4658 J. Phys. Chem. A 2009, 113, 4658–4664

The Elimination of Redundant Constraints in Surprisal Analysis of Unimolecular Dissociation and Other Endothermic Processes[†]

F. Remacle[*,‡,§,||] **and R. D. Levine**[*,‡,⊥]

The Fritz Haber Research Center for Molecular Dynamics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel, Département de Chimie, B6c, Université de Liège, B4000 Liège, Belgium, and Department of Chemistry and Biochemistry and Department of Molecular and Medical Pharmacology and Crump Institute for Molecular Imaging, The University of California, Los Angeles, Los Angeles, California 90095

Received: December 29, 2008; Revised Manuscript Received: February 1, 2009

It is well understood that energy rich polyatomic molecules do not dissociate promptly because the number, P, of their energy states far exceeds the number, N, of the decay channels. In the simplest RRK theory, the fraction N/P is the probability of dissociation. We discuss the distribution of the decay rates of maximal entropy and conclude that it is governed by at most N linearly independent constraints, $N < P$, or, more typically, $N \ll P$. This mathematical requirement already severely restricts the number of constraints. Beyond it however, on physical grounds, one constraint, or at most a few, may be dominant. We discuss why just one constraint, or a few but less than N, can be sufficient to describe the product state distribution.

1. Introduction

A characterization of a distribution by the formalism of maximum entropy[1] is possible in a variety of physical situations. These include ordinary thermal equilibrium[2] but also a great variety of systems which are neither large[3] nor have many degrees of freedom. In all such applications, what one seeks to do is to determine a distribution that is subject to constraints that the distribution should satisfy and whose entropy is maximal. It can be shown that this distribution if it exists is unique. Even though the distribution itself is unique, it can be written down in more than one way. Of course, these alternative ways are equivalent. Some of these ways correspond to imposing constraints that are not linearly independent. If so, the number of constraints can be reduced or equivalently some of the constraints are redundant. In this paper, we discuss the elimination of redundant constraints with special reference to physicochemical situations where the origin of the redundancy can be demonstrated.

We emphasize that the notion of linear dependency of constraints is not quite identical to the mathematical notion of linear dependency. The notion that we need depends also on the state of the system. Toward making this point in a physical context, we begin with a well-studied example: ordinary chemical equilibrium at a given temperature. The minimal set of constraints is clear.[4,5] It is energy and the number of atoms of each chemical element. (A more pedantic statement is the conservation of the number of atoms of each isotope of each chemical element.) Instead of maximizing the entropy, one can, equivalently, minimize the free energy at a given temperature and the number of atoms of each chemical element. As a concrete example, take a dilute gaseous mixture of hydrocarbons. We then need just the conservation of carbon and of hydrogen. In the language of free energy, two chemical potentials suffice to specify the composition of the system. To see this, take the case of normal octane, n-C_8H_{18}. Because the system is in chemical equilibrium, the particular reaction $n\text{-}C_8H_{18} \rightleftharpoons 8C + 18H$ is also at equilibrium. Of course, this equilibrium is very much to the left but it is a reaction at equilibrium because at equilibrium all possible reactions are at equilibrium. Therefore, the chemical potential of n-octane necessarily satisfies $\mu_{n\text{-}C_8H_{18}} = 8\mu_C + 18\mu_H$. The same considerations apply to all other hydrocarbons. Therefore the chemical potentials μ_C and μ_H suffice to determine the chemical potentials of all hydrocarbons that are present at chemical equilibrium. Textbooks often make life simple for the undergraduate student by arguing that at equilibrium the concentration of each and every species, including n-octane, is conserved. There are then numerous constraints, as many as the number of different species that are present in the mixture that is in chemical equilibrium. These constraints are valid but at the same time redundant. Each and every one of the many Lagrange multipliers that correspond to the many species is a linear combination of the far fewer Lagrange multipliers that correspond to the elements. The general expression of this linear dependence is

$$\mu_{\text{species}\,n} = \sum_{\text{elements}\,i} a_{ni}\mu_i \qquad (1)$$

where a_{ni} is the stoichiometric coefficient, namely, how many atoms of element i are present in one molecule of species n. Because each atom is also a chemical species, the number of different species is equal to or, typically, much larger than the number of different elements in the mixture. This is further discussed in the Supporting Information.

It is possible to object that there are several isomers all with the chemical formula C_8H_{18} and therefore that each isomer has exactly the same number of C and of H atoms. This is evidently correct, yet it does not follow that at chemical equilibrium all

---

[†] Part of the “George C. Schatz Festschrift”.

[*] Joint corresponding authors. E-mail: fremacle@ulg.ac.be (F.R.); rafi@fh.huji.ac.il (R.D.L.).

[‡] The Hebrew University of Jerusalem.

[§] Université de Liège.

[||] Directeur de Recherches, FNRS (Belgium).

[⊥] The University of California, Los Angeles.

