The Journal of Chemical Physics

RESEARCH ARTICLE | FEBRUARY 13 2025

Transport coefficients from equilibrium molecular dynamics

Special Collection: Molecular Dynamics, Methods and Applications 60 Years after Rahman

Paolo Pegolo ; Enrico Drigo ; Federico Grasselli ; Stefano Baroni

J. Chem. Phys. 162, 064111 (2025)
https://doi.org/10.1063/5.0249677

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Transport coefficients from equilibrium molecular dynamics

Cite as: J. Chem. Phys. 162, 064111 (2025); doi: 10.1063/5.0249677
Submitted: 19 November 2024 • Accepted: 17 January 2025 •
Published Online: 13 February 2025

Paolo Pegolo,[1] Enrico Drigo,[2] Federico Grasselli,[3,4] and Stefano Baroni[2,5,a)]

AFFILIATIONS

[1] COSMO—Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
[2] SISSA—Scuola Internazionale Superiore di Studi Avanzati, 34136 Trieste, Italy
[3] Department of Physics, Informatics and Mathematics, Università degli Studi di Modena e Reggio Emilia, 41125 Modena, Italy
[4] CNR Nano S3, 41125 Modena, Italy
[5] CNR-IOM—Istituto Officina Materiali, DEMOCRITOS SISSA Unit, 34136 Trieste, Italy

Note: This paper is part of the JCP Special Topic on Molecular Dynamics, Methods and Applications 60 Years After Rahman.
[a)] **Author to whom correspondence should be addressed:** baroni@sissa.it

ABSTRACT

The determination of transport coefficients through the time-honored Green–Kubo theory of linear response and equilibrium molecular dynamics requires significantly longer simulation times than those of equilibrium properties while being further hindered by the lack of well-established data-analysis techniques to evaluate the statistical accuracy of the results. Leveraging recent advances in the spectral analysis of the current time series associated with molecular trajectories, we introduce a new method to estimate the full (diagonal as well as off-diagonal) Onsager matrix of transport coefficients from a single statistical model. This approach, based on the knowledge of the statistical distribution of the Onsager-matrix samples in the frequency domain, unifies the evaluation of diagonal (conductivities and viscosities) and off-diagonal (e.g., thermoelectric) transport coefficients within a comprehensive framework, significantly improving the reliability of transport coefficient estimation for materials ranging from molten salts to solid-state electrolytes. We validate the accuracy of this method against existing approaches using benchmark data on molten cesium fluoride and liquid water and conclude our presentation with the computation of various transport coefficients of the Li_3PS_4 solid-state electrolyte.

I. INTRODUCTION

Understanding and controlling heat, mass, and charge transport in materials is key to advancing diverse scientific and technological applications, from planetary sciences to energy conversion, thermal management, and electronic device performance. The theory of transport is essentially a theory of hydrodynamic variables, i.e., of long-wavelength components of (current) densities of conserved extensive variables.[1,2]

The foundations of transport theory were laid in 1905 by Einstein in his milestone paper on Brownian motion,[3] where he first established the relation between a particle's diffusivity and its velocity autocorrelation function. The feasibility of exploiting this relation to estimate a molecular self-diffusion constant is as old as molecular dynamics (MD) itself.[4] Building on Onsager's foundational works on irreversible phenomena,[5] in the late 1950s, Einstein's relation was extended to general transport phenomena by Green,[6] Kubo[7] (GK), and, much in the spirit of Einstein himself, by Helfand.[8] Following the way opened by Rahman's seminal work introducing MD,[4] a number of papers documenting the computation of several transport coefficients started to appear in the decade to follow.[9,10]

Despite the long history of transport theory, significant progress is still ongoing in the field, often prodded by the impetus to overcome major technical difficulties in the numerical simulations of transport coefficients. Notably, the introduction of the concept of *gauge invariance* of transport coefficients rectified a common misconception due to the non-uniqueness of the microscopic representation of the densities of the conserved quantities being transported,[11] proving that the Onsager matrix is independent of such

representation,[12–15] thus enabling the simulation of heat and charge transport from first principles.[12,13,16] These advances notwithstanding, estimating transport coefficients via GK-MD requires molecular trajectories at least an order of magnitude longer than those for equilibrium properties,[17] and a reliable assessment of the resulting statistical accuracy remains a considerable challenge. In this paper, we review and further refine recent efforts aimed at overcoming these difficulties.

To estimate transport properties from molecular simulations, non-equilibrium MD[18] (NEMD) is sometimes preferred to GK (equilibrium) MD possibly because it serves as a direct representation of non-equilibrium processes and it does not require the evaluation and implementation of often cumbersome expressions for the currents. However, NEMD is also much more susceptible than GK-MD to finite-size errors, which negatively affect the accuracy and reliability of the results,[19–21] and a rigorous assessment of the statistical errors affecting them is even more challenging in this case. For these reasons, in what follows, we focus exclusively on GK-MD, with the primary objective of addressing and resolving the long-standing challenges associated with the implementation of a rigorous and accurate data analysis strategy.

The GK theory expresses transport coefficients through integrals of time correlation functions of the currents.[2,7] Equilibrium MD provides a mean to sample the current time series, but computing the integral of their correlation functions is far from trivial.[22] Numerous techniques have been devised to cope with the poor performance of the direct evaluation and subsequent integration of the current time series,[23] ranging from truncation techniques[24] to analytic integration of a fit of the correlation function[25] and to mean square displacement-like (Einstein–Helfand) approaches.[8] A common issue of all these methods is the intrinsic difficulty in estimating statistical uncertainties,[15] which requires averaging over multiple independent realizations or, equivalently, by dividing a single long trajectory into separate blocks. In the last decade, all these issues have been addressed exploiting the statistical properties of the time series and their power spectral densities, resulting in what we will refer to as *spectral methods*.[17,26,27]

These methods aim to estimate either *diagonal* transport coefficients—such as electrical and thermal conductivity and bulk and shear viscosity[17,26]—or *off-diagonal* ones that describe coupled phenomena, such as the thermoelectric effect.[27] Estimating diagonal coefficients is effectively addressed using *cepstral analysis*,[17,28] a technique adapted from speech recognition that leverages the positive-definiteness of power spectral densities. Although cepstral analysis efficiently estimates diagonal transport coefficients from short current time series,[17] it faces limitations: it cannot be applied to mixed (i.e., off-diagonal) transport coefficients,[27] and it may introduce a bias when the underlying spectrum exhibits prominent features at low frequencies.[17,29] To overcome these issues for off-diagonal coefficients, a specialized Bayesian regression procedure was then introduced and combined with cepstral analysis, enabling a more efficient computation of the Seebeck coefficient.[27] While successful, this method is computationally intensive and partially inherits the limitations of cepstral analysis.

In this article, we introduce an all-encompassing estimation technique, allowing for a complete characterization of transport coefficients. By generalizing spectral estimation, our method allows one to predict the entire Onsager matrix of transport coefficients from a single model, providing a comprehensive solution to the limitations of previous approaches. This paper is organized as follows: Sec. II is devoted to a compact overview of transport theory and GK linear response theory. Section III describes the statistical properties of the Onsager matrix estimator, the different spectral methods employed, and the new numerical tool capable of simultaneously handling the entire Onsager matrix. In Sec. IV, we showcase numerical benchmarks on the electrical, thermal, and thermoelectric transport coefficients of a molten salt and the shear viscosity of liquid water. Finally, Sec. V is devoted to applying the proposed method to compute the electrical and thermal conductivity of the paradigmatic solid-state electrolyte (SSE) Li_3PS_4, as well as its ionic Seebeck coefficient.

II. TRANSPORT THEORY

Onsager theory. When an external perturbation displaces the density of a conserved quantity, X—such as energy, volume, or the number of molecules of each molecular species, $\{E, V, N_i\}$—out of equilibrium, currents naturally arise to restore it. In perturbation theory, a linear relationship is assumed between the induced fluxes (i.e., macroscopic averages of current densities), $\boldsymbol{\jmath}$, and the thermodynamic forces, i.e., the gradients of the variables conjugated to the conserved densities, $\boldsymbol{f} = \nabla \frac{\partial S(X)}{\partial X}$, with $S(X)$ being the microcanonical entropy.[2] In this approximation, one can describe transport phenomena, such as heat, charge, and momentum transport, as well as coupled effects, such as thermoelectric transport. In a series of seminal papers, Lars Onsager introduced a solid theoretical foundation for irreversible processes under the hypothesis of microscopic reversibility.[5] Onsager's phenomenological equations provide a compact representation of this framework,[2]

$$\langle \hat{\boldsymbol{\jmath}} \rangle_{\mathrm{ne}} = \boldsymbol{L}\boldsymbol{f}. \tag{1}$$

Here, $\boldsymbol{L}$ is the matrix of the Onsager coefficients, which describe the linear response couplings between the thermodynamic forces and the induced fluxes. Note that throughout this article, random variables are denoted with a hat and their expected values are denoted with a pair of angled brackets, $\langle\cdot\rangle$. The notation $\langle\cdot\rangle_{\mathrm{ne}}$ represents a non-equilibrium average in the presence of an external perturbation.[7] The Onsager matrix has dimensions $M\ell \times M\ell$, where M is the number of distinct conserved quantities being transported and ℓ is the number of equivalent components of each flux, such as different Cartesian directions.

