \\
\mathcal{C}_2 &= 2K'(\tilde{n}_O - \tilde{n}_C) - 4\tilde{n}_C - 2\tilde{n}_O + 1, \\
\mathcal{C}_3 &= -2\tilde{n}_C.
\end{aligned} \tag{37}
$$

While analytical solutions do exist for cubic equations, they possess multiple branches—some of which are complex—depending on tedious combinations of the values of $\mathcal{C}_0$, $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ (Press et al. 2007). Since it is difficult to determine a priori which solution branch $\tilde{n}_{CH_4}$ is described by, we elect to solve the cubic equation using standard, canned numerical routines for solving polynomial equations. The other mixing

ratios can be obtained via

$$
\begin{aligned}
\tilde{n}_{H_2O} &= \frac{2\tilde{n}_O(K_2'\tilde{n}_{CH_4}^2 + 2\tilde{n}_{CH_4} + 1)}{1 + K'\tilde{n}_{CH_4} - 2\tilde{n}_O}, \\
\tilde{n}_{CO} &= K'\tilde{n}_{CH_4}\tilde{n}_{H_2O}, \\
\tilde{n}_{C_2H_2} &= K_2'\tilde{n}_{CH_4}^2.
\end{aligned} \tag{38}
$$

For completeness, we note that when acetylene is absent ($K_2' = 0$), the solution can be easily written down,

$$
\tilde{n}_{CH_4} = \frac{-\mathcal{C}_2 + (\mathcal{C}_2^2 - 4\mathcal{C}_1\mathcal{C}_3)^{1/2}}{2\mathcal{C}_1}. \tag{39}
$$

Notice how the coefficient $\mathcal{C}_0$, and thus $K_2'$, controls the extent to which the mixing ratio of methane is described by a quadratic versus cubic equation. Physically, we expect that at low temperatures ($K' \ll 1$), the mixing ratio of acetylene is negligible. In this limit ($K'$, $K_2' \ll 1$), we have $\tilde{n}_{CH_4} \approx 2\tilde{n}_C$ and $\tilde{n}_{H_2O} \approx 2\tilde{n}_O$. These asymptotic solutions explain the relatively simple behavior of the mixing ratios at low temperatures, as seen in Figure 3. It also offers an easy explanation for methane and water switching roles as the dominant molecule when the carbon-to-oxygen ratio is exactly unity, as noted by Kopparapu et al. (2012), since $\tilde{n}_{CH_4}/\tilde{n}_{H_2O} \approx \tilde{n}_C/\tilde{n}_O$.

Figure 3 shows the mixing ratios of methane, water, carbon monoxide, and acetylene as functions of $K'$. Our analytical model produces the following, salient trends.

1. When the atmosphere has a solar abundance of elements, water is always more abundant than methane (Burrows & Sharp 1999; Moses et al. 2013a). At low temperatures, methane dominates carbon monoxide as the carrier of carbon (Prinn & Barshay 1977; Barshay & Lewis 1978); this trend reverses at high temperatures (Burrows & Sharp 1999; Lodders & Fegley 2002).

2. When the atmosphere is carbon-rich, methane is the dominant molecule but has to compete with acetylene in some circumstances (Madhusudhan 2012; Moses et al. 2013a; Venot et al. 2015). Water is the dominant oxygen carrier only at low temperatures, superceded by carbon monoxide at high temperatures (Madhusudhan 2012; Moses et al. 2013a, 2013b).

3. Cold atmospheres are always methane-rich at the expense of carbon monoxide, regardless of the C/O (Madhusudhan 2012). The abundance of water is essentially constant across C/O (Madhusudhan 2012).

4. Hot atmospheres exhibit more complex behavior, in that they are methane-poor and water-rich when C/O < 1 (Madhusudhan 2012; Moses et al. 2013a). For C/O > 1, they become methane-rich and water-poor (Madhusudhan 2012; Moses et al. 2013a) with methane dominating carbon monoxide as the carrier of carbon when C/O becomes sufficiently larger than unity (Madhusudhan 2012). When C/O is large enough, acetylene overtakes methane as the dominant carrier of carbon (Madhusudhan 2012).