10.1021/jp811463h CCC: $40.75

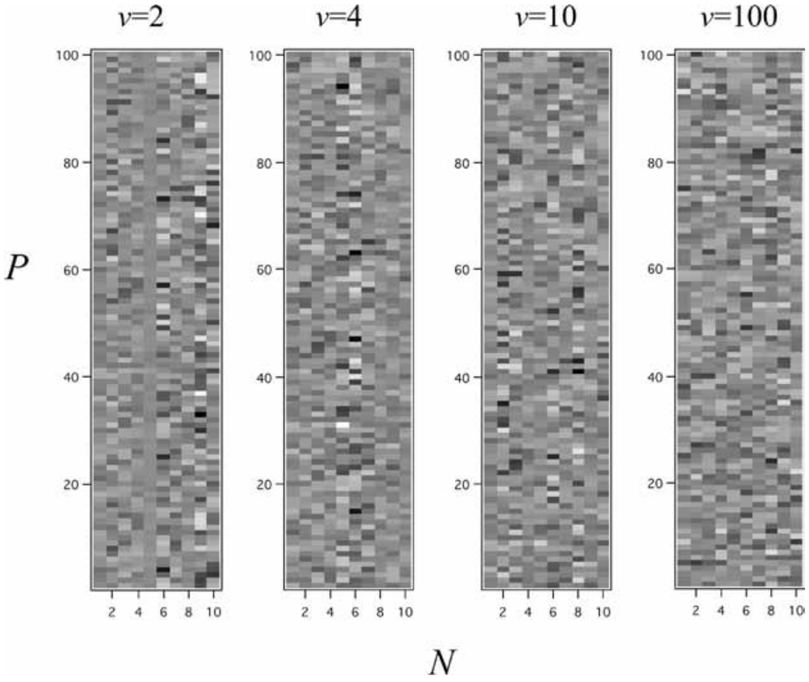

Figure 1. Matrices **X** for different values of the number of degrees of freedom ν ($N = 10$ and $P = 100$). The higher the value of ν, the more localized about the mean is the distribution of the rates; see text for more discussion. The results are shown as a heat map, meaning that the shade of gray indicates the value from negative values, lighter, to positive, darker. It requires a close examination to detect trends in such maps, and this is why the mathematical analysis that we present is useful.

of these isomers are equally likely. That different isomers can have different concentrations is actually one of the successes of the approach that seeks the distribution of maximal entropy subject to constraints. The point is that the concentration of any species has a prefactor that we prefer to call "the prior".[6,7] This is the number of quantum states of that species. As Mayer and Mayer[5] emphasize, at thermal equilibrium, the partition function is the number of effective quantum states at the temperature T. Different isomers have the same values for the constraints, but they differ in the value of their partition function. This is further discussed in the Supporting Information.

Equation 1 shows that, at thermal equilibrium, but not necessarily otherwise, there is a linear dependence among the Lagrange multipliers such that few multipliers (=whose number is the number of elements) suffice to linearly determine the multipliers for all of the species. In the Supporting Information, we show how to transcribe this linear dependence into an equivalent linear dependence for the constraints.

There is a second reason, well beyond the sheer number of constraints, why we care. The conservation of the number of molecules of each species is certainly conserved at equilibrium. However, it will not be conserved for a finite displacement from equilibrium. On the other hand, under nonrelativistic conditions, the conservation of the number of elements holds whether the system is or is not in chemical equilibrium.

In this article, we aim to generalize eq 1 with special reference to the situations exemplified by unimolecular dissociation and other endothermic processes where the number of reaction channels is smaller or typically much smaller than the number of quantum states. It is the case that this situation is at the heart of the Lindemann mechanism of unimolecular reactions as initially quantified by the RRK theory; see, e.g., section 11.5 of ref 8. The same idea is contained in the more recent RRKM theory, section 11.6 idem. The energy rich molecule has many states. In a dissociation state (=a channel), energy in excess of some threshold energy E_0 is confined in a particular degree of freedom known as the reaction coordinate. At some energy E that is above E_0, there are necessarily as many or more states than dissociating states.

The paucity of channels is also the rule in endothermic reactions because, by assumption, some of the energy of the system needs to be spent on overcoming the thermodynamic barrier to reaction.

There are other physicochemical systems with this characteristic meaning that, when you label events by two indices, one index has a range lower than the other index. We will specifically assume that there is a rectangular, $P \times N$ matrix, **X**, such that its columns are labeled by the index of lower range (N) and its rows by the index of wider range (P). The matrix elements of **X** carry both indices. Our task is to characterize the matrix **X**. In the example of unimolecular dissociation, **X** is the matrix of the amplitudes for channel specific dissociation for different states of the energy rich molecule. Moore and coworkers[9] have discussed this matrix for the dissociation of CH_2O, and we[10−12] and others[13] have also examined the statistical properties of such matrices starting with the fluctuations in spectral intensities.[14] Other attempts to examine an eigenchannel type decomposition of scattering amplitudes include ref 15. Simple model computations are reviewed in ref 16; see also refs 17−19. Propensity rules relating states and channels are discussed in refs 20 and 21 and references therein with important earlier work by Troe and Quack.[22]

There is a rich literature dealing with random matrix theory for characterizing the positions and widths of energy levels; see, for example, ref 23. We emphasize that we do not assume or imply that the matrix **X** belongs to a matrix ensemble. The entries in **X** as considered here are the, typically computed, amplitudes of dissociation of specific quantum states through particular channels. Our purpose is to delineate the limitations imposed by the physics and the mathematics of the problem.

In quantum chemistry, there is a different matrix **X** that is the matrix relating the states of physical interest to the basis