Green–Kubo linear response theory. Later, Green and Kubo expanded on Onsager's theory, demonstrating that each element of the Onsager matrix can be systematically computed using dynamical linear response theory at equilibrium.[6,7] In the GK formulation, the Onsager coefficients are obtained as time integrals of the correlation functions of the fluxes,

$$L_{ab} = \frac{V}{k_{\mathrm{B}}T}\int_0^\infty \langle \hat{\jmath}_a(t)\hat{\jmath}_b(0)\rangle \mathrm{d}t, \tag{2}$$

where V and T are the system's volume and temperature, respectively, and k_{B} is the Boltzmann constant. Unlike Eq. (1), here the angled brackets stand for the ensemble *equilibrium* average. The subscript a acts as a multi-index, labeling both the flux type and

their Cartesian direction. All these quantities evolve according to the unperturbed Hamilton equations of motion.

While the simple form of (2) is conceptually appealing, computing Onsager coefficients, in practice, is far from straightforward. Direct computation requires a model for the dynamics able to return a tractable expression of the fluxes and their time evolution. This is possible, under suitable approximations of weak anharmonicity, for solids not too far from mechanical equilibrium.[30–32] In more general cases, numerical simulations are needed to sample the time series of the fluxes, avoiding the need for such approximations. Extracting meaningful information from these time series relies on exploiting their statistical properties.

III. DATA ANALYSIS

Issues with time integration. The naïve evaluation of the GK integrals of Eq. (2) requires long equilibrium MD simulations to gather enough data for accurately resolving the long-time tails of the correlation functions.[15,17] While modern machine learning interatomic potentials (MLIPs) enable significantly longer, high-quality, MD simulations than in the past, an accurate estimate of the error bars of Onsager coefficients remains challenging. In fact, at times larger than the largest flux correlation time, the numerical sample of the time correlation function becomes a zero-mean stochastic process whose integral is a diffusive process whose variance increases linearly with time. Statistical consistency is restored through computational and labor intensive block analysis[2,17,27] using the so-called Helfand–Einstein formula,[8]

$$L_{ab}(\mathcal{T}) = \frac{V}{k_B T}\int_0^{\mathcal{T}} \left(1 - \frac{t}{\mathcal{T}}\right)\langle j_a(t) j_b(0)\rangle \mathrm{d}t, \tag{3}$$

which, in the large-$\mathcal{T}$ limit, is equivalent to the GK formula, Eq. (2), but is affected by a smaller statistical uncertainty,[15] thus mitigating the issue. This technical challenge is particularly pronounced when the transport coefficient of interest is inherently small as in the case of ionic conductivity in SSEs. Besides ionic conductivity, GK integration completely fails in the case of thermal conductivity in multicomponent systems.[26] Thermal conductivity, defined as the ratio between heat flux and temperature gradient in the absence of mass fluxes, is highly sensitive to statistical noise because its expression involves differences between similarly sized positive terms, each with its own independent uncertainty.[26,33–36] For solids, the issue can be partly mitigated by filtering out oscillations in the correlation function caused by normal modes and choosing the integration limit according to the so-called "first dip" criterion—when the denoised autocorrelation function first reaches zero.[37] However, this approach is not directly applicable to other physical systems and does not fully resolve the issue of obtaining reliable error bars.

A. Spectral methods

An entire class of methods is based on the observation that the Onsager coefficients in Eq. (2) can be expressed in terms of the zero-frequency value of the corresponding current cross power spectral density (PSD), defined as the Fourier transform of the appropriate current correlation functions,

$$L_{ab} = \frac{V}{2k_B T}\mathcal{S}_{ab}(0), \tag{4}$$

$$\mathcal{S}_{ab}(\omega) = \int_{-\infty}^{\infty} C_{ab}(t) e^{i\omega t} dt, \tag{5}$$

$$C_{ab}(t) = \langle j_a(t) j_b(0)\rangle. \tag{6}$$

The advantage of doing so is that PSD values at various frequencies can be estimated simultaneously with minimal computational effort by applying Fast-Fourier-Transform techniques, while the smoothness of the PSD can be leveraged to reduce the error of their estimates, in particular, in the zero-frequency limit, as explained below.[17]

The Wiener–Khintchine theorem. The PSD can be conveniently estimated, instead of from Eq. (4), by leveraging the Wiener–Khintchine theorem,[38,39] which states that in the long-time limit, it is the expectation of the products of the Fourier transforms of time series,

$$\mathcal{S}_{ab}(\omega) = \langle \widehat{\mathcal{S}}_{ab}(\omega)\rangle, \tag{7}$$

$$\widehat{\mathcal{S}}_{ab}(\omega) = \tilde{j}_a(\omega)\tilde{j}_b^*(\omega), \tag{8}$$

$$\tilde{j}(\omega) = \frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} j(t) e^{i\omega t} dt. \tag{9}$$

The PSD sample expressed in Eq. (8) is usually called the (cross-) *periodogram*. Its statistical properties are determined by those of the flux time series.

Fluxes as Gaussian processes. Under the reasonable hypothesis that *spatial* autocorrelations of locally conserved currents are short-ranged,[40] the corresponding macroscopic fluxes can be viewed as sums of nearly independent, identically distributed, stochastic variables. By the central limit theorem, the distribution of these fluxes is thus Gaussian. Stationarity, in turn, implies that the corresponding correlation function only depends on the time lag between the two fluxes, $C_{ab}(t_1, t_2) = C_{ab}(t_1 - t_2)$ [see Eq. (6)], and that the Fourier transforms of the flux time series, Eq. (9), are uncorrelated for different frequencies: $\langle \tilde{j}_a(\omega)\tilde{j}_b^*(\omega')\rangle = 0$ for $\omega \neq \omega'$.

Since the collection of various fluxes, $\{j\}$, forms a stationary multivariate Gaussian process (MVGP) and their Fourier transforms, Eq. (9), also constitute a MVGP that is uncorrelated across different frequencies, the cross-periodogram in Eq. (8) corresponds to a Wishart process.[41] A Wishart matrix is a random matrix that generalizes the concept of chi-square distribution to multidimensional normal variates.[42] More specifically, let us consider M thermodynamic fluxes (energy, charge, momentum, etc.), each of which is sampled ℓ independent times (for isotropic systems, ℓ is at least three, one per Cartesian direction). The cross periodogram for fluxes a and b is the average over independent samples,

$$\widehat{\mathcal{S}}_{ab}(\omega) = \frac{1}{\ell}\sum_{\alpha=1}^{\ell} \widehat{\mathcal{S}}_{ab}^{\alpha}(\omega), \tag{10}$$

where $\widehat{\mathcal{S}}_{ab}^{\alpha}$ is the product of the discrete cosine Fourier transforms of the flux time series $\hat{j}_a^{\alpha}$ and $\hat{j}_b^{\alpha}$. At each frequency, the cross periodogram matrix is a Wishart random matrix[41] with ℓ degrees of freedom and with the *scale matrix* given by the PSD matrix,

$$\widehat{\boldsymbol{\mathcal{S}}}(\omega) \sim \boldsymbol{\mathcal{W}}_M(\boldsymbol{\mathcal{S}}(\omega), \ell). \tag{11}$$

The probability density function for the Wishart distribution has a closed form, given by

$$p_{\mathcal{W}}(\boldsymbol{W}) = \frac{(\det \boldsymbol{W})^{(\ell-M-1)/2} e^{-\mathrm{Tr}\,(\boldsymbol{S}^{-1}\boldsymbol{W})/2}}{2^{\frac{\ell M}{2}} (\det \boldsymbol{S})^{\ell/2} \Gamma_M\left(\frac{\ell}{2}\right)}, \tag{12}$$

where $\Gamma_n(x)$ is the multivariate gamma function[43] and the scale matrix $\boldsymbol{S}$ is an $M \times M$ positive definite matrix.

The diagonal elements of the spectral matrix correspond to transport coefficients, such as thermal and electrical conductivity, as well as bulk and shear viscosities. The off-diagonal elements describe the coupling between different thermodynamic fluxes, such as thermoelectric effect, captured by coefficients such as Seebeck's. It is possible to prove that diagonal elements of the PSD matrix are chi-square distributed,[17,26] while off-diagonal elements follow the variance-gamma distribution.[27,44,45]

In Secs. III B and III C, we review some existing spectral methods that separately address diagonal and off-diagonal transport coefficients, leveraging the statistical properties of Wishart matrices. In Sec. III D, we then introduce a novel method that simultaneously targets the entire spectral matrix and therefore all transport coefficients, thus overcoming several limitations of the existing approaches.

