Instabilities in Complex Mixtures with a Large Number of Components

Richard P. Sear[1,*] and José A. Cuesta[2,†]

[1]*Department of Physics, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom*
[2]*Grupo Interdisciplinar de Sistemas Complejos (GISC), Departamento de Matemáticas, Universidad Carlos III de Madrid,
Avenida de la Universidad 30, 28911 Leganés, Spain*

Inside living cells are complex mixtures of thousands of components. It is hopeless to try to characterize all the individual interactions in these mixtures. Thus, we develop a statistical approach to approximating them, and examine the conditions under which the mixtures phase separate. The approach approximates the matrix of second-virial coefficients of the mixture by a random matrix, and determines the stability of the mixture from the spectrum of such random matrices.

Mixtures are not always simple, well-characterized, and made up of two or three components. The mixtures of biomacromolecules inside living organisms contain thousands of different macromolecules, and the oil extracted from wells by the petroleum industry also contains many different hydrocarbons and related compounds. This gives us two main problems: (i) the number of components is so large that the data we have are inadequate to characterize all the components, (ii) even if we were able to precisely characterize each component, then comprehending, and calculating with, this mass of data would be difficult. The sheer complexity of the mixture overwhelms us. An analogous problem afflicted nuclear physics 50 years ago. Large nuclei, such as ^{235}U, are complex many-body systems with complex spectra. Nuclear physicists were faced with energy spectra with so many energy levels that comprehending and predicting them directly was impossible. Starting with Wigner [1–4], they resorted to statistical methods, and replaced the complex and unknown Hamiltonian matrix of a nucleus with a Hamiltonian matrix whose elements were random variables. Using this rather drastic approximation they were able to reproduce the statistical properties of energy spectra, such as the probability distribution function of the level spacings. Subsequently, random matrices have been applied in many areas of physics [5,6].

Here, we will apply a statistical theory to complex mixtures. We start by noting that, at the simplest level, the interactions between two components i and j affect the free energy, and hence potentially drive a phase transition, via their second-virial coefficient B_{ij}. These second-virial coefficients form a symmetric matrix, of course, and the eigenvalues of this matrix describe the change in the excess free energy when the density is perturbed. For an $N \gg 1$ component mixture we need a huge number, of order N^2, virial coefficients to specify the mixture, but typically we are not interested in knowing each individual value but we are interested in how the system reacts to density perturbations, because if the free

energy change when the density is perturbed becomes negative, the system is unstable and will undergo a phase transition. The parallels between our situation and that faced by nuclear physicists 50 years ago are obvious, and so we adopt their solution: we replace the matrix of, unknown, second-virial coefficients of some mixture by a random matrix. For large N, the eigenvalues of this matrix depend on two variables only: the mean and standard deviation of its elements, which we can vary to fit a specific mixture or just vary to explore the generic features of the phase behavior of such a mixture. Once we are using random matrices, the fact that we have $N \gg 1$ components is a help, not a hindrance.

The Helmholtz free energy per unit volume, f, of an N component mixture truncated after the second-virial coefficient terms is

$$f = \sum_{i=1}^{N} \rho_i(\ln\rho_i - 1) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \rho_i B_{ij}\rho_j, \qquad (1)$$

where ρ_i is the number density of component i. The N densities form a row matrix $\rho = (\rho_1 \rho_2 \ldots \rho_N)$. We use units such that the thermal energy $k_B T = 1$. Here, we will not calculate complete phase diagrams with the densities and compositions of coexisting phases. We will calculate the limits of stability, spinodals, where the system becomes unstable with respect to density perturbations. Thus, we will be able to determine only qualitative features of the phase behavior, such as whether or not a phase transition occurs and whether the transition is demixing of the components or their condensation. Stability of the mixture requires that f be convex. Convexity requires that the second order term, $\delta^2 f$, in an expansion of f in powers of $\delta\rho$, be positive for any small perturbation $\delta\rho$. For our N component mixtures $\delta\rho$ is a row matrix of length N.

Now, $\delta^2 f = \frac{1}{2}\delta\rho(\mathsf{P} + \mathsf{B})\delta\rho^T$, where B is the matrix of second-virial coefficients and P is a diagonal matrix with the ith diagonal element equal to $1/\rho_i$: it contains the ideal or perfect gas contributions to the free energy

change. Of course, any $\delta\rho$ can be expressed as the sum of eigenvectors of $\mathsf{P} + \mathsf{B}$, and so the requirement $\delta^2 f > 0$ implies that all eigenvalues of $\mathsf{P} + \mathsf{B}$ must be positive. The mixture becomes locally unstable when the lowest eigenvalue becomes zero. Stability is determined only by the lowest eigenvalue. However, if each component has more or less the same mobility, then the decay of small density modulations can be described as the decay of the set of eigenvectors of $\mathsf{P} + \mathsf{B}$, with each eigenvector component of the density modulation decaying at a rate proportional to its eigenvalue.