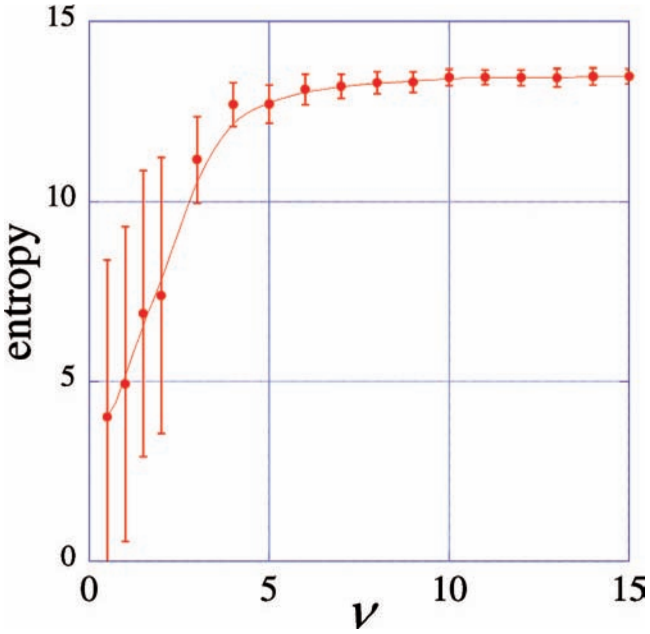

Figure 2. The entropy, in natural units, computed, points, as an average for 100 matrices **X** generated by a sampling as discussed in the text, shown vs the parameter ν that determines the distribution of the rates. See also Figure 1. As discussed in the text, the sampled matrices **X** for a given value of ν can vary, particularly so for lower values of the parameter ν. For each value of ν, the variance of the distribution of the values of the entropy is also shown and it is seen how the variance strongly decreases as ν increases. As the fluctuations in the possible value of a width diminish, all of the widths converge toward more similar values and the entropy of the distribution of the widths tends to a higher asymptotically constant value. The curve is a smooth line through the computed points. Note that the rates are made dimensionless by dividing by their sum and so the trace of rates is ensured to equal N. The results in Figure 3 are for the same case shown here and similarly exhibit the trend toward uniformity as ν increases.

states used to describe the system. The latter must be equal or larger than the physical number of states. The physics is different, but the rectangular nature of the matrix is the same. Aspects of quantum chemistry are indeed being discussed recently from an information theoretic point of view.[24−27] For a maximum entropy analysis of such matrices **X**, see ref 28.

Having reduced the sufficient number of constraints to the rank N of the matrix $\mathbf{X}^T\mathbf{X}$, we ask what is the necessary or minimal number of constraints or, in mathematical terms, can the de facto rank be even lower than N. This is a more subtle question because it is not possible to reproduce $\mathbf{X}^T\mathbf{X}$ by fewer eigenvectors than its rank. However, say that we just require a leading approximation. This will be possible if one eigenvalue of $\mathbf{X}^T\mathbf{X}$ is significantly larger than all others, or it may be more than one but fewer than N. These are sometimes called the principal components. The number of such components is the effective rank, as discussed in section 5. Figure 3 provides a concrete example.

Representing the product state distribution is closely related to the effective rank. Therefore, the two topics are discussed in the same section below. In principle, it takes N eigenvectors of $\mathbf{X}^T\mathbf{X}$—and these are equivalent to N constraints—to exactly describe the product state distribution. Say, however, we just want a leading approximation. Then, the number of constraints equals the effective rank. Often, as we show by example, it can be that only one constraint will be sufficient. We emphasize here that, as will be shown by the example below, the reason why one or a few constraints suffice is tied to the physics of the problem, as discussed for example in refs 20, 29, and 30.

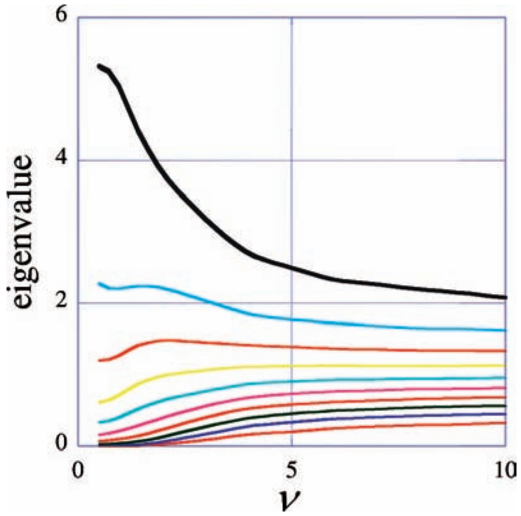

Figure 3. The 10 eigenvalues as a function of the number ν of the degrees of freedom for a system with $N = 10$ and $P = 100$. Averaged over 100 sampling of the matrix **X**. The dominant eigenvalue is shown as a thicker black curve. As ν increases, all of the eigenvalues lump together as expected, since the diagonal elements of **X** become more similar; see text. The results in Figure 2 are for the same case shown here. The behavior is as expected on theoretical grounds, but we have verified that it remains the same for other systems by increasing the number P of states or by sampling over 1000 matrices and (see text) by increasing the dispersion of the diagonal elements of $\mathbf{X}^T\mathbf{X}$.

2. Maximal Entropy

X is a matrix whose dimensions are the number, P, of states, the row labels, times the number, N, of channels, the column labels. When **X** represents amplitudes, the entries are often necessarily complex numbers. We then treat the real and complex parts separately, thereby doubling the number of columns. The mean over the entries in each column is zero. Otherwise, we enforce **X** to be mean centered. The matrix $\mathbf{X}\mathbf{X}^T$, dimensions P times P, where $\mathbf{X}^T$ is the transpose matrix, is mathematically the matrix of the covariances of the states.

$$\begin{aligned}(\mathbf{X}\mathbf{X}^T)_{pq} &= \sum_{n=1}^{N}(\mathbf{X})_{pn}(\mathbf{X})_{qn} \\ &\equiv \mathbf{x}_p\cdot\mathbf{x}_q^T\end{aligned} \tag{2}$$

The variances are the diagonal elements. In the second line of eq 2, we wrote the covariance between states p and q as the scalar product of two vectors where each such vector is a row of the matrix **X**. Each of the N columns of **X** is a reading for a different probe of the state. Therefore, we can regard **X** not as a matrix but as a sample of N readings of a random vector $\mathbf{X}_n$ of P components. Each column of **X** is a particular sample of values for $\mathbf{X}_n$ and is a column vector, $\mathbf{X}_n \equiv (X_{1n}, X_{2n},..., X_{Pn})^T$.