B. Diagonal elements

As previously mentioned, diagonal elements of the periodogram matrix are distributed as chi-square variates; namely, for a (discrete) frequency ω_k, $k = 1, \ldots, N-1$,

$$\widehat{\mathcal{S}}^k_{aa} = \frac{1}{\ell} \mathcal{S}_{aa}(\omega_k) \widehat{\xi}_k \tag{13}$$

where $\widehat{\xi}_k$ is a chi-square random variable with ℓ degrees of freedom,[46] $\widehat{\xi}_k \sim \chi^2_\ell$. Given the positive definiteness of χ^2 variables, one can take the logarithm of Eq. (13) to turn multiplicative noise into additive, thus enabling the use of low-pass filters to reduce the noise in the resulting signal. This approach lies at the foundation of *cepstral analysis*.[17,26] More specifically, the log-periodogram is equal to the logarithm of the true power spectrum plus a white, but non-Gaussian, noise term,

$$\widehat{\mathcal{L}}^k_{aa} \equiv \log\, \widehat{\mathcal{S}}^k_{aa} = \log \mathcal{S}_{aa}(\omega_k) + \log \frac{1}{\ell} \widehat{\xi}^k. \tag{14}$$

Cepstral analysis leverages the fact that the underlying spectrum is a smooth function of frequency, thus featuring a limited number of inverse Fourier coefficients, and that the noise is white so that a low-pass filter applied to the periodogram would eliminate most of it without affecting the spectral signal. The *cepstrum* is defined as the inverse Fourier transform of the log-periodogram,

$$\begin{aligned} \widehat{\mathcal{C}}^n_{aa} &= \frac{1}{N} \sum_{k=0}^{N-1} \widehat{\mathcal{L}}^k_{aa} e^{-2\pi i \frac{nk}{N}}, \\ \mathcal{C}^n_{aa} &= \frac{1}{N} \sum_{k=0}^{N-1} \log \mathcal{S}_{aa}(\omega_k) e^{-2\pi i \frac{nk}{N}}. \end{aligned} \tag{15}$$

The low-n (low-*quefrency*, in cepstral jargon)[28] cepstral coefficients are associated with the smooth log-spectrum plus low-quefrency noise, while the high-n coefficients correspond to pure noise. Under this assumption, an estimate of the log-spectrum at zero frequency can be reconstructed using a few P cepstral coefficients, those attributed to the true log-spectrum,

$$\widehat{\mathcal{L}}^0_{aa} = \widehat{\mathcal{C}}^0_{aa} + 2 \sum_{n=1}^{P} \widehat{\mathcal{C}}^n_{aa}. \tag{16}$$

The mean and variance of the reconstructed log-spectrum are given by[17,47]

$$\begin{aligned} \mathcal{L}^0_{aa} &= \log\, \mathcal{S}^0_{aa} + \psi(\ell) - \log\, \ell, \\ \mathrm{var}\, \mathcal{L}^0_{aa} &= \psi'(\ell/2) \left[\frac{4P-2}{N} + \frac{4P(1-P)-2}{N^2} \right], \end{aligned} \tag{17}$$

where ψ and ψ' are the digamma and trigamma functions, respectively. The value of P is a measure of the spectrum smoothness and is independent of N, allowing for systematic reduction of the cepstral estimator's variance by increasing the length of the flux time series.[17] A different statistical model corresponds to each value of P, i.e., to the number of retained cepstral coefficients. Estimating P means selecting among these model the best one at reproducing the zero-frequency value of the spectrum. This task can be addressed using model selection criteria, with the Akaike Information Criterion (AIC) being a widely used one. The AIC aims to balance a model's accuracy and simplicity. The AIC statistics is defined as

$$\mathrm{AIC}\,(P) = 2 \max_c \log \mathfrak{L}(P, c|\mathcal{C}) - 2P, \tag{18}$$

whose minimum with respect to P selects the optimal number, $P^\star$, of cepstral coefficients c to retain according to the AIC. Here, $\mathfrak{L}$ is the Laplace (i.e., Gaussian) approximation of the likelihood function of the cepstral coefficients,[17]

$$2 \log \mathfrak{L}(P, c|\mathcal{C}) = -\left(c^0 - \widehat{\mathcal{C}}^0_{aa}\right)^2 - \sum_{n=1}^{P} \left(c^n - \widehat{\mathcal{C}}^n_{aa}\right)^2 - \sum_{n=P+1}^{N-1} \left(\widehat{\mathcal{C}}^n_{aa}\right)^2. \tag{19}$$

The true cepstral coefficients with $n > P^\star$ are assumed to be zero. This approximation has proven to be very effective, especially for liquids,[17,26,48–51] but it can bias the cepstral estimate when the PSD has pronounced features at low frequency. Such cases include the thermal conductivity of polar solids (e.g., magnesium oxide[17]) and low-temperature Li-based SSEs,[29,52,53] as well as the viscosity of near-freezing liquids.[54]

A ready-to-use software package to perform cepstral analysis of multivariate time series is documented in Ref. 82.

C. Off-diagonal elements

In addition to this limitation, the cepstral analysis relies on the diagonal elements of the Wishart matrix to be positive. Therefore, it cannot be applied to off-diagonal matrix elements, thus preventing the study of coupled transport coefficients, such as Seebeck's. This specific case was addressed by considering the far more general method of Bayesian inference, exploiting the explicit knowledge of the probability density of the off-diagonal periodograms.[27] In the case of coupled heat and charge transport, $\boldsymbol{W}$ defined in Eq. (12) is a 2×2 matrix whose diagonal elements are the charge and heat frequency-dependent PSDs, while the off-diagonal entry

is the thermoelectric periodogram, which is known to follow the variance-gamma ($v\Gamma$) distribution, whose probability density is

$$p_{v\Gamma}(W_{qc}) = \frac{|W_{qc}|^{\frac{\nu-1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{2^{\nu-1}\pi\left(1-\rho^2\right)}\left(W_{qq}W_{cc}\right)^{\frac{\nu+1}{4}}} \times \mathcal{K}_{\frac{\nu-1}{2}}\left(\frac{|W_{qc}|}{\left(W_{qq}W_{cc}\right)^{1/2}\left(1-\rho^2\right)}\right) \times \exp\frac{\rho W_{qc}}{\left(W_{qq}W_{cc}\right)^{1/2}\left(1-\rho^2\right)}. \tag{20}$$

Here, Γ is the Euler gamma function,[55] $\mathcal{K}$ is the modified Bessel function of second kind,[56] ρ is the correlation coefficient, bounded by $-1 < \rho < 1$, and q and c label heat and charge, respectively.

The mere knowledge of the probability distribution, as given in Eq. (20), could be leveraged for practical purposes only if we could repeatedly sample the value at a given frequency (e.g., $\omega = 0$) a large number of times. However, this is not feasible when dealing with a single trajectory. Instead, we have access to periodogram values at numerous distinct frequencies, which become statistically independent in the limit of long trajectories. Therefore, we can leverage the assumed regularity of the various PSDs underlying the periodogram to fit the value at $\omega = 0$, which, as discussed, is proportional to the sought transport coefficient. This is obtained by fitting to the data a parametric model for the ω-dependent PSD functions. In Ref. 27, the diagonal PSDs, W_{qq} and W_{cc}—whose low-frequency limits are the heat and charge Onsager coefficients—were estimated using cepstral analysis, while the correlation coefficient $\rho(\omega)$ was modeled with a flexible spline, whose parameters are defined by the function's values at a set of predetermined frequencies (the knots). The Bayesian method was tested by computing the Seebeck coefficient of molten salts, revealing that, to achieve the same accuracy, it requires trajectories about one order of magnitude shorter than those needed by GK integration.[27]

The Bayesian method can also be applied to diagonal coefficients if the likelihood function is appropriately adjusted though it is more computationally demanding than cepstral analysis. By avoiding both the Gaussian approximation in the likelihood function and the assumption of null large-n cepstral coefficients, this method circumvents the bias in transport coefficient estimates often associated with cepstral analysis. Generally, careful consideration is needed to balance the minimal computational cost of cepstral analysis—along with its potential bias—against the greater reliability but higher computational expense of the Bayesian approach.