The eigenvalues of $\mathsf{P} + \mathsf{B}$ form the row matrix γ. For simplicity we will assume that in the mixture all components are present in equal amounts $\rho_i = \rho_T/N$, $i = 1, \ldots, N$, where ρ_T is the total density. Then $\mathsf{P} = (N/\rho_T)\mathsf{I}$, where I is the N by N unit or identity matrix, and the eigenvectors of $\mathsf{P} + \mathsf{B}$ are equal to those of B. The eigenvalues of $\mathsf{P} + \mathsf{B}$ are those of B, which form a row matrix λ, shifted by N/ρ_T, i.e.,

$$\gamma = \lambda + (N/\rho_T)u, \qquad (2)$$

where $u = (1, \ldots, 1)$.

Stability is determined by the sign of the lowest eigenvalue of $\mathsf{P} + \mathsf{B}$, $\gamma_{\min}$. It requires $\gamma_{\min} > 0$, and the spinodal is reached when $\gamma_{\min} = 0$. From Eq. (2) it follows that the total density ρ_T at the spinodal is $\rho_{\rm sp} = -N/\lambda_{\min}$, with $\lambda_{\min}$ the lowest eigenvalue of B.

Now, B is a random matrix, and for simplicity we choose its elements B_{ij}, $i \leq j$, as independent random variables with mean b and standard deviation σ. Since Wigner's pioneering work [1,2], the problem of characterizing the spectrum of such a random matrix has received a great deal of attention (see [6] for a review). In our particular case, there are two theorems that fully describe the spectrum of B.

The first theorem is due to Arnold [7], and states that the density of rescaled eigenvalues, $x \equiv \lambda/2\sigma N^{1/2}$, of B converges in probability, as $N \to \infty$, to

$$W(x) = \begin{cases} \frac{2}{\pi}\sqrt{1 - x^2}, & \text{if } |x| \leq 1, \\ 0, & \text{if } |x| > 1. \end{cases} \qquad (3)$$

This is known in the literature as Wigner's semicircle law. When $b \neq 0$, we also require a second theorem, due to Füredi and Komlós [8]. Under the further assumption that $|B_{ij}| \leq K$ for all $i, j = 1, \ldots, N$, this theorem states that if $b > 0$ $(b < 0)$, then the highest (lowest) eigenvalue is asymptotically Gaussian distributed with a mean $Nb + (b' - b) + \sigma^2/b + O(N^{-1/2})$ and variance $2\sigma^2$, and the remaining $N - 1$ eigenvalues follow Wigner's semicircle law. For this result we allow the diagonal elements, B_{ii}, to have a mean b', which differs from the mean of the off-diagonal elements b. So, for $b \neq 0$ there is an additional eigenvalue which may lie outside the semicircular distribution given by Eq. (3), whereas if $b = 0$, Eq. (3) gives the complete distribution. This lone eigenvalue contributes a

negligible amount to the density of eigenvalues W but can determine the limit of stability.

All this can be seen in numerically calculated spectra, for $N = 25$, in Fig. 1. For $|b| \gtrsim \sigma/N^{1/2}$ the probability density function, $p(\lambda)$, for the eigenvalues clearly exhibits a lone, Gaussian-distributed, eigenvalue, and for all b there is a clear semicircle. To generate Fig. 1 we used Gaussian-distributed values of the B_{ij} but the distribution of the B_{ij} is irrelevant for large N. The theorems of Arnold and of Füredi and Komlós [7,8] do not depend on the exact distribution of the virial coefficients.