In the example of an energy rich molecule at a given energy, the vector $\mathbf{X}_n$ is the coupling of channel n to the P states. The row vector of N components $\mathbf{x}_p$ as defined by eq 2 is the channel specific amplitudes for dissociation of state p, $p = 1, 2,..., P$. The matrix $\mathbf{X}\mathbf{X}^T$ is then the rate matrix often denoted by $\mathbf{\Gamma}$.[31] It is off diagonal because states are indirectly coupled among themselves due to their coupling to the continuum, cf. eq 2. When we diagonalize the rate matrix, these off diagonal elements tend to keep the eigenvalues apart. This is further discussed in section 5.

A multivariate normal (or Gaussian) distribution is a distribution of maximal entropy subject to given values for the mean and the covariances (including the variances), altogether $P + P^2$ constraints, of the random vector $\mathbf{X}_n$. The probability density for each particular set of components of $\mathbf{X}_n$ is normal

$$P(\mathbf{X}_n) \equiv f(X_{1n}, X_{2n}, \ldots, X_{Pn}) \propto \exp\left(-\frac{1}{2}\mathbf{X}_n^T \mathbf{\Lambda} \mathbf{X}_n\right) \tag{3}$$

The matrix $\mathbf{\Lambda}$, where $(\mathbf{\Lambda})_{pq} = \lambda_{pq}$, is the matrix of P^2 Lagrange multipliers. On physical grounds, the components of the random vector $\mathbf{X}_n$ have zero mean and so the distribution is symmetric about zero.

Our technical objective is to reduce the number of Lagrange multipliers (=the number of constraints) down not only to N^2 but all the way to N, where by physical reasoning $N < P$ and more typically $N = P$. The very fact that this can be done means that the matrix $\mathbf{\Lambda}$ is singular and as such has no inverse. Therefore, the distribution (eq 3) cannot be written in the canonical form for a normal distribution of a random vector $\mathbf{X}_n$

$$P(\mathbf{X}_n) = (1/(2\pi)^{P/2}|\mathbf{\Sigma}|^{1/2}) \exp\left[-\frac{1}{2}\mathbf{X}_n^T \mathbf{\Sigma}^{-1} \mathbf{X}_n\right] \tag{4}$$

The reason is that when the columns of $\mathbf{X}$ are mean centered the covariance matrix is P by P and is given as $\mathbf{X}\mathbf{X}^T$, cf eq 2. However, the rank of the P by N, $P > N$, matrix $\mathbf{X}$ is N or lower and so the matrix $\mathbf{X}\mathbf{X}^T$ must be singular. The inverse matrix that appears in eq 4 cannot be the inverse of the covariance, $\mathbf{X}\mathbf{X}^T$.

At this point, we ask for the probability of joint N different samples, meaning the probability of $\mathbf{X}_1$ times the probability of $\mathbf{X}_2$ times the,..., times the probability of $\mathbf{X}_N$

$$P(\mathbf{X}) \equiv P(\mathbf{X}_1)\, P(\mathbf{X}_2) \ldots P(\mathbf{X}_N) \propto \prod_{n=1}^{N} \exp\left(-\frac{1}{2}\mathbf{X}_n^T \mathbf{\Lambda} \mathbf{X}_n\right) \tag{5}$$

Since each of the exponents in eq 5 is a number, the product of exponentials can be written in the more conventional form

$$P(\mathbf{X}) = \exp\left(-\frac{1}{2}\sum_{n=1}^{N} \mathbf{X}_n^T \mathbf{\Lambda} \mathbf{X}_n\right) \tag{6}$$

Using eq 2 above, this last result can also be written in a suggestive alternative form

$$\begin{aligned} P(\mathbf{X}) &\propto \exp\left(-\frac{1}{2}\sum_{pq} \lambda_{pq} (\mathbf{X}\mathbf{X}^T)_{pq}\right) \\ &= \exp\left(-\frac{1}{2}\sum_{pq} \lambda_{pq} \mathbf{x}_p \cdot \mathbf{x}_q^T\right) \\ &= \exp\left(-\frac{1}{2}\sum_{pq} \lambda_{pq} \sum_{k}^{N} x_{pk} x_{qk}\right) \end{aligned} \tag{7}$$

where the $\mathbf{x}_p$'s are the rows of the matrix $\mathbf{X}$. Recall that for a dissociating molecule each $\mathbf{x}_p$ is the set of channel specific amplitudes for state p, $p = 1, 2, \ldots, P$.

In eq 7, the sum over λ_{pq} is over P^2 terms and here is where we need to reduce the number of Lagrange multipliers by seeking a transformation to linearly independent constraints. The required transformation begins by expressing the number that is the exponent in eq 6 or 7 as a trace

$$\begin{aligned} P(\mathbf{X}) &\propto \exp\left(-\frac{1}{2}\sum_{n=1}^{N} \mathbf{X}_n^T \mathbf{\Lambda} \mathbf{X}_n\right) \\ &= \exp\left(-\frac{1}{2} Tr(\mathbf{X}^T \mathbf{\Lambda} \mathbf{X})\right) \end{aligned} \tag{8}$$

Then, introducing an orthogonal matrix $\mathbf{U}$ that diagonalizes the matrix Λ of Lagrange multipliers

$$\mathbf{U}^T \Lambda \mathbf{U} = \boldsymbol{\lambda} \tag{9}$$

so that

$$\begin{aligned} P(\mathbf{X}) &\propto \exp\left(-\frac{1}{2} Tr(\mathbf{X}_T \mathbf{\Lambda} \mathbf{X})\right) \\ &= \exp\left(-\frac{1}{2} Tr(\mathbf{X}^T \mathbf{U}\mathbf{U}^T \mathbf{\Lambda} \mathbf{U}\mathbf{U}^T \mathbf{X})\right) \\ &= \exp\left(-\frac{1}{2} Tr(\mathbf{Z}^T \boldsymbol{\lambda} \mathbf{Z})\right) \end{aligned} \tag{10}$$