These trends are in agreement with the numerical calculations of equilibrium chemistry presented in Madhusudhan (2012) and Moses et al. (2013a), but with one exception. We have assumed $K_2'/K' = 10^{-2}$; higher values would produce the unphysical result that acetylene dominates carbon monoxide,

even at low temperatures (not shown). The cold model in the lower panel of Figure 3 shows an overabundance of acetylene, which is in disagreement with Figure 2 of Madhusudhan (2012). This discrepancy arises from the fact that $K_2'/K'$ is not a constant and must possess a (steep) temperature dependence.

To investigate this discrepancy further, we stop treating $K'$ and $K_2'$ as free parameters and instead relate them to temperature and pressure via the Gibbs free energy tabulated in the JANAF database (http://kinetics.nist.gov/janaf/). Our implementation of this procedure is described in Heng & Lyons (2015). In Figure 4, we recalculate the models in Figure 3. We see that the basic trends previously discussed are preserved, although the curves display quantitative differences as expected. The mixing ratios versus C/O match surprisingly well even at a quantitative level. The previous result regarding acetylene is verified to be an artifact of assuming $K_2'/K'$ to be constant. We further verified that acetylene becomes dominant over methane only for carbon-rich atmospheres with $T \geqslant 2000\,\mathrm{K}$ (not shown), in agreement with Madhusudhan (2012).

Overall, it is surprising how well our model is able to reproduce the main trends of mixing ratios versus the carbon-to-oxygen ratio. It is surprising because this rich variety of behavior originates from the chemical analogue of geometry—it is merely stoichiometric book-keeping (Lodders & Fegley 2002). The dependence of the normalized equilibrium constants on temperature is a distraction if all one seeks is to understand these trends in a qualitative sense.

## 5. SUMMARY

We have presented a unified, novel, self-consistent formalism for understanding the atmospheric chemistry of exoplanets from the viewpoint of an astrophysicist. In doing so, we addressed ambiguities associated with the equilibrium constant and obtained a novel derivation of the Arrhenius equation. We also generalized previous work on analytical models of systems in chemical equilibrium with carbon, hydrogen, and oxygen and showed that they reproduce several key trends published in the literature and computed using more sophisticated numerical calculations. We anticipate that such models are useful for inclusion in retrieval models of exoplanetary atmospheres to maintain their chemical plausibility as a first approach (Benneke 2015).

## APPENDIX A
## CHEMICAL KINETICS

For completeness, we restate the formalism concerning chemical kinetics.

### A.1. Evolution Equations

Unlike Gibbs free energy minimization, chemical kinetics is the treatment of a network of reactions as a system of mass conservation equations. The evolution of the reaction $X_1$ is described by the partial differential equation,

$$\frac{1}{a_1}\left(\frac{\partial n_{X_1}}{\partial t} - K_{zz}\frac{\partial^2 n_{X_1}}{\partial x^2}\right) = \mathcal{P} - \mathcal{L}n_{X_1}^{a_1} - \mathcal{J}_{X_1}. \qquad (40)$$

The importance of the $1/a_1$ factor cannot be over-stated: it allows for the reaction rates of reactants and products with different stoichiometric coefficients to be placed on the same

footing (Johnston 1966; Steinfeld et al. 1989). The production and loss rates are

$$\mathcal{P} = n_{Z_1}^{b_1} n_{Z_2}^{b_2} k_r,$$
$$\mathcal{L} = n_{X_2}^{a_2} k_f. \qquad (41)$$

The reaction rate associated with photochemistry is given by $\mathcal{J}_{X_1}$, which generally depends on $n_{X_1}$.

The diffusion coefficient ($K_{zz}$) is used to *mimic* advection, convection, and turbulence and subsume their collective influence into a single free parameter. Generally, advection, convection, and turbulence hardly resemble diffusion in any rigorous way—one often argues that these processes operate on scales that are so small, compared to the characteristic atmospheric length scale of interest, that it "looks" like diffusion. The use of $K_{zz}$ is rigorous and exact only for molecular diffusion. Notwithstanding, the inclusion of a diffusion coefficient allows us to treat situations with disequilibrium chemistry induced by atmospheric motion or mixing without resorting to a full-blown, three-dimensional calculation.

