26/12/2024 03:41

(Article begins on next page)

Theory and numerical simulation of heat transport in multi-component systems

Riccardo Bertossa,[1] Federico Grasselli,[1] Loris Ercole,[1, *] and Stefano Baroni[1, 2, †]

[1] *SISSA – Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, 34136 Trieste, Italy*
[2] *CNR – Istituto Officina dei Materiali, SISSA, 34136 Trieste*

The thermal conductivity of classical multi-component fluids is seemingly affected by the intrinsic arbitrariness in the definition of the atomic energies and it is ill-conditioned numerically, when evaluated from the Green-Kubo theory of linear response. To cope with these two problems we introduce two new concepts: a *convective invariance* principle for transport coefficients, in the first case, and *multi-variate cepstral analysis*, in the second. A combination of these two concepts allows one to substantially reduce the noise affecting the estimate of the thermal conductivity from equilibrium molecular dynamics, even for one-component systems.

The transport properties of macroscopic systems are determined by the dynamics of *conserved currents*, *i.e.* of the long-wavelength components of the currents associated to the densities of conserved extensive variables [1, 2]. Let J^i be the macroscopic average of the i-th conserved current, which from now on we dub a (conserved) *flux*. In the case of heat transport in an M-component fluid, the relevant conserved quantities are the total energy and the total number (or mass) of molecules of each independent component. As the total-mass flux is the total momentum, which is also a constant of motion, the number of relevant conserved fluxes is reduced from $M+1$ to M: energy, which we label as the zero-th, and $M-1$ *convective* fluxes, which can be identified with *any* independent linear combinations of the molecular mass or number fluxes. In the linear regime, the relevant fluxes are linear combinations of the corresponding conjugate *affinities*, F^i, defined as the gradients of the intensive variables conjugate to the conserved ones. These are the inverse temperature for the energy and the chemical potential divided by the temperature for the molecular numbers. The resulting Onsager relations [3] read:

$$J^i = \sum_{j=0}^{M-1} \Lambda^{ij} F^j, \qquad (1)$$

where, in order to simplify the notation, the vector character of fluxes and affinities has been overlooked or their Cartesian indices are incorporated in the suffixes. The Green-Kubo (GK) theory of linear response [2, 4, 5] states that the Λ matrix in Eq. (1) can be expressed in terms of the time correlation functions of the various *flux processes*, $\mathcal{J}^i$, which are phase-space functions, as:

$$\Lambda^{ij} = \frac{V}{k_B} \int_0^\infty \left\langle \mathcal{J}^i(t)\mathcal{J}^j(0) \right\rangle dt, \qquad (2)$$

where k_B is the Boltzmann constant, V is the system's volume, $\langle \cdot \rangle$ indicates a canonical average, and where the system is in thermodynamic *equilibrium* [2]. Here and

in the following, processes and their samples will be denoted by by calligraphic letters, as in "$\mathcal{X}$", while their average properties by roman ones, as in "X". For instance, $J^i = \langle \mathcal{J}^i \rangle$, which vanishes in the absence of perturbations. The heat conductivity is defined as the ratio between the heat flux, J^Q, and the temperature gradient that generates it: $J^Q = -\kappa \nabla T$, in the absence of net particle flow. The heat flux is defined as $\mathcal{J}^Q = \mathcal{J}^0 - \sum_{s=1}^M h^s \mathcal{J}^s$, where $\mathcal{J}^0$ and $\mathcal{J}^s$ are the energy and particle-number flux samples of all molecular species, respectively, and h^s are the corresponding partial enthalpies [6]. The energy flux is defined in terms of atomic positions, $\mathbf{R}_n$, velocities, $\mathbf{V}_n$, energies ϵ_n as [2, 7]:

$$\mathcal{J}^0 = \frac{1}{V}\left[\sum_n \mathbf{V}_n \epsilon_n + \sum_{n,m}(\mathbf{R}_n - \mathbf{R}_m)\mathbf{F}_{nm} \cdot \mathbf{V}_n \right] \qquad (3)$$

where $\mathbf{F}_{nm} = -\partial\epsilon_m/\partial\mathbf{R}_n$ and n runs over all the atoms.

In solids and one-component fluids, energy is the only conserved quantity relevant to heat transport because the convective fluxes either vanish or do not contribute to energy transport [2]. In these cases, the thermal conductivity is basically given by $\kappa = \Lambda^{00}/T^2$. As the energy flux, Eq. (3), is obtained via the continuity equation from the energy density, which is ill-defined at the atomic scale, it has long been feared that no quantum-mechanical expression for the heat conductivity could be obtained from first principles. This apparent conundrum was solved only recently by the introduction of a *gauge invariance* principle for transport coefficients, according to which different energy densities integrating to the same total energy give rise to fluxes that differ one from the other by the total derivative of a bounded vector, which does not contribute to the value of Λ^{00} in Eq. (2) [2, 8]. The situation is not nearly as clear when one considers the intrinsic indeterminacy in the definition of the atomic energies. When the energies of all the atoms of a same species, say the s-th, are shifted by a same *self-energy*, δ^s, not depending on the atomic environment (*i.e.* $\epsilon_n \to \epsilon_n + \delta^{s(n)}$, where

n is an atomic index), it is to be expected that all the transport properties remain unchanged. For instance, in a quantum-mechanical simulation, the heat conductivity cannot depend on whether atomic cores contribute to the definition of the atomic energy, as they would in an all-electron calculation, or not, as they would when using pseudo-potentials. In the latter case, the energy of isolated atoms would depend on the specific form of pseudo-potential adopted, which is to a large extent arbitrary, but the heat conductivity in all cases should not. When the atomic energy of the s-th species is shifted as above, the energy flux is modified as: $\mathcal{J}^0 \rightarrow \mathcal{J}^0 + \sum_s \delta^s \mathcal{J}^s$. When only one atomic component is present, $\mathcal{J}^s$ is proportional to the total, conserved, momentum, which can be assumed to vanish identically, so that Λ^{00} is independent of atomic self-energies. A slight generalization of this argument allows one to arrive at the same conclusion in the one-component molecular case [2, 9].