The new diagonal matrix $\boldsymbol{\lambda}$ has at most N nonzero entries along its diagonal. This is because the P by P covariance matrix $\mathbf{X}\mathbf{X}^T$, eq 2, is singular, since $\mathbf{X}$ is a P by N rectangular matrix where $N < P$. Thus, $\mathbf{X}\mathbf{X}^T$ is necessarily of rank equal or lesser than N. The same is true for the rank of the N by N matrix $\mathbf{X}^T\mathbf{X}$. This matrix has the same nonzero eigenvalues as the bigger P by P matrix $\mathbf{X}\mathbf{X}^T$. The latter matrix has also $N-P$ eigenvalues that are all zero.[17] In linear algebra, the equality of nonzero eigenvalues of $\mathbf{X}^T\mathbf{X}$ and of $\mathbf{X}\mathbf{X}^T$ is sometimes known as the Froebenius−Schur lemma.[32] On physical grounds, the columns of the matrix $\mathbf{X}$ have zero mean. This ensures that the matrix $\mathbf{X}^T\mathbf{X}$ can have a rank that equals N. Had we mean centered the rows of $\mathbf{X}$, the rank would have been at most $N - 1$.

When the rank of the N by N matrix $\mathbf{X}^T\mathbf{X}$ is N, all of the N eigenvalues are positive. We can therefore define a nonsingular diagonal matrix $\boldsymbol{\lambda}_+$ by retaining only the positive eigenvalues of $\boldsymbol{\lambda}$. This matrix is invertible, and the N by N covariance matrix is $\boldsymbol{\lambda}^{-1}$

$$\mathbf{\Sigma} = \boldsymbol{\lambda}_+^{-1} = \mathbf{U}^T \mathbf{X}^T \mathbf{X} \mathbf{U} = \mathbf{Z}^T \mathbf{Z} \tag{11}$$

When the rank of $\mathbf{X}^T\mathbf{X}$ is lower than N, we can still define $\boldsymbol{\lambda}_+$ by retaining only the positive eigenvalues of $\boldsymbol{\lambda}$. The size of this square matrix is the rank of $\mathbf{X}^T\mathbf{X}$.

With the covariance given by eq 11, the normalized distribution, cf. eq 4, is of the form

$$\begin{aligned} P(\mathbf{X}) &= (1/(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2})^N \exp\left[-\frac{1}{2}\sum_{n=1}^{N} \mathbf{Z}_n^T \mathbf{\Sigma}^{-1} \mathbf{Z}_n\right] \\ &= (1/(2\pi)^{N/2}|\mathbf{\Sigma}|^{1/2})^N \exp\left[-\frac{1}{2} Tr(\mathbf{Z}^T \Sigma^{-1} \mathbf{Z})\right] \end{aligned} \tag{12}$$

Equation 12 defines a normal multivariate distribution, and as such, its entropy is well-known to be given by

$$H(\mathbf{z}) = \frac{1}{2}\ln|\Sigma| + \frac{1}{2}N\ln(2\pi e) = \frac{1}{2}\ln((2\pi e)^N \prod_{n=1}^{N} \lambda_{+n}) \tag{13}$$

where $|\boldsymbol{\Sigma}|$ is the determinant and the N λ_{+n}'s are the positive eigenvalues of the P by P covariance matrix whose rank is N, $N < P$.

With eqs 12 and 13, we have achieved the desired reduction in the number of Lagrange multipliers and have expressed the entropy in terms of the remaining N. As a technical result, we identified the values of the Lagrange multipliers with the eigenvalues of the N by N matrix $\mathbf{X}^T\mathbf{X}$. Note that this is not the same matrix as the, far larger in dimensions, rate matrix $\boldsymbol{\Gamma} = \mathbf{X}\mathbf{X}^T$ which mathematically is the covariance matrix, cf. eq 2. On physical grounds, it is clear that for unimolecular dissociation the number P of quasibound states exceeds by orders of magnitude the number of internal states N at the transition state. The reduction to no more than N Lagrange multipliers is therefore significant.

3. Entropy and Entropy Deficiency

We computed the entropy on the assumption that, in the absence of any constraints except normalization, the maximal entropy is for a uniform distribution. This means that the expression for the entropy is

$$H(\mathbf{z}) = -\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f(\mathbf{z})] \tag{14}$$

and this holds whether the distribution $f(\mathbf{z})$ is normal or not.

Next, we consider the possibility of additional constraints. Say that we denote the normal distribution by $f_N(\mathbf{z})$. Let there be a multivariate distribution $f(\mathbf{z})$ that has the same mean and covariance as $f_N(\mathbf{z})$. From the form of the exponent of the normal distribution, it follows that

$$\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f_N(\mathbf{z})] = \int \mathrm{d}\mathbf{z}\, f_N(\mathbf{z}) \ln[f_N(\mathbf{z})] \tag{15}$$

This allows us to show explicitly that the multivariate distribution $f(\mathbf{z})$ either has a lower entropy or it is the normal distribution itself. First, compute the entropy of the multivariate distribution $f(\mathbf{z})$

$$\begin{aligned} H(\mathbf{z}) &= -\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f(\mathbf{z})] \\ &= -\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f(\mathbf{z})/f_N(\mathbf{z})] - \int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f_N(\mathbf{z})] \\ &= -\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f(\mathbf{z})\, f_N(\mathbf{z})] + \\ &\qquad \{-\int \mathrm{d}\mathbf{z}\, f_N(\mathbf{z}) \ln[f_N(\mathbf{z})]\} \end{aligned} \tag{16}$$

The second third line follows from eq 15. If therefore there are additional constraints that we have not yet identified, the entropy deficiency,[33] the difference between the entropy of the normal distribution and of the unknown multivariate distribution (of the same mean and variance),

$$\begin{aligned} DH(\mathbf{z}) &\equiv \{-\int \mathrm{d}\mathbf{z}\, f_N(\mathbf{z}) \ln[f_N(\mathbf{z})]\} - \{-\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f(\mathbf{z})]\} \\ &= \int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) \ln[f(\mathbf{z})/f_N(\mathbf{z})] \end{aligned} \tag{17}$$

needs to be finite. It can be easily shown that the entropy deficiency has to be non-negative. The intuitive argument is that if we are adding one or more constraints the value of the entropy must go down from what it is for the normal multivariate distribution because, among all distributions of the same mean and variance, it is the distribution of maximal entropy.