D. Direct estimation of the Wishart matrix

Since the form of Eq. (12) is known, it seems reasonable to directly estimate the distribution parameters of the spectral matrix similar to the approach used for off-diagonal Onsager coefficients. The Bayesian approach presented in Ref. 27 relies on extensive Monte Carlo Markov chains (MCMCs) to sample the posterior distribution. Despite ensuring accurate estimation of the parameter values and their complete distribution, MCMCs are computationally expensive. In this work, rather than performing a full-fledged Bayesian MCMC analysis, we adopt a frequentist approach. Specifically, we fit a model by minimizing the negative log-likelihood (NLL) function. The sampled periodogram matrix, $\boldsymbol{W}$, serves as the input, while the spectral matrix is modeled using an appropriate scale matrix, $\boldsymbol{S}$. Given that the scale matrix is positive definite, its elements are not all independent. A practical way of enforcing positive definiteness is constructing a model for the upper triangular matrix $\boldsymbol{C}$ such that $\boldsymbol{S} = \boldsymbol{C}\boldsymbol{C}^{\mathsf{T}}$. Therefore, the log-likelihood function, modulo some constants in $\boldsymbol{C}$, becomes

$$\log \mathfrak{L} = \frac{\ell - M + 1}{2} \log\det \boldsymbol{W} - \frac{1}{2}\operatorname{Tr}\left(\left(\boldsymbol{C}\boldsymbol{C}^{\mathsf{T}}\right)^{-1}\boldsymbol{W}\right) - \ell \log\det \boldsymbol{C}. \tag{21}$$

$\boldsymbol{C}$ is known as the Cholesky decomposition of $\boldsymbol{S}$.[57] Similar to the off-diagonal Onsager coefficients, there are no strong physical priors for modeling $\boldsymbol{C}$ using a specific functional form. Therefore, a convenient approach is to express $\boldsymbol{C}(\omega)$ as a set of spline functions, one for each element in the upper triangle of $\boldsymbol{C}$. The parameter optimization process involves minimizing the NLL with respect to the model parameters, thereby identifying the best parameter set that fits the data. A schematic illustration of this approach is sketched in Fig. 1

The selection of the number of parameters to be used in the Wishart NLL estimation is guided by the AIC. Depending on the empirical shape of the periodogram, one can determine where to place the spline nodes, ranging from $\omega = 0$ to the Nyquist frequency of the time series.[27] A straightforward approach is to use equispaced nodes. However, since the ultimate goal of transport coefficient estimation is to obtain a value at $\omega = 0$, it may be advantageous to use different node spacings, such as logarithmic spacing. This allows for better sampling of the low-frequency region, where the most significant part of the spectrum is often located, using fewer parameters. This strategy is particularly useful when the spectrum exhibits abrupt changes at low frequencies, such as zero-frequency peaks or dips. In some cases, the focus may shift from estimating transport coefficients at $\omega = 0$ to resolving finite frequencies with higher accuracy, such as when estimating infrared spectra. In these scenarios, resolving details at specific frequencies becomes more important than capturing the behavior at zero frequency. For such applications, using ad hoc spacing or simply maintaining equispaced nodes may be more appropriate to accurately represent the spectral features across the frequency range.

A final consideration concerns uncertainty estimation. While minimizing the NLL is computationally more efficient than Bayesian regression, it provides only an approximation of the full distribution

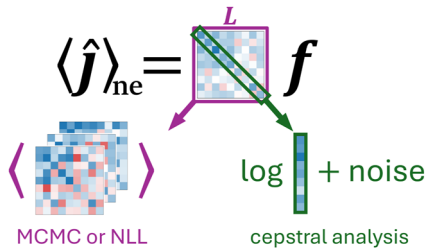

FIG. 1. Schematic of the spectral estimation of transport coefficients. While the cepstral method offers a simple and efficient approximation for diagonal Onsager coefficients, it may lack accuracy in more complex cases. In contrast, modeling the entire spectral matrix provides a comprehensive approach, yielding all transport coefficients—both diagonal and off-diagonal—simultaneously.

of the parameters, and therefore an approximation of their variance, and, hence, their uncertainty. In an NLL estimate, the uncertainty is obtained from the inverse Hessian of the minimized log-likelihood with respect to the parameters. As empirically shown in Sec. IV E, this approximation holds well in the context of transport coefficient estimation with spline models. Indeed, in the large-sample limit, the posterior distribution of the parameters converges to a Gaussian, as proved by the Bernstein–Von Mises theorem.[58] In the simple condition of independent samples, as it is the case for different frequencies of the periodogram matrix, the theorem remarks that the likelihood is the distribution of the sum of the elements in the dataset which, by the central limit theorem, is Gaussian.

IV. VALIDATION

Let us now validate the Wishart estimator against results in the literature obtained using cepstral analysis. In particular, we consider heat and electric conductivity, as well as the Seebeck coefficient, of molten CsF whose transport coefficients were computed from Bayesian regression analysis in Ref. 27. Additionally, we consider the case of liquid water's shear viscosity.[54]

For heat and charge currents in a two-component material, the Onsager equations read[40]

$$\begin{pmatrix} \langle \hat{\boldsymbol{j}}_c \rangle_{\mathrm{ne}} \\ \langle \hat{\boldsymbol{j}}_q \rangle_{\mathrm{ne}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma} & -\boldsymbol{K}_{12} \\ \boldsymbol{K}_{12} e & -\boldsymbol{L}_{qq} \end{pmatrix} \begin{pmatrix} \boldsymbol{E} \\ \dfrac{\nabla T}{T} \end{pmatrix}, \tag{22}$$

where $\boldsymbol{j}$ is the Cartesian current vector as in $\boldsymbol{j} = \left\{\boldsymbol{j}_c, \boldsymbol{j}_q\right\}$. The electric flux, $\boldsymbol{j}_c$, is computed as $\boldsymbol{j}_c = \sum_s q_s \boldsymbol{j}_s$, where q_s is an integer multiple of the elementary charge, i.e., the oxidation state associated with chemical species s, and $\boldsymbol{j}_s$ is the number flux of species s, $\boldsymbol{j}_s = V^{-1} \sum_{I \in s} \boldsymbol{v}_I$. It should be noted that using integer charges for describing transport processes is neither an arbitrary choice nor an approximation: it is a fundamental requirement of quantum mechanics owing to topological properties of the ground state wavefunction of electronic insulators.[16,61–63] The heat flux, $\boldsymbol{j}_q$, is computed as $\boldsymbol{j}_q = \boldsymbol{j}_e - \sum_s h_s \boldsymbol{j}_s$, with $\boldsymbol{j}_e$ being the energy flux, h_s being the molecular enthalpy,[64–66] and $\boldsymbol{E}$ being the local electric field. The Onsager matrix, $\boldsymbol{L}$, described in Eq. (22) has rank 6, and its blocks $\boldsymbol{\sigma}$, $\boldsymbol{K}_{12}$, and $\boldsymbol{L}_{qq}$ are 3×3 Cartesian tensors, whose entries are proportional to some Onsager coefficients.

The Onsager coefficients defined in Eq. (22) can be computed using the GK theory as

$$\boldsymbol{\sigma} = \frac{V}{k_{\mathrm{B}} T} \int_0^\infty \langle \hat{\boldsymbol{j}}_c(t) \otimes \hat{\boldsymbol{j}}_c(0) \rangle \mathrm{d}t, \tag{23}$$

$$\boldsymbol{K}_{12} = \frac{V}{k_{\mathrm{B}} T} \int_0^\infty \langle \hat{\boldsymbol{j}}_c(t) \otimes \hat{\boldsymbol{j}}_q(0) \rangle \mathrm{d}t, \tag{24}$$

$$\boldsymbol{L}_{qq} = \frac{V}{k_{\mathrm{B}} T} \int_0^\infty \langle \hat{\boldsymbol{j}}_q(t) \otimes \hat{\boldsymbol{j}}_q(0) \rangle \mathrm{d}t. \tag{25}$$

The symbol $\otimes$ indicates the external product between Cartesian vectors as in $(\boldsymbol{u} \otimes \boldsymbol{v})_{\alpha\beta} = u_\alpha v_\beta$.

A. Electrical conductivity

In anisotropic media, the Cartesian components of the fluxes are independent but not equivalent. Therefore, in conducting systems, the resulting electric conductivity tensor $\boldsymbol{\sigma}$, defined in Eq. (23), is itself a Wishart distributed matrix. In isotropic liquids or solids, $\mathrm{Tr}\,\boldsymbol{\sigma}$ is proportional to a χ^2_ℓ random variable equivalent to the transport coefficient measured in experiments.[52] Here, $\ell = 6 = 2 \times$ (3 equivalent Cartesian components). We analyze electrical conductivity of molten CsF at 1400 K and ambient pressure, modeled with a semi-empirical Fumi–Tosi force field.[67–69] Details can be found elsewhere[27] and in the Materials Cloud repository associated with the present article. Wishart NLL estimate yields all the transport coefficients at the same time. We compare the estimate of σ with individual χ^2-NLL and cepstral estimates. The results are all compatible with one another. The spectrum estimated with the different methods is shown in Fig. 2(a).