In multi-component fluids the convective fluxes of individual species are not identically vanishing. As a consequence, the expressions for $\mathcal{J}^0$ and $\mathcal{J}^Q$ do not coincide and the Onsager coefficient Λ^{0i} are affected by the spurious atomic self-energies, as described before. The Green-Kubo integrals Λ^{Qi} (see Eq. (2)) computed from the heat flux, instead, would remain unaffected because the self-energy contributions to the partial enthalpies cancel their contribution to the energy flux. It seems therefore that spurious self-energy effects can be disposed of, at the price however of computing partial enthalpies, a rather cumbersome task [10], which had better be avoided. In practice, heat conductivities are usually measured in conditions of vanishing mass transport, where the average J^Q and J^0 do coincide, making one hope that the computation of partial quantities (energies or enthalpies, see Ref. 11) can indeed be bypassed. To see this more formally, we impose that the convective fluxes in Eq. (1) vanish, and solve for the energy flux. This is best achieved by partitioning the Λ matrix into a 1×1 energy block and an $(M-1) \times (M-1)$ convective block and by performing a block inversion. The resulting expression for the heat conductivity is:

$$\kappa = \frac{\bar{\Lambda}^0}{T^2}, \qquad (4)$$

where $\bar{\Lambda}^0 = 1/\left(\Lambda^{-1}\right)^{00} = \Lambda^{00} - \sum_{i,j=1}^{M-1} \Lambda^{0i} \left(\Lambda_{M-1}^{-1}\right)^{ij} \Lambda^{j0}$ is the inverse of the energy block of Λ^{-1}, that is to say the *Schur complement* of the convective block (SCCB) in Λ [12]. Using standard matrix manipulations and the bilinearity of Λ^{ij} with respect to the fluxes, it is straightforward to verify that the SCCB, and hence the heat conductivity, is invariant with respect to the addition of *any* linear combinations of convective fluxes to the energy flux: $\mathcal{J}^0 \rightarrow \mathcal{J}^0 + \sum_{m=1}^{M-1} c^m \mathcal{J}^m$, while the whole Λ matrix is not. We dub this remarkable property the *convective invariance* of heat conductivity in multi-component systems. An important consequence of convective invariance

is the independence of the heat conductivity on atomic self-energies, thus solving the first of our problems. Another important consequence is that the heat conductivities computed from the heat or energy fluxes coincide, thus dispensing us from the task of computing partial enthalpies to evaluate the former.

Having thus cleared the way, we now move to evaluating Eq. (4) from equilibrium molecular dynamics (EMD). In order to be specific and streamline the discussion, we specialize Eq. (4) to the two-component case: $\kappa = \left(\Lambda^{00} - \left(\Lambda^{10}\right)^2 \big/ \Lambda^{11}\right) \big/ T^2$. This expression is very sensitive to the statistical errors affecting the matrix elements appearing therein, because it is the difference of two positive numbers whose magnitude may be comparable, and because the errors affecting each of them may be large and difficult to estimate [13–16].

In order to cope with the latter problem, we have generalized to multivariate processes our *cepstral analysis* approach to evaluating transport coefficients from EMD [2, 17]. Cepstral analysis [18] is a technique, commonly used in signal analysis and speech recognition, to process the power spectrum of a time series, leveraging its smoothness and the statistical properties of its samples. In the one-component case, according to Eqs. (1-2) a transport coefficient is proportional to the zero-frequency value of the power spectrum of the appropriate flux: $\kappa \propto S(\omega = 0)$, where $S(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} C(t) dt$, and $C(t) = \langle \mathcal{J}(t)\mathcal{J}(0) \rangle$ is the flux time auto-correlation function. The Wiener-Kintchnine theorem [19] states that $S(\omega)$ is asymptotically proportional to the expectation of the squared modulus of the truncated Fourier transform of the flux sample: $S(\omega) = \lim_{\tau \to \infty} \langle \mathcal{S}_\tau(\omega) \rangle$, where $\mathcal{S}_\tau(\omega) = \frac{1}{\tau}|\tilde{\mathcal{J}}_\tau(\omega)|^2$ is the *sample spectrum* and $\tilde{\mathcal{J}}_\tau(\omega) = \int_0^\tau \mathcal{J}(t) e^{i\omega t} dt$. In the long-time limit, $\mathcal{S}_\tau(\omega)$ is a process whose values are independent for $\omega \neq \omega'$ and individually distributed as $\mathcal{S}_\tau(\omega) = S(\omega)\xi(\omega)$, where $\xi(\omega) \sim \frac{1}{2}\chi_2^2$, χ_2^2 being a chi-square variate with two degrees of freedom. The multiplicative nature of the noise affecting the sample spectrum suggests that the power of the noise can be reduced by applying a low-pass filter to its logarithm. In cepstral analysis this idea is leveraged to devise a consistent and asymptotically unbiased estimator for the the zero-frequency value of the flux power spectrum, which is proportional to the transport coefficient we are after [2, 17].

In the multi-component case, an EMD simulation samples M stationary stochastic processes, $\mathcal{J}^i$ ($i = 0, \cdots M-1$), one for each conserved flux, which can be thought of as different components of a same multivariate process. For such a process it is customary to define a cross time-correlation function, $C^{ij}(t) = \langle \mathcal{J}^i(t)\mathcal{J}^j(0) \rangle$, and a *cross power spectrum*, $S^{ij}(\omega) = \lim_{\tau \to \infty} \langle \mathcal{S}_\tau^{ij}(\omega) \rangle$, where $\mathcal{S}_\tau^{ij}(\omega) = \frac{1}{\tau}\tilde{\mathcal{J}}_\tau^i(\omega)^* \tilde{\mathcal{J}}_\tau^j(\omega)$ is the *sample cross-spectrum* of the multivariate process. The Onsager coefficients of Eq. (2) are proportional to the zero-frequency values of

the cross-spectrum, $S_0^{ij} = S^{ij}(\omega = 0)$: $\Lambda^{ij} = \frac{V}{2k_B} S_0^{ij}$. Eq. (4) shows that in order to evaluate the heat conductivity in the multi-component case one needs an efficient estimator for the SCCB in S_0. In analogy with Eq. (4), we indicate the SCCB in $S(\omega)$ and in $\mathcal{S}(\omega)$ as $\bar{S}^0(\omega)$ and $\bar{\mathcal{S}}^0(\omega)$, respectively, which in the following will be dubbed the *reduced (sample) spectrum*. Their zero-frequency values will be labeled as $\bar{S}_0^0$ and $\bar{\mathcal{S}}_0^0$, respectively. Mind the difference between the power spectrum, denoted by $S(\omega)$, and its sample, denoted by $\mathcal{S}(\omega)$.