Equation 14 has one feature that, at first sight, is disturbing. The probability density $f(\mathbf{z})$ has dimensions. Specifically, the very fact that the density is normalized

$$\int \mathrm{d}\mathbf{z}\, f(\mathbf{z}) = 1 \tag{18}$$

means that $f(\mathbf{z})$ has the dimension inverse to that of z^N. Therefore, $\exp(H(\mathbf{z}))$ has the dimension of z^N. To retain the interpretation of entropy as information, it is therefore necessary to reexamine the expression for the entropy of a continuous distribution.[34] Or one can seek an alternative interpretation for the expression (eq 14) of the entropy of a continuous distribution. This is possible by adapting the definition of the entropy as a measure of the number of typical sequences; e.g.,[35] see Shannon[36] for the original discussion. We recall that in the discrete case when an experiment is repeated many times some sequences of outcomes are possible in principle and yet are extremely rare. As a very simple example, consider tossing a biased coin that has a higher propensity to fall head ups. The sequence TTT,..., TTHT,... cannot be ruled out, but it is very atypical. The number of typical sequences of N outcomes is roughly $\exp(NH)$, where H is the entropy of the distribution of the two faces. One shows that for a continuous distribution $f(z)$ the dimension bearing number $\exp(NH)$ is the volume, in N dimensional space, within which most of the probability of observing N outcomes is contained. Therefore, $\exp(H)$ is the "length", of dimension z, where most of the probability is concentrated. For a Gaussian distribution, we expect this length to be proportional to the width σ and it is easy to verify that if we scale the length by σ then all Gaussian distributions have the same entropy. This interpretation of the entropy of a continuous distribution is important in that it plays a key role in proving the coding theorems of information theory.[35]

4. A Numerical Example: Rate Constants for Unimolecular Dissociation

As an illustration of our considerations, we return to the example of unimolecular dissociation, with P states and N channels, $N < P$. We generate a P by N matrix $\mathbf{X}$ of the transition amplitudes between states and channels. For simplicity, we take the amplitudes to be real. We draw the amplitudes in a given channel (=given column, see discussion of eq 2) from a Gaussian distribution of zero mean and a variance (=the partial rate for this channel) that is sampled from a chi-squared distribution with ν degrees of freedom; see ref 12 or refs 10 and 11 for additional details. Figure 1 shows typical matrices $\mathbf{X}$ using a "heat map" format, meaning that larger entries are shown in a darker shade. The amplitudes can be negative and by small, light shading, we mean negative entries so that about zero is gray. The matrices are displayed for a range of values of the parameter ν that governs the number of degrees of freedom.

The fluctuation of rates is narrower the higher the number of degrees of freedom ν. Explicitly, the result is eq 2.28 of ref 11, where y is the rate and the average rate is $\langle y \rangle$

$$\langle (y - \langle y \rangle)^2 \rangle = 2\langle y \rangle^2/\nu \tag{19}$$

Books usually cite the width of a chi-square distribution as increasing with ν. To derive eq 19, one must distinguish between the dimensionless variable chi-squared $\chi^2 = \nu y/\langle y \rangle$ and the rate; see also eq 2.25 of ref 11. Another possible pitfall is that $\langle y \rangle$ is the average rate and not the average amplitude. In our case where the amplitudes are mean centered, $\langle y \rangle$ is the variance of the distribution of the amplitudes.

For each matrix, one can compute the entropy of the matrix as discussed above and the results are given as a function of ν in Figure 2. The entropy increases with increasing value of ν because the fluctuation of the rates becomes narrower so that they are all about equal and so their distribution is more uniform. At a high value, the distribution is so narrow that the entropy tends to an asymptotically constant value. For the same value of ν, one can generate many matrices **X** and these can have somewhat different values for their entropies.[37] In our problem, this variability is because, on physical grounds, the matrices **X** have a number of columns smaller than the number of rows. Thus, sampling different matrices **X** can lead to different results. The large number of rows of **X**, 100 in our example, much higher in reality, suffices to sample the Gaussian distribution for the transition amplitudes to a given channel. However, the smaller number of columns, which is physically to be expected, is not sufficient to sample a chi-squared distribution for a small ν. The error bars show the variance of the distribution of the values of the entropy at a given value of ν. This variance is large for low values of ν because the distribution of rates is broad.

Complementary to the discussion of the entropy are the results for the eigenvalues of the N by N matrix $\mathbf{X}^T\mathbf{X}$. This matrix has the same nonzero eigenvalues as the bigger P by P rate matrix $\mathbf{XX}^T$. (The latter matrix has also $N-P$ eigenvalues that are all zero.) We find that as ν increases the eigenvalues tend to a common value. The asymptotic value of this value is the trace of $\mathbf{XX}^T$ divided by N, the number of channels.