B. Thermal conductivity

In multi-component systems, thermal conductivity is not proportional to any individual Onsager coefficient. The thermodynamic definition of the transport coefficient is the derivative of the heat flux with respect to the temperature gradient at null mass fluxes, $\mathcal{J}$,

$$\kappa_{\alpha\beta} = \left. \frac{\partial \langle \mathcal{J}_{q\alpha} \rangle_{\mathrm{ne}}}{\partial (\nabla T)_\beta} \right|_{\langle \mathcal{J} \rangle_{\mathrm{ne}} = 0}. \tag{26}$$

From the Onsager equations defined in (22), the thermal conductivity matrix, $\boldsymbol{\kappa}$, reads

$$\boldsymbol{\kappa} = \frac{1}{T} \left(\boldsymbol{L}_{qq} - \boldsymbol{K}_{12} \boldsymbol{\sigma}^{-1} \boldsymbol{K}_{12} \right). \tag{27}$$

In other words, the thermal conductivity tensor is the Schur complement of the convective block $\boldsymbol{\sigma}$ in the Onsager matrix. It can be proved that the Schur complement of a Wishart matrix is still a Wishart distributed tensor.[70] Therefore, even if in multi-component systems the entries of $\boldsymbol{\kappa}$ are not proportional to GK integrals, the estimator of the tensor is still Wishart distributed, it being the Schur complement of a Wishart matrix. Since molten CsF is an isotropic fluid, its thermal conductivity can be obtained from the (one-dimensional) Schur complement of the spectral matrix. In the context of Wishart-NLL estimation, κ results from Eq. (27), where each of the three coefficients in the right-hand side is an element of the estimated spectral matrix. Cepstral and χ^2-NLL estimates instead rely on the so called *reduced spectrum*,[26] i.e., the frequency-dependent Schur complement of the convective block of the periodogram matrix. In this case, the quantity is χ^2 distributed with $\ell - M + 1$ degrees of freedom.[70] Again, the results of the three methods are compatible within the estimated uncertainty, as shown in Fig. 2(b).

C. Seebeck coefficient

We compare the Wishart-NLL results with those obtained from the Bayesian estimate of S through the marginal distribution of off-diagonal transport coefficients,[27] where the frequency dependent Seebeck coefficient is expressed as

$$S(\omega) = \frac{\rho(\omega)}{T} \sqrt{\frac{L_{qq}(\omega)}{\sigma(\omega)}} \tag{28}$$

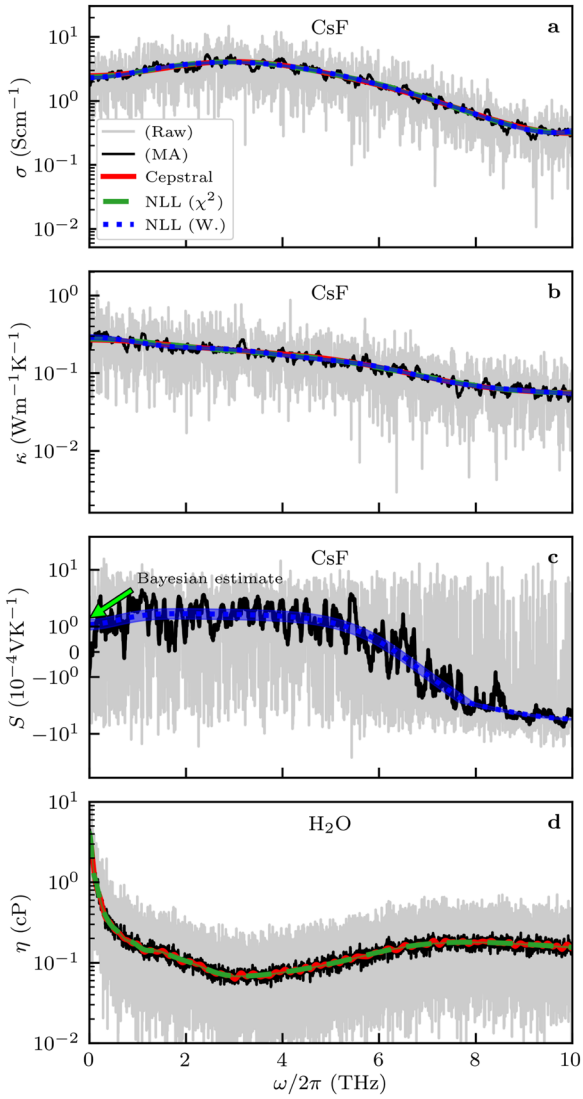

FIG. 2. Comparison of the NLL estimation of transport coefficients and the cepstral method. The label "MA" indicates a moving average.[59] Electric and heat flux time series for a semi-empirical force-field model of CsF at 1400 K and ambient pressure are taken from the work of Drigo *et al.*[27]; time series for the shear stress of H_2O at 300 K were made available by Malosso *et al.*[54] Note the log-scale (symlog scale[60]) in the *y*-axis for diagonal (off-diagonal) transport coefficients to magnify the comparison between the different methods.

and $L_{qq}(\omega)$ and $\sigma(\omega)$ are given by cepstral analysis, while $\rho(\omega)$ results from a MCMC run.[27] In general, the MCMC approach is computationally more expensive than the simple NLL minimization. Since when the Bayesian method was introduced, the off-diagonal coefficient is estimated relying on diagonal coefficients obtained with cepstral analysis,[27] to avoid incurring in the possible biases of cepstral analysis, we choose to validate the Wishart-NLL approach in a thermodynamic condition (high temperature, low pressure) where cepstral results are easy to obtain, not to introduce possible artificial sources of inconsistency. Indeed, the two methods yield compatible results, as shown in Fig. 2(c).

D. Shear viscosity

The same considerations made for electrical conductivity hold for shear viscosity η, defined as

$$\eta = \frac{V}{k_B T}\int_0^\infty \langle \widehat{\Xi}_s(t)\widehat{\Xi}_s(0)\rangle \mathrm{d}t, \tag{29}$$

where Ξ_s is any component element of the off-diagonal stress tensor $\{\Xi_{xy}, \Xi_{yz}, \Xi_{xz}\}$.[54] For shear viscosity, we compare our results with those on liquid water simulated with MLIP trained on SCAN level density-functional theory *ab initio* calculations.[54] Since viscosity does not couple with heat or electric transport and liquid water is isotropic, we compare χ^2-NLL results with cepstral ones only. The η PSD comparing the two methods is shown in Fig. 2(d).

E. Validation against MCMC sampling

We further validate the NLL scheme against Bayesian MCMC for the case of molten CsF. We take the estimated values of the $\boldsymbol{C}$ matrix obtained from NLL minimization as the starting point for MCMC sampling of the parameters' posterior distribution with an uninformative flat prior. The number of parameters selected by the AIC from NLL minimization is 7. The chains are initialized

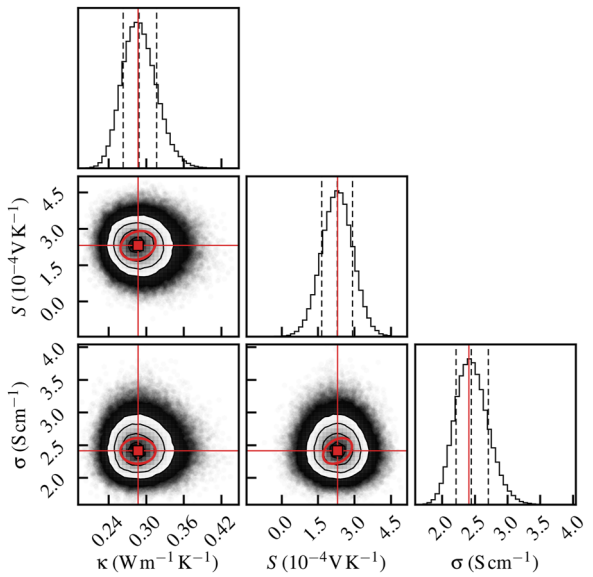

FIG. 3. Triangle plot illustrating the comparison between MCMC and NLL estimation of κ, S, and σ of molten CsF. In black, the posterior distribution obtained from MCMC sampling with a Wishart likelihood. Diagonal plots show histograms of the individual transport coefficients, with dashed vertical lines indicating the 0.16, 0.5, and 0.84 quantiles. Off-diagonal scatter plots show pairwise distributions with 2D histograms and the same quantile contours. Overlaid in red are the results from NLL estimation. Red squares and lines indicate the NLL-estimated values, while red ellipses represent the covariance matrix from the Laplace approximation. The overlap between the red ellipses and the innermost black contours suggests strong agreement, confirming the accuracy of the Laplace approximation.

with twice as much random walkers as the NLL results plus independent instances of standard Gaussian noise. MCMCs are run for 100 000 steps to ensure convergence, measured in terms of the estimated parameter autocorrelation time. The maximum correlation time results to be 380 steps, thus ensuring the total length of the chain to be more than 250 times longer than the correlation time. The sampled $\boldsymbol{C}$ matrix is mapped to the spectral matrix, and the results at zero frequency are shown in Fig. 3, displaying striking agreement with NLL ones, thus validating the NLL estimation of both average values and covariances.