In analogy to the uni-variate case, the reduced sample spectrum can be shown to be a process distributed like

$$\bar{\mathcal{S}}^0(\omega) \sim \bar{S}^0(\omega)\,\xi(\omega), \tag{5}$$

where the ξ's are independent identically distributed random variables, $\xi \sim \frac{1}{\nu}\chi_\nu^2$, χ_ν^2 being the chi-square distribution with $\nu = 2(\ell - M + 1)$ degrees of freedom, where M is the number of conserved fluxes and $\ell \geq M$ is the number of flux samples used to sample the spectrum. This means that the above formulas apply to an effective sample spectrum, defined as the average over ℓ independent samples of the spectrum. For instance, in an isotropic fluid one has one equivalent flux process per Cartesian component ($\ell = 3$); multiple flux samples can be created by either running multiple EMD trajectories for different initial conditions or, equivalently, by breaking a long trajectory into multiple segments. Eq. (5) shows that the reduced sample spectrum is an unbiased estimator of the reduced spectrum ($\langle \xi(\omega) \rangle = 1$). Unfortunately, this estimator is not consistent, in that its variance does not vanish when the length of the time series grows large, and is actually independent of it. In order to reduce the power of the noise and obtain a consistent estimator of the reduced spectrum, we apply a low-pass filter to its logarithm. To this end, one first performs a (inverse) Fourier transform of the logarithm of the reduced spectrum, and one retains a number of coefficients, P^*, equal to the estimated number of non-vanishing Fourier coefficients of the logarithm of the reduced spectrum. By doing so, the estimator of the heat conductivity we are after, $\mathcal{K}$, and its statistical uncertainty can be expressed as:

$$\mathcal{K} = \frac{V}{2k_B T^2}\exp\left[\mathcal{C}_0 + 2\sum_{n=1}^{P^*-1}\mathcal{C}_n - L_0\right]$$
$$\frac{\Delta\kappa}{\kappa} = \sigma_0\sqrt{\frac{4P^*-2}{N}}, \tag{6}$$

where the $\mathcal{C}$s are the (inverse) Fourier coefficients of the logarithm of reduced sample spectrum, $L_0 = \langle \log(\xi) \rangle = \psi(\ell - M + 1) - \log(\ell - M + 1)$, and $\sigma_0^2 = \langle \log(\xi)^2 \rangle - L_0^2 = \psi'(\ell - M + 1)$, ψ and ψ' being the di- and tri-gamma functions [20], respectively. A thorough derivation of the above formulas as well as a detailed description of the workflow for the analysis of the data produced by EMD
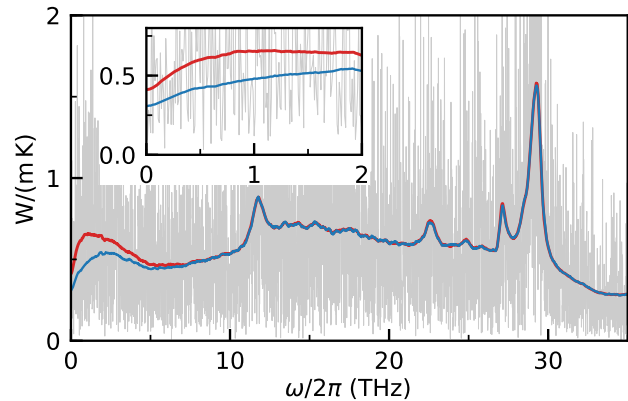

FIG. 1. Energy flux of a water-ethanol mixture. Gray: energy-flux sample power spectrum from a 100 ps trajectory. Red and blue: moving averages over a window of 0.2 THz of the energy-flux power sample spectrum and of the SCCB of the sample cross-spectrum, computed from a long (28 ns) trajectory. Inset: low-frequency region of the spectrum.

simulations are presented in the Supplemental Material (SM) [21].

In order to validate our methodology, we have computed the thermal conductivity of an equimolar water-ethanol mixture. Not aiming at an optimal description of the system, but just at a realistic benchmark, we used the simple OPLSAA [22] all-atom flexible force field. Classical EMD simulations were run with the LAMMPS package [23] at a temperature of ≈ 350 K and a density of 0.80 g/cm^3, corresponding to $800 + 800$ molecules in a cubic cell with an edge of 47.47 Å.