For the product $Z_1$, the evolution equation is

$$\frac{1}{b_1}\left(\frac{\partial n_{Z_1}}{\partial t} - K_{zz}\frac{\partial^2 n_{Z_2}}{\partial x^2}\right) = \mathcal{P}' - \mathcal{L}'n_{Z_1}^{b_1} - \mathcal{J}_{Z_1}, \qquad (42)$$

where the production and loss rates are

$$\mathcal{P}' = n_{X_1}^{a_1} n_{X_2}^{a_2} k_f,$$
$$\mathcal{L}' = n_{Z_2}^{b_2} k_r. \qquad (43)$$

### A.2. Why Photochemistry Is a Disequilibrium Effect

In the absence of atmospheric mixing ($K_{zz} = 0$), we may add the evolution equations for $X_1$ and $Z_1$ to obtain

$$\frac{1}{a_1}\frac{\partial n_{X_1}}{\partial t} + \frac{1}{b_1}\frac{\partial n_{Z_1}}{\partial t} = -\mathcal{J}_{X_1} - \mathcal{J}_{Z_1}. \qquad (44)$$

If we integrate this expression, we obtain

$$\frac{n_{X_1}}{a_1} + \frac{n_{Z_1}}{b_1} = -\int(\mathcal{J}_{X_1} + \mathcal{J}_{Z_1})dt + \mathcal{C}. \qquad (45)$$

If we do the same for all combinations of reactants and products, then we obtain

$$\frac{n_{X_1}}{a_1} + \frac{n_{X_2}}{a_2} + \frac{n_{Z_1}}{b_1} + \frac{n_{Z_2}}{b_2}$$
$$= -\int(\mathcal{J}_{X_1} + \mathcal{J}_{X_2} + \mathcal{J}_{Z_1} + \mathcal{J}_{Z_2})dt + \mathcal{C}', \qquad (46)$$

where $\mathcal{C}$ and $\mathcal{C}'$ are constants of integration.

This result informs us that photochemistry is an intrinsically disequilibrium effect because it allows the total number of particles in the system to vary with time. In its absence, the total number of particles is an invariant quantity.

### A.3. Producing Chemical Equilibrium in the Steady-state Limit

If we neglect atmospheric mixing and photochemistry, the steady-state limit of the evolution equations yields

$$n_{X_1}^{a_1} n_{X_2}^{a_2} k_f = n_{Z_1}^{b_1} n_{Z_2}^{b_2} k_r. \qquad (47)$$

Since this is identical to the setup in which we used to define our dimensional equilibrium constant ($K'_{eq}$), we conclude that our evolution equations correctly produce chemical equilibrium in the steady-state limit.

## APPENDIX B
## GIBBS FREE ENERGY FOR THE HYDROGEN ATOM

We use the Gibbs free energy associated with the hydrogen atom from the JANAF database. Here, we list it in units of $kJ \, mol^{-1} \, K^{-1}$, from 0 to 6000 K (in increments of 100 K) and at $P_0 = 1$ bar: 216.035, 212.450, 208.004, 203.186, 198.150, 192.957, 187.640, 182.220, 176.713, 171.132, 165.485, 159.782, 154.028, 148.230, 142.394, 136.522, 130.620, 124.689, 118.734, 112.757, 106.760, 100.744, 94.712, 88.664, 82.603, 76.530, 70.444, 64.349, 58.243, 52.129, 46.007, 39.877, 33.741, 27.598, 21.449, 15.295, 9.136, 2.973, $-3.195$, $-9.366$, $-15.541$, $-21.718$, $-27.899$, $-34.082$, $-40.267$, $-46.454$, $-52.643$, $-58.834$, $-65.025$, $-71.218$, $-77.412$, $-83.606$, $-89.801$, $-95.997$, $-102.192$, $-108.389$, $-114.584$, $-120.780$, $-126.976$, $-133.172$, $-139.368$. If we denote each of these numbers by $\tilde{G}_H$, then we have $\Delta \tilde{G}_0 = -2\tilde{G}_H$ for the net reaction in Equation (24). It follows that

$$K' = \frac{P}{P_0} \exp\left(-\frac{\Delta \tilde{G}_0}{\mathcal{R}_{univ} T}\right), \qquad (48)$$

where $\mathcal{R}_{univ} = 8.3144621 \, J \, K^{-1} \, mol^{-1}$ is the universal gas constant. See Heng & Lyons (2015) for more explanation on the unit conversion between $\Delta G_0$ and $\Delta \tilde{G}_0$.