V. APPLICATION TO SOLID-STATE ELECTROLYTES

Let us now consider the SSE Li_3PS_4 (LPS). Unlike molten salts or molecular liquids, SSEs exhibit significant diffusion of an ionic species (in this case, of Li ions), while maintaining a solid matrix formed by the other components. Transport coefficients estimation is, in fact, a more complicated task in SSEs than in liquids and solids. In liquids, the task is simplified by the inherent ergodic dynamics, which allow for short simulations, and by the typically small vibrational contributions to the spectral densities, which otherwise complicate the spectral estimation of transport coefficients.[17,29,53] In solids, while vibrational contributions severely affect the spectra, these effects can be managed using lattice dynamics methods in the quasi-harmonic regime. For systems exhibiting strong anharmonicity, it is still possible to attempt to filter out the vibrational contributions from the heat current time series.[37]

We obtain the spectral matrix using Wishart-NLL estimation. The resulting transport coefficients are given in Fig. 4 (electrical conductivity), Fig. 5 (thermal conductivity), and Fig. 6 (Seebeck coefficient). In particular, thermal conductivity can be estimated from a 2×2 Onsager matrix since LPS can be considered as a two-component material, where the diffusive Li^+ species move through a solid matrix made of PS_4^{3-} ions.[52] The results for electrical and thermal conductivity are in fair agreement with GK calculations already present in the literature.[52,53] Minor variations in σ are to be attributed to several differences in the approach of Ref. 52, namely, the original ML model is a Gaussian approximation potential rather than a NEP, the pseudopotentials for the reference calculations are different, and the system size is much larger here (6144 atoms instead of 768). Regarding the Seebeck coefficient, the estimated values are of the same order of magnitude as those of different molten salts at high temperature.[27] We remark that imperfect ergodicity and constraints on cell shape make simulations dependent on the phase in which the system is initialized. In Ref. 52, it was further observed that simulations starting in the nonconductive γ phase undergo a structural phase transition to a Li^+-conductive mixed α/β phase when $T \gtrsim 625$ K, which is most likely the reason why nonvanishing σ and S are observed in the present simulations. The same applies to the simulations analyzed in this work. Therefore, in Figs. 4–6,

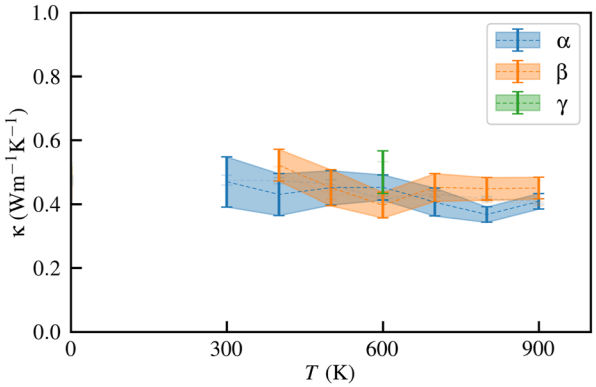

FIG. 5. Thermal conductivity of the three phases of LPS estimated through a Wishart model maximizing the likelihood function. Error bars are estimated in a stochastic fashion sampling parameters as multinormal variables centered on the maximum-likelihood results and with the inverse Hessian of the posterior distribution as the covariance matrix. In more transparent shades of the same colors and partially hidden by the present results, we show the values of κ taken from Ref. 53.

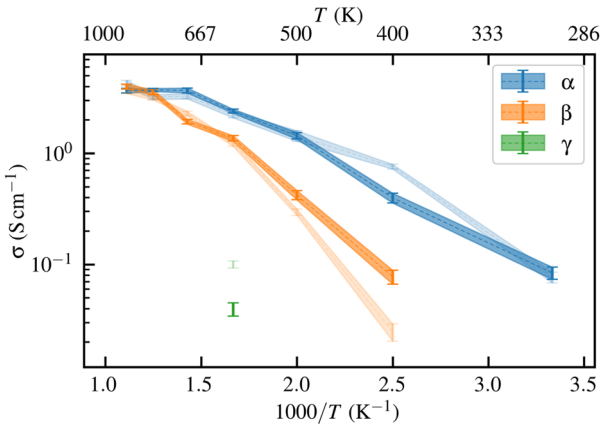

FIG. 4. Electrical conductivity of the three phases of LPS estimated through a Wishart model maximizing the likelihood function. Error bars are estimated standard deviations in the Laplace approximation for the posterior distribution. In more transparent shades of the same colors, we show the results from Ref. 52, which demonstrate fair agreement with our findings.

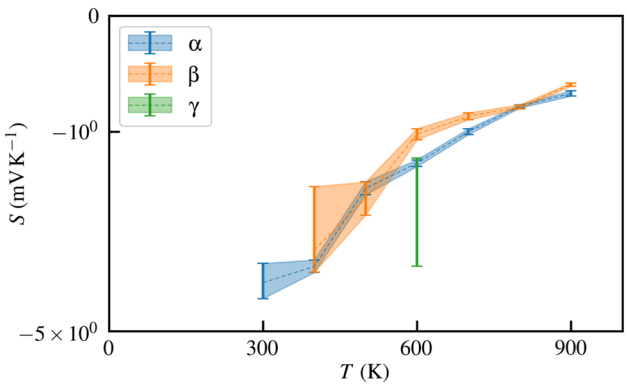

FIG. 6. Ionic Seebeck coefficient of the three phases of LPS estimated through a Wishart model maximizing the likelihood function. Error bars are estimated in a stochastic fashion sampling parameters as multinormal variables centered on the maximum-likelihood results and with the inverse Hessian of the posterior distribution as the covariance matrix. Note the symlog scale in the ordinates.

we label only the simulation at 600 K with "γ" as we verified that it remained in the γ phase throughout the production run.

VI. CONCLUSIONS

By leveraging the statistical properties of Wishart processes, we develop a one-shot Bayesian regression analysis for estimating the whole Onsager matrix from a single model. By minimizing the NLL function associated with a Wishart process, we overcome the numerical challenges due to the realization of MCMC.[27] The NLL approach is not devoted solely to diagonal transport coefficients but addresses all transport coefficients. We benchmark NLL against MCMC Bayesian regression and cepstral analysis, proving that the Laplace approximation is as accurate as preexisting spectral techniques in their working regimes.[17,27] We focused on the quality of the Gaussian hypothesis by validation against MCMC, pointing out that the NLL estimates and uncertainties are in good agreement with computationally expensive Monte Carlo simulations, in accordance with the Bernstein–Von Mises theorem. We exploit the newly introduced method to target the transport properties of a paradigmatic SSE in a range of temperatures, showing that having a high Seebeck coefficient at relatively low temperatures, LPS could be considered a good ionic thermoelectric material.

ACKNOWLEDGMENTS

We thank Cesare Malosso for sharing the shear stress time series of water and Davide Tisi and Alfredo Fiorentino for fruitful discussions. We express our gratitude to an anonymous reviewer of a previous work[27] for inspiring the idea of a comprehensive Bayesian regression analysis of the whole Onsager matrix.

This work was partially supported by the European Commission through the MAX Center of Excellence for supercomputing applications (Grant No. 101093374); the Italian MUR, through the PRIN project ARES (Grant No. 2022W2BPCK); and the Italian National Centre for HPC, Big Data, and Quantum Computing (Grant No. CN00000013), funded through the *Next generation EU* initiative. P.P. acknowledges funding from the Swiss National Science Foundation (SNSF) under Sinergia Project No. CRSII5_202296. F.G. acknowledges financial support from UNIMORE through the FAR 2024 project "Revolutionizing All-Solid-State Sodium Batteries with Advanced Computational Tools and Mixed Glass Former Effects," PI: Alfonso Pedone, CUP E93C24001990005, and from the EMPEROR Project, CUP E93C24001040001, sponsored by the National Quantum Science and Technology Institute (Spoke 5), Grant No. PE00000023, funded by the European Union – NextGeneration EU.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Paolo Pegolo: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Enrico Drigo**: Conceptualization (equal); Data curation (supporting); Formal analysis (supporting); Investigation (equal); Methodology (equal); Software (supporting); Validation (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Federico Grasselli**: Conceptualization (equal); Formal analysis (supporting); Investigation (equal); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Stefano Baroni**: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data and scripts that support the results within this paper are available in the Materials Cloud platform[71] at https://doi.org/10.24435/materialscloud:hf-ar. The SPORTRAN version used for data analysis is currently hosted on GitHub at https://github.com/ppegolo/sportran/tree/wishart and it will be part of a future release of the software.