The two theorems above permit us to describe, for large N, the lowest eigenvalue of B and hence the spinodal instability of our mixture. Let us define rescaled variables $\beta \equiv N^{1/2}b/\sigma$ and $\xi_{\rm sp} \equiv \rho_{\rm sp}\sigma/N^{1/2}$. For $\beta \lesssim -1$, the (rescaled) lowest eigenvalue, $x_{\min}$, which determines the spinodal is due to a lone eigenvalue; see the bottom curve in Fig. 1. This eigenvalue has a mean value $(\beta + \beta^{-1})/2$ and standard deviation $1/\sqrt{2N}$. As the latter goes to zero when $N \to \infty$, $x_{\min}$ is a self-averaging quantity. So, for large N we can take [9]

$$\langle \xi_{\rm sp} \rangle = \left\langle \frac{-1}{2x_{\min}} \right\rangle \qquad (4)$$

for the spinodal. Replacing $x_{\min}$ by its mean value,

$$\langle \xi_{\rm sp} \rangle = -1/(\beta + \beta^{-1}). \qquad (5)$$

The nature of the instability is described by the corresponding eigenvector. Füredi and Komlós show that this eigenvector is almost parallel to u [8], so it is a condensation instability, as the densities of all the components increase (or decrease) together according to the eigenvector: the instability looks like the incipient formation of one phase enriched in *all* the components coexisting with a phase depleted in all the components. Conversely, if
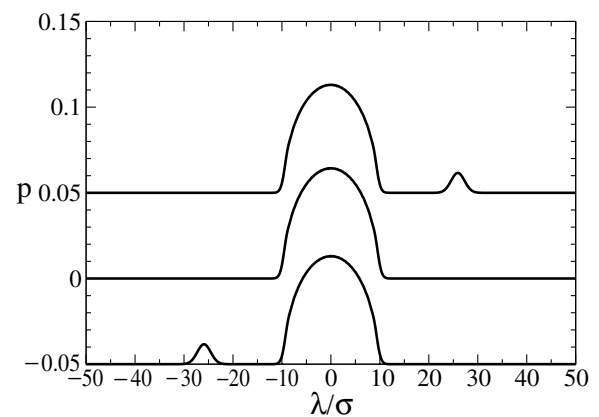

FIG. 1.   The probability density function for the eigenvalues of B, $p(\lambda)$. The three curves are, from bottom to top, for $b/\sigma = -1$, 0, and 1. The curve for $b/\sigma = -1$ is shifted down by 0.05, and that for $b/\sigma = 1$ is shifted up by 0.05, as otherwise their central semicircular parts are almost on top of each other.

$\beta \gtrsim -1$, then $x_{\min}$, and hence the spinodal, is determined by the semicircle. The lowest eigenvalue will be near the lower end of the semicircle. We then estimate the distribution of the lowest eigenvalue as follows.

Given the semicircle law, the expected number of (rescaled) eigenvalues between -1 and X will be given by $N \int_{-1}^{X} W(x)dx$, so $x_{\min}$ will be in the interval $[-1, X(N)]$, where $X(N)$ is defined by

$$ N \int_{-1}^{X(N)} W(x)dx = 1. \qquad (6) $$

Using Eq. (3), Eq. (6) becomes $\pi + 2\arcsin X + 2X\sqrt{1-X^2} = 2\pi/N$, whose solution is

$$ X(N) = -1 + m + \frac{1}{10}m^2 + \frac{11}{350}m^3 + O(m^4), \qquad (7) $$

with $m \equiv (1/2)(3\pi/2N)^{2/3}$. Now, for a given (large) N, $x_{\min}$ will be roughly distributed according to

$$ p_N(x_{\min}) = \begin{cases} NW(x_{\min}), & \text{if } -1 \leq x_{\min} \leq X(N), \\ 0, & \text{otherwise.} \end{cases} \qquad (8) $$

(Notice that $\int_{-\infty}^{\infty} p_N(x)dx = 1$, so it is a well defined probability density.) As $X(N) \to -1$ when $N \to \infty$, $x_{\min}$ is self-averaging in this case as well. We can then make use of Eq. (4) to determine the spinodal. Thus from Eqs. (7) and (8),

$$ \langle \xi_{sp} \rangle = \frac{1}{2} + \frac{3}{10}m + \frac{33}{140}m^2 + O(m^3). \qquad (9) $$

As for the eigenvector, we know that when there is a lone eigenvalue its corresponding eigenvector is almost parallel to u, so the eigenvector of any eigenvalue of the semicircle must be almost *orthogonal* to u (the matrix is symmetric). By continuity, this holds for $x_{\min}$ even when there is no lone eigenvalue. Then, roughly half of the components of the eigenvector are of one sign, while the rest are of the opposite sign. The instability is with respect to a density modulation in which about half the components are separating from the other half: the instability looks like *demixing* into two phases, each one enriched in some components and depleted in others.

The predictions of theory in both regimes, Eqs. (5) and (9), are compared with the results of numerical calculations in Fig. 2, for $N = 25$, 100, and 400. The crossover between demixing and condensation occurs for $\beta \simeq -1$ in all cases. This crossover can also be seen by looking at the angle, θ, the eigenvector of $x_{\min}$ makes with u. In Fig. 3, we have plotted the mean and standard deviation of $|\cos(\theta)|$, as a function of β, for $N = 25$. At around $\beta = -1$ the cosine drops and the standard deviation peaks, indicating that the instability eigenvector is switching over from being nearly parallel to u to being nearly perpendicular.