In Fig. 1 we report the energy-flux spectrum of our water-ethanol model. The gray line indicates the diagonal element of the raw sample spectrum, computed from a 100 ps trajectory, which is too noisy to be used as an estimator. Performing a moving average [24] of the sample spectrum would consistently reduce the noise, at the price however of requiring much longer trajectories (the blue line reports an average performed from a 28 ns trajectory) and introducing a bias that is difficult to evaluate and remove. Note that while at high frequency, where the spectrum is dominated by intra-molecular vibrations, the energy-energy diagonal element of the cross-spectrum (red line) and the SCCB coincide, in the low-frequency region, which is characterized by a strongly diffusive behavior, the two differ markedly and only the latter is meaningful to estimate κ. In Fig. 2 we report the thermal conductivity of the water-ethanol solution computed from Eq. (4) as a function of the upper limit of integration in Eq. (2), and as obtained from a 100 ps EMD trajectory. The spikes in the estimated conductivity result from the vanishing of the Λ^{11} denominator in Eq. (4), which is in turn due to the fact that the integrals of the correlation functions in Eq. (2) behave as random walks as soon as the integrand vanishes, eventually assuming any value.
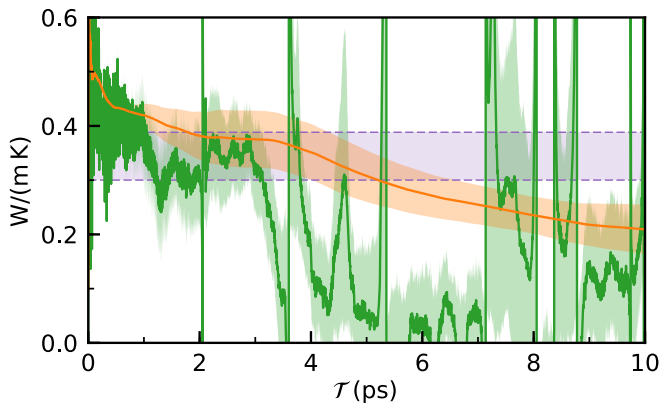
FIG. 2. Thermal conductivity of a water-ethanol solution, Eq. (4), as a function of the upper limit of integration in Eq. (2). The shaded areas indicate the estimated statistical error. Green: direct estimate from from Eqs. (4) and (2). Orange: estimate from the Einstein-Helfand relation, see main text. Purple: cepstral analysis estimate.
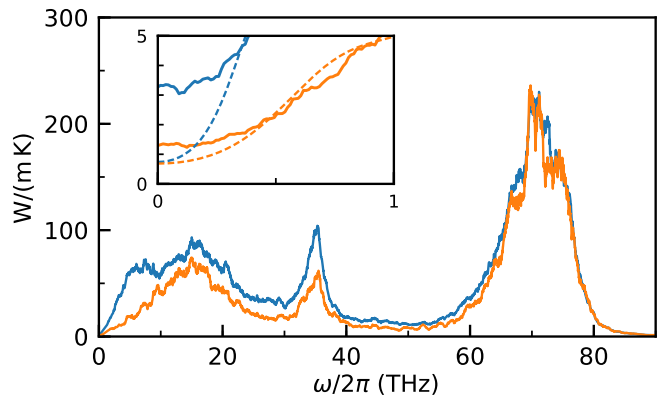

FIG. 3. Multi-variate analysis of *ab initio* water. The sample power spectrum of the DFT energy flux would be out-of-scale and numerically intractable. Blue: SCCB of a two-component analysis performed with the total momentum of the oxygen atoms as an inert flux. Orange: three-component analysis, where the third flux is the electronic adiabatic current. Both are filtered with a moving average of width 1 THz. In the inset, the spectrum obtained from cepstral analysis is displayed in dashed lines.

This behavior can be partially corrected by replacing the GK estimate of the Onsager coefficients through Eq. (2) with an equivalent one, based on the Einstein-Helfand relation [2, 25], which is statistically better behaved (orange line): $\Lambda^{ij} = \lim_{\tau \to \infty} \frac{V}{2k_B\tau} \left\langle \int_0^\tau \mathcal{J}^i(t)dt \int_0^\tau \mathcal{J}^j(t)dt \right\rangle$. Even so, the statistical accuracy that can be achieved with even longer trajectories is totally inadequate.

Cepstral analysis was performed over a 100 ps EMD trajectory using the `ThermoCepstrum` open-source code, which is freely available for download [26]. We set $\ell = 3$ (Cartesian components) and $M = 2$ (number of conserved fluxes), thus obtaining $L_0 = \psi(2) - \log(2) \approx -0.270$ and $\sigma_0^2 = \psi'(2) \approx 0.644$. The cutoff frequency [17] used for cepstral analysis, and ensuring convergence in κ, was $\omega^*/2\pi \approx 35$ THz. The number of cepstral coefficients $P^* \approx 45$ was estimated from the Akaike Information Criterion [17, 27]. The final estimate of the heat conductivity resulting from Eqs. (6) is $\kappa = 0.34 \pm 0.04$ W/(m K). In order to validate our statistical analysis, cepstral analysis was repeated for all the 100 ps extracted from a 28 ns long trajectory, confirming the normal distribution of the estimated conductivity and the value of the relative error. The small relative error (13%) achieved by analyzing trajectories as short as 100 ps shows that cepstral analysis opens the way to heat-transport simulations using *ab initio* EMD even in multi-component systems. Our analysis equally applies to systems with any number of components in any charge state, because the SCCB is invariant under any nonsingular linear trasformation of the convective fluxes. This implies that in binary molten salts and ionic fluids the electric current, which is a linear combination of mass/number fluxes, can be taken as a proxy of the convective fluxes. An application to molten sodium chloride is presented in the SM [21].

Multivariate cepstral analysis turns out to be instrumental in heat-transport simulations even for one-component systems. The recently discovered *gauge invariance* of heat conductivity [8] means that, while transport coefficients are largely independent of the detailed form of the energy flux, the flux power spectrum and the resulting statistical properties of the estimator do depend on it. The question then naturally arises of how to choose the form of the flux so as to optimize these statistical properties. In the case of *ab initio* water [8], for instance, the total power of the energy flux derived from density-functional theory (DFT) is so large as to make the statistical analysis of its spectrum at low-frequency intractable, due to large atomic self-energies [9] that cannot be easily defined and eliminated in *ab initio* simulations. A couple of *ad hoc* solutions to this problem have been devised [8, 9], leveraging gauge invariance to subtract from the energy flux a linear combination of "inert" fluxes not contributing to heat transport (such as the electronic flux or the mass flux of one of the two atomic species). While this remedy was effective and actually allowed one to get meaningful results from relatively short trajectories, the question remains of how to optimally choose this linear combination. Multivariate cepstral analysis provides the answer. The idea is to compute, along with the energy flux to be analyzed, a number of inert fluxes and treat all of them on a par as though they were conserved fluxes of a multi-component system: the total power of the SCCB will be considerably reduced, thus making the cepstral analysis feasible at low frequency. This is illustrated in Fig. 3, where we report the multi-variate power spectrum of *ab initio* (heavy) water, computed using the total momentum of the oxygen

atoms as an inert flux (blue), as well as this and the adiabatic electronic current as inert fluxes (orange). The simulation setting is the same as in Ref. 8. The spectrum of the bivariate SCCB is always larger than the trivariate one. Using cepstral analysis, the zero-frequency limits of these two spectra coincide, as they must, whereas their running averages do not and depend sensitively on the width of the frequency window. The heat conductivities estimated by cepstral analysis are $(0.74 \pm 0.16) \, \mathrm{W/(m\,K)}$ and $(0.69 \pm 0.18) \, \mathrm{W/(m\,K)}$, respectively, while the experimental value is $0.61 \, \mathrm{W/(m\,K)}$ [28].