APPENDIX A: MACHINE LEARNING INTERATOMIC POTENTIAL

We enlarge an existing dataset[52,53] for Li_3PS_4 developed in the context of Gaussian approximation potentials[72] to the large-data regime required for neural network MLIPs. The dataset augmentation is performed by means of the active learning scheme implemented in DP-GEN v0.12.0,[73] and it involves progressive inclusion of new structures sampled according to some criterion from MD simulations propagated by a committee of MLIPs trained on the same dataset with different randomly initialized weights. The criterion for inclusion in the dataset is the lack of relative agreement between the members of the committee, measured via the root mean squared deviation (RMSD) on the predicted forces. If this metric is above some system-dependent threshold, the configuration is considered out of sample and, thus, added to the dataset for the next iteration of the active learning procedure. Configurations where the RMSD of the forces predicted by the committee of MLIPs is below the threshold are already appropriately represented in the current version of the dataset and, therefore, are not added for the next iteration. We apply the scheme to reach around 10 000 structures of Li_3PS_4 in its different phases, with temperatures ranging from 300 to ~1000 K. The density functional theory calculations to compute total energy and forces for the structures in the resulting dataset are performed with QUANTUM ESPRESSO v7.2[74–77] at the PBEsol level.[78] A version 4 neuroevolution potential (NEP)[79] is then trained to reproduce total energies and forces for this dataset using GPUMD v3.9.4.[80] Specific details on calculation parameters can be found in the respective scripts in the Materials Cloud repository associated with this article.

APPENDIX B: SIMULATION DETAILS

MD simulations of Li_3PS_4 are performed with GPUMD v3.9.4.[80] All simulations are carried out in the canonical ensemble by means of the Bussi–Donadio–Parrinello thermostat.[81] The atomic

positions are initialized in the respective structural phase, α, β or γ, with an initial volume estimated by an equilibration run at the given temperature and ambient pressure in the NpT ensemble with a fully flexible simulation cell. Energy and mass fluxes are sampled every 0.5 fs, which is also the time step of Verlet integration.

REFERENCES

[1] L. P. Kadanoff and P. C. Martin, "Hydrodynamic equations and correlation functions," Ann. Phys. **24**, 419–469 (1963).

[2] S. Baroni, R. Bertossa, L. Ercole, F. Grasselli, and A. Marcolongo, "Heat transport in insulators from ab initio Green-Kubo theory," in *Handbook of Materials Modeling: Applications: Current and Emerging Materials* (Springer International Publishing, 2020), pp. 809–844.

[3] A. Einstein, "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen," Ann. Phys. **322**, 549–560 (1905).

[4] A. Rahman, "Correlations in the motion of atoms in liquid argon," Phys. Rev. **136**, A405–A411 (1964).

[5] L. Onsager, "Reciprocal relations in irreversible processes. I," Phys. Rev. **37**, 405–426 (1931); "Reciprocal relations in irreversible processes. II," **38**, 2265–2279 (1931).

[6] M. S. Green, "Markoff random processes and the statistical mechanics of time-dependent phenomena," J. Chem. Phys. **20**, 1281–1295 (1952); "Markoff random processes and the statistical mechanics of time-dependent phenomena. II. Irreversible processes in fluids," **22**, 398–413 (1954).

[7] R. Kubo, "Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems," J. Phys. Soc. Jpn. **12**, 570–586 (1957); R. Kubo, M. Yokota, and S. Nakajima, "Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance," *ibid.* **12**, 1203–1211 (1957).

[8] E. Helfand, "Transport coefficients from dissipation in a canonical ensemble," Phys. Rev. **119**, 1–9 (1960).

[9] C. Bruin, "Transport coefficients from molecular dynamics for a very small system," Phys. Lett. A **28**, 777–778 (1969).

[10] D. Levesque, L. Verlet, and J. Kürkijarvi, "Computer 'experiments' on classical fluids. IV. Transport Properties and time-correlation functions of the Lennard-Jones liquid near its triple point," Phys. Rev. A **7**, 1690–1700 (1973).

[11] S. Stackhouse, L. Stixrude, and B. B. Karki, "Thermal conductivity of periclase (MgO) from first principles," Phys. Rev. Lett. **104**, 208501 (2010).

[12] A. Marcolongo, "Theory and ab initio simulation of atomic heat transport," Ph.D. thesis, SISSA, 2014.

[13] A. Marcolongo, P. Umari, and S. Baroni, "Microscopic theory and quantum simulation of atomic heat transport," Nat. Phys. **12**, 80–84 (2016).

[14] L. Ercole, A. Marcolongo, P. Umari, and S. Baroni, "Gauge invariance of thermal transport coefficients," J. Low Temp. Phys. **185**, 79–86 (2016).

[15] F. Grasselli and S. Baroni, "Invariance principles in the theory and computation of transport coefficients," Eur. Phys. J. B **94**, 160 (2021).

[16] F. Grasselli and S. Baroni, "Topological quantization and gauge invariance of charge transport in liquid insulators," Nat. Phys. **15**, 967–972 (2019).

[17] L. Ercole, A. Marcolongo, and S. Baroni, "Accurate thermal conductivities from optimally short molecular dynamics simulations," Sci. Rep. **7**, 15835 (2017).

[18] M. P. Allen and D. J. Tildesley, "Nonequilibrium molecular dynamics," in *Computer Simulation of Liquids* (Oxford University Press, 2017).

[19] P. K. Schelling, S. R. Phillpot, and P. Keblinski, "Comparison of atomic-level simulation methods for computing thermal conductivity," Phys. Rev. B **65**, 144306 (2002).

[20] H. Dong, Z. Fan, L. Shi, A. Harju, and T. Ala-Nissila, "Equivalence of the equilibrium and the nonequilibrium molecular dynamics methods for thermal conductivity calculations: From bulk to nanowire silicon," Phys. Rev. B **97**, 094305 (2018).

[21] M. Puligheddu and G. Galli, "Atomistic simulations of the thermal conductivity of liquids," Phys. Rev. Mater. **4**, 053801 (2020).

[22] D. Nevins and F. J. Spera, "Accurate computation of shear viscosity from equilibrium molecular dynamics simulations," Mol. Simul. **33**, 1261–1266 (2007).

[23] R. E. Jones and K. K. Mandadapu, "Adaptive Green-Kubo estimates of transport coefficients from molecular dynamics based on robust error analysis," J. Chem. Phys. **136**, 154102 (2012).

[24] L. d. S. Oliveira and P. A. Greaney, "Method to manage integration error in the Green-Kubo method," Phys. Rev. E **95**, 023308 (2017).

[25] Y. Zhang, A. Otani, and E. J. Maginn, "Reliable viscosity calculation from equilibrium molecular dynamics simulations: A time decomposition method," J. Chem. Theory Comput. **11**, 3537–3546 (2015).

[26] R. Bertossa, F. Grasselli, L. Ercole, and S. Baroni, "Theory and numerical simulation of heat transport in multicomponent systems," Phys. Rev. Lett. **122**, 255901 (2019).

[27] E. Drigo, S. Baroni, and P. Pegolo, "Seebeck coefficient of ionic conductors from Bayesian regression analysis," J. Chem. Theory Comput. **20**, 6152–6159 (2024).

[28] B. P. Bogert, "The quefrency alanysis of time series for echoes: Cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking," in *Proceedings of the Symposium on Time Series Analysis* (Wiley, New York, 1963), pp. 209–243.

[29] P. Pegolo, S. Baroni, and F. Grasselli, "Temperature- and vacancy-concentration-dependence of heat transport in Li_3ClO from multi-method numerical simulations," npj Comput. Mater. **8**, 24 (2022).

[30] L. Isaeva, G. Barbalinardo, D. Donadio, and S. Baroni, "Modeling heat transport in crystals and glasses from a unified lattice-dynamical approach," Nat. Commun. **10**, 3853 (2019).

[31] M. Simoncelli, N. Marzari, and F. Mauri, "Unified theory of thermal transport in crystals and glasses," Nat. Phys. **15**, 809–813 (2019).

[32] A. Fiorentino, E. Drigo, S. Baroni, and P. Pegolo, "Unearthing the foundational role of anharmonicity in heat transport in glasses," Phys. Rev. B **109**, 224202 (2024).

[33] N. Galamba, C. A. Nieto de Castro, and J. F. Ely, "Equilibrium and nonequilibrium molecular dynamics simulations of the thermal conductivity of molten alkali halides," J. Chem. Phys. **126**, 204511 (2007).

[34] N. Ohtori, M. Salanne, and P. A. Madden, "Calculations of the thermal conductivities of ionic materials by simulation with polarizable interaction potentials," J. Chem. Phys. **130**, 104507 (2009).

[35] M. Salanne, D. Marrocchelli, C. Merlet, N. Ohtori, and P. A. Madden, "Thermal conductivity of ionic systems from equilibrium molecular dynamics," J. Phys.: Condens. Matter **23**, 102101 (2011).