5. Effective Rank and the Distribution of Product States

To provide a physical application for the role of an effective rank that is lower than the nominal rank N, we consider the time evolution and specifically the distribution of product states. The time evolution is governed by the entire Hamiltonian and not just by the rate matrix. This means that it depends also on the spacing of the energy levels; see, for example, ref 19. We therefore reiterate the approximation that the density of states of the energy rich polyatomic molecule is high enough that the states form a quasicontinuum. This is often realistic for molecules with more than a few atoms. It is then possible to restrict attention to a set of P states that are effectively degenerate. The Hamiltonian matrix of this isoenergetic set of states is $\mathbf{H} = E\mathbf{I} - i\mathbf{\Gamma}$, where $\mathbf{I}$ is the identity matrix. Diagonalizing this Hamiltonian is equivalent to diagonalizing the P by P rate matrix $\mathbf{\Gamma} = \mathbf{XX}^T$, a matrix whose rank N is much smaller than its dimension P. N is then identified as the number of rate constants that are not zero. All the other $P-N$ states are trapped.[17,38] In this section, we go beyond this important result to resolve the dynamics among the N outgoing channels. We argue that the effective rank is the number of channels that dominate the process. Of course, these N_{eff} channels are not necessarily the physical channels that we use to initially define the elements of the **X** matrix. Rather, these N_{eff} channels are each a coherent linear combination of zero-order channels. We explicitly identify these combinations below. These combinations determine the product state distributions, distributions that are defined over the zero-order channels. We therefore have a tool to understand the characterization of final states sometimes called surprisal analysis.[6,7]

Consider a vector **A**, of N components. For simplicity, take it to be real, but if it is complex, we could, just as for the matrix **X**, separate it into real and imaginary parts. We think of the components of **A** as the amplitudes to exit in the different N channels. The N eigenvectors of the matrix $\mathbf{X}^T\mathbf{X}$

$$\mathbf{X}^T\mathbf{X}\mathbf{z}_i = \lambda_i \mathbf{z}_i, \quad i = 1, 2, ..., N \tag{20}$$

are normalized and orthogonal, $\mathbf{z}_i^T \cdot \mathbf{z}_j = \delta_{i,j}$. There is an effective rank when the N eigenvectors separate into a group of N_{eff} vectors with eigenvalues λ_i, $i = 1,..., N_{\text{eff}}$, that are significantly larger in magnitude. Often there is only one such runaway eigenvalue in which case $N_{\text{eff}} = 1$. An example for the same system used to generate Figures 1 and 2 is shown in Figure 3. As is made clear by the evidence in Figure 3, whether there is an effective rank or not depends on the specific physics. As the system becomes more statistical as judged by the entropy becoming higher, compare Figures 2 and 3, the dominant eigenvalue merges with the others to get a uniform distribution of final quantum states (as discussed in section 5 below).

Why is it quite often that if there is an effective rank it is unity so that one eigenvector dominates? This must be a consequence of the theorem, often attributed to Rayleigh, that there is one eigenvalue between every two diagonal elements of the matrix. There can be special circumstances, which must be driven by the physics, that one or more diagonal elements of $\mathbf{X}^T\mathbf{X}$ are significantly bigger than the others. Then, we can have a higher effective rank. Otherwise, only one eigenvalue can run away. It does not have to happen, and as seen in Figure 3, it is a special case but it can happen when $\nu < 1$. We could make Figure 3 look more dramatic by introducing more spread in the diagonal elements of $\mathbf{X}^T\mathbf{X}$, but we intentionally chose to show a conservative example.

The eigenvectors of $\mathbf{X}^T\mathbf{X}$ form a complete orthonormal basis. However, when there is an effective lower rank that is lower than N, we might try the expansion

$$\mathbf{A} = \sum_{i=1}^{N_{\text{eff}}} \alpha_i \mathbf{z}_i, \quad N_{\text{eff}} = 1, 2, ... \tag{21}$$

When N_{eff} is small, the problem of describing the product state distribution is reduced to the determination of N_{eff} numbers α_i, where $i \le N_{\text{eff}} < N - 1$. Often N_{eff} can be just one in which case one constraint suffices to well characterize the product state distribution. Because the product state distribution is normalized, if there is only one dominant eigenvector, we do not need to know the corresponding coefficient α in eq 21. Since the eigenvectors are normalized, the result $\mathbf{A} = \mathbf{z}_{\text{dominant}}$ determines the distribution $|\mathbf{A}_n|^2 = |(\mathbf{z}_{\text{dominant}})_n|^2$, $n = 1, 2, ..., N$.

6. Concluding Remarks

We discussed the distribution of the coupling of a few outgoing channels to many more states of the system. The physics of the few vs the many was understood already in the early days of the kinetics of chemical reactions, practically as soon as one learned to count quantum states. The notion of fluctuations in the rates came much later. Currently, there are other systems of topical interest that can be characterized by similar considerations. In physical chemistry, we have, for example, ZEKE spectroscopy[39] where many, many Rydberg states are coupled to far fewer ionization channels.[40] The exceptional stability of the high Rydberg states is governed by similar considerations.[41] Other cases of slow electron emission are reviewed in ref 42. A recent example of charge transfer is the rate of coupling of charged quantum dots to the channels outgoing to the electrodes. The fluctuations in this case are reviewed in ref 43. However, one can also go outside of physical chemistry proper. A much more complex system with similar abstract characteristics is a cell. The membrane has a few receptors. When stimulated, these can induce signaling pathways where many proteins are phosphorylated.[44,45] Or, different but usually few mutations can induce an extensive signaling response; see, e.g., ref 46. The common feature is that few channels are coupled to many states.

For all of these systems, we show two aspects. Mathematically, the number of constraints necessary to exactly reproduce the output is smaller than what can be expected and the number necessary to provide a leading approximation can be even smaller, possibly just one.

Acknowledgment. The work of F.R. and R.D.L. is partially supported by the EC FET Nano-ICT project MOLOC.

Supporting Information Available: Description of the constraints at thermal and chemical equilibrium. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Jaynes, E. T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, U.K., 2004.

(2) Tribus, M. *Thermodynamics and Thermostatics: An Introduction to Energy, Information and States of Matter, with Engineering Applications*; D. Van Nostrand Company: New York, 1961.