We conclude by noticing that the combination of the newly devised convective invariance and multivariate cepstral analysis, besides providing fresh theoretical insight in transport phenomena, will hopefully broaden the scope of heat-transport simulations to complex multi-component fluids, as well as provide new tools to make *ab initio* simulations possible.

---

* Present address: Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

† baroni@sissa.it

[1] L. P. Kadanoff and P. C. Martin, Ann. Phys. **24**, 419 (1963); D. Foster, *Hydrodynamic fluctuations, broken symmetry, and correlation functions* (Benjamin, 1975).

[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*, edited by W. Andreoni and S. Yip (Springer International Publishing, Cham, 2018) pp. 1–36, 2nd ed., arXiv:1802.08006 [cond-mat.stat-mech].

[3] L. Onsager, Phys. Rev. **37**, 405 (1931); Phys. Rev. **38**, 2265 (1931).

[4] M. S. Green, J. Chem. Phys. **20**, 1281 (1952); J. Chem. Phys. **22**, 398 (1954).

[5] R. Kubo, J. Phys. Soc. Jpn. **12**, 570 (1957); R. Kubo, M. Yokota, and S. Nakajima, J. Phys. Soc. Jpn. **12**, 1203 (1957).

[6] S. R. D. Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (Dover Publications, 1984).

[7] J. H. Irving and J. G. Kirkwood, J. Chem. Phys. **18**, 817 (1950).

[8] A. Marcolongo, P. Umari, and S. Baroni, Nature Phys. **12**, 80 (2016); L. Ercole, A. Marcolongo, P. Umari, and S. Baroni, J. Low Temp. Phys. **185**, 79 (2016).

[9] A. Marcolongo, L. Ercole, and S. Baroni, in preparation.

[10] P. G. Debenedetti, J. Chem. Phys. **86**, 7126 (1987); R. V. Abteilung and C., Phys. Rev. A. **35**, 3487 (1987); P. Sindzingre, C. Massobrio, and G. Ciccotti, Chem Phys. **129**, 213 (1989).

[11] R. H. Trimble and J. M. Deutch, J. Stat. Phys. **3**, 149 (1971).

[12] D. E. Crabtree and E. V. Haynsworth, Proceedings of the American Mathematical Society **22**, 364 (1969).

[13] N. Galamba, C. a. Nieto de Castro, and J. F. Ely, J. Chem. Phys. **126**, 204511 (2007).

[14] N. Ohtori, M. Salanne, and P. A. Madden, J. Chem. Phys. **130**, 104507 (2009).

[15] M. Salanne, D. Marrocchelli, C. Merlet, N. Ohtori, and P. A. Madden, J. Phys. Condens. Matter **23**, 102101 (2011).

[16] S. Bonella, M. Ferrario, and G. Ciccotti, Langmuir **33**, 11281 (2017).

[17] L. Ercole, A. Marcolongo, and S. Baroni, Sci. Rep. **7**, 15835 (2017).

[18] B. P. Bogert, J. R. Healy, and J. W. Tukey, in *Proceedings of the Symposium on Time Series Analysis* (1963) pp. 209–243; D. G. Childers, D. P. Skinner, and R. C. Kemerait, Proceedings of the IEEE **65**, 1428 (1977).

[19] N. Wiener, Acta Math. **55**, 117 (1930); A. Khintchine, Math. Ann. **109**, 604 (1934).

[20] E. W. Weisstein, "Polygamma functions," (), from MathWorld – A Wolfram Web Resource http://mathworld.wolfram.com/PolygammaFunction.html.

[21] See Supplemental Material at [URL will be inserted by publisher] for *i)* a full derivation of the formulas on multivariate cepstral analysis; *ii)* the application of the latter to a simulation of a molten salt (NaCl); *iii)* a complete workflow on cepstral analysis of EMD data.

[22] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[23] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).

[24] E. W. Weisstein, "Moving average," (), from MathWorld – a Wolfram Web Resource http://mathworld.wolfram.com/MovingAverage.html.

[25] E. Helfand, Phys. Rev. **119**, 1 (1960).

[26] L. Ercole and R. Bertossa, "`ThermoCepstrum`: a code to estimate transport coefficients from the cepstral analysis of a multi-variate current stationary time series," https://github.com/lorisercole/thermocepstrum (2017–2018).

[27] H. Akaike, *Information theory and an extension of the maximum likelihood principle, in 2nd International Symposium on Information Theory* (edited by B. N. Petrov and F. Cski, 1972) pp. 267–281; IEEE Trans. Autom. Control **19**, 716 (1974).

[28] N. Matsunaga and A. Nagashima, Journal of Physical and Chemical Reference Data **12**, 933 (1983).

Supplemental Material to
"Theory and numerical simulation of heat transport in multi-component systems"

Riccardo Bertossa,[1] Federico Grasselli,[1] Loris Ercole,[1] and Stefano Baroni[1, 2, *]

[1] SISSA – Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, 34136 Trieste, Italy
[2] CNR – Istituto Officina dei Materiali, SISSA, 34136 Trieste
(Dated: May 16, 2019)

S1 – MULTIVARIATE CEPSTRAL ANALYSIS

Practical numerical equilibrium molecular dynamics (EMD) simulations provide access to *discrete* samples of several flux process, *i.e.* to finite multivariate *time series.* Let us then suppose to have ℓ independent such samples, of length N: $\{^p\mathcal{J}_n^i\}$ ($p = 1, \ldots \ell$; $i = 0, \ldots M-1$; $n = 0, \ldots N-1$). For instance, the three Cartesian components of a current in an isotropic medium may be thought of as different samples of a same process, and so would different segments of a same long time series. Stationarity implies that $\langle \mathcal{J}_n^i \rangle$ does not depend on n and that $\langle \mathcal{J}_n^i \mathcal{J}_m^j \rangle$ only depends on $n-m$. We further assume that $\langle \mathcal{J}_n^i \rangle = 0$ and that $\langle \mathcal{J}_n^i \mathcal{J}_0^j \rangle$ is an even function of n, which is the case when $\mathcal{J}^i$ and $\mathcal{J}^j$ have the same signature under time-reversal. We define the *cross-periodogram* as the average discrete-frequency spectrum of the ℓ samples:

$$\mathcal{S}_k^{ij} = \frac{1}{\ell} \sum_{p=1}^{\ell} \frac{\epsilon}{N} \left(^p\tilde{\mathcal{J}}_k^i \right)^* \, ^p\tilde{\mathcal{J}}_k^j, \tag{S1}$$

where

$$^p\tilde{\mathcal{J}}_k^i = \sum_{n=0}^{N-1} \, ^p\mathcal{J}_n^i \mathrm{e}^{2\pi i \frac{kn}{N}} \tag{S2}$$

is the discrete Fourier transform of the p-th time series and ϵ its sampling time. The Wiener-Khintchine theorem [1] states that $\mathcal{S}_k^{ij}$ is an (asymptotically) unbiased estimator of the cross-spectrum $S^{ij}(\omega) = \int_{-\infty}^{+\infty} \mathrm{e}^{i\omega t} \langle \mathcal{J}^i(t)\mathcal{J}^j(0)\rangle dt$, *i.e.* $S^{ij} = \lim_{N\to\infty}\langle \mathcal{S}^{ij}\rangle$. In the large-$N$ limit the real and imaginary parts of $\tilde{\mathcal{J}}_k^i$ are normal deviates that are uncorrelated amongst themselves and for $k \neq k'$. Therefore, for each discrete frequency k, the cross-periodogram is a random matrix distributed as a complex Wishart deviate [2]: $\mathcal{S}_k \sim \mathcal{CW}_M\big(S(\omega_k), \ell\big)$, where, in the literature on time-series analysis, the variance matrix $S(\omega_k)$ is called the cross power spectrum. We have the following notable property of Wishart deviates: *the Schur complement of each of the diagonal blocks of a Wishart deviate is also a Wishart deviate* [3]. More explicitly, let $\mathcal{S} \sim \mathcal{CW}_M\big(S, \ell\big)$ be a Wishart matrix of rank M and $\bar{\mathcal{S}}$ the Schur complement of rank L of any of its diagonal blocks of rank $M-L$; one then has:

$$\bar{\mathcal{S}} \sim \mathcal{CW}_L\big(\bar{S}, \ell-M+L\big). \tag{S3}$$

The Schur complement of the convective-current block (SCCB) of the cross-periodogram, $\bar{\mathcal{S}}^0$, (dubbed the *reduced periodogram*) has rank $L = 1$. Therefore, since the Wishart distribution for rank-one matrices is a χ^2 distribution, we conclude that, for each frequency k,

$$\bar{\mathcal{S}}_k^0 \sim \bar{S}_k^0 \xi_k, \tag{S4}$$

where the ξ's are independent identically distributed (iid) random variables, $\xi \sim \frac{1}{\nu}\chi_\nu^2$, χ_ν^2 being the chi-square distribution with $\nu = 2(\ell - M + 1)$ degrees of freedom. Strictly speaking, this distribution holds for $k \notin \{0, \frac{N}{2}\}$ because for these two frequencies the Fourier components of the flux time series are real, and $\nu = \ell - M + 1$. For the purposes of the following discussion, we overlook this exception, thus making an overall error $\sim \mathcal{O}(1/N)$, which vanishes in the $N \to \infty$ limit that is assumed throughout our statistical analysis. Moreover for clarity we included a factor ℓ/ν in the SCCB, that is needed to obtain the correct mean value, see Ref. 4.

Equation (S4) shows that the SCCB of the cross-periodogram (that we dub *reduced* periodogram) is an unbiased estimator of the SCCB of the cross-spectrum (the *reduced* spectrum), which is proportional to κ. Unfortunately, this estimator is not consistent, in that its variance does not vanish as $N \to \infty$ (actually, it is independent of N). A consistent estimator is obtained by applying a low-pass filter to the *logarithm* of the reduced periodogram, thus introducing a bias, which however can be made arbitrarily small as $N \to \infty$. Eq. (S4) shows that

$$\log\big(\bar{\mathcal{S}}_k^0\big) - \log\big(\bar{S}_k^0\big) \sim \log(\xi) \tag{S5}$$

are iid stochastic variables whose expectation and variance are: $L_0 = \langle \log(\xi)\rangle = \psi(\ell - M + 1) - \log(\ell - M + 1)$ and $\sigma_0^2 = \langle \log(\xi)^2\rangle - L_0^2 = \psi'(\ell - M + 1)$, ψ and ψ' being the di- and tri-gamma functions [5], respectively. In order to implement such a filter, we define the energy-flux *cepstrum* as the inverse Fourier transform of $\log \bar{\mathcal{S}}^0$:

$$\mathcal{C}_n = \frac{1}{N} \sum_{k=0}^{N-1} \log\big(\bar{\mathcal{S}}_k^0\big) \mathrm{e}^{-2\pi i \frac{kn}{N}}. \tag{S6}$$

According to a generalized central-limit theorem for the Fourier transform of iid stochastic variables [6], for $N \to \infty$ the *cepstral coefficients*, $\mathcal{C}_n$, are iid normal variates

whose expectation value is

$$\langle \mathcal{C}_n \rangle = \frac{1}{N} \sum_{k=0}^{N-1} \left[\langle \log(\xi) \rangle + \langle \log(\bar{S}_k^0) \rangle \right] e^{-2\pi i \frac{kn}{N}}$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} \left[L_0 + \langle \log(\bar{S}_k^0) \rangle \right] e^{-2\pi i \frac{kn}{N}}$$

$$= L_0 \delta_{n0} + C_n,$$

where $C_n = \frac{1}{N} \sum_{k=0}^{N-1} \log(\bar{S}_k^0), e^{-2\pi i \frac{kn}{N}}$, and the variance, that can be computed in a similar way, is $\langle \mathcal{C}_n^2 \rangle - \langle \mathcal{C}_n \rangle^2 = \frac{1}{N} \sigma_0^2$. The assumption that S is a smooth function implies that the number of significant cepstral coefficients $\mathcal{C}_n$ in the cepstral expansion of Eq. (S6) is hopefully *small*, allowing us to implement here a low-pass filter. We thus indicate by P^* a small integer such that

$$C_n \approx 0 \quad \text{for} \quad P^* \leq n \leq N - P^*. \quad \text{(S7)}$$