[36] S. Bonella, M. Ferrario, and G. Ciccotti, "Thermal diffusion in binary mixtures: Transient behavior and transport coefficients from equilibrium and nonequilibrium molecular dynamics," Langmuir **33**, 11281–11290 (2017).

[37] F. Knoop, M. Scheffler, and C. Carbogno, "*Ab initio* Green-Kubo simulations of heat transport in solids: Method and implementation," Phys. Rev. B **107**, 224304 (2023).

[38] N. Wiener, "Generalized harmonic analysis," Acta Math. **55**, 117–258 (1930).

[39] A. Khintchine, "Korrelationstheorie der stationären stochastischen prozesse," Math. Ann. **109**, 604–615 (1934).

[40] P. C. Martin, "Sum rules, Kramers-Kronig relations, and transport coefficients in charged systems," Phys. Rev. **161**, 143–155 (1967).

[41] M.-F. Bru, "Wishart processes," J. Theor. Probab. **4**, 725–751 (1991).

[42] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population," Biometrika **20A**, 32–52 (1928).

[43] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," Ann. Math. Stat. **35**, 475–501 (1964).

[44] K. Pearson, G. B. Jeffery, and E. M. Elderton, "On the distribution of the first product moment-coefficient, in samples drawn from an indefinitely large normal population," Biometrika **21**, 164–193 (1929).

[45] S. D. Peddada and D. S. P. Richards, "Proof of a conjecture of M. L. Eaton on the characteristic function of the Wishart distribution," Ann. Probab. **19**, 868–874 (1991).

[46] Here, ℓ is the number of independent Gaussian samples used to compute the periodogram. In the simple case of a single MD run of an isotropic material, where there are three independent Cartesian components, using discrete cosine transforms implies $\ell = 3$; using full Fourier transforms, $\ell = 6$, where the additional factor of two accounts for real and imaginary parts of the flux in frequency domain.[82]

[47] P. Pegolo, "Charge and heat transport in ionic conductors," Ph.D. thesis, SISSA, 2023.
[48] A. Marcolongo, L. Ercole, and S. Baroni, "Gauge fixing for heat-transport simulations," J. Chem. Theory Comput. **16**, 3352–3362 (2020).
[49] V. Rozsa and G. Galli, "Solvation of simple ions in water at extreme conditions," J. Chem. Phys. **154**, 144501 (2021).
[50] D. Tisi, L. Zhang, R. Bertossa, H. Wang, R. Car, and S. Baroni, "Heat transport in liquid water from first-principles and deep neural network simulations," Phys. Rev. B **104**, 224202 (2021).
[51] P. Pegolo, S. Baroni, and F. Grasselli, "Self-interaction and transport of solvated electrons in molten salts," J. Chem. Phys. **159**, 094116 (2023).
[52] L. Gigli, D. Tisi, F. Grasselli, and M. Ceriotti, "Mechanism of charge transport in lithium thiophosphate," Chem. Mater. **36**, 1482–1496 (2024).
[53] D. Tisi, F. Grasselli, L. Gigli, and M. Ceriotti, "Thermal conductivity of Li_3PS_4 solid electrolytes with *ab initio* accuracy," Phys. Rev. Mater. **8**, 065403 (2024).
[54] C. Malosso, L. Zhang, R. Car, S. Baroni, and D. Tisi, "Viscosity in water from first-principles and deep-neural-network simulations," npj Comput. Mater. **8**, 139 (2022).
[55] E. W. Weisstein, Gamma Function. From MathWorld–A Wolfram Web Resource, https://mathworld.wolfram.com/GammaFunction.html.
[56] E. W. Weisstein, Bessel Function of the Second Kind. From MathWorld–A Wolfram Web Resource, https://mathworld.wolfram.com/BesselFunctionoftheSecondKind.html.
[57] J. Nash, *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation* (Routledge, 1990), pp. 84–93.
[58] A. W. van der Vaart, *Asymptotic Statistics*, *Cambridge Series in Statistical and Probabilistic Mathematics* (Cambridge University Press, 1998).
[59] E. W. Weisstein, Moving Average From MathWorld–A Wolfram Web Resource, https://mathworld.wolfram.com/MovingAverage.html.
[60] J. B. W. Webber, "A bi-symmetric log transformation for wide-range data," Meas. Sci. Technol. **24**, 027001 (2012).
[61] P. Pegolo, F. Grasselli, and S. Baroni, "Oxidation states, Thouless' pumps, and nontrivial ionic transport in nonstoichiometric electrolytes," Phys. Rev. X **10**, 041031 (2020).
[62] R. Resta, "Faraday law, oxidation numbers, and ionic conductivity: The role of topology," J. Chem. Phys. **155**, 244503 (2021).
[63] P. Pegolo, S. Baroni, and F. Grasselli, "Topology, oxidation states, and charge transport in ionic conductors," Ann. Phys. **534**, 2200123 (2022).
[64] P. G. Debenedetti, "The statistical mechanical theory of concentration fluctuations in mixtures," J. Chem. Phys. **87**, 1256–1260 (1987).
[65] P. G. Debenedetti, "Fluctuation-based computer calculation of partial molar properties. I. Molecular dynamics simulation of constant volume fluctuations," J. Chem. Phys. **86**, 7126–7137 (1987).
[66] P. G. Debenedetti, "Fluctuation-based computer calculation of partial molar properties. II. A numerically accurate method for the determination of partial molar energies and enthalpies," J. Chem. Phys. **88**, 2681–2684 (1988).
[67] F. G. Fumi and M. P. Tosi, "Ionic sizes and born repulsive parameters in the NaCl-type alkali halides—I: The Huggins-Mayer and Pauling forms," J. Phys. Chem. Solids **25**, 31–43 (1964).
[68] M. P. Tosi and F. G. Fumi, "Ionic sizes and born repulsive parameters in the NaCl-type alkali halides—II: The generalized Huggins-Mayer form," J. Phys. Chem. Solids **25**, 45–52 (1964).
[69] J. Wang, Z. Sun, G. Lu, and J. Yu, "Molecular dynamics simulations of the local structures and transport coefficients of molten alkali chlorides," J. Phys. Chem. B **118**, 10196–10206 (2014).
[70] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics* (Springer Science & Business Media, 1999), pp. 90–92.
[71] L. Talirz, S. Kumbhar, E. Passaro, A. V. Yakutovich, V. Granata, F. Gargiulo, M. Borelli, M. Uhrin, S. P. Huber, S. Zoupanos, C. S. Adorf, C. W. Andersen, O. Schütt, C. A. Pignedoli, D. Passerone, J. VandeVondele, T. C. Schulthess, B. Smit, G. Pizzi, and N. Marzari, "Materials cloud, a platform for open computational science," Sci. Data **7**, 299 (2020).
[72] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," Phys. Rev. Lett. **104**, 136403 (2010).
[73] Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang, and E. Weinan, "DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models," Comput. Phys. Commun. **253**, 107206 (2020).
[74] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, "Quantum ESPRESSO: A modular and open-source software project for quantum simulations of materials," J. Phys.: Condens. Matter **21**, 395502 (2009).
[75] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni *et al.*, "Advanced capabilities for materials modelling with Quantum ESPRESSO," J. Phys.: Condens. Matter **29**, 465901 (2017).
[76] P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. De Gironcoli, P. Delugas, F. Ferrari Ruffino *et al.*, "Quantum ESPRESSO toward the exascale," J. Chem. Phys. **152**, 154105 (2020).
[77] I. Carnimeo, F. Affinito, S. Baroni, O. Baseggio, L. Bellentani, R. Bertossa, P. D. Delugas, F. F. Ruffino, S. Orlandini, F. Spiga, and P. Giannozzi, "Quantum ESPRESSO: One further step toward the exascale," J. Chem. Theory Comput. **19**, 6992–7006 (2023).
[78] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, "Restoring the density-gradient expansion for exchange in solids and surfaces," Phys. Rev. Lett. **100**, 136406 (2008).
[79] Z. Fan, Z. Zeng, C. Zhang, Y. Wang, K. Song, H. Dong, Y. Chen, and T. Ala-Nissila, "Neuroevolution machine learning potentials: Combining high accuracy and low cost in atomistic simulations and application to heat transport," Phys. Rev. B **104**, 104309 (2021).
[80] Z. Fan, Y. Wang, P. Ying, K. Song, J. Wang, Y. Wang, Z. Zeng, K. Xu, E. Lindgren, J. M. Rahm *et al.*, "GPUMD: A package for constructing accurate machine-learned potentials and performing highly efficient atomistic simulations," J. Chem. Phys. **157**, 114801 (2022).
[81] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," J. Chem. Phys. **126**, 014101 (2007).
[82] L. Ercole, R. Bertossa, S. Bisacchi, and S. Baroni, "SporTran: A code to estimate transport coefficients from the cepstral analysis of (multivariate) current time series," Comput. Phys. Commun. **280**, 108470 (2022).