For $\beta < -1$, as $N \to \infty$, the instability, which is with respect to condensation, approaches that of a single com-

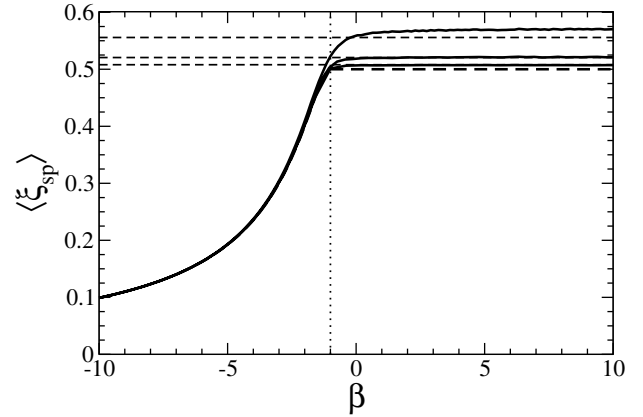FIG. 2. The mean reduced total density at the spinodal, $\langle \xi_{sp} \rangle \equiv \langle \rho_{sp} \rangle \sigma/N^{1/2}$, as a function of the reduced mean second virial coefficient, $\beta \equiv N^{1/2}b/\sigma$. The solid curves are the numerically calculated mean density of the spinodal for $N = 25$, 100, and 400 (from top to bottom). The short-dashed horizontal lines are the predictions of Eq. (9) for these values of N (again from top to bottom). The $N \to \infty$ limit of Eq. (5) for $\beta < -1$, and of Eq. (9) ($\langle \xi_{sp} \rangle = 1/2$) for $\beta > -1$, is plotted as a long-dashed curve. For $\beta < -1$, it lies on top of the numerically calculated values and so is not visible. The crossover between condensation and demixing at $\beta = -1$ is marked by a vertical dotted line.

ponent system with second-virial coefficient b. Such a system becomes unstable at a spinodal density $\rho_{sp}b = -1$, and Eq. (6) yields this result as $N \to \infty$. For $\beta > -1$, $\langle \xi_{sp} \rangle \to 1/2$, as $N \to \infty$, and so the total number density at the spinodal diverges as $N^{1/2}$. The mixture becomes stable with respect to demixing at all densities as $N \to \infty$—also consistent with the mixture behaving as a single component system. For finite but large N, demixing requires either high densities or a broad distribution of the second-virial coefficients, i.e., large σ. In
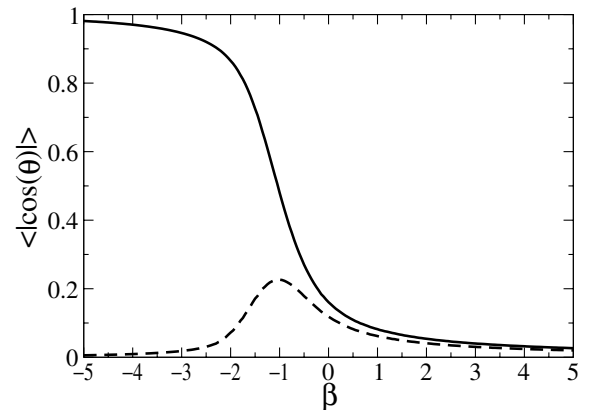

FIG. 3. The mean (solid curve) of the absolute value of the cosine of the angle θ between the eigenvector with the smallest eigenvalue and the vector with all its elements equal to 1, $\langle |\cos(\theta)| \rangle$, and its standard deviation (dashed curve), as a function of the reduced mean second virial coefficient, β.

Fig. 2, we see that even for the smallest N of 25, the prediction of our theory for demixing ($\beta > -1$) is quite accurate. We also see that for condensation, our theoretical prediction is on top of the numerical curves for all three values of N, and that the density of the condensation instability is insensitive to N.

Our analysis relies on several simplifying assumptions. We have employed a second-virial coefficient approximation, have studied a mixture of components with equal densities, and have taken the virial coefficients to be independent random variables: $\langle B_{ij}B_{ik}\rangle = \langle B_{ij}\rangle^2$, $j \neq k$. All three assumptions can be relaxed.