Unfortunately, the expectation value of the inverse Fourier transform of $\mathcal{C}_n$ does not represent $\log(\bar{S}_k^0)$, because of the additional L_0 factor:

$$\left\langle \sum_{n=0}^{N-1} \mathcal{C}_n e^{2\pi i \frac{kn}{N}} \right\rangle = L_0 + \sum_{n=0}^{N-1} C_n e^{2\pi i \frac{kn}{N}}.$$

Since we are interested in the $k = 0$ component and Eq. (S7), we must consider the stochastic variable

$$\mathcal{L}^0 = -L_0 + \mathcal{C}_0 + 2 \sum_{n=1}^{P^*-1} \mathcal{C}_n, \quad \text{(S8)}$$

whose expectation value and variance are

$$\langle \mathcal{L}^0 \rangle = \log(\bar{S}_0^0)$$
$$\text{var}(\mathcal{L}^0) = \frac{4P^* - 2}{N} \sigma_0^2. \quad \text{(S9)}$$

$\mathcal{L}^0$ is thus an *unbiased estimator* for the logarithm of the SCCB in the zero-frequency cross-spectrum of the conserved fluxes and, hence, of the thermal conductivity we are after. At fixed P^*, the variance of $\mathcal{L}^0$ vanishes in the $N \to \infty$ limit, thus making the estimator consistent. Of course, any finite choice of P^* introduces a bias, which is a decreasing function of P^*, while the statistical error is an increasing function of it. Its optimal value is the one which makes the bias of the order of the statistical error. By adopting this value in the $N \to \infty$ limit both the bias and the statistical error can be made arbitrarily small.

An estimate of the "optimal" value for P^* can be obtained with any of the many *model selection* techniques available in the statistics literature [7]; in our applications we choose to adopt the Akaike information criterion (AIC) [8, 9], even if other more sophisticated, and possibly more efficient, methods could be devised.
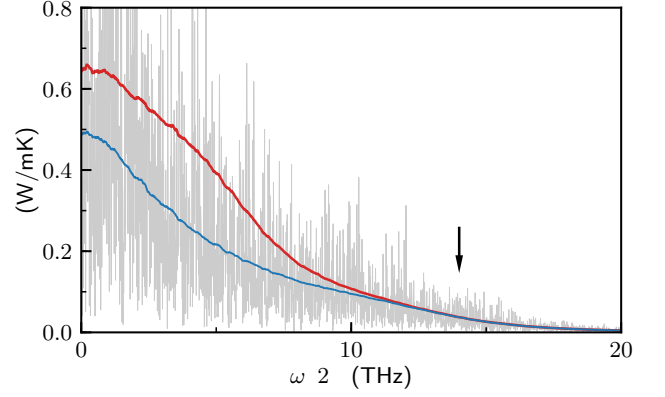

FIG. S1. Energy flux of molten NaCl. Gray: heat flux periodogram from a 100 ps trajectory segment. Red and blue: moving average over a window of 0.2 THz of the energy-flux periodogram, and moving average of the reduced periodogram $\bar{\mathcal{S}}^0(\omega_k)$, respectively, computed over the whole trajectory (34 ns). The vertical arrow indicates the value of the Nyqvist frequency used to perform the cepstral analysis.

S2 – APPLICATION TO A MOLTEN SALT

In this section we present the results obtained for molten sodium chloride (NaCl), paradigmatic of a two-component (neutral) mixture of charged particles. As for the water-ethanol mixture, classical MD simulations were run with the LAMMPS package [10] in the NVE ensemble. For NaCl we used the Fumi-Tosi [11] classical force field, and ran simulations at a temperature of $\approx 1400\,\text{K}$ and a density of $1.29\,\text{g/cm}^3$, corresponding to 864 formula units in a cubic cell with an edge of 40.21 Å, and a simulation time step of 1 fs. The cepstral analysis was again performed over blocks of 100 ps. The energy flux is computed via the standard expression adopted in the LAMMPS package. A 34 ns long trajectory was employed to extract many 100 ps blocks, in order investigate the statistics of the resulting estimates for the transport coefficients. In Fig. S1 we show the (moving average of the) energy flux periodogram, $\mathcal{S}_k^0$ (red), along with the reduced periodogram $\bar{S}_k^0$ (blue), whose expectation value is proportional to the thermal conductivity in conditions of zero charge current (see Eq. (3) in the main manuscript), according to

$$\kappa = \frac{V}{2k_B T^2} \bar{S}^0(\omega = 0). \quad \text{(S10)}$$

We notice that the red and blue curves substantially differ at low frequencies, thus giving a different zero-frequency value, and demonstrating the importance of employing Eq. (S10) in place of $\kappa = \frac{V}{2k_B T^2} S^{00}(0)$, when there is more than one conserved flux. From a standard Green-Kubo analysis of the whole trajectories (34 ns) we obtained $\kappa_{\text{ref}} = (0.485 \pm 0.005)\,\text{W/mK}$, that can be con-
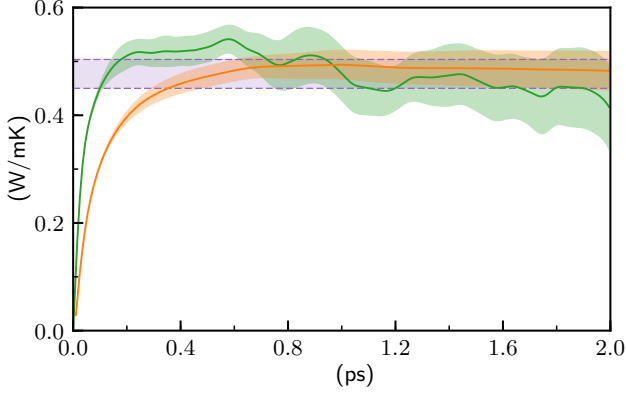
FIG. S2. Thermal conductivity estimator for molten NaCl, as a function of the upper limit of time-integration. The shaded ares surrounding the lines indicate one standard deviation, as estimated by a standard block analysis over all the trajectory segments. Green: direct time-integration according to Eqs. (2-3) in the main text. Orange: direct time-integration of the Einstein-Helfand integrals. Purple: cepstral analysis estimate from the reduced periodogram.