## REFERENCES

Allen, M., & Yung, Y. L. 1981, JGR, 86, 3617
Atkins, P. W., & de Paula, J. 2006, Physical Chemistry (8th ed.; New York: Freeman)
Barshay, S. S., & Lewis, J. S. 1978, Icar, 33, 593
Benneke, B. 2015, arXiv:1504.07655v1
Blecic, J., Harrington, J., & Bowman, M. O. 2015, arXiv:1505.06392v1
Burrows, A., & Sharp, C. M. 1999, ApJ, 512, 843
Ciesla, F. J., & Charnley, S. B. 2006, in Meteorites and the Early Solar System II, ed. D. S. Lauretta & H. Y. McSween, Jr. (Tucson, AZ: Univ. Arizona Press)
DeVoe, H. 2015, Thermodynamics and Chemistry, second edition, sixth version (http://www.chem.umd.edu/thermobook; 1st ed. by Prentice-Hall)
Eisenberg, D., & Crothers, D. 1979, Physical Chemistry with Applications to the Life Sciences (California: Benjamin/Cummings)
Fegley, B., & Lodders, K. 1996, ApJL, 472, L37
Fermi, E. 1936, Thermodynamics (New York: Dover)
Gail, H.-P., & Sedlmayr, E. 2014, Physics and Chemistry of Circumstellar Dust Shells (New York: Cambridge Univ. Press)
Glassman, I., Yetter, R. A., & Glumac, N. G. 2015, Combustion (5th ed.; Massachusetts: Elsevier)
Heng, K., & Lyons, J. R. 2015, arXiv:1507.01944v1
Hu, R., Seager, S., & Bains, W. 2012, ApJ, 761, 166
Hu, R., Seager, S., & Bains, W. 2013, ApJ, 769, 6
Jacobson, M. Z. 2005, Fundamentals of Atmospheric Modeling (New York: Cambridge Univ. Press)
Johnston, H. S. 1966, Gas Phase Reaction Rate Theory (New York: Ronald Press Company)
Klotz, I. M., & Rosenberg, R. M. 2008, Chemical Thermodynamics: Basic Concepts and Methods (7th ed.; New Jersey: Wiley)
Kopparapu, R. K., Kasting, J. F., & Zahnle, K. J. 2012, ApJ, 745, 77
Line, M. R., & Yung, Y. L. 2013, ApJ, 779, 3
Lodders, K. 2003, ApJ, 591, 1220
Lodders, K., & Fegley, B. 2002, Icar, 155, 393
Madhusudhan, N. 2012, ApJ, 758, 36
Moore, W. J. 1972, Physical Chemistry (4th ed.; Englewood Cliffs, NJ: Prentice-Hall)
Moses, J. I., Madhusudhan, N., Visscher, C., & Freedman, R. S. 2013a, ApJ, 763, 25
Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, ApJ, 737, 15
Moses, J. I., Line, M. R., Visscher, C., et al. 2013b, ApJ, 777, 34
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: the Art of Scientific Computing (3rd ed.; New York: Cambridge Univ. Press)
Prinn, R. G., & Barshay, S. S. 1977, Sci, 198, 1031
Slater, J. C. 1939, Introduction to Chemical Physics (New York: McGraw-Hill)
Smith, W. R., & Missen, R. W. 1982, Chemical Reaction Equilibrium Analysis: Theory and Algorithms (New York: Wiley)
Steinfeld, J. I., Francisco, J. S., & Hase, W. L. 1989, Chemical Kinetics and Dynamics (Englewood Cliffs, NJ: Prentice-Hall)
Swendsen, R. H. 2012, An Introduction to Statistical Mechanics and Thermodynamics (New York: Oxford Univ. Press)
Upadhyay, S. K. 2006, Chemical Kinetics and Reaction Dynamics (Dordrecht: Springer)
van Zeggeren, F., & Storey, S. H. 1970, The Computation of Chemical Equilibria (New York: Cambridge Univ. Press)
Venot, O., Hébrard, E., Agúndez, M., Decin, L., & Bounaceur, R. 2015, A&A, 577, A33
Visscher, C., & Moses, J. I. 2011, ApJ, 738, 72
Yung, Y. L., & DeMore, W. B. 1999, Photochemistry of Planetary Atmospheres (New York: Oxford Univ. Press)
Zahnle, K., Marley, M. S., Freedman, R. S., Lodders, K., & Fortney, J. J. 2009, ApJL, 701, L20