(3) Levine, R. D. *Physica E* **2001**, *9* (3), 591–599.

(4) Fowler, R. *Statistical Mechanics*, 2nd ed.; Cambridge University Press: London, 1961.

(5) Mayer, J. E.; Mayer, M. G. *Statistical Mechanics*; Wiley: New York, 1961.

(6) Levine, R. D. *Molecular Reaction Dynamics*; The University Press: Cambridge, U.K., 2005.

(7) Levine, R. D.; Bernstein, R. B. *Acc. Chem. Res.* **1974**, *7*, 393–400.

(8) Steinfeld, J. I.; Francisco, J.; Hase, W. *Chemical Kinetics and Dynamics*; Prentice-Hall: Upper Saddle River, NJ, 1989.

(9) Miller, W. H.; Hernandez, R.; Moore, C. B.; Polik, W. F. *J. Chem. Phys.* **1990**, *93* (8), 5657–5666.

(10) Levine, R. D. *Ber. Bunsen-Ges. Phys. Chem. Chem. Phys.* **1988**, *92* (3), 222–227.

(11) Levine, R. D. *Adv. Chem. Phys.* **1988**, *70*, 53–95.

(12) Remacle, F.; Levine, R. D. *J. Phys. Chem.* **1996**, *100*, 7962–7971.

(13) Hase, W. L.; Cho, S. W.; Lu, D. H.; Swamy, K. N. *Chem. Phys.* **1989**, *139* (1), 1–13.

(14) Alhassid, Y.; Levine, R. D. *Phys. Rev. Lett.* **1986**, *57*, 2879–2882.

(15) Aldegunde, J.; Aoiz, F. J.; Saez-Rabanos, V.; Kendrickd, B. K.; de Miranda, M. P. *Phys. Chem. Chem. Phys.* **2007**, *9* (43), 5794–5808.

(16) Remacle, F.; Levine, R. D. *ACH—Models Chem.* **1997**, *134* (5), 613–617.

(17) Remacle, F.; Munster, M.; Pavlov-Verevkin, V. B.; Desouter-Lecomte, M. *Phys. Lett. A* **1990**, *145*, 265–268.

(18) Remacle, F.; Levine, R. D. *Phys. Lett. A* **1996**, *211* (5), 284–288.

(19) Remacle, F.; Levine, R. D. *Mol. Phys.* **1996**, *87*, 899.

(20) Lorquet, J. *J. Phys. Chem. A* **2007**, *111*, 8050.

(21) Lorquet, J. *Int. J. Quantum Chem* **2008**, *108*, 1629.

(22) Quack, M.; Troe, J. *Ber. Bunsen-Ges. Phys. Chem.* **1974**, *78*, 240.

(23) Ullah, N. *Matrix Ensembles in the Many-Nucleon Problem*; Clarendon Press: Oxford, U.K., 1987.

(24) Morales, J.; Martinez, T. J. *J. Phys. Chem. A* **2001**, *105* (12), 2842–2850.

(25) Nalewajski, R. F. *J. Math. Chem.* **2008**, *43* (2), 780–830.

(26) Nalewajski, R. F. *J. Math. Chem.* **2008**, *44* (2), 419–445.

(27) Nalewajski, R. F.; Parr, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (16), 8879–8882.

(28) Levine, R. D. *J. Stat. Phys.* **1988**, *52* (5−6), 1203–1220.

(29) Schatz, G. C.; Ross, J. *J. Chem. Phys.* **1977**, *66* (3), 1037–1053.

(30) Schatz, G. C.; Ross, J. *J. Chem. Phys.* **1977**, *66* (7), 2943–2958.

(31) Levine, R. D. *Quantum Mechanics of Molecular Rate Processes*; Dover: Mineola, NY, 1999.

(32) Bodewig, A. *Matrix Calculus*; North Holland: 1959.

(33) Bernstein, R. B.; Levine, R. D. *J. Chem. Phys.* **1972**, *57* (1), 434–445.

(34) Dinur, U.; Levine, R. D. *Chem. Phys.* **1975**, *9*, 17–27.

(35) Ash, R. A. *Information Theory*; Dover: Mineola, NY, 1990.

(36) Shannon, C. *Bell Syst. Tech. J.* **1948**, *27*, 379, 623.

(37) Ahmed, N. A.; Gokhale, D. V. *IEEE Trans. Inf. Theory* **1989**, *35* (3), 688–692.

(38) Remacle, F.; Levine, R. D. *Phys. Lett. A* **1996**, *211*, 284.

(39) Held, A.; Schlag, E. W. *Acc. Chem. Res.* **1998**, *31* (8), 467–473.

(40) Remacle, F.; Levine, R. D. *J. Chem. Phys.* **1996**, *104* (4), 1399–1414.

(41) Remacle, F.; Levine, R. D. *ACH—Models Chem.* **1997**, *134* (5), 619–634.

(42) Campbell, E. E. B.; Levine, R. D. *Annu. Rev. Phys. Chem.* **2000**, *51*, 65–98.

(43) Alhassid, Y. *Rev. Mod. Phys.* **2000**, *72* (4), 895–968.

(44) Amit, I.; Wides, R.; Yarden, Y. *Mol. Syst. Biol.* **2007**, 3.

(45) Janes, K. A.; Albeck, J. G.; Gaudet, S.; Sorger, P. K.; Lauffenburger, D. A.; Yaffe, M. B. *Science* **2005**, *310* (5754), 1646–1653.

(46) Skaggs, B. J.; Gorre, M. E.; Ryvkin, A.; Burgess, M. R.; Xie, Y.; Han, Y.; Komisopoulou, E.; Brown, L. M.; Loo, J. A.; Landaw, E. M.; Sawyers, C. L.; Graeber, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (51), 19466–19471.

JP811463H