sidered as a reference. Furthermore, for each of the 100 ps segments we computed $\bar{\mathcal{S}}_k^0$, and obtained an estimate $\bar{S}^0(0)$ and the thermal conductivity κ, according to Eq. (S10). We set $\ell = 3$ (Cartesian components) and $M = 2$ (number of conserved fluxes) to define the theoretical statistical distributions necessary for the cepstral analysis. As in the case of the water-ethanol mixture, the parameters are $L_0 = 1 - \gamma - \log 2 \approx 0.270$ and $\sigma_0^2 = \pi^2/6 - 1 \approx 0.644$. For the cepstral analysis we employed the cutoff frequency $f^* \approx 14\,\text{THz}$. The estimate of κ has been verified to not depend on the specific value of f^*. The number of cepstral coefficients P^* has been estimated by the optimization of the AIC [8, 9] for each segment independently, and the average number of coefficients is: $P_A^* = 5$. Finally, the relative error on κ estimated by the second of Eqs. (S11) is 6.5%. The average value of thermal conductivity of NaCl, computed over all the blocks from cepstral analysis, is $(0.487 \pm 0.002)\,\text{W/mK}$, in perfect agreement with the reference values obtained above.

S3 – WORKFLOW

In a nutshell, our cepstral method requires the following list of operations, summarized in this Section for clarity:

1. Extract ℓ independent samples of N (equally spaced) steps of the time series for the energy flux $\mathcal{J}_n^0$, $n = 1, \ldots, N$, and for a complete set of independent particle fluxes $\mathcal{J}_n^m$, $m = 1, \ldots, M - 1$.

2. Compute the discrete Fourier transforms of these fluxes, $\tilde{\mathcal{J}}^i(\omega_k)$, the cross-periodogram $\mathcal{S}^{ij}(\omega_k)$, $i, j = 0, \ldots, M - 1$, and the SCCB of the cross-periodogram (with the factor ℓ/ν) for each frequency, $\bar{\mathcal{S}}^0(\omega_k)$, where $k = 1, \ldots, N$.

3. Calculate the logarithm of the SCCB of the cross-periodogram, $\log \bar{\mathcal{S}}^0(\omega_k)$.

4. Compute the inverse discrete Fourier transform of the result to obtain the cepstral coefficients $\mathcal{C}_n$.

5. Retain the first P^* cepstral coefficients, the optimal number P^* being computed via the Akaike information criterion [4, 9].

6. Evaluate the thermal conductivity and its statistical uncertainty estimated as:

$$
\begin{aligned}
\mathcal{K} &= \frac{V}{2k_B T^2}\exp\left[\mathcal{C}_0 + 2\sum_{n=1}^{P^*-1}\mathcal{C}_n - L_0\right] \\
\frac{\Delta\kappa}{\kappa} &= \sigma_0\sqrt{\frac{4P^* - 2}{N}},
\end{aligned}
\tag{S11}
$$

where L_0 and σ_0 are properties of the χ^2 distribution, as discussed before.

The relative error in the conductivity results from the estimate of the absolute error in its logarithm. P^* depends in general on the cutoff (Nyqvist) frequency, f^*, used to analyze the spectrum, while the final value of κ and its statistical uncertainty are largely independent of it [4, 9].

This workflow of operations is implemented in the `ThermoCepstrum` code [12], for which we provide an example folder at `https://github.com/lorisercole/thermocepstrum`.

---

* baroni@sissa.it
[1] N. Wiener, Acta Math. **55**, 117 (1930).
[2] N. Goodman, Ann. Math. Stat. **34**, 152 (1963); **34**, 178 (1963); A real Wishart deviate $\mathcal{W}_M(S, \ell)$ is defined as the distribution of the random matrix $\mathcal{W}^{ij} = \frac{1}{\ell}\sum_p {}^p\mathcal{X}^i \, {}^p\mathcal{X}^j$, where ${}^p\mathcal{X}^i$ $(p = 1, \ldots, \ell)$ are ℓ independent samples of an M-variate normal deviate of covariance $\langle\mathcal{X}^i \, \mathcal{X}^j\rangle = S^{ij}$. In the straightforward extension to complex Wishart matrices $\mathcal{CW}_M(S, \ell)$, the covariance matrix is Hermitian, $S^{ij} = \langle(\mathcal{X}^i)^* \, \mathcal{X}^j\rangle = (S^{ji})^*$ rather than symmetric as in the real case.
[3] This property is demonstrated in the real case, *e.g.*, in M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics (Springer Texts in Statistics)* (Springer-Verlag, New York, 1999), Propositions 7.5 p. 90, and 7.9, p. 92. The generalization to the complex case is trivial.
[4] 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*,

edited by W. Andreoni and S. Yip (Springer International Publishing, Cham, 2018) pp. 1–36, 2nd ed., arXiv:1802.08006 [cond-mat.stat-mech].

[5] E. W. Weisstein, "Polygamma functions," From MathWorld – A Wolfram Web Resource `http://mathworld.wolfram.com/PolygammaFunction.html`.

[6] T. W. Anderson, *The Statistical Analysis of Time Series* (Wiley-Interscience, 1994); M. Peligrad and W. B. Wu, Ann. Prob. **38**, 2009 (2010).

[7] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging* (Cambridge University Press, 2008).

[8] H. Akaike, *Information theory and an extension of the maximum likelihood principle, in 2nd International Symposium on Information Theory* (edited by B. N. Petrov and F. Cski, 1972) pp. 267–281; IEEE Trans. Autom. Control **19**, 716 (1974).

[9] L. Ercole, A. Marcolongo, and S. Baroni, Sci. Rep. **7**, 15835 (2017).

[10] S. Plimpton, J. Comput. Phys. **117**, 1 (1995).

[11] F. G. Fumi and M. P. Tosi, J. Phys. Chem. Solids **25**, 31 (1964); J. Phys. Chem. Solids **25**, 45 (1964).

[12] L. Ercole and R. Bertossa, "ThermoCepstrum: a code to estimate transport coefficients from the cepstral analysis of a multi-variate current stationary time series," `https://github.com/lorisercole/thermocepstrum` (2017–2018).