Many properties of random matrices, such as the distribution of their eigenvalues (properly scaled), are *self-averaging* [6]. A property R of an N by N matrix A_N is said to be self-averaging if $R(A_N)$ converges in probability, as $N \to \infty$, to $\langle R(A_N)\rangle$, where $\langle \rangle$ denotes an average over some ensemble of matrices A_N. This simply means that, for large N, almost all matrices A_N have the same value of R. In particular, the lowest (scaled) eigenvalue x_{min} of our random matrices B is a self-averaging property of these matrices and hence so is the value of the density at the spinodal. The distribution function of the spacing of the eigenvalues of random matrices is also self-averaging, and experimental data on nuclei show that the spectra of nuclei far from their ground state are, approximately, self-averaging [4,5]. The good agreement between theory and experiment in the study of nuclei *relies* on both the model and the experimental system having a self-averaging property. As emphasized by Dyson and others [2–5], the theoretical prediction is obtained via statistical methods, but this is then compared to the results for a *single* experimental system, e.g., a ^{235}U nucleus.

So, are the thermodynamic properties of complex mixtures, such as those found inside living cells, self-averaging? We know of no experiments explicitly addressing this question, but self-averaging does have measurable consequences. If the mixtures of proteins in cytoplasms are self-averaging, then we would expect the thermodynamics of the cytoplasms of different bacteria cultured in the same environment to be rather similar, even though the amino-acid sequences of many of their proteins will be rather different. For example, if the mixtures of proteins inside, say, *E. coli* and *Salmonella* are extracted and subjected to a potentially destabilizing perturbation such as increased salt levels, then, if bacterial cytoplasms are self-averaging, we would expect the two extracts to become unstable at very similar amounts of added salt.

Quite generally, to develop theories for very complex systems, specified by very large numbers of parameters, there seems little alternative to statistical approaches. By statistical approaches we mean those that have parameters which, instead of being definite numbers which are put

into the model, are random variables taken from a probability distribution function which is put into the model. The cytoplasm of bacteria such as *E. coli* is a very complex system: it is a mixture of thousands of different types of rather complex biomacromolecules, mostly protein but also RNA, DNA, polysaccharides, etc. [10,11,13]. We cannot obtain nor do we want details of all the interactions of these molecules, but we do want to understand and to be able to predict the collective properties of this mixture, such as its osmotic pressure, where it becomes unstable and so on. In these circumstances, statistical approaches, such as the one described here, are the only means of making predictions. They rely on self-averaging occurring in the experimental system, and so, in order to establish their validity, we need to know whether or not the complex mixtures found inside living cells have self-averaging thermodynamic properties, something that experiments will have to assess.

---

*Electronic address: r.sear@surrey.ac.uk
†Electronic address: cuesta@math.uc3m.es

[1] E. P. Wigner, Ann. Math. **53**, 36 (1951).
[2] E. P. Wigner, SIAM Rev. **9**, 1 (1967).
[3] F. J. Dyson, J. Math. Phys. (N.Y.) **3**, 140 (1962).
[4] C. E. Porter, *Statistical Theory of Spectra: Fluctuations* (Academic Press, New York, 1965).
[5] T. Guhr, A. Müller-Groeling, and H. A. Weidenmüller, Phys. Rep. **299**, 189 (1998); P. J. Forrester, N. C. Snaith, and J. J. M. Verbaarschot, J. Phys. A **36**, R1 (2003).
[6] M. L. Mehta, *Random Matrices* (Academic Press, San Diego, 1991).
[7] L. Arnold, J. Math. Anal. Appl. **20**, 262 (1967); Z. Wahrscheinlichkeitstheor. verw. Geb. **19**, 191 (1971).
[8] Z. Füredi and J. Komlós, Combinatorica **1**, 233 (1981).
[9] For every finite (but large) N there is a tiny fraction of matrices with $x_{min} \geq 0$; these have to be excluded from the average. When $N \to \infty$, this fraction vanishes and the average becomes well defined.
[10] F. C. Neidhardt, in *E. coli and S. typhimurium: Cellular and Molecular Biology*, edited by F. C. Neidhardt *et al.* (American Society for Microbiology, Washington, DC, 1987).
[11] *E. coli* has 4385 proteins [12], most of which function inside the cytoplasm of the *E. coli* cell.
[12] The amino-acid sequences of the proteins of *E. coli* can be downloaded, e.g., from the European Bioinformatics Institute (http://www.ebi.ac.uk/proteome). *E. coli* was sequenced by F. R. Blattner *et al.*, Science **277**, 1453 (1997).
[13] R. P. Sear, J. Chem. Phys. **118**, 5157 (2003).